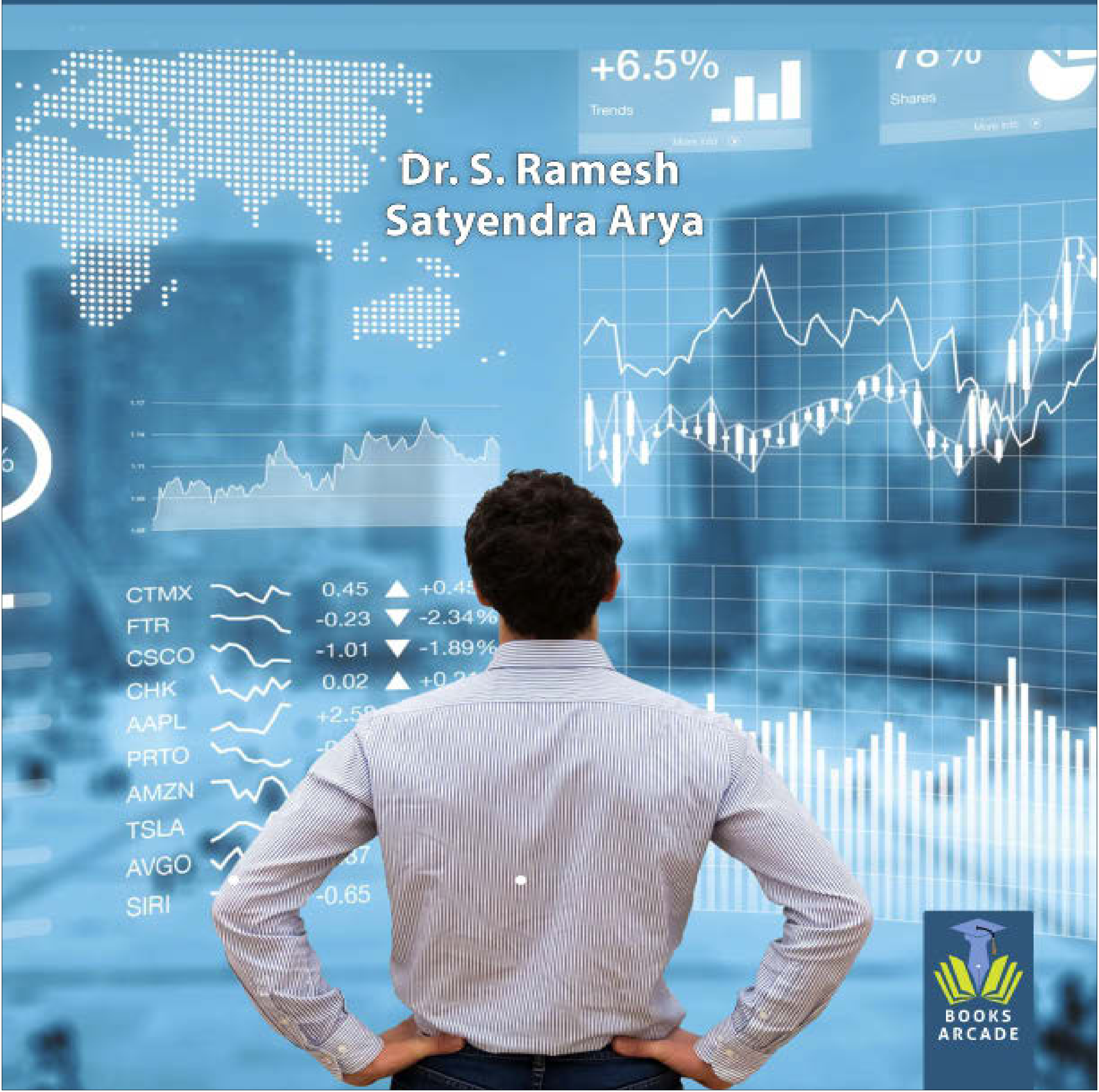
PRINCIPLES BUSINESS FORECASTING

Dr. S. Ramesh
Satyendra Arya



BOOKS ARCADE

Principles Business Forecasting

Principles Business Forecasting

Dr. S. Ramesh
Satyendra Arya

BOOKS ARCADE
KRISHNA NAGAR, DELHI

Principles Business Forecasting

Dr. S. Ramesh
Satyendra Arya

BOOKS ARCADE

Regd. Office:
F-10/24, East Krishna Nagar, Near Vijay Chowk, Delhi-110051
Ph. No: +91-11-79669196, +91-9899073222
E-mail: info@booksarcade.co.in, booksarcade.pub@gmail.com
Website: www.booksarcade.co.in

Year of Publication 2023

International Standard Book Number-13: 978-81-19199-74-7

CONTENTS

CHAPTER 1

PREDICTIVE ANALYTICS FOR EFFECTIVE MANAGEMENT DECISION MAKING: A FORECASTING APPROACH

Dr. S. Ramesh, Professor
Department of Commerce, JAIN (Deemed-to-be University), Bangalore, India
Email Id- ramesh.srinivasan@jainuniversity.ac.in

ABSTRACT:

Forecasting is a crucial component of management decision-making processes as it allows organizations to plan and prepare for the future. Forecasting involves the use of past and present data to predict future events, trends, and outcomes. Effective forecasting requires a systematic approach that involves gathering and analyzing relevant data, selecting appropriate forecasting methods, and communicating the results to decision-makers. Managers must also consider factors that may impact the accuracy of the forecasts, such as changes in the economic or political environment, shifts in consumer behavior, and technological advancements.

KEYWORDS:

Data Analysis, Decision-Making, Forecasting, Management, Trends.

INTRODUCTION

Businesses of today are always seeking for ways to increase their financial performance. Due to the dynamic nature of markets, business forecasters have been crucial to the success of these entrepreneurial endeavors by presenting top management with the most precise projections available. The goal of a business prediction is to use statistical analysis and subject-matter expertise to create reliable projections that will eventually direct subsequent corporate planning efforts. In order to give senior management with business decision options that are supported by sound statistical research, a skilled business analyst or forecaster must concentrate on identifying and monitoring the important business indicators that have an influence on sales and revenue [1].

Technology advancements have completely changed how we analyses information and create business and economic projections. These developments in forecasting theory and practice are a direct result of the complexity and competitiveness of global business growing. A solid knowledge basis is crucial since complexity raises the risk involved in business choices. Nowadays, forecasting is a tool used by businesses of all sizes to make financial and operational choices. Few managers are acquainted with the range of methods that have been developed over the previous several decades, despite the fact that most managers are aware of the need for better forecasting. They depend on knowledgeable, experienced people. Professionals may use advanced data analysis methods for predicting with the use of personal computers.

Methodologies for forecasting have been around since the eighteenth century. Regression analysis serves as one illustration. The subject of forecasting has substantially increased as a result of recent innovations like the Box-Jenkins and neural networks. Forecasting experts must master the use of these tools as more sophisticated and complicated methodologies are created, just as their managers must get a fundamental understanding of these forecasting options. Most people are interested in using these methodologies in their own work in a practical way, even if there may be some respect for the theoretical frameworks. This book offers forecasting approaches that academics and forecasting experts may utilize to provide management information for making decisions.

Forecasting's goal is to provide managers information that will make decision-making easier. Almost every organization, whether it be public or private, works in a dynamic and unpredictable environment with incomplete foresight. Organizations require a forecasting process that enables them to accurately and promptly anticipate the future since forecasting is a crucial component of the planning and control system. Making the proper judgments and being able to predict future changes are key components of effective business leadership. Even though there may still be some degree of uncertainty in these business choices, forecasting may be utilized as a tool to help. The general goal of top management is to make judgments based on anticipating economic elements that are important for strategic planning and execution. Forecasters may lessen the level of ambiguity around a business choice even if they cannot predict the future with absolute precision [2].

Managers must make strategic choices in every area of the business structure, from production and inventory to buying, accounting, marketing, finance, people, and services, in order to compete in the global economy. Many chief executives of multinational corporations (MNCs) are aware of how crucial foreign markets are to the expansion of their companies. Several of these businesses have a lengthy history of doing business internationally. Others are attempting to profit from the altered economic climate brought forth by globalization.

The structure and scale of these enterprises suggest that businesses operating internationally have a significant chance of making financial advantages. Several MNCs generate more than half of their revenues outside of their home nations. For instance, companies like Dow Chemical, Coca-Cola, Honeywell, and Eastman Kodak have long been present on the global market and continue to generate large profits from their activities abroad. With operations in more than 200 countries and a diversified staff of around 90,500 people, The Coca-Cola Company generated more than 71 percent of its revenues from sales outside of the United States in 2007. Forecasting is a crucial tool in making business choices for these companies. By providing products or services that target certain niche markets, smaller businesses are also breaking into the global market. Recognizing market opportunities for their goods and making accurate demand projections in these areas are essential to their capacity to thrive in these markets.

DISCUSSION

When addressing the state of the economy as a whole, policymakers and economists both have to make strategic choices. Making sound business choices has depended on generating accurate macroeconomic predictions. It may be argued that the shorter time intervals between actions and their effects, as well as the speeding up of globalization and technological progress, have made forecasting more difficult [3]. Growth, stagnation, or recession the three most basic macroeconomic questions have an impact on almost everything significant to a company. After all, the predicted need for higher production greatly influences investments in plant, equipment, acquisitions, inventories, systems, employees, and training. Accurate projections are necessary for each and every one of these choices.

Every functional area of company may benefit from the use of forecasting, which is a strong instrument. Forecasting is used by production managers to direct their inventory management and production plan. Companies with several product lines are worried about minimizing labor and material costs. Moreover, the manufacturing process is greatly influenced by trends, the availability of materials, manpower, and plant capacity. In light of new product lines, new markets, and ambiguous demand circumstances, production managers need frequent short-term predictions of product demand as well as long-term demand projections.

Marketers need projections to help them make choices, just as production managers do. Making decisions on marketing strategy and advertising plans and expenditures requires accurate estimates of the market's size and characteristics, including market share, pricing trends, sources of competition, and market demography. The projection might also take into account inventories, sales income, and product demand [4]. A crucial component of product research is forecasting. Marketers create their projections using both qualitative and quantitative methods. The Delphi method, sales force estimates, customer intention surveys, and the jury of executive opinion are examples of qualitative forecasting techniques. These qualitative forecasting techniques are quick, affordable, and adaptable. These methods have drawbacks because they rely on subjective judgments, which introduce biases, uncertainties, and inconsistencies into the forecast. The time-series method or causal models are the quantitative methods used to forecast market conditions. Later chapters of this book discuss this methodology.

Service sector industries such as financial institutions, airlines, hotels, hospitals, sport and other entertainment organizations all can benefit from good forecasts. Finance and accounting departments make use of forecasting in a number of areas. Financial forecasting allows the financial manager to anticipate events before they occur, particularly the need for raising funds externally. The most comprehensive means of financial forecasting is to develop a series of pro forma, or projected, financial statements. Based on the projected statements, the firm is able to estimate its future levels of receivables, inventory, payables, and other corporate accounts as well as its anticipated profits and borrowing requirements. Cash flow and rates of revenue and expense projections are critical to making business decisions.

In addition, speculation in the asset markets requires the use of effective forecast. The airlines, whether large or small, can benefit from good forecast of the load factor, fleet management, fuel and other cost projections. In the hotel and entertainment industries, accurate projection of hotel occupancy rates, for example, have implications for all the other guest services offered. Hospitals have long used forecasting tools to determine the use of emergency room personnel, and cost projections. In the sport industry, forecasts are used for ticket sales for any sporting event. Revenue projections are made based on the performance of a team during a year or years. Figure 1 illustrate the Different Demand of Forecasting Technique.

Figure 1: Illustrate the Different Demand of Forecasting Technique.

The use of forecasts in human resource departments is also critical when making decisions regarding the total number of employees a firm needs. This has implications for the resources of the firm and the need for training of employees. Such forecasts as the number of workers in functional areas, the nature of the workforce (i.e., part-time versus full-time), trends in absentee-ism and lateness, and productivity can be helpful in resource planning and management decisions.

Forecasts are used in the public sector in making decisions in the macro- economy. Economic policy is based, in part, on forecast of important economic indicators. Projections of the GNP, employment, rate of inflation, industrial production, and expected revenues from personal and corporate income taxes all depend on good forecasts. Government uses these forecasts to guide monetary and fiscal policy of the country. Among the many uses of forecasts, population (or demographic) forecasts play an important role in planning government expenditures on health care, social insurance, and infrastructure [5].

The above examples of how forecasts are used in the various business and economic activities are by no means exhaustive. This simply indicates the significance and breadth of forecasting in decision making. Forecasting as a tool in planning has received a great deal of attention in recent decades. Part of this increased attention is the need to operate successfully in a dynamic global market that is changing constantly. Secondly, with technological advances in computers and quick access to firm-generated data, organizations are looking at ways to improve their decision-making processes. Furthermore, methodological improvements in forecasting have expanded the ability of managers in the private and the public sectors to effectively use these tools in making timely business and economic decisions. How to incorporate these developments into the firm's decisions is both an art and a science. Figure 2 illustrate the need of an effective budget system in the company.

Today, firms have a wide range of forecasting methodologies at their disposal. These vary from intuitive predictions to extremely advanced quantitative approaches. Each of these strategies has its pros and limits. To employ them correctly is an art. A manager must choose a certain technique based on both personal experience and professional judgment. The skill of predicting is in knowing when it is necessary and knowing how to combine qualitative and quantitative data. This paper addresses forecasting methods that may support management choices made with common sense.

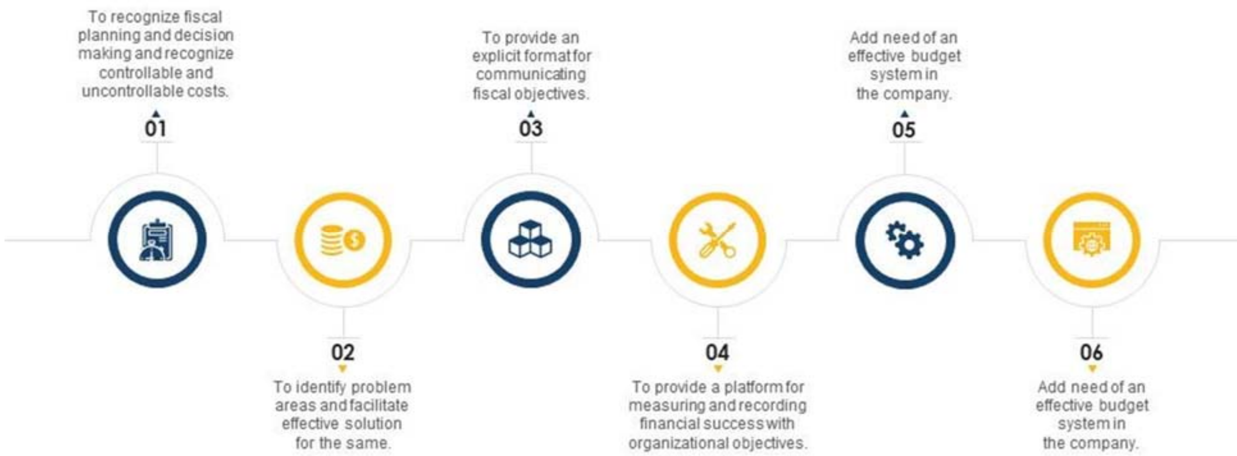


Figure 2: Illustrate the need of an effective budget system in the company.

The scientific foundations of model construction include the science of predicting. As in any other branch of science, scientists start by offering the most straightforward explanation for a phenomenon. The model is often regarded as a suitable instrument for future prediction if it accurately captures the circumstances of the actual world and its outcomes match observable phenomena. On the other side, scientists adopt more sophisticated models if the basic model is unable to adequately capture or explain the observed occurrence. The number of assumptions that must be made in a model increases with its complexity [6].

Simple models have been employed by economists to identify data patterns, which have subsequently been utilized to make future predictions. An economist may use an economic theory or model, which is a collection of definitions and presumptions, to explain certain kinds of occurrences. An economic theory outlines the process through which certain economic variables interact, usually in the form of a set of equations. For instance, according to the theory of consumer choice, customer preferences, income levels, the cost of the item in question, as well as the cost of comparable products and services, all influence how much of a given thing people would purchase. According to this notion, when a good's price increases, fewer people will normally buy it. There are theories in macroeconomics that suggest the overall amount of investment is influenced by the interest rate. These theories specifically suggest that higher interest rates would deter investment in real capital development (investment). We must ascertain the accuracy of these theories' predictions (forecasts) of economic events in order to assess their utility. To capture the effect of the many factors on the model under these circumstances, multivariate models are utilized [7].

A company must have a systematic forecasting process that can be swiftly implemented and changed in order to consistently provide accurate projections as required. The forecaster benefits by following a procedure that is based on the scientific method, just as in any other scientific activity. The instructions that make up the language of the scientific method are descriptions of sequences of acts or procedures that are precise enough for anybody to follow. These guidelines are referred to as "operational definition." An operational definition should include a list of all steps required to repeatable produce the same outcomes.

Forecasting procedures might be simple or complex. It starts when an organization's management needs a response to a management decision. For instance, they can inquire as to whether a product's enhancements would result in a material rise in demand. In other circumstances, the management could request a prediction if they need to commit a significant number of resources to a project or if a situation in the business environment indicates that a choice has to be made. A forecast's opening statement is a question. The forecaster is in charge of making sure that the forecast's goals are spelled out in detail [8]. It is crucial to understand why the prediction is required and how the outcomes will be utilized. Depending on the kind of issue and the forecast's duration, organizations create unique predictions. For instance, opposed to production projections that are weekly or monthly, capital budget estimates are long-term in nature.

After the choice is taken to create a prediction, a theoretical model that addresses management concerns must be created. The link between the model's many variables may be elaborated on using the theoretical framework of the model. It also enables the division of impacts into internal and external elements. The term "internal factors" refers to those variables that the company may influence. Price, product quality, product attributes, marketing and advertising expenses, and logistics are a few examples (distribution facilities). Elements that are external to the company operate beyond the firm's control. The interest rate, the rate of inflation, income, employment, and the exchange rate in global trade are examples of exogenous influences. In addition, we may choose a trend or regression model for this assignment if management is interested in a long-term trend projection. The moving average, exponential smoothing, or Box-Jenkins models, for example, may be used in the analysis if management is interested in a short-term prediction, i.e., weekly or monthly forecasts.

The forecaster is now prepared to acquire data that supports the analysis's conceptual framework after determining the theoretical model. The information may originate from company records or from other sources.

The kind and quality of the data being collected should be carefully considered. Unless it is essential to their everyday operations, businesses often do not gather disaggregated data. The company may need to obtain such data when highly disaggregated data are required to produce a projection. It could take more time and money to do so, which would make forecasting more difficult. When presented with a situation like this, the forecaster must weigh all of the potential outcomes before making a prediction [9].

Data analysis should not be seen as only a mechanical procedure. The forecaster should get quite familiar with the characteristics of the company and the sector it represents before calculating the forecasting model.

The market structure in which the business works, the sources of competition, the company's position within the industry, etc., are all examples of information that aids in forecasting and should be thoroughly examined. As a result, the forecaster may see the predicted model in a dynamic way. Changes may be made at this point if a need to reassess the findings arises. Models should be examined for validity and reliability. To evaluate the model's accuracy, forecast and actual outcomes should be compared. At this point, evaluation acts as a control procedure. Sometimes, the dependability

The model's reliability and validity increase with the addition of new variables, changes to the time frame, or adjustments to the data's periodicity. The forecast should be modified appropriately using the feedback loop. Presenting the findings to management is the last step in the forecasting process. The management of the company seeking the prediction must realize that, even if the duty of giving a forecast may be finished, the process of fine-tuning the forecast has not. To be clear, a good prediction is dynamic rather than static. The management would benefit from developing a process for regularly assessing its projections and from being aware that unforeseen market changes can affect the predicted model's underlying assumptions, necessitating a new estimate. Forecasters in the information age have access to a broad variety of datasets that may be quickly accessed for use in predicting. Aggregate and disaggregate data for private and public sector organizations is sourced from various databases [10], [11].

CONCLUSION

Forecasting is a critical tool for management decision-making. It allows organizations to anticipate future events, trends, and outcomes, enabling them to plan for and respond to potential risks and opportunities. Effective forecasting requires a systematic approach that involves gathering and analyzing relevant data, selecting appropriate forecasting methods, and considering factors that may impact the accuracy of the forecasts. By using forecasting techniques, managers can make informed decisions about future investments, resource allocation, and strategic planning, enabling their organizations to stay ahead of the competition. Overall, forecasting is a vital component of effective management decision-making, and organizations that prioritize it are more likely to achieve long-term success.

REFERENCES

[1] S. Seyedzadeh, F. P. Rahimian, I. Glesk, and M. Roper, “Machine learning for estimation of building energy consumption and performance: a review,” *Visualization in Engineering*. 2018. doi: 10.1186/s40327-018-0064-7.

[2] P. Norouzian-Maleki, H. Izadbakhsh, M. Saberi, O. Hussain, M. Jahangoshai Rezaee, and N. GhanbarTehrani, “An integrated approach to system dynamics and data envelopment analysis for determining efficient policies and forecasting travel demand in an urban transport system,” *Transp. Lett.*, 2022, doi: 10.1080/19427867.2020.1839716.

[3] A. S. Ahmad *et al.*, “A review on applications of ANN and SVM for building electrical energy consumption forecasting,” *Renewable and Sustainable Energy Reviews*. 2014. doi: 10.1016/j.rser.2014.01.069.

[4] V. Trubnikov, A. Meynkhard, K. Shvandar, O. Litvishko, and V. Titov, “Medication market performance analysis with help of analytic hierarchy processing,” *Entrep. Sustain. Issues*, 2020, doi: 10.9770/jesi.2020.8.1(60).

[5] A. Paulauskaite-Taraseviciene, V. Raudonis, and K. Sutiene, “Forecasting municipal solid waste in Lithuania by incorporating socioeconomic and geographical factors,” *Waste Manag.*, 2022, doi: 10.1016/j.wasman.2022.01.004.

[6] A. G. Fayoumi and A. F. Hajjar, “Advanced learning analytics in academic education: Academic performance forecasting based on an artificial neural network,” *Int. J. Semant. Web Inf. Syst.*, 2020, doi: 10.4018/IJSWIS.2020070105.

[7] A. Naseem, S. T. H. Shah, S. A. Khan, and A. W. Malik, “Decision support system for optimum decision making process in threat evaluation and weapon assignment: Current status, challenges and future directions,” *Annual Reviews in Control*. 2017. doi: 10.1016/j.arcontrol.2017.03.003.

[8] K. P. Lin, P. F. Pai, and Y. J. Ting, “Deep belief networks with genetic algorithms in forecasting wind speed,” *IEEE Access*, 2019, doi: 10.1109/ACCESS.2019.2929542.

[9] S. M. Abilov and A. A. Kaigorodtsev, “Qualimetric methods in assessing the effectiveness of management decisions in universi-ties,” *Humanit. Soc. Sci. Rev.*, 2019, doi: 10.18510/hssr.2019.7529.

[10] H. N. Perera, B. Fahimnia, and T. Tokar, “Inventory and ordering decisions: a systematic review on research driven through behavioral experiments,” *International Journal of Operations and Production Management*. 2020. doi: 10.1108/IJOPM-05-2019-0339.

[11] M. Lei, S. Li, and S. Yu, “Demand forecasting approaches based on associated relationships for multiple products,” *Entropy*, 2019, doi: 10.3390/e21100974.

CHAPTER 2

UNVEILING DATA PATTERNS FOR IMPROVED FORECASTING: A COMPARATIVE ANALYSIS OF FORECASTING TECHNIQUES

Dr Yavana Rani.S, Associate Professor
Department of Decision Sciences, CMS Business School, JAIN Deemed to-be University, Bangalore, India
Email Id- dr.yavanarani@cms.ac.in

ABSTRACT:

Data patterns and the choice of forecasting techniques are two important aspects of time series analysis. In this context, data patterns refer to the underlying structure of the time series, such as trend, seasonality, and cycles. Understanding these patterns is crucial for selecting appropriate forecasting techniques, which can accurately predict future values of the time series. There are several techniques available for forecasting time series data, including simple moving averages, exponential smoothing, and ARIMA models. The choice of technique depends on the type of data pattern present in the time series, as well as other factors such as the amount of available data and the desired level of accuracy.

KEYWORDS:

Cycles, Data Patterns, Forecasting Techniques, Seasonality, Time Series Analysis.

INTRODUCTION

Forecasting in real-world business is a combination of science and art. It is a science in that accurate forecasting will always be improved by the proper use of advanced statistical methods. Empirical evidence seldom, if ever, accurately predicts the outcome of an art. The user must pick amongst many relationships to find the equations that will provide predictions that are as accurate as possible. This book does not claim that prediction error can be totally eliminated; rather, it demonstrates how to reduce it. A number of forecasting techniques may be used to this end. These approaches will often complement one another rather than be antagonistic. Predicting the course of change alone may work in certain cases.

For instance, a model that could successfully forecast the direction of the stock market the next day would be very useful and lucrative, even without knowing how much it would increase or decrease. Despite several attempts, no such model has ever been successfully built, and the objective is likely to remain elusive. On the other hand, because monthly movements in the CPI have been negative just about 1% of the time during the previous 40 years, a model that predicted the direction but not the size of change in the consumer price index (CPI) the next month would be essentially meaningless [1]. There are several forecasting methods, and not all of them are based on exact statistical methods. The finest predictions may sometimes come from educated judgement, like when "insiders" have access to corporate information that no one else has. Surveys may provide helpful information on predictions for the whole economy, certain businesses or enterprises, or specific sectors. These techniques should be used if they increase predicting accuracy. These are just briefly covered in this work since there is no reliable method to determine how much educated judgements or survey approaches have improved prediction accuracy. The emphasis of this article is on demonstrating how statistical and econometric techniques may be utilized to build forecasting models and reduce prediction errors. Originally, structural equations were used to construct the majority of economic projections; more recently, time-series analysis has been more successfully used. Both systems' advantages and drawbacks in terms of producing the best projections are noted.

The focus of this book is on useful business forecasting; it is not a theoretical treatise. Because of this, the number of theorems and proofs which are also present in many other texts will be reduced to a minimal, with the majority of the content focusing on real-world forecasting instances. This work will specifically show how statistical theory often has to be modified to account for those issues that come up in real empirical estimate. It is not appropriate to disparage or disregard techniques for modifying the models to improve predicted accuracy since they are a crucial component of real-world business forecasting [2]. Probability theory and other mathematical techniques are used in statistics to help people make better decisions when there is uncertainty present. In addition to being extensively employed in economics, statistical theory and findings are also relevant to a broad range of other fields, including as sociology, agriculture, astronomy, biology, and physics.

When there is doubt, whether it stems from chance or a lack of knowledge about the real connection behind the test, statistics are used to offer solutions. We know the underlying probability distribution and, thus, the percentage of straights that will be dealt in a poker hand over the long run, but we do not know what the following hand will reveal, to demonstrate the first example. For example, we probably don't understand the real underlying link between capital expenditure and interest rates, or between inflation and unemployment rates, or between changes in the value of the dollar and net exports economic situations in which the underlying probability distribution is understood are uncommon [3]. Applying statistical and mathematical techniques to the study of economic data in order to support or challenge economic ideas is known as econometrics. The application of structural equations allows for the proof or denial of a certain hypothesis. Comparatively, the use of statistical techniques with economic data to get parameter estimates without mentioning a specific theory is growing.

DISCUSSION

Time-series analyses, and integrated autoregressive moving-average (ARIMA) models are one such method. There is no effort to assume an underlying theory; instead, these models simply correlate a given economic measure with its own lagged values after adjusting for trend and seasonal influences. Although while statistical or econometric approaches are often used in economic forecasting, this need not always be the case. Some predictions don't use any math at all; for instance, surveys or polls might provide accurate forecasts without using any econometric approaches. These projections, however, are not included in this book. The majority of the examples will only apply to predictions that make use of statistical techniques.

No prediction is ever accurate, and predictions of what will happen in the future almost always include mistakes. Everybody who has ever made an effort at prediction is aware of it. On the other hand, forecasting may be quite helpful if it yields more accurate results than other approaches to making predictions about the future. So, the relevant test for any prognosis is not whether the findings are correct in comparison to the alternatives, but rather how accurate they are. The acceptable response is always "Compared to what?" similar to the classic Henny Youngman one-liner "How's your wife?"

The distinction between the art of predicting and the science of statistics and econometrics is underlined throughout the whole book. The majority of theorems and proofs in those domains are based on improbable assumptions about the distribution of the error components, and they also assume that the method used to generate the data is constant for both the sample period and the forecast period. Indignantly referred to as ad hoc adjustment and unworthy of the name of econometrics, adjusting models to get better projections when these assumptions are not fulfilled is a common practice [4]. From 1940 to 1970, theoretical improvements in statistical and econometric processes received the majority of focus, but systematic approaches to forecasting adjustments to increase their accuracy received little attention. The emphasis gradually shifted to developing techniques that produced useful forecasts even if they did not follow the theoretical procedures developed in earlier decades as macroeconomic models failed to foresee any of the major changes in the economy in the 1970s.

The fundamental problems in forecasting, as opposed to econometrics, have been categorized by Clements and Hendry 1 in Predicting Economic Time Series, a reference book that is suggested for persons with more advanced mathematical abilities. "The features of the real-world forecasting venture that give rise to the sources of forecast error induce a fairly radical departure from the literature on "optimal" forecasting and at the same time help to explain why some seemingly ad hoc procedures work in practice," they write [5].

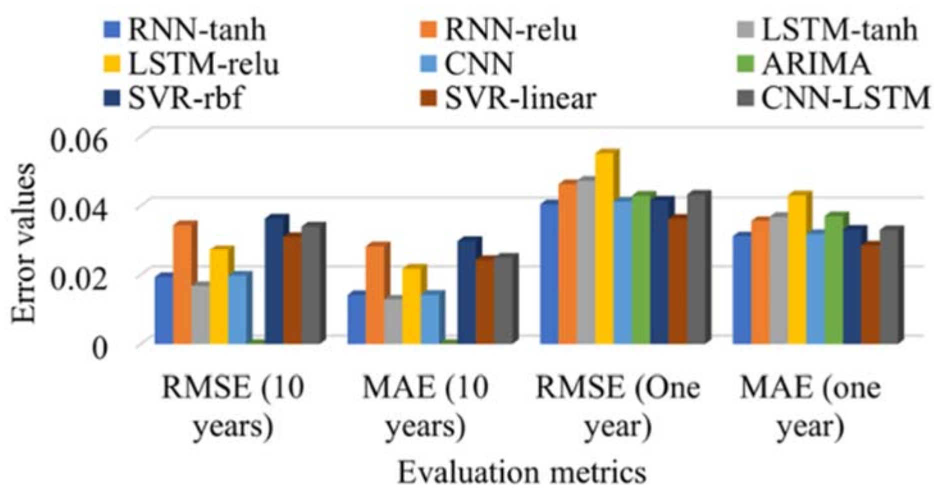


Figure 1: Illustrate the Comparative analysis of forecasting methods based on evaluation metrics for different datasets with different sample sizes.

Compared to Clements and Hendry, this method is much less rigorous mathematically. Also, the discussion of forecasting accuracy begins with structural models and then moves to time-series analysis, contrary to the procedure that they (and others) use. Yet the methodology in which real-world practical forecasting is approached is very much in the spirit of their approach. While the method of least squares is used for the vast majority of the examples, the reader should always keep in mind that the assumptions of the classical linear model seldom hold in practical business forecasting. Figure 1 illustrate the Comparative analysis of forecasting methods based on evaluation metrics for different datasets with different sample sizes.

In many cases, the underlying data generating function has shifted, the variables are not normally distributed, the residuals are not independent, and the independent variables are not known at the time of the forecast. Even more important, repeated rerunning of regression equations, and the substitution of different empirical data series that measure the same theoretical concept, often help to improve forecasting accuracy but are outside the constructs of the classical linear model. For this reason, the statistical estimates generated under those assumptions must be interpreted carefully, and often with a degree of skepticism. It is too crude to say that what makes the best forecasts is “whatever works,” but that is closer to the spirit of the present approach than the method of choosing rigorous statistical techniques that minimize the root mean square error or other similar measures in the sample period but generate suboptimal forecasts [6].

Sometimes structural econometric models provide better forecasts, and sometimes the results from ARIMA models with no economic structure are better. In certain cases, the combination of these methods will generate better forecasts than either method separately. Far from being relegated to the criticism of ad hoc adjustments, changing the model during the forecast period will invariably produce better results than a “pure” unadjusted model, provided it is done properly.

As Newbold and Granger have written, “the evaluation criteria employed should be as demanding as possible since the object ought to be self-criticism rather than self-congratulation”. The principal aim should be to build a forecasting model that will generate the smallest forecasting error, not necessarily maximise the goodness-of-fit statistics over the sample period. The reader should always keep in mind that any forecasting model, no matter how sophisticated the underlying statistical techniques, must perform better than forecasts generated by random variables or naive methods. That means always checking whether the model provides better results than other methods that are available – including naive models, surveys, and qualitative judgments. A naive model generally assumes that the level or rate of change of the variable to be predicted this period will be the same as last period, or the change this period will be the same as the average change over an extended time period.

For a time, series without any significant trend, such as the Treasury bill rate, a naive model might state that the bill rate this month will be the same as it was last month. For a time series with a significant trend, the naive model would usually be couched in terms of percentage changes. For example, a naive model might state that the percentage change in the S&P 500 stock prices index next month will equal the percentage change last month, or it might equal the average percentage change over the past 480 months. A more sophisticated type of non-structural model incorporates regression equations using lagged values of the variable that is to be predicted. If more complicated modelling techniques cannot generate forecasts that beat these naive models, the model building attempt is presumably not worthwhile [7]. For people engaged in industry and finance, where having more accurate forecasts than your competitors will materially improve profitability, forecasts are useful if they provide results that are more accurate than the competition's.

A model that accurately predicted the direction of change in the stock market the next day 60 percent of the time would be tremendously valuable even though it would be wrong almost half the time – regardless of the methodology used to develop those predictions. In a similar vein, calculations by this author have shown, in some semi-annual polls of economists published in the Wall Street Journal, over 50 percent of the forecasts incorrectly predicted the direction interest rates would change over the next six months. Hence any model that could even predict the direction in which interest rates would move over the next several months would significantly improve the current status of forecasting financial markets.

And the option not to anticipate at all involves throwing up the towel and declaring that any departures from historical patterns can never be expected. Only if the variable in issue constantly expanded at the same pace and was never exposed to external shocks would that be the case. Yet even if events are genuinely unexpected a (an oil shock, a war, a wildcat strike, a factory disaster) forecasting models can astill give important assistance showing how to get back on track. Almost everyone who has a managerial or executive position in business or finance makes predictions about the future. These estimates will never be exact, but if the practitioner uses strong statistical approaches and has the flexibility to modify the predictions when actual occurrences deviate from expected values, they are likely to be much improved.

Forecasters are humbled by their work. This does not necessarily imply that those who chose predicting as a career are modest; in fact, the contrary is more likely to be true. Yet, unlike economic theories, which can endure for decades before anybody can ever confirm whether they are true or helpful, forecasters typically learn fast whether or not their ideas are valid. It is sometimes stated that forecasting economic variables is not a helpful activity since highly visible projections of the entire economy or financial markets have built a very lacklustre track record over the previous 30 years. The majority of consensus predictions for real GDP, inflation, and interest rates haven't been substantially more accurate than those made by naïve models, in fact. Others have drawn the conclusion that predicting models do not perform very well in light of these findings. But, before coming to that conclusion, we need look at the root of these forecasting blunders. Imagine, for instance, if the majority of analysts believed that interest rates would rise as a result of impending inflation.

In anticipation of this, the Federal Reserve tightened policy to such an extent that inflation plummeted and, by the time six months had passed, interest rates actually declined. Whilst I don't claim that always happens, this is a plausible theory. It is thus necessary to classify the main causes of these mistakes before starting our investigation of how to decrease forecasting errors. Others can be lowered using a number of techniques that will be covered in this book, while others may be intractable [8]. The theory of economic forecasting is pretty well established when the econometric model and the process producing the model agree in a world that doesn't change and when the underlying facts are reliable and accurate. In these situations, the forecast period's root mean square forecasting error shouldn't be more than what the sample period data suggest.

This is a rare occurrence; most projections call for the act to establish a shop three blocks away. Before, a recession was associated with a stronger dollar on a macroeconomic level; now, it is associated with a lower dollar. Once American Can changed its business to provide financial services, entirely new issues started to affect its corporate profitability. While structural changes may be clear when put in such blunt words, they often take place in a more subtle way. Macroeconomists predicted that the US economy would develop at a slower pace from 1997 to 1999, between 2-2 12%, but real growth stayed close to 4% each year. Forecasters anticipated that full employment would lead to increased interest rates, declining stock values, and slower GDP, but this did not materialise. There were some fundamental changes in the economy, at least in hindsight. For one reason, greater inflation was no longer caused by full employment. Additionally, the technology revolution raised capital investment and productivity growth more quickly. But, even after a number of years, the consensus prediction was unable to predict this change.

Even after the best model has been estimated, certain terms may still be ignored. Model misspecification may result from the model builder's ignorance. These variables may often be expectational variables for which no data are available. For instance, economists agree that the predicted rate of inflation, a quantity that cannot be easily observed, influences bond rates. A business may discover that although lowering prices by 5% would not result in any competitive backlash, doing so by 10% would result in rivals matching the reduced pricing. The variable that would be absent in this scenario is the threshold at which rivals would react, which itself is subject to change over time.

Moreover, the underlying model could not be linear. The acquisition of capital goods with lengthy lifespans as opposed to computers and cars tends to expand more quickly when the rate of capacity utilisation is high than when it is low, according to one reasonably simple and often reported example. Although if interest rates are low, financing is freely accessible, stock prices are growing quickly, sales are thriving, and profits are surging, capital investment on long-lived assets is often slow at the start of a business cycle upturn. Even though borrowing rates are higher and growth is slower, once businesses reach their maximum capacity, they are more inclined to boost this kind of capital investment. This issue may be improved to some degree by adding variables that cause the equation to be nonlinear; I'll talk about one such case later. For instance, investment could increase more quickly when the rate of capacity utilisation is higher than a certain threshold (let's say 85%) than when it is lower.

A simple rule of thumb, however, will often lead to model misspecification since a given level of capacity utilisation would affect investment differently depending on the average age of the capital stock. An effort to determine the precise pace of capital expenditure there are three main sources of data that are often utilised to estimate forecasting models. The Federal Reserve Board of Governors, the Bureau of Economic Analysis (BEA), and the Bureau of the Census are among the federal institutions that produce the majority of macroeconomic statistics. Individual businesses generate financial market information on the sales and profitability of certain companies. The Conference Board index of consumer attitudes and the National Association of Purchasing Managers index of business conditions in the manufacturing sector are two examples of the many industry associations and private sector organizations that produce indexes of consumer and business sentiment, as well as measures of economic activity that fall into the intermediate category [9].

The vast majority of macroeconomic or industrial statistics are collected by sampling, which implies that only a very tiny portion of all transactions are measured, with the exception of particular data based on prices provided in financial markets. Data collecting techniques are sometimes insufficient, even when considerable effort is made to count every participant. Every individual in the US is meant to be tallied at the decennial census, however statisticians generally agree that the reported population in major cities is much lower than the actual population, with many of the uncounted persons being illegal immigrants. Hence, certain inaccuracies persist even in this most thorough data gathering endeavour that is designed to count everyone. The assumption that mistakes from smaller samples are comparatively bigger is plausible.

Almost all government-collected and -provided macroeconomic and industrial data series have been updated. The aforementioned issuing organisations make an effort to provide monthly or quarterly statistics as soon as practicable after the period has concluded. The term "advance" or "preliminary" data is often used to describe these releases. These results are then typically amended during the next several months. They are then changed once again using yearly benchmarks, and these modifications often take seasonal changes into account. Lastly, they are updated once more using five-year censuses of the manufacturing, service, and agriculture industries.

Moreover, revisions may be made to some of the more extensive statistics due to methodological changes, like the GDP and the CPI. The data that the federal government prepares and releases often undergoes significant modifications. This may occur when preliminary data, which are published soon after the relevant time period has finished, are based on a small sample and are later corrected when more complete data become available. Other times, seasonal elements change throughout time. The updated data accurately reflects this new information.

That since a stronger currency attracts foreign investment, interest rates will be lower than they otherwise would be. Each of these hypotheses can be independently tested, but without additional modifications, they are worthless for forecasting either the dollar's value or interest rates since they reach opposing results. This is a sign of a bigger forecasting issue, where an in isolation, a single theory could provide solid empirical findings, but since the variables it holds constant are in actuality constantly changing, it might be worthless for predicting.

These illustrations provide a sense of the difficulties in creating a useful forecasting model. Several of the examples feature many variables that must be predicted concurrently in conjunction with one another. Yet, model designers often err by failing to recognise the false correlation generated by similar trends in a number of the time series, even in circumstances when the independent variables are genuinely known beforehand and are therefore truly exogenous.

It is very difficult to create forecasting models using econometrics. "Econometric modelling is an activity that should not be done for the first time," as Clive Granger put it. To create effective forecasting models, practise is required. The foundation for estimating regression equations may come from statistical theories, but every modeller must still address the following practical issues: Where do I obtain the data? How should the data be gathered and organised?

There are hundreds, if not thousands, of data sources, as well as several graphical programmes and efficient econometric tools. Yet, this author and his colleagues have discovered that providing model builders with an exhaustive list of potential sources does not assist. This essay takes a more eclectic approach, giving the information sources and software programmes that have served us well in our own research. Nowadays, the Internet is a great resource for finding a lot of pertinent information. The addresses in this book come with the accompanying warning: These regularly change, and sometimes new connections are not provided. Nonetheless, even if the locations have changed, it should be simple to locate them by simply explaining the relevant data sources [10].

It is advisable to review the data for flaws before utilizing it for any kind of analysis. Spending a lot of effort developing and testing a model only to discover the underlying data are inaccurate is one of the most frustrating experiences. The data will often be indicated as being seasonally adjusted or unadjusted, however when the model is being used for forecasting, differing techniques of seasonal adjustment might generate a range of issues. Also, it is always advisable to look for outlier observations and missing or incomplete data, both of which might significantly skew the model's predictions used to calculate the association between income and a variety of factors, including geography, family size, race and age.

CONCLUSION

A comparative analysis of forecasting techniques is essential for businesses and organizations to make informed decisions about future trends and developments. Different forecasting techniques have their strengths and weaknesses, and it is crucial to choose the appropriate method based on the nature of the data, the purpose of the forecast, and the level of accuracy required. The accuracy of a forecasting technique is influenced by several factors, including the quality of the data, the selection of appropriate variables, the time horizon, the level of detail, and the forecasting model's complexity. Some forecasting methods, such as quantitative models, rely heavily on statistical analysis and historical data to generate forecasts. Other techniques, such as qualitative methods, involve expert judgment and subjective assessments to forecast future events.

REFERENCES

[1] I. Naim and T. Mahara, “Comparative Analysis of Univariate Forecasting Techniques for Industrial Natural Gas Consumption,” *Int. J. Image, Graph. Signal Process.*, 2018, doi: 10.5815/ijigsp.2018.05.04.

[2] S. Sood and K. Jain, “Comparative Analysis of Techniques for Forecasting Tourists’ Arrival,” *J. Tour. Hosp.*, 2017, doi: 10.4172/2167-0269.1000285.

[3] P. Ozoh, S. Abd-Rahman, J. Labadin, and M. Apperley, “A Comparative Analysis of Techniques for Forecasting Electricity Consumption,” *Int. J. Comput. Appl.*, 2014, doi: 10.5120/15426-3841.

[4] S. Raizada and J. R. Saini, “Comparative Analysis of Supervised Machine Learning Techniques for Sales Forecasting,” *Int. J. Adv. Comput. Sci. Appl.*, 2021, doi: 10.14569/IJACSA.2021.0121112.

[5] S. KumarSharma and V. Sharma, "Comparative Analysis of Machine Learning Techniques in Sale Forecasting," *Int. J. Comput. Appl.*, 2012, doi: 10.5120/8429-2198.

[6] A. A. Mir, M. Alghassab, K. Ullah, Z. A. Khan, Y. Lu, and M. Imran, "A review of electricity demand forecasting in low and middle income countries: The demand determinants and horizons," *Sustainability (Switzerland)*. 2020. doi: 10.3390/SU12155931.

[7] K. NainSukhia, A. Ashraf Khan, and M. Bano, "Introducing Economic Order Quantity Model for Inventory Control in Web based Point of Sale Applications and Comparative Analysis of Techniques for Demand Forecasting in Inventory Management," *Int. J. Comput. Appl.*, 2014, doi: 10.5120/18856-7385.

[8] Z. H. Kilimci *et al.*, "An improved demand forecasting model using deep learning approach and proposed decision integration strategy for supply chain," *Complexity*, 2019, doi: 10.1155/2019/9067367.

[9] S. Motepe, A. N. Hasan, and R. Stopforth, "Improving Load Forecasting Process for a Power Distribution Network Using Hybrid AI and Deep Learning Algorithms," *IEEE Access*, 2019, doi: 10.1109/ACCESS.2019.2923796.

[10] J. Liu, L. Ding, X. Guan, J. Gui, and J. Xu, "Comparative analysis of forecasting for air cargo volume: Statistical techniques vs. machine learning," *J. Data, Inf. Manag.*, 2020, doi: 10.1007/s42488-020-00031-1.

CHAPTER 3

MACROECONOMIC INDICATORS AND THEIR IMPACT ON BUSINESS FORECASTS: AN ANALYSIS OF THE RELATIONSHIP BETWEEN ECONOMIC GROWTH AND CORPORATE PERFORMANCE

Dr S.Yogananthan, Adjunct Faculty
Department of Decision Sciences, CMS Business School, JAIN Deemed to-be University, Bangalore, India
Email Id- dr.s_yogananthan@cms.ac.in

ABSTRACT:

The Macro economy and Business Forecasts are important factors for decision-making in business, economics, and government policy. The macroeconomic encompasses the overall performance of the economy, including inflation, employment, GDP, and interest rates. Business forecasts, on the other hand, are predictions of future business conditions, such as sales, revenue, and profits. Forecasts are essential for businesses as they help them make informed decisions about investment, hiring, and production. Accurate macroeconomic and business forecasts can also aid government policymakers in creating policies that stimulate economic growth and stability.

KEYWORDS:

Business Forecasts, Decision-Making, Economics, Government Policy, Inflation, Macro Economy.

INTRODUCTION

The Congressional Budget Office often uses statistics from the Federal Revenue Service to assess how certain tax law changes will affect people at different levels of income distribution, such as whether a given tax reduction would mostly benefit the "rich" or the "poor." Data from consumer surveys show how much money is spent on various products and services at various income levels. For instance, economists could wish to look at the spending patterns of a certain set of customers to ascertain their income, savings, and consumption patterns (the proportions of money spent on food, rent, vehicles, vacations, etc.) in a given month, say June 1995. Similar surveys may be used to compare the kind of things that customers in, say, New York City and Denver, Colorado, buy. Individual businesses utilise cross-section analysis to learn more specifically about who purchases their items in supermarkets and department shops. Based on age, income, eating habits, cigarette and alcohol use, parental health history, and other characteristics, data on personal health gathered at a particular period may be used to identify the types of people who are most at risk for developing certain illnesses [1].

Cross-section data are often used in econometric research. For instance, researchers could be curious to know how various consumer groups responded to tax changes in the past. Cross-section statistics have been used by economists to identify the causes of a country's general growth rate, including governmental policies, the saving and investment ratio, population education, and many other variables. Financial advisers would be interested in calculating the likelihood that a municipal bond issuance will default based on the municipality's per capita income, age, sex, and racial characteristics, the projects the money will be used for, the current tax base and growth in the base, and other factors. The use of cross-section data to forecast different occurrences has many additional practical applications, some of which will be discussed later in this book.

Panel data are cross-sectional data that have been examined using the same sample at many points in time. For instance, the June 1995 statistics may have been flawed since, on average, people only purchase a new automobile every four years, making that month out of the ordinary. Consequently, questions regarding income, savings, and consumption might be asked of the same persons in January 1997 and at subsequent times. These people's spending habits might be monitored over a longer period of time to help identify how much is saved at various income levels, whether upper-income people spend a bigger percentage of their income on housing, transportation, or medical care, or a variety of other goods [2].

Another question that may be answered using panel data is whether people who began smoking cigarettes at an early age continued to smoke throughout their lives, while others who started smoking later found it easier to stop. This data might also be used to examine if a rise in cigarette excise taxes has a longer-term or shorter-term impact on smoking reduction. Even if they are estimating sales for a specific sector or firm, forecasters that utilise time-series data often use government statistics.

These projections will depend on the state of the US and global economies, unless they are fully influenced by technology. There are 70 agencies that provide US economic statistics, according to the major US government data search engine (see section 2.2). Nevertheless, the Bureau of Economic Analysis (BEA), the Bureau of the Census, the Bureau of Labor Statistics (BLS), and the Board of Governors of the Federal Reserve System are the primary data sources for the majority of economic forecasting requirements (Fed). The Federal Revenue Agency's Statistics of Income Division, the Department of Agriculture's Economic Research Service, and the Department of Energy's Energy Information Administration are a few other crucial government data sources. The discussion at this stage will be restricted to the first four agencies since this is a small book on predicting as opposed to the sources of government data.

The BEA, which is a division of the Commerce Department, produces the National Income and Product Accounts (NIPA). BEA calculates and publishes data on GDP in current dollars and adjusted for inflation, consumption, and personal and corporate income. Moreover, BEA provides thorough information on employment and personal income by precise industry categorization for each state, county, and city. The BEA gathers and analyses data that has been gathered by several other governmental organisations. The Bureau of the Census, which is a division of the Department of Commerce, gathers the majority of the series that are used as inputs for NIPA.

The decennial count of every person in the nation is undoubtedly the Census Bureau's most well-known accomplishment, yet it only makes up a tiny portion of its overall work. Census is the publisher of the majority of the monthly statistics on economic activity that the government issues. These reports contain information on wholesale and retail sales, inventories, housing starts and building activity, exports and imports, manufacturing shipments, orders, and inventories. The statistics in the census publications mentioned below are all accessible on a monthly basis, in contrast to the majority of NIPA numbers (apart from consumption and income), which are quarterly data [3]. Moreover, Census releases the Quarterly Report for Manufacturing Companies, which offers information on all significant balance sheet and income statement elements for all significant manufacturing sectors, organised by asset size. The BLS, which is a division of the Labor Department, releases data on wages, prices, employment, unemployment, productivity, and labour costs on a monthly basis.

DISCUSSION

The most immediate influence on financial markets comes from the BLS statistics. The most frequently followed economic indicators by financial markets are the producer pricing index and consumer price index (PPI and CPI), as well as the Employment and Earnings Report, which provides information on employment, unemployment, and pay rates. The BLS also gathers monthly information on employment and unemployment in states and metropolitan areas [4]. The Fed is the fourth important source of government statistics. The majority of its publications, as might be anticipated, deal with monetary issues, such as the money supply, bank assets and liabilities, interest rates, and foreign exchange rates. The Fed does, however, also publish data on industrial output and capacity utilisation for the whole economy as well as for each specific manufacturing sector.

The majority of the important figures that economists rely on are gathered and published in a monthly report by the Council of Economic Advisers named, fittingly enough, Economic Indicators. It costs $55.00 annually from the Government Printing Office and includes little over 500 series of economic data as of 2002. Economic Indicators does not include a lot of historical data since it is intended to provide the most current statistics. While monthly and quarterly statistics are only provided for recent years, this may be found in the yearly editions of the Economic Report of the President, a good source for annual government data.

The Commerce Department's Survey of Current Business includes complete NIPA tables and a few additional series, although it now has far less data since budget constraints removed thousands of items from its tables. You may get all the information in that publication by going to the BEA homepage at www.bea.doc.gov and following the instructions. In order to get data for GDP by industry, state personal income, and a variety of regional economic statistics, CD-ROMs may be ordered from the BEA order desk for a fee ranging from $20.00 to $35.00. Viewers of this website may also get specific NIPA accounts tables and information for certain US states.

For those who need access to government statistics right away, the Commerce Department offers STAT-USA, a service that delivers all important economic reports minutes after the relevant federal department releases them. It costs at least $175 a year to subscribe in order to get the info right now. For individuals who regularly monitor the data and rely on those figures to make financial market judgments, it is well worth the effort, but for those who are just developing a quarterly or yearly model, time is often not of the concern. Figure 1 illustrate the Leading Indicators in economics [5].

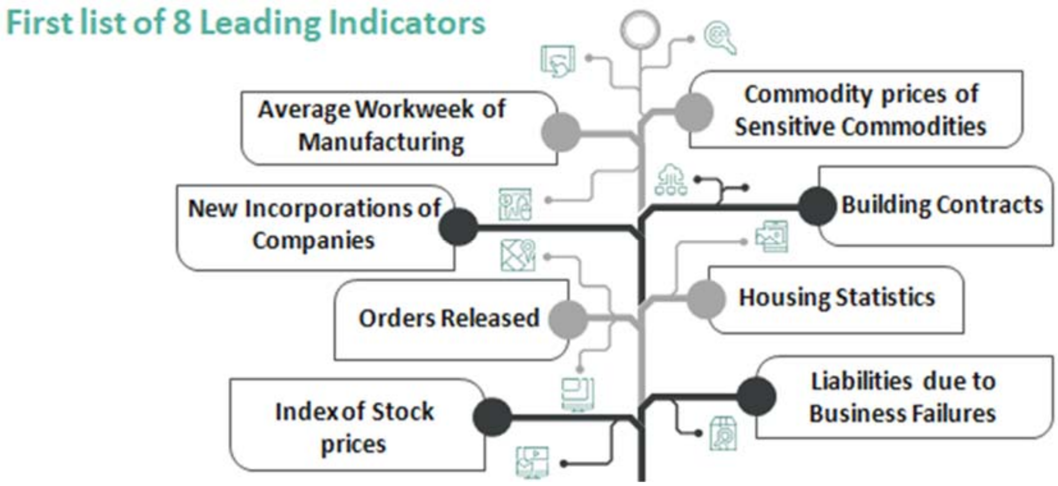


Figure 1: Illustrate the Leading Indicators in economics.

The Statistical Abstract of the United States, which is published yearly by Census, is the ideal place to start for "one-stop" shopping for people who are unfamiliar with the range of government statistics. The majority of the 1500 tables of data in it relate to the US economy, and it also includes a list of websites. We have only examined US government data up to this point, but a rising number of commercial apps use data from other countries. Most other industrialised nations have central sources of data comparable to the US, which is helpful for people who are creating particular models of other nations. The majority of the time, however, model builders will want general summary statistics for a wide range of nations. These statistics could be used, for instance, to identify which nations have the highest potential for growth, the best locations for new factories, or the highest risk of currency depreciation.

For broad information about the world, there are two sources. The first is the Organization for Economic Cooperation and Development (OECD), which has its headquarters in Paris but also maintains a presence in Washington, DC, where the majority of its statistical reports are kept. It releases a number of data volumes, mostly NIPA numbers and fundamental information on employment, output, and pricing. The majority of the data series are available exclusively for the 29 OECD nations and are published on a monthly or quarterly basis [6].

Around 170 nations are covered by the International Monetary Fund (IMF), which is based in Washington, DC, however practically all of the series are yearly. As might be anticipated, its series primarily focus on financial information, balance of payments data, and exchange rates, with few information on actual output, production, prices, and employment. Visit Statistics Canada's website at www.statcan.ca to find specific statistics for Canada. The European Central Bank's website, ecb.int, provides data for Europe. The Economist magazine, which can be viewed at www.economist.com, includes significant economic series for important nations in each weekly edition for people who wish to stay current on foreign statistics at a reasonably low cost. Certain statistics are accessible to all users, while the majority are only available to subscribers of the publication's print edition.

For individuals curious about demographic and socioeconomic statistics, the Census Bureau maintains extensive databases for 227 nations and regions. The University of Chicago's Center for Research on Security Pricing is the go-to source. Yet, individuals who simply want to evaluate a small number of organisations or need data for a little period of time are probably not in need of that enormous database. A free CD-ROM containing daily stock market data for up to 15 years will be sent by Worden Brothers, which may be reached. They expect customers would update the data at $1/day, but even if they don't, the CD-ROM will provide a tonne of historical information on certain companies.

Accessing the Web is your best option. One handy resource with more than 50 databases that provide specific business data is Hoover's Online. Public Register's Annual Report Service, which offers free annual reports for more than 3,600 businesses, is one of such databases. All reports that must be submitted by public corporations are available in the Securities and Exchange Commission's EDGAR file; most individuals wouldn't require them, but it may be a useful resource.

The Conference Board and University of Michigan are responsible for conducting the two important surveys. The Conference Board is open to having their data shared and offers a lot of it for no cost or a little price. On the other side, the University of Michigan worries that few people would subscribe if they release the findings of their poll. Yet, as all wire services publish their stories shortly after they are published, the information may be accessed secondhand from a variety of sources. The Conference Board, however, offers a considerably better client experience. This author has conducted empirical research and discovered just a little difference between the two series [7]. The National Association of Home Builders and the National Association of Realtors conduct the major surveys. These surveys include information on the number of houses sold, the average price, building attributes, and customer attitudes towards the possibility of making a purchase in the near future. In both instances, the aggregate statistics are provided for free, but the particular reports' data must be paid for.

Published by the National Association of Purchasing Managers is the most well-known survey. It is published monthly and is based on surveys regarding shipping, production, employment, delivery times, and particularly prices paid and received that were completed by roughly 250 buying managers. There are also a number of regional purchasing managers' indexes that are produced, most notably for Chicago and New York, but the national survey is far more commonly used since it is widely believed to be more accurate. The American Iron & Steel Institute, Association for Manufacturing Technology previously Association of Machine Tool Builders, American Petroleum Institute, Electronics Industry Association, Dataquest Gartner for computer shipments and revenues, and Semiconductor Industry Association are some of the major associations that will make their summary data accessible for free or at a low cost.

Many people who create models want to be able to access whole history series of quarterly or monthly data without having to manually enter it in. Basically, there are three options. The first step is to copy and paste each series from the Web; this part covers the main sources of data from the Internet. Second, each government agency has the option to order CDs or discs. Finally, by buying a complete database from a commercial vendor, you may hire someone else to execute the labor-intensive work for you. Haver Analytics assembles the datasets utilised in combination with EViews. Similar databases are offered by other commercial providers, although at somewhat higher costs. The information cited in this essay, unless otherwise noted, was either directly gathered by the author or may be found in the Haver Analytics database. With the exception of a few international interest and currency rates, the basic Haver database exclusively contains American data.

Both printed and CD-ROM versions of comprehensive foreign statistics are available for purchase from the OECD and IMF. A whole volume might be written on the part on gathering data from the Internet. The aim is not to cover all or even most of the online resources for economic data, however. Its purpose is to provide a thorough but yet manageable directory for locating the majority of the data that are likely to be helpful in creating econometric models. The obvious option is to go straight to that website for individuals who are aware of the data series they need and the public or private sector source for that data. Many detailed data websites on the Web are advised if you don't know who releases the data or are unsure about the series you require [8].

In order of increasing generality, below are several websites that merge many datasets. Would be guaranteed to be understood. Nevertheless, it simply implies that the likelihood of the next event happening would be known in advance. The sample mean and variance would, however, become closer to the population mean and variance if more tests were conducted. Even more crucially, the underlying probability distribution remains constant and each observation is independent. The likelihood of success on the subsequent spin at the roulette wheel is independent of how many times in a row you have won or lost, providing the wheel is not "fixed."

When the underlying probability distribution is unknown, the other kind of forecasting model is used, which is more relevant to business forecasting. For instance, we believe that as income increases, consumers spend more, and when real interest rates fall, firms invest more. They are obviously fair theories and are buttressed by economic theory. However, consider all the factors we don’t know: how much consumption will change when income changes, the time lag, other factors that affect income, the fact that the observations are not independent most people are creatures of habit, and the fact that we don’t know what income will be in the future. Even more significant, the link between consumption and income may fluctuate for the same people depending on the economic climate. They may be more optimistic or more pessimistic; they may have recently moved into a larger home and need more furniture, their children may be approaching college age, and a host of other factors.

Over the past century, a large amount of statistical literature has been devoted to the issue of the “best” methods of estimating empirical relationships. The majority of these articles are related to the method of least squares. However, almost all of the tests and relationships are based on assumptions that do not exist in the typical practical business forecasting environment. The major problems can be briefly summarized as follows:

a) The data are not normally distributed.
b) The residuals are not all independent the forecasting error in this period is often closely connected with the error last period.

c) The independent variables are supposed to be known at the time of forecast, which is generally not the case.
d) The data are sometimes inaccurate and subject to substantial revision.
e) Finally, and most important, the underlying data generation function may have shifted during the sample period, or even more damaging during the forecast period.

In spite of all these drawbacks, the vast majority of economic forecasting models are estimated using least squares, and the examples given in this book will follow this approach. However, emphasis will be placed on adjusting for the fact that the classical least squares criteria often do not occur. For this a simple example can be used to illustrate this point of why the sample variance must be adjusted by the degrees of freedom. One can always connect two points with a straight line. The mean value is the average of these two points.

The variance is supposed to be the dispersion around the line connecting these two points, but there isn't any variance: the line connects the two points exactly, leaving no residual. Similarly, a plane can always be drawn through three points, aand so on. The argument is the same as we move into n dimensions. The more variables that are contained in the equation, the more likely it is that the n- dimensional line will connect all the points, even if the relationship doesn't explain anything. Thus an unbiased estimate of the true variance must be calculated by adjusting for the degrees of freedom [9].

The square root of the sample period variance is known as the standard deviation, which is the more common measure used in statistical parlance. The comparison of the estimated mean to its standard deviation indicates whether that mean is statistically significantly different from some reassigned value, usually zero. The mean and variance are the two sample statistics most often used to describe the characteristics of the underlying probability distributions. They are anot the only ones. Statistics books generally refer to the methods of "moments," which show that the mean and variance are only the first and second moments aof a long list of characteristics that describe various probability distributions [10], [11].

CONCLUSION

Macro economy and Business Forecasts are critical tools for decision-making in various industries, including business, economics, and government policy. While accurate forecasts can help organizations make informed decisions about investment, hiring, and production, developing forecasts can be challenging due to the complexity and unpredictability of the macroeconomic and business conditions. Nevertheless, careful analysis of macroeconomic and business forecasts can help reduce risk and improve the chances of success. As such, it is crucial for businesses and governments to continually monitor and update their forecasts to make timely and effective decisions that align with their goals and objectives.

REFERENCES

[1] S. J. Cho, C. Y. Chung, and J. Young, "Study on the relationship between CSR and financial performance," *Sustain.*, 2019, doi: 10.3390/su11020343.

[2] M. Yunis, A. Tarhini, and A. Kassar, “The role of ICT and innovation in enhancing organizational performance: The catalysing effect of corporate entrepreneurship,” *J. Bus. Res.*, 2018, doi: 10.1016/j.jbusres.2017.12.030.

[3] I. Dzingai and M. B. Fakoya, “Effect of corporate governance structure on the financial performance of johannesburg stock exchange (JSE)-listed mining firms,” *Sustain.*, 2017, doi: 10.3390/su9060867.

[4] E. López-Zapata, F. E. García-Muiña, and S. M. García-Moreno, “Analysing the relationship between diversification strategy and firm performance: The role of the economic cyclew,” *Cuad. Gest.*, 2019, doi: 10.5295/cdg.170738el.

[5] C. E. Hou, W. M. Lu, and S. W. Hung, “Does CSR matter? Influence of corporate social responsibility on corporate performance in the creative industry,” *Ann. Oper. Res.*, 2019, doi: 10.1007/s10479-017-2626-9.

[6] M. Inekwe, F. Hashim, and S. B. Yahya, “CSR in developing countries – the importance of good governance and economic growth: evidence from Africa,” *Soc. Responsib. J.*, 2021, doi: 10.1108/SRJ-10-2019-0336.

[7] M. Romano, A. Cirillo, C. Favino, and A. Netti, “ESG (Environmental, social and governance) performance and board gender diversity: The moderating role of CEO duality,” *Sustain.*, 2020, doi: 10.3390/su12219298.

[8] Q. T. Tran, “Economic policy uncertainty and corporate risk-taking: International evidence,” *J. Multinatl. Financ. Manag.*, 2019, doi: 10.1016/j.mulfin.2019.100605.

[9] T. T. H. Phan, H. X. Tran, T. T. Le, N. Nguyen, S. Pervan, and M. D. Tran, “The relationship between sustainable development practices and financial performance: A case study of textile firms in Vietnam,” *Sustain.*, 2020, doi: 10.3390/SU12155930.

[10] J. P. Tinambunan, T. Irawan, and T. Andati, “Implementation of Good Corporate Governance on Financial Performance Coal Companies Moderated by Capital Structure,” *Int. J. Res. Rev.*, 2021, doi: 10.52403/ijrr.20210613.

[11] K. Heenetigala and A. F. Armstrong, “The Impact of Corporate Governance on Firm Performance in an Unstable Economic and Political Environment: Evidence from Sri Lanka,” *SSRN Electron. J.*, 2012, doi: 10.2139/ssrn.1971927.

CHAPTER 4

EXPLORING EFFECTIVE STRATEGIES FOR DATA COLLECTION AND ANALYSIS IN FORECASTING: A COMPARATIVE STUDY OF QUANTITATIVE AND QUALITATIVE APPROACHES

Dr Ravishankar S Ulle, Assistant Professor
Department of Decision Sciences, CMS Business School, JAIN Deemed to-be University, Bangalore, India
Email Id- dr.ravishankarulle@cms.ac.in

ABSTRACT:

Data collection and analysis are crucial components of forecasting. Accurate forecasts rely on the availability of high-quality data and the ability to analyze that data effectively. Data collection involves gathering relevant information from various sources and transforming it into usable data for analysis. Data analysis involves using statistical and mathematical methods to identify patterns, trends, and relationships in the data. This can include techniques such as regression analysis, time-series analysis, and machine learning algorithms.

KEYWORDS:

Data Collection, Data Analysis, Forecasting, Mathematical Methods, Regression Analysis, Statistical Analysis, Time-Series Analysis.

INTRODUCTION

Sometimes it is useful to find out how much distributions deviate from the normal distribution by looking at the third and fourth moments, known as scenes and kurtosis. For example, a distribution might be "lopsided" with the peak value far away from the middle, which is skewers. The tails might be too "fat," which is kurtosis. Also, the distribution could have more than one peak. However, for practical purposes in most practical statistical work including but not limited to economics the mean and variance are the only tools that are used to describe the shape of the probability distribution. That is because the normal distribution, which is the most important distribution for statistical work, is completely defined by its mean and variance one of the major aims of this book is to explain how to build a forecasting model that will minimize forecast error.

As will be seen in numerous examples, independent variables that appear to be highly correlated with the dependent variable in the sample period often show a much smaller correlation in the forecast period. Nonetheless, in a brief statistical review it is useful to indicate the tests used to determine which variables are statistically significant, aand how well the equation fits, over the sample period. We want to determine if the parameter estimates the coefficients in the model are significantly different from zero, and also what proportion of the total variance of the dependent variable is explained by the regression equation. The statistical significance of each coefficient is determined by dividing the value of each coefficient by its standard error. If the residuals are normally distributed, the parameter estimates will generally be statistically significant from zero at the a95% probability level if this ratio is 2 or greater, and at the 99% level if this ratio is 2.7 or greater [1].

The proportion of the variance of the dependent variable explained by the equation is known as R-squared. It is sometimes thought that the higher the R-squared, the more accurate the forecasts will be; but as will be shown throughout this book, that is often not the case. Nonetheless, virtually every model builder looks at the values of R-squared in determining which equation to choose, and to a certain extent I will follow that general practice. After determining the correlation coefficient, the model builder wants to know whether this correlation is significantly different from zero at some designated level, usually the 95% probability level. One could easily test whether the parameter estimate is significantly different from some other value, but most of the time researchers want to determine whether the coefficient is significantly different from zero.

Estimating the regression equation yields an estimate of the intercept and the slope coefficient b; the least squares algorithm also supplies estimates of the variances of the estimated values of ab. The significance level is determined by taking the ratio of the coefficient to its standard error, which is the square root of the variance. In everyday terms, this means the standard error serves as a measure of the dispersion of the probability distribution of that coefficient around its mean value. If the standard error is small relative to the coefficient, the probability is high that the actual value will be close to the estimate; if the standard error is large relative to the coefficient, the actual value could be just about anything, and the estimated value is not very useful. If the error term is normally distributed, then we can determine whether the coefficient is significantly different from some desired test value, generally zero [2].

Some actual numerical examples are provided later. For now, consider the case where the coefficient is 0.70 and the standard error is 0.30. Also assume that the error term is normally distributed. We have already noted that one rule of thumb almost taken for granted in most of the empirical articles in economics states that if this ratio is greater than 2, the variable is significantly different from zero; or, in short, significant. For the practicing econometrician, that is the rule used to show that your results are meaningful. Perhaps a better level of significance could be found, but this result is so ingrained in statistics that we will continue to use it. The graphs shown in this text are produced by the EViews software program, which is used throughout this book. Just as there are hundreds if not thousands aof sources of data, there are many different software programmers written for the Pc that can be used to estimate regressions and build models. However, for our purposes, the list can quickly be narrowed down to a few names.

DISCUSSION

The programmed should be primarily designed for economic model building, which means including an efficient simulation capability as well as estimating regression equations. It should be simple to generate transformations of the variables, including lags, percentage changes, and ratios, and it should also be easy to add dummy variables and estimate nonlinear equations. The program should also contain a full battery of standard tests to determine whether the various parameters are statistically significant. It should also permit easy data aentry and exit and be compatible with existing large-scale economic databases[3].

Other programmers satisfying all these criteria include SAS, SPSS, PCGIVE, aand RATS. Minitab and Excel are widely used for spreadsheet forecasting but are not so useful for building models. In this author's experience, the modelling capabilities of Views are easier to use than those found in competing programs. The examples and printouts in this text are based on Views; other programs generally have similar formats and provide essentially the same information and the following text identifies some of the standard terms that appear along with each regression equation to show the reader what is expected. For most of the equations in this book, an abbreviated form is used to convey the most important statistical information.

The sample period is given along with the number of observations, in case any years were skipped because of missing data. In this case, there are 50 years from 1949 through 1998 inclusive, so no data are missing. From time to time it might be advisable to omit one or more observations if it appeared to be far out of line with the rest of the information. Alternatively, data might be missing for one or more observations. R-squared is the percentage of the variance of the dependent variable explained by the equation. Adjusted R-squared (often called RSQ in this text) is R adjusted for degrees of freedom; in this case it is 0.89. The standard error of the regression equation is 0.0076, or 0.76%. That means that approximately two times out of three, the sample period error for predicting the wage rate was less than 0.76%, compared with an average change of 5.6% (noted below) (noted below). The sum of squares residual is the standard error squared multiplied by the degrees of freedom; it does not add very much information [4].

It is a test for the autocorrelation of the residuals. If no autocorrelation is present, the DW statistic is s2. If this statistic is less than about 1.4, the residuals are serially correlated. When that happens, the t-ratios and overstated, so the equation will usually not predict as well as indicated by the sample period statistics. The DW of 2.07 indicates there is no autocorrelation in this equation. The mean dependent variable is 0.056, which means the average annual change ain wage rates over the sample period was 5.6%. The SD dependent variable line show the standard error of the variable around its mean, which is 2.3%.

Since several terms are significant, the overall equation must be significant in aany case; for this reason, the aF-ratio is not used very often. While it is possible ato estimate an equation where none of the t-ratios is greater than 2 but the overall F-statistic was significant, that would mean cluttering the equation with individual terms that are not significant, which would ordinarily generate a very poor forecasting equation. A brief note on the number of significant digits. The actual programme for EViews generally shows six or seven numbers. I have reduced this clutter to show two or three significant figures, which makes more sense economically. Figure 1 illustrate the Data analysis process diagram.

A typical graph, showing the actual values, those estimated by the equation, and the residuals, is the top half of this figure shows the actual values of changes in wage rates compared with the estimated values calculated by the regression equation; these are also called simulated or fitted values. The bottom half shows the residuals, defined as the actual minus the fitted values. The largest error occurs in 1989, when wages rose far less than would be predicted by the equation; almost as large a discrepancy occurred in 1992 in the other direction. The construction of econometric models is often based on economic theory [5].

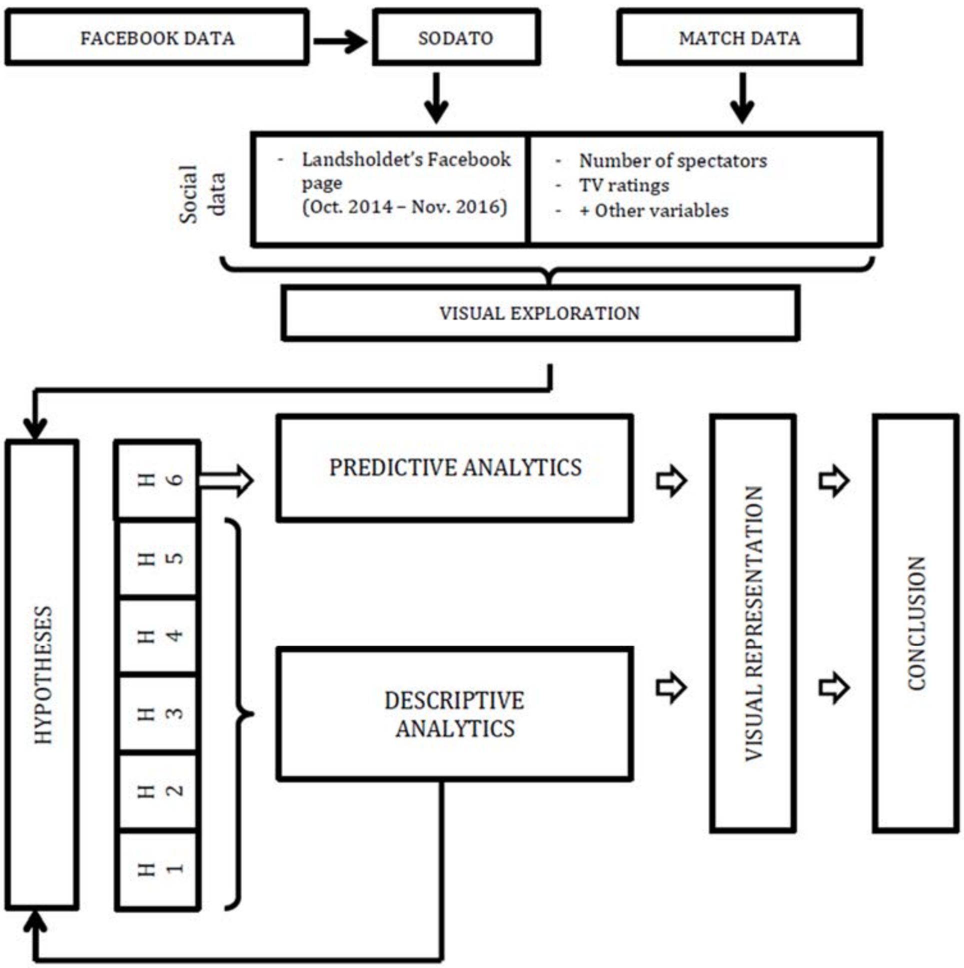


Figure 1: Illustrate the Data analysis process diagram.

However, in virtually all cases, the researcher looks at the underlying data in order to form some opinion of how the variables are correlated, and whether the correlation is improved when the independent variables are lagged. There are three principal methods of displaying time-series data. Line graphs usually show two or more series graphed against time. Scatter diagrams have all the sample period points for one variable on the y-axis and the other variable on the x-axis. Bar graphs are often utilized to describe the characteristics of a single series; the most common use in this text is histograms, where either the original series or the residuals from a regression equation can be checked for normality and other statistical properties. Bar graphs can be used for multiple variables, either on a side-to-side basis or stacked. Sometimes pie charts are used as graphical aids, but these are usually for a snapshot of events at some given time and are not ordinarily utilized with time-series data.

The well-known comment about lies, damn lies, and statistics, variously attributed to Benjamin Disraeli and Mark Twain among others, summarizes show many people view graphical analysis. The same data can tell completely different stories depending on how they are presented. To see this, consider the simple relationship between consumption and disposable income, both in constant dollars. Surveys are the primary source of government data. Respondents occasionally fail to return their forms. So what should be done? Interpolating the data based on the businesses that did return their forms is a logical solution.

Businesses that didn't return their forms had unusual circumstances that would have significantly changed the data. Although the initial data are seriously flawed, the issue is eventually resolved once more complete numbers are available. The approach is modified occasionally. In October 1999, the Bureau of Economic Analysis (BEA), the organisation that compiles GDP and related figures, decided to include software purchased by businesses as part of investment; previously, it had been treated as an intermediate good and excluded from GDP. This resulted in a thorough data revision that increased the average growth rate of the previous decade by an average of 0.4%. This change was appropriate and timely because software had begun to play a bigger role in the economy as a whole [6].

Another significant example is how, in the middle of the 1990s, the method for calculating the rate of inflation was modified. Therefore, the same modifications to each of the CPI's individual components would lead to a 0.7% decrease in the overall inflation rate. These adjustments reflected the higher quality of many consumer durables, a shift in consumer behaviour from department stores to discount malls, and changes in market baskets that contained a higher proportion of less expensive goods. The majority of economists concurred that these adjustments were necessary and, in many cases, long overdue. The rate of inflation had been overstated by an average of 1.1% annually, according to a commission led by Michael Boskin, a former chairman of the Council of Economic Advisers.

Figure 2: Illustrate the data collection methods.

When new information and methodologies are made available, the statisticians of the federal government cannot be blamed for including them in their data releases. It would be a grave mistake to omit these changes, in fact. However, the emergence of preliminary data that are subsequently significantly revised raises important issues in the development and assessment of forecasting models. It has occasionally significantly affected policy choices, at least in the past. For instance, the recession of 1990–1991 was one of the key instances of preliminary data being misinterpreted.

When that recession began, BEA initially claimed it was relatively mild, with a drop in real GDP of only 2%. Figure 2 illustrate the data collection methods.

Based on the data that were initially released, the Fed assumed that the recession was not very severe and cautiously eased. It probably would have eased much more quickly if it had been aware of the true extent of the real GDP decline. In fact, after the recovery did not progress as expected in 1991, the Fed did lower short-term interest rates by the end of 1992 to unusually low levels, and the economy did eventually recover. The Fed tightened so much in 1994 as a result of the increased inflationary pressures, though, that real growth fell to 1% in the first half of 1995. The economic effects of those false data didn't entirely go away until the second half of that year.

The most accurate forecast would have indicated that the economy is in worse shape than the government reports suggest, that the Fed will initially fail to loosen policy enough and that the economy will take longer than expected to recover. As a result, the Fed will eventually need to loosen policy more than usual, which will result in interest rates being much lower than anyone else anticipates two years from now. At that point, inflationary expectations will rise, and the Fed will be forced to tighten policy once more. Naturally, no one did, and no one could have been expected to have predicted such a series of events [7].

But economists came under heavy fire for failing to accurately predict the decline in interest rates, overestimating the size of the initial rebound, and misjudging the severity of the recession. Following that recession, no forecaster received praise, but it is reasonable to assume that errors would have been less with more precise data. The Nixon Administration's implementation of wage and price controls came to an end in May 1974. As a result, that month saw a record increase in the producer price index (PPI). The Bureau of Labor Statistics' (BLS') seasonal adjustment programme assumed for the following few years that the PPI always increased sharply in May, causing the May PPI data to significantly deviate from the unseasonally adjusted data while remaining essentially unchanged. The obvious course of action in this situation may have been to ignore those data, but it is unclear what approach the forecaster should take. Using no May data to run regression equations? utilising data that hasn't been adjusted for season? Treating May 1974 as a dummy variable, say by treating it as 1 during that time and 0 elsewhere? All of these are feasible, but none is recommended. Of course, the federal government is not the only one that updates its data. For many different reasons, businesses frequently restate their earnings. When they ship the goods, they record sales, but if they aren't bought, write-offs must be made. Large one-quarter writeoffs are occasionally caused by reorganisations or the sale of divisions.

Other times, accounting mistakes are to blame. The majority, if not all, attempts to forecast stock prices based on reported company earnings are hampered by the changes and inconsistencies in these data, despite analysts' best efforts to take these anomalies into account. There will never be a perfect answer to the data revisions problem. Furthermore, it makes no sense to criticise government statisticians for producing the most precise projections based on imperfect information and the dynamic nature of the economy. However, at this point it would be appropriate to make a few observations about data revisions [8]. The data available at the time the forecasts were released, rather than the most recent data revisions, should be considered when evaluating forecast accuracy. This implies, for instance, that depending on which set of "actual" data is used, an effort to assess the forecasting accuracy of macroeconomic forecasts made many years ago will produce very different results.

Some forecast error, but by no means all of it, is caused by the presumptions of changes in fiscal and monetary policy that are based on the preliminary data released by the government. These assumptions sometimes seem to have been incorrect in later revisions of the data. It is helpful and appropriate to use dummy or truncated variables in the regression equation when estimating a structural model over a long period of time. A methodological change in the CPI that started in 1994, for instance, can be explicitly entered as a new variable; any such change would have had a value of 0 prior to 1994.

This is frequently cited as one of the main explanations for why macroeconomic modelling in the 1970s and 1980s failed. It has been argued that making economic predictions based on historical data is impossible because people change their behaviour in response to past experiences, which causes them to respond differently to similar phenomena in the future. Although it was developed by Oskar Morgenstern in 1928, this idea is not particularly new. It is commonly referred to as the Lucas Critique. Formally, we can say that this idea has changed during the sample period or between the sample and forecast periods by stating that the model's underlying data generation process has changed. I bring up the origins of this idea to emphasise that it has a much longer history than the notion that the short-term trade-off between inflation and unemployment was primarily the result of ill-managed monetary policy in the 1950s and 1960s [9]–[11].

CONCLUSION

Data collection and analysis are essential components of forecasting. Collecting relevant and accurate data from various sources and analyzing it using statistical and mathematical techniques are crucial in producing reliable forecasts. Careful planning and attention to detail are required to ensure that the data collected is of high quality and suitable for analysis. Effective data analysis can help identify patterns, trends, and relationships in the data, allowing for informed decision-making. Ultimately, the accuracy and usefulness of forecasting depend on the quality of data collection and analysis.

REFERENCES

[1] A. Abadie, A. Diamond, and J. Hainmueller, "Comparative Politics and the Synthetic Control Method," *Am. J. Pol. Sci.*, 2015, doi: 10.1111/ajps.12116.

[2] A. Pudianti, A. Suprapti, and J. A. Syahbana, "The Tools Of Qualitative Approach To Measure Rural Transformation: The Case Of Yogyakarta Rural Village," *Tataloka*, 2019, doi: 10.14710/tataloka.21.1.11-22.

[3] R. R. Alford, G. King, R. O. Keohane, and S. Verba, "Designing Social Inquiry: Scientific Inference in Qualitative Research.," *Contemp. Sociol.*, 1995, doi: 10.2307/2076556.

[4] F. P. J. Marques, C. Mont'Alverne, and I. B. Mitozo, "A empresa jornalística como ator político: Um estudo quanti-qualitativo sobre o impeachment de Dilma Rousseff nos editoriais de Folha e Estadão," *Obs.*, 2018, doi: 10.15847/obsobs12320181166.

[5] P. Flores, M. Alvarez, and J. Torres, "Identifying Difficulties of Software Modeling Through Class Diagrams: A Long-Term Comparative Analysis," *IEEE Access*, 2022, doi: 10.1109/ACCESS.2022.3157290.

[6] S. Fainshmidt, M. A. Witt, R. V. Aguilera, and A. Verbeke, "The contributions of qualitative comparative analysis (QCA) to international business research," *Journal of International Business Studies*. 2020. doi: 10.1057/s41267-020-00313-1.

[7] L. Muñoz-Pascual, C. Curado, and J. Galende, "The triple bottom line on sustainable product innovation performance in SMEs: A mixed methods approach," *Sustain.*, 2019, doi: 10.3390/su11061689.

[8] M. Lederle, J. Tempes, and E. M. Bitzer, "Application of Andersen's behavioural model of health services use: A scoping review with a focus on qualitative health services research," *BMJ Open*. 2021. doi: 10.1136/bmjopen-2020-045018.

[9] F. Nwaiwu, M. Duduci, F. Chromjakova, and C. A. F. Otekhile, "Industry 4.0 concepts within the czech sme manufacturing sector: An empirical assessment of critical success factors," *Bus. Theory Pract.*, 2020, doi: 10.3846/btp.2020.10712.

[10] M. Dixon-Woods, S. Agarwal, D. Jones, B. Young, and A. Sutton, "Synthesising qualitative and quantitative evidence: A review of possible methods," *Journal of Health Services Research and Policy*. 2005. doi: 10.1258/1355819052801804.

[11] A. Karunia, H. Hetika, and R. Harjanti, "Kajian Strategi Pengembangan Usaha Mikro, Kecil Dan Menengah (UMKM)," *J. REP (Riset Ekon. Pembangunan)*, 2020, doi: 10.31002/rep.v5i1.2301.

CHAPTER 5

COMPARATIVE ANALYSIS OF SMOOTHING TECHNIQUES IN FORECASTING: A STUDY OF MOVING AVERAGE, EXPONENTIAL SMOOTHING, AND HOLT-WINTERS METHODS

Dr Vinoth Kumar.V, Assistant Professor
Department of Decision Sciences, CMS Business School, JAIN Deemed to-be University, Bangalore, India
Email Id- dr.vinothkumar_v@cms.ac.in

ABSTRACT:

Forecasting with smoothing techniques involves using mathematical models to analyze time series data and make predictions about future trends. This method relies on the assumption that future values will be similar to past values, with some degree of variation. One popular technique is exponential smoothing, which involves calculating a weighted average of past observations, with greater weight given to more recent data. Another approach is moving average smoothing, which involves taking the average of a fixed number of past observations.

KEYWORDS:

Exponential Smoothing, Forecasting, Moving Average Smoothing, Smoothing Techniques, Time Series Data, Weighted Average.

INTRODUCTION

The Lucas Critique is, despite being a very well-known example, merely a special case of evolving expectations. Based on what has happened in the past, economic agents frequently alter their behaviour patterns. That is not just accurate on a large scale. As competitors enter and leave the industry, sales growth for individual companies will be significantly impacted. Depending on how their competitors respond, businesses will either raise or lower their prices.

Depending on recent changes to the bankruptcy laws, borrowers may have a higher or lower rate of loan default they learned from their past behavior and changed it accordingly in the future. However, there are still a lot of economic connections that hold over a wide range of diverse experiences. Ceteris paribus, consumers will spend more if their income increases; however, they will consume more if they believe the change to be permanent as opposed to temporary. Capital spending will decline as interest rates rise. The amount of net exports will decrease as the value of the currency rises. The stock price will increase if a company's profit growth rate picks up speed. Similar cases where structural relationships hold up in the face of shifting expectations abound [1].

An adjustment in one sector of the economy may occasionally lead to adjustments in other ones that are consistent with prior performance. In the second half of the 1990s, the US economy provided a significant illustration of this. The price/earnings ratio of the stock market doubled despite the fact that bond yields were roughly the same in 1995 and 2000 due to a significant shift in expectations regarding future profit growth. Few forecasters were successful in foreseeing that change. On the other hand, consumer and capital spending were impacted in a way that was consistent with prior historical experience as a result of the increase in stock prices and the decline in the cost of equity capital. Additionally, the faster rate of capital stock growth brought on by an increase in the capital spending to GDP ratio boosted productivity growth, lowering the rate of inflation and further lowering interest rates. That in turn increased real growth to such a degree that the federal budget position changed from being in deficit to being in surplus, which in turn increased equity prices. It was difficult to foresee the shift in the stock market, but given that shift, it was simpler to foresee stronger economic growth. However, when the stock market fell, the opposite happened: less capital spending, slower productivity, and a return to deficit financing. In the mid-1970s, Fed Chairman Arthur Burns used monetary policy to counteract the recessionary effects of higher oil prices, leading to unusually low real interest rates. In contrast, in the early 1980s, Chairman Paul Volcker refused to accommodate the further increase in oil prices, leading to unusually high real interest rates.

No model based on data estimated through 1979 could have foreseen the dramatic rise in real interest rates that began in late 1980. In retrospect, it is obvious that one could choose a set of economic variables to track this pattern, but that is not the point. The Blue Chip consensus estimate for the six-month commercial paper rate for 1981 in July 1980 was 8.7%; the actual value was 14.8% [2]. This is one of the most obvious policy changes to have ever affected the US economy. What lessons can be gleaned from this experience by forecasters? Short-term interest rate fluctuations are largely, if not entirely, influenced by the Federal Open Market Committee's decisions. Because of this, predicting short-term interest rates today consists mainly of educated guesses about the next move the Fed will make. But eventually we discover another lesson. If the Fed maintains short-term rates below equilibrium for an extended period of time, inflation and interest rates will eventually increase; however, if it maintains short-term rates above equilibrium, inflation and interest rates will eventually decline.

In this instance, a model that captured this underlying relationship would be helpful in the long run but would be very ineffective in the short run when it came to forecasting interest rates. A prediction that inflation and interest rates would begin to fall in 1982 and initiate the biggest bull market in history would have been especially useful. The few forecasters who did correctly foresee that development, however, were little-trusted. The short-term effects of unexpected policy changes or exogenous changes cannot be predicted, not even by the best econometric models. However, after these modifications have taken place, properly constructed models ought to be able to provide helpful insights into what will transpire over the long term. The decision to lower real per capita government spending during President Clinton's term was made by senior administration officials, though it took place over a number of years [3].

DISCUSSION

The budget deficit was subsequently changed to a budget surplus, which—as was already mentioned above—was one of the factors contributing to a nearly unprecedented rise in the price/earnings ratio of the stock market. Government officials carry out changes of this nature. Others, like energy shocks, wars, and natural disasters, have nothing to do with policy but still have an impact on the economy. They will not be included in any forecasts unless anticipated.

However, if they were anticipated, drastic measures would be taken to prevent or reverse these developments.

The majority of consumers and businesses did not significantly change their behaviour patterns after the first energy shock because they saw it as a once-in-a-lifetime experience. As a result, oil imports kept rising, which eventually led to another increase in oil prices. Attitudes significantly altered following the second energy shock. With predictions that oil prices would reach $100 per barrel by the end of the twentieth century, the majority of people now anticipated that significant price increases would continue on a regular basis. Figure 1 illustrate the Forecasting with Smoothing Techniques [4].

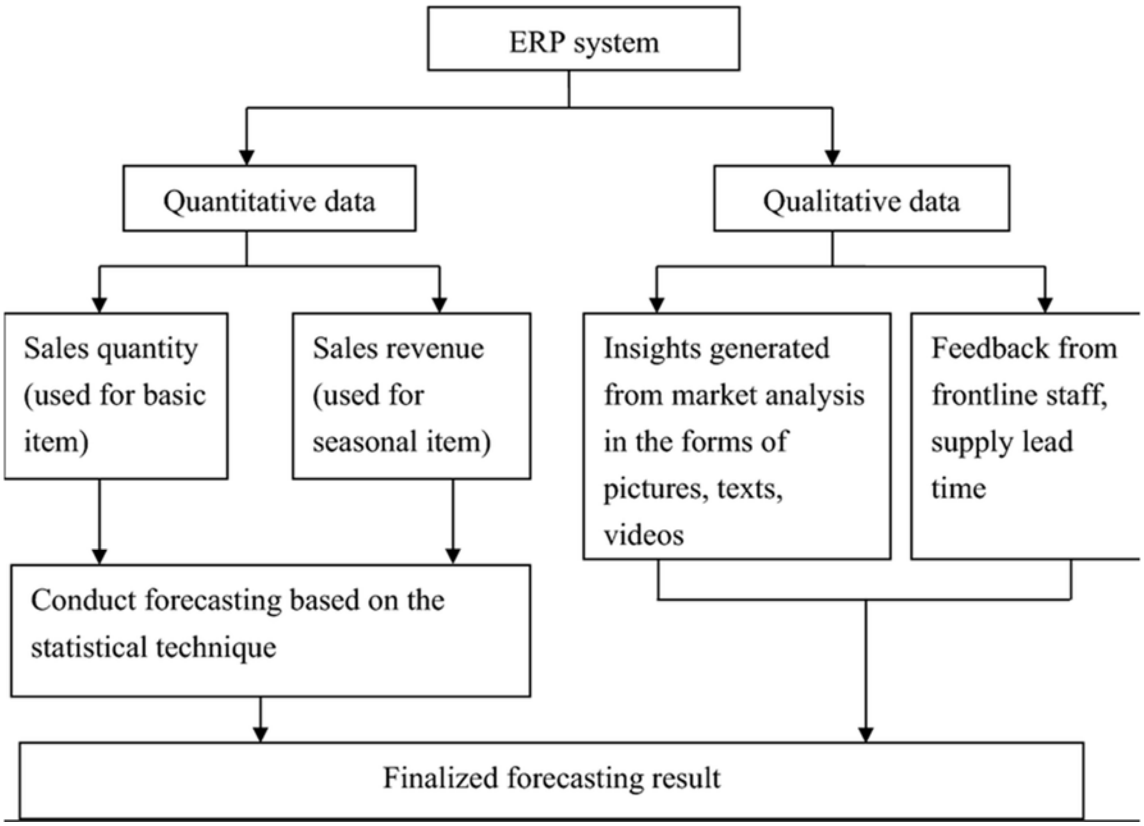


Figure 1: Illustrate the Forecasting with Smoothing Techniques.

As a result, both consumers and businesses began using less energy, increasing their purchases of fuel-efficient cars, and increasing their construction of fuel-efficient buildings. These initiatives were effective enough to cut down on oil imports, which caused a more than 50% decrease in energy prices in 1986. They were lower in real terms in 1998 than they were in 1972, just before the first energy shock, at that time [5].

The assumption regarding energy prices had an impact on all economic predictions made in the 1980s, whether they were accurate or not. Although conditional vs. unconditional, point vs. interval, and alternative scenarios weighted by probabilities are some of the alternative types of forecasts that were discussed, this example shows the value of these types of forecasts. Producing alternate forecasts based on various oil price scenarios higher, stable, or lower would have been an appropriate course of action for many businesses. Businesses would have been more prepared for the sudden drop in crude oil prices by more than half in 1986 if more weight had been given to the lower-price scenario when prices began to gradually decline in the mid-1980s as the global energy glut grew.

There is little to be gained by highlighting how forecasts are flawed when they are unable to account for unexpected exogenous shocks, many of which would not have happened if they had been correctly predicted. When these shocks do occur, models that accurately predict their effects can still be very helpful in predicting what will come next. In some instances, models intended for forecasting produce much larger errors than would be suggested by the sample period statistics because some of the independent variables are assumed to be exogenous when they are not. A variable is technically considered exogenous if its value is not determined by an economic model but affects how the endogenous variables are determined. There are different levels of exogeneity, though, in real life. Only a small number of factors like the state of the weather and military spending are exogenous always. The majority of the time, policy variables also include some endogenous elements.

After the collapse of the Thai baht, Korean won, Indonesian rupiah, and Malaysian ringgit in the latter half of 1997, US net exports declined dramatically in 1998 and the first half of 1999. As a result, manufacturing production rose much more slowly than total GDP; whereas during boom years, production usually rises faster than overall GDP. North Carolina has the highest proportion of workers in manufacturing, so its growth rate fell sharply after the collapse of those currencies.

A model that linked growth in North Carolina employment to the value of the dollar (among other variables) would show a high correlation. However, a forecast made in 1997 would have been inaccurate if it had assumed the values of those currencies would remain stable. In such a case, the model would appear to work well, but forecasts of the North Carolina economy would be far off the mark. In this case, the equations might have continued to work well in the sense of high correlations and low standard errors, but the forecasts would have been poor because of the inability to predict the exogenous variables [6]. In the past, monetary policy used to be treated as exogenous, although this error is made far less often today. Even in the days before Paul Volcker, the Fed routinely tightened monetary policy as inflation increased. Thus assuming that monetary policy variables were exogenous and would not change invariably led to forecast errors that were much larger than expected from the sample period statistics.

Analyses of macroeconomic models undertaken many years ago by this author showed that the single biggest source of error in multi-period forecasting was caused by using the lagged dependent variable on the right-hand side of the equation. If current consumption were estimated as a function of lagged consumption, for example, an error made one quarter could distort all the forecasts from that point forward. I will discuss a variety of methods to overcome that difficulty; now that this error has been well documented, it does not occur when most people think of forecasts, they think of point estimates. For example, sales will rise 12% next year, the Dow will climb to 12,000 a year from now, the Federal Open Market Committee will vote to boost the Federal funds rate 25 basis points at its next meeting, the price of oil will climb 20% over the next six months, and so on.

While it is true that point estimates are the most common type of forecasts, there are many other ways in which forecast results can be presented. Sometimes a range for the predicted variable is more appropriate; other times, different probabilities are assigned to alternative scenarios. Sometimes the penalties associated with being too high or too low are equal; at other times, the loss function is asymmetric. In this section, I discuss some of the more common types of alternative forecasts.

Suppose a company has a limited amount of excess manufacturing capacity. If sales grow less than 5% per year, the company will be better off using its existing facilities. If sales grow more than 5% per year, it will be better off building a new plant. In this case, the point estimate for sales growth is not as important as the probability that sales growth will exceed 5%. A similar case might be made for advertising budgets. If a firm thinks a a$1 million expenditure on advertising will boost sales by at least $5 million, it awill decide to go ahead and spend the money. It doesn't matter so much whether the increase in sales is $6 or $10 million, but if it is $4 million, the expenditure awill not be made [7].

A company may have a loan covenant with the bank stating that if cash reserves drop below a certain level, the loan will be called. That level might be correlated with the assumption of increased profitability, so a decline in profits would trigger the loan call. In that case, the key forecast is whether company profits have risen or not, rather than the precise amount they would increase. Forecasts can be either absolute or conditional.

Some examples of absolute, or unconditional forecasts are: real GDP will grow 4% next year, the Republicans will retain (or regain) majority control of Congress, and company sales will rise aat least 15% per year over the next decade. However, many forecasts are issued aon a conditional basis: real GDP will grow 4% next year if the Fed does not tighten, the Republicans will be the majority party in Congress if they also capture the Presidency, and sales will grow if competitors do not double their capital spending and advertising budgets.

The choice of which type of forecast is appropriate will depend largely aon how the results are to be used. A speculator in financial markets wants to know whether prices will rise or fall, not whether they will rise or fall under certain circumstances. An automobile dealer wants to know what lines of vehicles will sell most quickly, so he can optimize his ordering procedure. A pharmaceutical company wants to know how rapidly a new drug will be adopted.

Conversely, conditional forecasts can often be quite useful. Firms might want to determine how fast sales are likely to grow under normal business conditions, using those results as guidelines for rewarding superior performance. If sales are then affected by some exogenous event, guidelines can be adjusted accordingly. Forecasts of production planning might be determined based on the assumption that materials are delivered on time, compared with what might happen if a strike occurred. The most common way of delivering conditional forecasts is by using alternative scenarios, as discussed next.

A forecast that sales will rise 8% if the economy booms, rise 6% if real growth remains sluggish, and fall 2% if there is a recession may appear to be an excuse to avoid offering a firm forecast at all. However, that is not always true. In many cases, firms need to be prepared to take appropriate action if the economy falters even if the probability of that occurring is relatively low. Based on the historical forecasting record of macroeconomists, it would appear that recessions were not predictable. Consider the case of a lending institution involved in sub-prime auto loans. As long as the economy remains healthy, the vast majority of these loans will be repaid; if a recession strikes, the loss rate will rise enough to put the company out of business. Prudence might dictate less risky loans; but if the company is too picky, it will lose business to competitors and won't make enough loans to stay in business.

In this case the most appropriate procedure would be to assess the probability of a recession occurring next year. If it were only 5%, then the lending institution would continue to expand its sub-prime loan portfolio. On the other hand, if it were to rise to 25%, some trimming would be in order. Note that in this case the probability of an actual downturn the following year is well below a50%, yet some adjustment in corporate strategy is warranted. The alternative-scenario method of forecasting can also be used for long-range planning, since long-term economic forecasts are generally little more than trend extrapolations in any case. The company might discover that the probability of meeting its stated goal of a 15% annual gain in sales and earnings would occur only if the most optimistic macroeconomic forecast, with a probability of only 10%, were met. The company could then make plans to move into faster-growing areas of the economy or, alternatively, trim its ambitious long-term goals to more realistic levels [8].

So far we have been assuming that a forecast error of a+8% carries the same penalty as an error of a-8%. Often, however, that is not the case. For many companies, if sales increase faster than expected, that is fine; but if they don't, disaster strikes. I have already described such a situation for a sub-prime auto lending company. The same general type of argument could be applied to municipal bonds; as long as the community tax base grows above a certain rate, the interest and principal will be repaid, but if it dips below that rate, the issuing authority will default on the bonds.

In many companies, the rewards for exceeding the plan are substantial: bonuses, promotions, and larger budgets for next year. Similarly, the penalties for failing to meet planned targets are severe, including loss of employment. In a situation of that sort, many planners will set targets below their predicted level, so they will appear to have exceeded their goals. Eventually, management may catch on to this trick and fire all the planners, which is another risk. Nonetheless, the percentage of plans that are exceeded compared with the percentage that are not met strongly suggests that corporate planners are well aware of the asymmetric loss function managers may face a similar dilemma. If they beat the benchmark averages Dow Jones Industrials, S&P 500, or NASDAQ composite index they are handsomely rewarded; investors will switch their assets into those funds, and salaries and bonuses rise. If their performance falls short of the gains posted by the major averages, they will lose customers and possibly their own jobs.

This is not just a hypothetical example. The so-called January effect occurs because many money managers aggressively buy growth stocks early in the years or the previous December and, if they can show substantial gains, lock in those gains and buy the equivalent of index funds for the rest of the year. In the same vein, very few money managers who are already ahead of the average for the first three quarters of the year would take risks in the fourth quarter that would jeopardize their hefty year-end bonuses. So far we have not specified how many time periods in the future are being predicted. That can make a great deal of difference in the way a model is formulated. In models used to forecast only one period ahead, it might well be appropriate to use the lagged value of the variable that is being predicted.

Interest rates in the next period might very well depend on rates this period, as well as on other variables such as the inflation rate, growth rate, unemployment rate, value of the currency, budget surplus or deficit, and other relevant variables. However, suppose the model is used to predict interest rates on a monthly basis for the next 12 months. In this case, the forecasts for interest rates later ain the year would depend on "lagged" values of interest rates that were not known at the time of forecast. For example, suppose the forecast made at the beginning of March for interest rates depends on the level of interest rates in January and February. As the year progresses, the forecast for interest rates in June would depend on their level in April and May, which are not yet known [9].

For this reason, using the lagged dependent variable for multi-period forecasts causes serious difficulties that do not exist for the single-period forecast. That does not rule out the use of lagged dependent variables on a-priori basis, but it does raise a red flag. One of the tenets of the classical linear model, as will be shown in the next chapter, is that the values of all the independent variables are known at the time of forecast. Obviously, that is not the case when the lagged dependent variable is used in multi-period forecasting. Hence it is advisable to use a different approach when multi-period forecasts are required. To a certain extent, the difference between short- and long-run forecasts can be viewed as the difference between single- and multi-period forecasting. However, whereas short-term forecasts are more generally concerned with deviations from trends, long-run forecasts are often designed to predict the trend itself. As a result, different methods should be used [10], [11].

CONCLUSION

Forecasting with smoothing techniques is a powerful tool for predicting future trends based on historical data. The use of techniques such as exponential smoothing and moving average smoothing allows analysts to make accurate predictions about a variety of variables, including sales, stock prices, and weather patterns. Forecasting with smoothing techniques is an essential component of many industries and can help organizations make informed decisions about their future plans and strategies. By carefully analyzing historical data and applying appropriate smoothing techniques, analysts can make reliable predictions about future trends and make better decisions for their organizations.

REFERENCES

[1] D. A. Adeyinka and N. Muhajarine, "Time series prediction of under-five mortality rates for Nigeria: comparative analysis of artificial neural networks, Holt-Winters exponential smoothing and autoregressive integrated moving average models," *BMC Med. Res. Methodol.*, 2020, doi: 10.1186/s12874-020-01159-9.

[2] V. Chaurasia and S. Pal, "Application of machine learning time series analysis for prediction COVID-19 pandemic," *Res. Biomed. Eng.*, 2022, doi: 10.1007/s42600-020-00105-4.

[3] S. Hansun, “New estimation rules for unknown parameters on holt-winters multiplicative method,” *J. Math. Fundam. Sci.*, 2017, doi: 10.5614/j.math.fund.sci.2017.49.2.3.

[4] S. Jere, B. Kasense, and O. Chilyabanyama, “Forecasting Foreign Direct Investment to Zambia: A Time Series Analysis,” *Open J. Stat.*, 2017, doi: 10.4236/ojs.2017.71010.

[5] H. Shih and S. Rajendran, “Comparison of Time Series Methods and Machine Learning Algorithms for Forecasting Taiwan Blood Services Foundation’s Blood Supply,” *J. Healthc. Eng.*, 2019, doi: 10.1155/2019/6123745.

[6] R. Bakri, U. Data, and A. Saputra, “Marketing Research : The Application of Auto Sales Forecasting Software to Optimize Product Marketing Strategies,” *J. Appl. Sci. Eng. Technol. Educ.*, 2019, doi: 10.35877/454ri.asci1124.

[7] R. Bakri, U. Data, and N. P. Astuti, “Aplikasi Auto Sales Forecasting Berbasis Computational Intelligence Website untuk Mengoptimalisasi Manajemen Strategi Pemasaran Produk,” *J. Sist. Inf. BISNIS*, 2019, doi: 10.21456/vol9iss2pp244-251.

[8] N. Barman and M. B. Hasan, “Comparison of Forecasting Techniques for Short-term and Long-term Real Life Problems,” *Dhaka Univ. J. Sci.*, 2017, doi: 10.3329/dujs.v65i2.54523.

[9] C. J. Lynch and R. Gore, “Short-range forecasting of COVID-19 during early onset at county, health district, and state geographic levels using seven methods: Comparative forecasting study,” *J. Med. Internet Res.*, 2021, doi: 10.2196/24925.

[10] V. Chaurasia and S. Pal, “COVID-19 Pandemic: Application of Machine Learning Time Series Analysis for Prediction of Human Future,” *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3652378.

[11] Oviliani Yenty Yuliana and Tanti Octavia, “RANCANG SISTEM INFORMASI PERSEDIAAN BAHAN BAKU TERKOMPUTERISASI PT. KPL,” *J. Manaj. dan Kewirausahaan*, 2001.

CHAPTER 6

ADAPTIVE FILTERING FOR TIME SERIES FORECASTING: A COMPARATIVE STUDY OF KALMAN FILTER AND PARTICLE FILTER

Dr Abhinav Tiwari, Assistant Professor
Department of Decision Sciences, CMS Business School, JAIN Deemed to-be University, Bangalore, India
Email Id- dr.abhinav_tiwari@cms.ac.in

ABSTRACT:

Adaptive filtering is a powerful signal processing technique that has gained widespread attention in the field of forecasting. It is a data-driven approach that adjusts the weights of a filter in real-time based on the characteristics of the input signal. The goal of adaptive filtering is to estimate the underlying process by reducing noise and extracting meaningful information from the input data. Adaptive filtering algorithms are widely used in time-series forecasting applications, where the objective is to predict the future values of a signal based on its past behavior. By analyzing the historical data and updating the filter parameters in real-time, adaptive filtering can provide accurate predictions even when the underlying process is highly non-stationary or exhibits complex dynamics.

KEYWORDS:

Adaptive Filtering, Filter Parameters, Data-Driven, Real-Time, Signal Processing, Time-Series Forecasting.

INTRODUCTION

One of the principal goals of short-term forecasting, and one that has been emphasized by time-series analysis, is to remove the trend from time-series variables so the underlying properties of the series may be properly examined. If company sales have been growing an average of 12% per year, the challenge in short-term forecasting is to indicate how much sales next year will deviate from that trend. Long-range forecasters, on the other hand, might want to determine show many years it will take for the trend growth in sales to diverge from that s12% average gain.

The difference is analogous to the split responsibilities of the COO, who asks “How are we doing?”, and the CEO, who asks “Where are we heading?” as a function of income, prices as a function of unit labor costs, or interest rates as a function of the inflation are all likely to have serially correlated residuals. Because consumer spending patterns, for example, change slowly over time, the number of independent observations is probably far less than the sample period data would indicate. Consequently, the standard errors are significantly understated.

Virtually all statistical and econometric tests are based on the underlying assumption that the residuals are normally distributed. Often, however, that is not the case. That is another reason why the calculated goodness-of-fit statistics overstate the robustness of the equation [1]. The “law of large numbers” indicates that as the sample size increases, all distributions with a finite variance tend to approach the normal distribution. However, that is scant comfort to those who must deal with relatively small samples. Furthermore, some financial market data do not have bounded data; ain particular, percentage changes in daily stock prices are not normally distributed. Every once in a while, an unexpected event will cause a much larger change than could be expected from past history – especially in financial markets. Such distributions are colloquially referred to as “fat tails.” Estimates based on the assumption of a normal distribution when that is not the case are likely to generate disappointing forecasts.

Spurious correlation may destroy the usefulness of any model for forecasting, even if the sample period statistics appear to provide a remarkably accurate fit. Many studies have shown that series that actually have no correlation s– because they were generated from random number series – can provide highly significant goodness-of-fit statistics if enough alternative regressions are calculated. This problem has become particularly virulent in the PC era, where it is a simple matter to run hundreds if not thousands of regression equations very quickly.

The problem of “data mining” has also run rampant because of quick and inexpensive computing power. This issue always represents somewhat of a dilemma. One does not want to test only one or two versions of any given equation. After all, the theory may not be precisely specified; and even if the long run determinants are well determined, the lag structure and adjustment process may not be known. Empirical approximations of theoretical concepts may not be precise, so it is logical to try several different measures of the concept in question. Also, research results are often improved when alternative specifications were tried because the first attempt did not produce reasonable results [2].

Yet having provided all these reasons for diligent research, it is much more likely that econometricians and statisticians will “torture the data until they confess” instead of failing to calculate the necessary minimum number of regressions. Sometimes the equation fits very well during the sample period, and the goodness-of-fit statistics hold even in the forecast period, yet the equation generates very poor forecasts because the values of the independent variables are anot known. For example, sales growth for a particular company or individual product line is likely to change if competitors react to an erosion of their market share.

At the macroeconomic level, financial markets certainly will react differently to anticipated and unanticipated changes in policy. Consumers are likely to alter their spending patterns based on what they think will happen in the future as well as changes in current and lagged income and monetary conditions. It is not very helpful to develop theories that produce optimal forecasts under severely stylized sets of assumptions that are rarely encountered in the real world. Practical business forecasting invariably consists of two interrelated steps: use of standard statistical theory that has been developed based on restrictive assumptions, followed by modification of that theory to improve actual forecasting accuracy. These two steps cannot be considered in isolation. Thus, even in this introductory chapter, I have pointed out some of the major pitfalls that commonly occur in forecasting models. Further details will be provided throughout the text [3].

DISCUSSION

Robust economic theories, which have been verified by sophisticated econometric methods, do not generate accurate forecasts unless they are further modified. Economic theory says that the riskless long-term interest rate is related to the underlying growth rate of the economy, the Federal budget deficit ratio, and the expected rate of inflation. Econometrics can be used to test this theory. However, it cannot be used for forecasting unless, in addition, we can find an accurate way to predict the expected rate of inflation. Essentially the same comments could be made for forecasting the stock market, foreign exchange rates, or commodity prices. Since inflationary expectations are not formed in a vacuum, they could presumably be tied to changes in economic and political variables that have already occurred. So far, no one has been very successful at this.

The price of oil is tied to the world demand and supply for oil, which can perhaps be predicted accurately by econometric methods, using the geopolitical situation of Saudi Arabia vis-à-vis the US and other major powers as a major factor in the forecast. However, world economic hegemony cannot be predicted econometrically and probably cannot be predicted very well with any method so this is not a useful forecasting model. Certainly no one in the early 1980s publicly predicted the fall of the Berlin Wall by the end of the decade. Historically, the growth rate for PCs, modems, and other high-tech equipment can be accurately tracked over the sample period by identifying the time when major innovations were introduced and matching their performance to various growth curves. In the future, since the timing of such innovations is unknown, such a set of regression equations would not serve as a useful forecasting model [4].

Economic theory says that the value of the dollar depends on relative real interest rate differentials; the higher the real rate in the US, the more likely it is that the dollar will appreciate that a stronger dollar will attract capital from abroad, hence resulting in a lower interest rate than would otherwise occur. Both of these theories can be verified separately, but unless further adjustments are made they are useless for predicting either the value of the dollar or interest rates, since they lead to opposite conclusions. This is indicative of a larger problem in forecasting, where an individual theory may provide robust empirical results in isolation but may be useless for forecasting because the factors that are being held constant in the theory are in fact always changing.

These examples provide a flavour of the problems of building a practical forecasting model. Many of the examples involve interrelationships between several variables that must be predicted simultaneously. However, even in the cases where the independent variables are actually known ahead of time, and in that sense are truly exogenous, model builders often go astray by failing to realize the spurious correlation introduced by common trends in several of the time series [5]. Using econometrics to build forecasting models is deceptively difficult. As Clive Granger has put it, “econometric modelling is an activity that should not be attempted for the first time.”

It takes prac the same data in a scatter diagram, which reinforces that conclusion. The residuals of that equation; when put on a different scale, it is more easily seen that the errors in predicting consumption with this simple equation may be as much as $200 billion per year. Figure 1 illustrate the Time series prediction using adaptive learning algorithm.

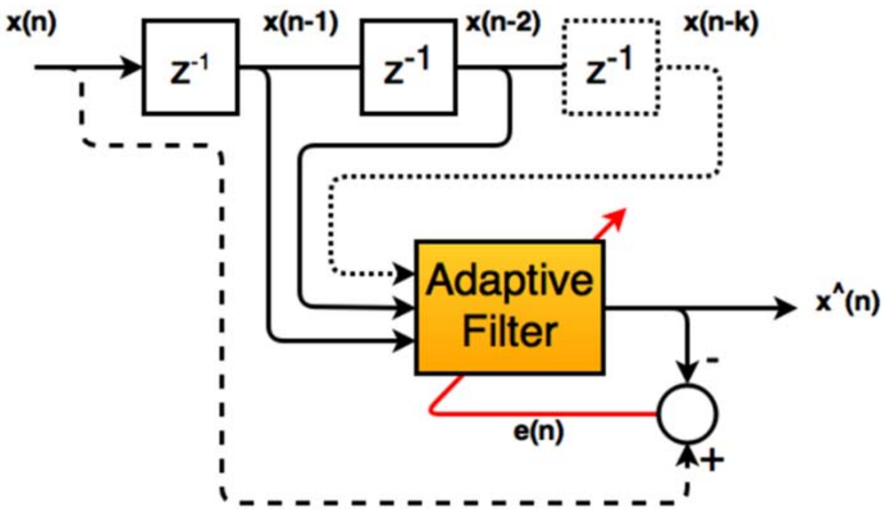


Figure 1: Illustrate the Time series prediction using adaptive learning algorithm.

One could claim that, without reference to further benchmarks, we don't know whether $200 billion is a "large" error or not. Some further comparison is warranted. In 1999, real consumer spending in the US was about $6,000 billion, and over the past 10 years had grown at an average annual rate of 3.5% per year. Hence a naive model that said the growth rate in 2000 would continue at 3.5% would predict an increase of about $210 billion. In fact the actual increase, based on preliminary data, was $316 billion, for an error of $106 billion. Seen in that light, a $200 billion error is abnormally large, since it is almost double the error generated by a naive model. Finally, the actual and forecast values for the percentage changes in each of these variables; which makes it obvious that while income is an important determinant of consumption, it is hardly the only one.

The lines in the top part of this graph show the actual percentage change in consumption compared with the percentage changes that are estimated by the regression equation, which in this case simply states that percentage changes in consumption are a function of percentage changes in income plus a constant term. The line in the bottom part of this graph, which is on a different scale, plots the residuals, or the differences between the actual and estimated values of the dependent variable [6]. Another illustration is the annual correlation between the Federal funds rate and the rate of inflation as determined by the consumer price index (CPI). Typically, we observe that when the inflation rate fluctuates, the displays a scatter plot of annual data for the funds rate and the inflation rate for the years 1955 through 1998 (prior to 1955, there are no data available for the funds rate). There is no doubt that the series are positively correlated, albeit imperfectly. The regression line as calculated using least squares is shown by the solid line. Because the slope of the regression line is marginally below unity, the funds rate is marginally positive when the inflation rate is zero.

In, a line graph is used to display the same two variables. The funds rate was barely higher than the inflation rate from 1955 to 1980. The difference between the funds rate and the inflation rate was significantly larger between 1981 and 1989, signalling a change in Federal Reserve policy. The scatter diagram does not clearly illustrate this, whereas the line graph does [7]. Figure 2 illustrate the Adaptive Filter Design.

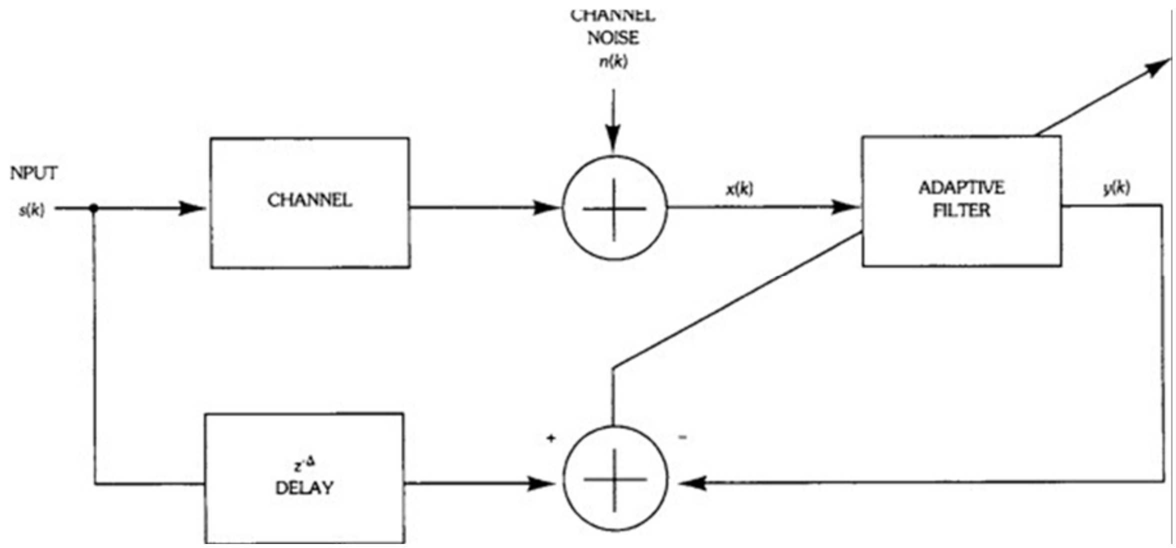


Figure 2: Illustrate the Adaptive Filter Design.

The residuals are supposed to be normally distributed in accordance with the classical linear model's underlying assumptions. A quick test to determine whether this is the case is to look at the residuals' histogram. We examine the equation's residuals, in which the Federal funds rate is a function of the inflation rate. I have discovered that persuading students to double-check the data before using it is one of the most difficult tasks when teaching forecasting courses. Even when information is gathered from reliable sources, mistakes do occur.

The beginning and ending dates aren't always exactly as listed, and the series can occasionally be corrupted. Even if the data are error-free, one or two outliers may cause the entire modelling process to be distorted. If you don't check beforehand, you won't notice this until your regression estimates give unrealistic sample period estimates or poor forecasts. Series that are supposed to be in comparable units on occasion are not; one is in millions while the other is in thousands [8]. Most government data are seasonally adjusted, but most business data are not, with the exception of financial markets. Therefore, an adjustment procedure is needed if you plan to combine the two types of data. Part III will go into greater detail on this subject, but for now, let's quickly review some of the main seasonal adjustment techniques, along with their benefits and drawbacks.

Seasonal patterns are common in economic time-series data. The majority of the time, government data have already been seasonally adjusted, but individual company data are frequently not. Without first accounting for seasonal factors, attempts to use these data for modelling will typically produce subpar outcomes.

Sales increase around Christmastime, more people go to the beach in the summer, snow shovel sales increase during the winter, broiler (chicken) prices peak around July 4th week, the unemployment rate for construction workers increases during the winter, and so on. If these patterns are regular, it is best to take the common seasonal factors out; otherwise, one risked having a correlation that was based more on seasonal factors than underlying economic trends. The famous example of this is when an economist correlated seasonally unadjusted consumer spending with unadjusted money supply data. Because both variables rise significantly in the fourth quarter, an erroneously high correlation was found. The claim made by some wag was that this economist "discovered that the money supply causes Christmas."

The variable for December, for example, would be 1 for that month and 0 elsewhere, and so on. Suppose one calculated a regression for unseasonally adjusted department store sales on dummy variables for each month of the year. The results of that regression would show a very strong correlation, but the equation would not have explained anything other than the fact that sales data from department stores increase before Christmas and Easter and decrease between the months of February and July. An equation of that type would not contain any pertinent information, but the fit would be high [9].

Retailers typically want to know if sales this year, or this holiday season, will be better or worse than usual, adjusted for the general growth trend. The remaining data series is more likely to resemble a random variable and thus more closely satisfy the fundamental statistical criteria and tests after the trend and seasonal factors have been removed. The statistical findings are thus more likely to offer a realistic assessment of how accurate the forecasts will be. Although the odds are improved, this does not ensure that the outcomes will be helpful. Once the data have been successfully entered into EViews or a comparable program, it is very easy to create a histogram for each variable and ensure that erroneous observations won't dominate any potential regression equations. Take your time; you won't regret it.

Technically, only the residuals need to conform to the standard statistical requirements. Outliers should be treated with dummy variables or excluded if there are any because otherwise, the results of the regression will be heavily influenced by them. I'll demonstrate what happens if outliers are ignored later. Let's say that a measurement deviates from the mean by five standard deviations. The probability that the variable actually has a normal distribution is only about one in a million. However, since the sum of squares is being minimized, practically speaking, such a residual would have a weight 25 times greater than an observation that deviates from the mean by one standard deviation. That one outlier would dominate the regression equation in a sample with 20 to 50 observations, effectively making the regression just fit that point.

The quarterly percentage changes in auto sales are depicted in a histogram. It is obvious that the changes are not evenly distributed. Given the significant kurtosis and fat tails, the likelihood that this series is normally distributed is less than 10-6. Given that, the next concern is whether there is a strong economic justification for those outliers. Major steel strikes caused the first pair, and major auto strikes caused the other two. Strike periods should therefore be treated differently. The best course of action in this situation is to treat auto strikes as a dummy variable; in other circumstances, outliers should be completely excluded. One common practise among model builders is to perform a large number of regression equation calculations for a specific dependent variable before selecting the equation with the highest. For a number of reasons, that practise is frequently suboptimal.

First of all, changing the dependent variable's form (for example, from level to percentage change) may lower R2, but it may also lower the standard error. It's important to avoid comparing bicycles and apples. In some cases, an equation with a lower R2 will produce forecasts that are a lot more accurate. The goodness-of-fit statistics are always overstated in such cases. Particularly, many time-series data-based models have autocorrelated residuals. Forecast errors frequently exceed what the equation predicts when that occurs because the t-ratios and R2 statistics are overstated [10].

The Durbin-Watson (DW) statistic is used to examine the residuals' first-order autocorrelation. DW can be between 0 and 4; when it is close to 2, there is no autocorrelation; when it is below 1.4, there is positive autocorrelation in the residuals. The Cochrane-Orcutt transformation is a widely used technique for autocorrelation correction. However, a transformed equation frequently produces subpar predictions, particularly for multi-period forecasting. The R2 statistic is based on the levels form of the dependent variable, which results in a highly irrational estimate of how accurate the forecasts will be. If DW is very low, this transformation is almost the same as using a first-difference equation.

The residuals could also be heteroscedastic, which indicates that a few very extreme outliers frequently dominate them. The goodness-of-fit statistics are overstated if the residuals are heteroscedastic, but that isn't the main issue. The coefficients in least-squares regressions are frequently distorted during normal times because they give the most weight to extreme observations. In light of this, unless the same unusual circumstances arise, such an equation would not be the best choice for forecasting. However, this X assumption is not true in the majority of common time-series equations. For instance, consumption affects income as well as income affects consumption because if consumers increase their spending, more output will be produced, which in turn increases income. At the time of the forecast, current income is also unknown. Bond prices and stock prices have a positive correlation, but when stock prices decline, there may be a "flight to quality," which raises bond prices. The price of gasoline affects how much is consumed, but OPEC may decide to raise prices if consumption increases significantly.

The identification problem, also referred to as simultaneity bias or the identification problem, is a significant issue when developing forecasting models and is covered in more detail later in this book. But initially, the simpler presumption that Y has no influence on X is used to develop the theory and operating principles. Unbiased estimates of the mean and variance, m and s 2, were given in the previous chapter. The consistency and effectiveness of these estimates are also desired. Unbiasedness means that the expected value of the variable is equal to the population mean. It is absolutely one desired property of statistical estimators, or parameter estimations. However, it is not the only one.

Consistency means that the error diminishes as the sample size increases. One would certainly expect that a sample size of 100 would have a smaller standard error than a sample size of 10. This term is similar in most cases to asymptotic unbiasedness, which means the bias falls to zero as the size of the sample increases. There are a few odd probability distributions where the two terms are not the same, but for our purposes they may be considered equivalent. While presumably no researcher wants a “biased” estimate, consistency is actually more important to statisticians than bias. Small sample sizes (often in the range of less than 20 observations) generally do not give robust estimates anyhow, but it is critical that as the sample size grows, the error diminishes [11].

Otherwise there is no reason to suppose that the researcher is zeroing in on the correct value. Efficiency is the other important criterion; that means the estimate in question has a smaller variance than any other estimate for a given sample size. Sometimes efficiency is more important than unbiasedness. Consider the case where a mugger has attacked you and there are two witnesses. The actual height of the mugger is 5 ft 10in. The first witness thinks the mugger was 5 ft 10in, but isn't sure; his height could have been anywhere between 5 ft 2in and 6 ft 6in.

That is an unbiased estimate, but not very useful. The other says his height was between 5 ft 8in and 5 ft 9in, whereas in fact it turns out to be 5 ft 10in. That is a biased estimate but more useful. Quite often it is the case that two variables X and Y will appear to have a very high correlation, but when one calculates a regression equation that includes a third variable Z, the partial correlation between X and Y will disappear. That means the correlation between the Y and the residuals based on a regression between X and Z is zero. For example, the number of marriages might be quite highly correlated with the number of drunken drivers arrested, but when population is added to the equation, that correlation disappears, since it only reflects a common trend [12], [13].

CONCLUSION

Adaptive filtering is a powerful technique for time series forecasting that allows models to adjust to changing data patterns over time. It can be used to extract hidden patterns and trends from data and make accurate predictions about future trends. Adaptive filtering algorithms are widely used in various fields such as finance, engineering, and economics. Some of the most popular adaptive filtering algorithms used for time series forecasting include Kalman filtering, particle filtering, and neural networks.

REFERENCES

[1] N. Amor, A. Meddeb, S. Marrouchi, and S. Chebbi, "A comparative study of nonlinear Bayesian filtering algorithms for estimation of gene expression time series data," *Turkish J. Electr. Eng. Comput. Sci.*, 2019, doi: 10.3906/elk-1809-187.

[2] S. A. Gadsden, M. Al-Shabi, and S. R. Habibi, "Estimation Strategies for the ConditionMonitoring of a Battery Systemin a Hybrid Electric Vehicle," *ISRN Signal Process.*, 2011, doi: 10.5402/2011/120351.

[3] H. Lakehal, M. Ghanai, and K. Chafaa, "BBO-Based State Optimization for PMSM Machines," *Vietnam J. Comput. Sci.*, 2022, doi: 10.1142/S2196888822500026.

[4] M. W. Khan, N. Salman, A. Ali, A. M. Khan, and A. H. Kemp, "A comparative study of target tracking with Kalman filter, extended Kalman filter and particle filter using received signal strength measurements," in *Proceedings of 2015 International Conference on Emerging Technologies, ICET 2015*, 2016. doi: 10.1109/ICET.2015.7389222.

[5] V. M. M, J. S N, and Y. T, "An Analysis of the Kalman, Extended Kalman, Uncented Kalman and Particle Filters with Application to DOA Tracking," *Signal Image Process. An Int. J.*, 2017, doi: 10.5121/sipij.2017.8603.

[6] M. Kiani and S. H. Pourtakdoust, "State estimation of nonlinear dynamic systems using weighted variance-based adaptive particle swarm optimization," *Appl. Soft Comput. J.*, 2015, doi: 10.1016/j.asoc.2015.04.029.

[7] X. Lai, C. Qin, W. Gao, Y. Zheng, and W. Yi, "A state of charge estimator based extended kalman filter using an electrochemistry-based equivalent circuit model for lithium-ion batteries," *Appl. Sci.*, 2018, doi: 10.3390/app8091592.

[8] S. Kaviani, H. Salarieh, A. Alasty, and M. Abediny, "Comparison of nonlinear filtering techniques for inertial sensors error identification in INS/GPS integration," *Sci. Iran.*, 2018, doi: 10.24200/sci.2017.4328.

[9] M. Chaabane, M. Mansouri, H. Nounou, M. Nounou, and A. Ben Hamida, "Enhanced particle filter for states and parameters estimation in structural health monitoring applications," *J. Civ. Struct. Heal. Monit.*, 2016, doi: 10.1007/s13349-016-0171-x.

[10] F. Yang, Y. Xing, D. Wang, and K. L. Tsui, "A comparative study of three model-based algorithms for estimating state-of-charge of lithium-ion batteries under a new combined dynamic loading profile," *Appl. Energy*, 2016, doi: 10.1016/j.apenergy.2015.11.072.

[11] T. Ljouad, A. Amine, and M. Rziza, "A hybrid mobile object tracker based on the modified Cuckoo Search algorithm and the Kalman Filter," *Pattern Recognit.*, 2014, doi: 10.1016/j.patcog.2014.04.003.

[12] K. Hehdly, M. Laaraiedh, F. Abdelkefi, and M. Siala, "Cooperative localization and tracking in wireless sensor networks," *Int. J. Commun. Syst.*, 2019, doi: 10.1002/dac.3842.

[13] N. Daroogheh, N. Meskin, and K. Khorasani, "Ensemble Kalman Filters for State Estimation and Prediction of Two-Time Scale Nonlinear Systems with Application to Gas Turbine Engines," *IEEE Trans. Control Syst. Technol.*, 2019, doi: 10.1109/TCST.2018.2870044.

CHAPTER 7

PREDICTING THE FUTURE: SIMPLE REGRESSION MODELS FOR FORECASTING TRENDS AND PATTERNS

Vaibhav Goutham Suresh, Assistant Professor
Department of General Management, CMS Business School, JAIN Deemed to-be University, Bangalore, India
Email Id- vaibhavsuresh@cms.ac.in

ABSTRACT:

Forecasting with simple regression involves using a linear relationship between two variables to make predictions about the future. Simple regression involves using one independent variable to predict the dependent variable. This method is often used in economics, finance, and marketing to forecast trends and patterns. Simple regression is based on the assumption of a linear relationship between the variables, and may not be appropriate if the relationship is more complex. Additionally, other factors such as external events or changes in the underlying data can also affect the accuracy of the predictions made using simple regression.

KEYWORDS:

Dependent Variable, Independent Variable, Linear Relationship, Prediction, Simple Regression, Scatter Plot.

INTRODUCTION

X and Y are uncorrelated, but when variable Z is added to the equation, both Y and Z became significant because they are negatively correlated with each other. For example, we might find a very low simple correlation between capital spending and interest rates, but when the growth in output is added to the regression, that term is significantly positive, while interest rates become significantly negative. That is because, in a statistical sense, high interest rates are usually associated with low growth, and vice versa. It is a simple matter to calculate the covariance matrix for all the variables used ain a given regression, but that doesn't impart much useful information because the variables are generally of different magnitudes (e.g., some are interest rates, some are in billions of dollars, some are percentage changes, and so on) [1].

However, this defect can be easily remedied by transforming the covariance matrix into the correlation coefficient matrix using the formula, which is repeated here Errors of measurement. Data reported by the government are based on incomplete information. Missing observations tend to be interpolated in such a way that the data are smoothed. Also, the same biases may occur systematically in the sampling programmed for different time periods.

This is probably the most serious problem. Positive correlation of residuals often signifies one or more significant variables are missing. In some cases, such as expectation variables, these missing variables cannot be measured precisely. Sometimes the equation is piecewise linear, which means the underlying structural relationship has shifted during the sample period.

The equation is linear in both periods but one or more of the coefficients has changed. In other cases, the coefficients vary with the phase of the business cycle. Sometimes one or one or more of the independent variables should be raised to some power other than unity. That could mean an exponential power such as squared, cubed, etc., or ait could indicate an inverse correlation, where the form of the independent variable should be 1/X instead of X. The effect of habit. Even if the data are correct and the equation is correctly specified, economic decisions are often based on habit, so that the error term in this period after taking into account all the relevant variables really is correlated with the error term in the previous period. Sometimes this information can be used to help improve forecast accuracy, but using it runs the risk of generating poor forecasts whenever habits do change.

The presence of autocorrelation does not affect unbiasedness or consistency, but it does affect efficiency. That means the standard errors that are calculated using OLS are understated, and hence the significance levels of the individual terms (and often the entire equation) are overstated. Results that appear to be significant actually are not emphasize that the presence of autocorrelation in the residuals does not necessarily mean the parameter estimates are incorrect; it simply means that the sample period statistics will probably understate errors during the forecast period [2].

In working with time-series data, it is often the case that quarterly data will indicate significant autocorrelation, while the same equation estimated with "catch up" after one quarter, but that is only because they depend on the actual lagged value. If the fitted lagged value were to be used, the simulated values would drift ever-further away from the actual value. This is perhaps an extreme example, but it illustrates the point well. Using the lagged dependent variable on the right-hand side of the equation will often result ain an equation with apparently superb sample period fits, but it will be useless for forecasting because the lagged dependent variable also has to be predicted.

Even for single-period forecasting, the equation given above would miss virtually every turning point, and hence would be useless for actual forecasts. Forecasting errors often arise when trying to predict Y if the most important independent variable is Y-1, which means relying primarily on that variable for predicting Y. By definition, Y-1 never turns down ahead of time. Thus relying on the lagged dependent variable means missing almost all the turning points. Furthermore, aY-1 will not only rise during the first period that Y fell, but will continue to rise because the forecast will continue to contain erroneous feedback from the variable that failed to turn around. The more periods that are predicted with this equation, the worse the forecasts.

Thus it seldom if ever pays to put the lagged dependent variable on the right-hand side of an equation that will be used for multi-period forecasting. If that variable continues to be very significant in spite of all other changes that have been made, switch to percentage first differences, or other methods that eliminate the trend, many of which are discussed in detail in the next chapter estimating data from a panel survey of consumers, whose income ranges all the way from (say) $10,000 to $1,000,000. Let us assume for these purposes that the same factors govern consumption at all levels of income, so the various consumption functions are similar. In that case, the standard deviation for the a$1 million consumers would be 100 times as large as the standard deviation for the $10,000 consumers. A few rich consumers would therefore dominate the sample in terms of the statistical tests [3].

DISCUSSION

The straightforward solution to this problem is to scale the results so that s (say) a 5% error for the rich gets the same weight as a 5% error for the poor. The simplest case stems from the fact that most time series increase over time (consumption, production, employment, prices, etc). (Consumption, production, employment, prices, etc.). If the level of the dependent variable is, say, ten times as great at the end of the sample period as it was at the beginning, then the error term is also likely to be ten times as great. However, if this variable is correlated with an independent variable (income in the consumption function) with the same general trend, the residuals probably awill not be heteroscedastic. Even if heteroscedasticity remains, this might have the net result of giving more recent observations greater weight, which in many cases is a good idea anyhow.

This can perhaps best be illustrated by looking at financial data. On Monday, October 19, 1987, the Dow Jones Industrial Average plunged 508 points, or 22% a decline almost twice as much as the next largest percentage drop (including Black Tuesday in 1929). If all percentage changes are treated equally in a statistical sense, the results will be biased in the sense that the market will be shown to fall more on the 19th of each month, or each Monday, or each October. In macroeconomic data, price equations might be dominated by energy shocks, thereby neglecting the importance of other key variables such as unit labor costs, capacity utilization, monetary policy, and so on [4].

These problems can be quite severe in the sense that not only are the goodness-of-fit statistics overstated, but the parameter estimates themselves will become biased, and forecasts based on these estimates will invariably generate the wrong answers. Case Study 3 illustrates how this can happening White's correction, or the Newey–West correction. There is nothing wrong ain adjusting the standard errors and it should be done if heteroscedasticity is found to be present, but most of the time it does not make much difference. By experimenting, the reader can quickly determine that using these adjustments generally has much less effect on the parameter estimates than using the AR(1) adjustment, or changing from levels to changes.

Another method of diminishing heteroscedasticity is to use weighted least squares or put the dependent variable in ratio form. Weighted least squares (WLS)10 is somewhat arbitrary in the sense the model builder must choose which weights to use; one common method is to use one of the independent variables that has the same trend or scale factor as the dependent variable. In the consumption function, for example, that would be income, whether time series or cross-section data were being used. One could also take the ratio of consumption/income as the dependent variable; the impact of these changes is discussed in the following chapter. In most cases, using WLS does not change the parameter estimates very much either. Figure 1 illustrate the Linear Model Forecasting.

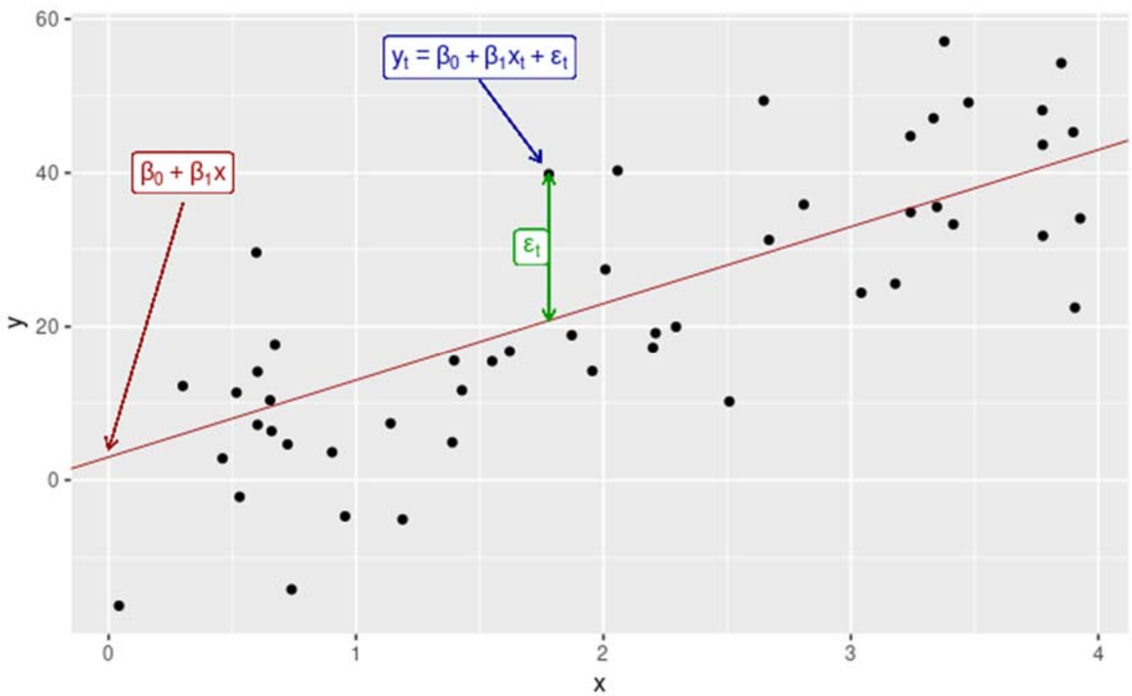


Figure 1: Illustrate the Linear Model Forecasting.

If one of the values of a time series has an extreme outlying value, thus leading to heteroscedasticity, it cannot be ignored; otherwise that one value will dominate the results, and the equation will eventually be reduced to fitting these one or two extreme points. One way to proceed is usually to treat this with a dummy variable. Another is simply to disregard the errant observations completely. This process, often known as masking, consists of automatically excluding any observation where the error term is larger than reassigned multiple of the standard error of the overall equation. One common rule of thumb is to exclude observations whose error term is more than three times the standard error.

In many cases, omitting outliers will generate parameter estimates that will produce forecasts with smaller errors than if the outliers were included. However, simply omitting all observations that cannot be explained can become a dangerous procedure. In particular, one should determine whether those outliers were caused by a specific exogenous development. For example, electric power usage would rise more during extremely cold winters or hot summers, insurance claims would rise dramatically after a hurricane, and entertainment expenditures for a given city – especially one of moderate size would rise sharply after the local baseball team wins the World Series. At first glance it might appear these outliers should be discarded, but in fact they can usefully be correlated with the indicated exogenous development. Only in cases where outlying values do not appear to be related to any realistic independent variable should they be excluded from the equation. The previous chapter discussed the issues of autocorrelation and heteroscedasticity in the residuals, and illustrated the standard statistical adjustments that are used when these problems arise. It also pointed out the possible pitfalls of building forecasting models when those conditions are present [5].

Multicollinearity occurs when two or more of the independent variables in the regression equation are very highly correlated. Unlike autocorrelation and heteroscedasticity, there are no specific tests for multicollinearity, but it can be aeven more serious because it distorts the values of the coefficients themselves, rather than affecting only the goodness-of-fit statistics. The usual problem is that while the sum of the coefficients of highly collinear variables is close to the true underlying value, the individual coefficients contain significant errors. Thus aunless the relationship between these values is exactly the same in both the sample and forecast periods, predictions from such a model are likely to contain serious mistakes.

The problems of multicollinearity stem from two major sources: different variables that are highly collinear, and lagged values of the same variable, which will be highly collinear if that series contains strong trends. The treatment of these is quite different in the first case, variables should be combined, or the strong common trend should be removed by using first differences, percentage changes, ratios, or weighted least squares. In the second case, several lagged values of the same variable should be combined into only a few terms by the use of distributed lags. Both these methods are considered in detail. Note that the overall goodness-of-fit statistics in the first four equations are all virtually the same when the components of income are disaggregated [6].

That ought to be a tipoff something is wrong, for theory suggests that at least the short-term marginal propensity to consume from volatile components of income is smaller than from stable components. The fact that the estimated value of the coefficients for TR and YOTH are greater than unity looks suspicious right away the negative signs on W and TR are clearly inappropriate. When multicollinearity is eliminated by taking percentage changes, wages are the most important variable, and transfers are much less important because the major cyclical component of transfers is tied to the business cycle, and rises awhen other income declines. Later we will see that this consumption function ais still seriously incomplete because no monetary variables are included.

Also note that the standard errors become enormous when extreme multicollinearity is present. That doesn't always happen; but when it does, that is an obvious hint this condition exists. In that case, the logical choice is to drop one or more of the variables. There are a few tests that suggest multicollinearity is present, but they are not discussed here nor are they included in EViews because (i) they do not provide any additional information that cannot be gleaned from the correlation matrix and comparison of the sizes of the standard errors, and (ii) unlike with autocorrelation and heteroscedasticity, there is no simple way to fix the problem. There are some tests known as "complaint indicators," which tell you that multicollinearity is present, but not what to do about it [7].

If an equation with extreme multicollinearity is used for forecasting, the results will contain very large errors if there is even a tiny change in the relationship between the multicollinear independent variables, because the coefficients have been blown up to unrealistically high values. Ordinarily, if the relationship between the independent variables changes a little bit, the forecast error will be quite small. Hence it is generally a poor idea to generate forecasts using an equation with extreme multicollinearity. In fact, the underlying equation might or might not be multiplicative.

There are no a-priori rules for determining when an equation is linear and when it is log-linear. In a log-linear equation, the elasticity's remain the same over the entire sample period. That may or may not be an appropriate assumption.

A linear demand curve means that, at relatively low prices, the demand is inelastic, so an increase in price will boost total revenues; while at relatively high prices the demand is elastic, so a further increase in price will reduce total revenues. The log-linear demand curve assumes that the price elasticity is constant along the entire length of the curve. On a-priori basis there is no way to determine which assumption is better. In the airline equation, both the levels and the logarithm equations give non-significant results for the price terms, so the issue cannot be decided with these equations.

In that case, the function is neither linear nor log-linear, and must be estimated using nonlinear techniques, a method discussed later in this chapter. Production functions are generally thought to be log-linear, with constant elasticities of substitution. It is often assumed a certain percentage increase in costs results in the same percentage increase in prices whether the economy is in a boom or a recession. As a matter of fact, the change in the markup factor is probably due more to monetary policy and expectations than to the phase of the business cycle, leading to a complicated nonlinear relationship that usually is not estimated directly (i.e., a linear approximation is used by including monetary factors separately). But here again there is no conclusive empirical evidence that using logarithms is better or worse [8].

In many cases, the empirical evidence does not permit one to choose between linear and log-linear equations. If the theory provides strong reasons to expect constant elasticities, use the logarithmic formulation; otherwise use the linear form. For series with strong trends, the results generally do not differ very much. The logarithm form is often preferred because, as noted above, the coefficients are elasticities, making comparison easier if one is working with equations involving hundreds of commodities, countries, companies, or individuals.

As noted using percentage changes is virtually the same as using first differences of logarithms, although there is one slight difference: in calculating percentage changes, there is some ambiguity about whether the denominator should be the current period, the previous period, or some average of the two periods. By taking differences of logarithms, this ambiguity is resolved. In the more usual case, the changes in the variances are approximated by the changes in one of the trend variables (in the airline travel function, the obvious choice would be income). The equation is then divided through by that variable. In this case, weighted least squares has some similarity to ratios, but the results are usually closer to the OLS equations than the ratio equation.

There are few differences between the OLS and WLS equations because all of the independent variables have significant trends: income rises and relative price falls. When some of the independent variables do not have any trends, such as percentage changes or interest rates, WLS often improves the coefficients of these trendless terms, in which case the forecasting accuracy of the equation generally improves. Most of the time, however, there is not much difference between OLS and WLS estimates.

Any time one is calculating regressions using time series with strong trends – whether they are components of aggregate demand and income, individual demand and supply functions, production functions, money supply, stock prices, or any other variable that grows over time – the original set of equations, based on the relevant theory, will usually show positive autocorrelation of the residuals, and will usually suffer from multicollinearity as well. These maladies could be due to a number of different factors, but most of the time the culprit is the strong common trend.

Both the sample period statistical results and the forecasting properties of the equation are likely to be unsatisfactory unless these problems are resolved. Using the lagged dependent variable on the right-hand side of the equation – or using an autocorrelation adjustment – will provide a "quick fix" in the sense that the sample period statistical tests will appear to be better, but often the multi-period forecasting record will become worse. It is better to use one of the methods mentioned here to eliminate the common trends. by random fluctuations, which usually means the absolute values of the parameter estimates are biassed down. If monthly or quarterly data are being used, one logical choice is to try annual percentage changes (i.e., this month or quarter over the same month or quarter a year ago). For forecasting with multiequation models, it is better for the lagged variable to be an average over the past year than simply a year ago, in order to avoid spurious cycles in forecasting more than one year out.

If the annual percentage change method does not work, consider either ratios or detrending each series. These methods often do not solve the problem of autocorrelation. Yet while an AR(1) adjustment will superficially solve that problem, it generally will not improve forecasting accuracy for multi-period predictions and often makes the errors larger [9]. For variables without trends interest rates, inflation rates, foreign exchange rates of the dollar, etc. levels equations are preferred. There may still be some autocorrelation, but that is best handled by improving the specification of the equation rather than by moving to percentage

So far we have looked at the problem of multicollinearity as it applies to two or more independent variables with strong trends. For illustrative purposes we used annual data. However, an even more common cause of multicollinearity occurs when quarterly or monthly data are used and the theory suggests several lagged values of one or more of the independent variables. For example, consumption depends on lagged as well as current income. Because of multicollinearity, the estimated coefficients in regression equations will generally be nonsensical if an entire string of lagged variables is entered in a single equation. Yet theory does not tell us precisely how long the lag will be, nor what shape the distribution will take: whether 90% of the reaction will take place in the first time period, or whether it will be spread over several years.

In most key macroeconomic equations consumption, investment, exports, interest rates, wages and prices, etc. economic choices depend on lagged as well as current variables. The problem is obviously more important the shorter the time period considered: lagged values are more important for quarterly and monthly data than for annual data. At the industry level, changes in shipments, new orders, and inventories depend on what has happened in the past as well as the present. Only in cross-section data are lags generally considered unimportant. Since the product of two parameters cannot be estimated with linear methods, it is generally assumed that b1 is unity. That is essentially the same as dividing by b1, which would not affect any of the coefficients except the constant term, whose value is generally unimportant. A formal exposition involves Lagrangian multipliers and more algebra than is appropriate here. However, on a heuristic level, we can think of fitting a quadratic (or a cubic) as an approximation of a more complicated lag structure that exists in the real world. Such an exercise is little more than curve fitting; but as pointed out earlier, theory doesn't tell us the length or the shape of the actual lag structure, even if it does suggest the variables and the approximate magnitudes of the coefficients that should be expected.

The PDL method is useful because it sharply reduces the number of degrees of freedom that are used in the estimation of the equation, it reduces multicollinearity, and it reduces the probability that one or two outliers will determine the shape of the estimated lag distribution. The major drawback to this method, as will be seen, is that for variables with strong trends, it is often difficult to determine empirically how long the lag should be, and what order the polynomial should be. Two of the most common examples of the use of PDLs in macroeconomics are the consumption function, and the lag between capital appropriations and actual capital expenditures. In the case of the consumption function, Friedman estimated a 17-year lag on income to approximate permanent income, although his original work was done before the concept of PDLs were used in econometrics. The regressions of capital spending on appropriations by Almon was the seminal use of PDLs in econometrics.

Some textbooks suggest starting with a high-order polynomial and then dropping the insignificant terms; but in real life it does not make much sense to experiment with any polynomial higher than a cubic unless one has independent information that would suggest a more complicated lag structure. Don't expect to be able to pinpoint the precise length of the lag structure. The R2 will vary hardly at all for adjacent lags [10]. In long lags, the coefficients often have the wrong sign for a few months or quarters, which doesn't make any economic sense. However, it is often difficult to get rid of these erroneous signs without compromising the rest of the equation. My advice would be if the coefficients with the wrong sign are tiny, leave them in. "Tiny" in this context means a t-ratio with an absolute value of less than 0.5. 4 Often, the tendency is to keep adding lags as long as 2 keeps increasing.

However, that generally tends to make the lag longer than is the case in the underlying population. As pointed out, the adjusted R2 will keep increasing as long as the t-ratios of each additional term are greater than 1. However, in the case of PDL, this rule of thumb should be modified so that additional lag terms are not added unless the t-ratios of those additional terms exceed. Since the average length of business cycles used to be 4–5 years, it is sometimes the case, when using time-series data before the 1980s that long lags may simply be picking up a spurious correlation with the previous cycle; this is known as the echo effect. Since business cycles happen less frequently these days, splitting the sample period and estimating the equations starting in 1982 should reduce if not eliminate any such effect. If the maximum length of lag is reduced in the more recent sample period, then the earlier correlation probably does represent an echo effect [11].

CONCLUSION

Simple regression models can be useful for forecasting in certain situations. They are relatively easy to implement and interpret, and they can provide valuable insights into the relationship between a dependent variable and one or more independent variables. However, the usefulness of simple regression models for forecasting depends on several factors, such as the quality and availability of data, the stability of the relationship between variables over time, and the presence of outliers or other data anomalies. In some cases, more advanced modeling techniques may be necessary to account for these factors and produce more accurate forecasts. Overall, simple regression models can be a valuable tool for forecasting, but their limitations must be carefully considered and addressed in order to produce reliable results.

REFERENCES

[1] H. Rahman, I. Selvarasan, and A. Jahitha Begum, "Short-term forecasting of total energy consumption for India-a black box based approach," *Energies*, 2018, doi: 10.3390/en11123442.

[2] A. A. Elshayeb, A. A. Ahmed, M. A. El Siddig, and A. A. El Hussien, "Prevalence of current patterns and predictive trends of multidrug-resistant Salmonella Typhi in Sudan," *Ann. Clin. Microbiol. Antimicrob.*, 2017, doi: 10.1186/s12941-017-0247-4.

[3] H. Zhang and Y. Chen, "Analysis and Application of Grey-Markov Chain Model in Tax Forecasting," *J. Math.*, 2021, doi: 10.1155/2021/9918411.

[4] F. Ziel, "Quantile regression for the qualifying match of GEFCom2017 probabilistic load forecasting," *Int. J. Forecast.*, 2019, doi: 10.1016/j.ijforecast.2018.07.004.

[5] G. Dudek, "Generalized regression neural network for forecasting time series with multiple seasonal cycles," *Adv. Intell. Syst. Comput.*, 2015, doi: 10.1007/978-3-319-11310-4_73.

[6] F. Jiang, Z. Zhao, and X. Shao, "Modelling the COVID-19 infection trajectory: A piecewise linear quantile trend model*," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 2022, doi: 10.1111/rssb.12453.

[7] Suhartono Suhartono, "A Comparative Study Of Forecasting Models For Trend And Seasonal Time Series Does Complex Model Always Yield Better Forecast Than Simple Models ," *J. Tek. Ind.*, 2005.

[8] M. Stokes, P. van Leeuwen, and J. Ozanne-Smith, "The use of injury surveillance databases to identify emerging injury hazards.," *Int. J. Inj. Contr. Saf. Promot.*, 2005, doi: 10.1080/17457300512331342180.

[9] S. Guritno, "a Comparative Study of Forecasting Models for Trend and Seasonal Time Series Does Complex Model Always Yield Better Forecast Than Simple Models," *J. Tek. Ind.*, 2005.

[10] P. H. Hsieh, "A Study of Models for Forecasting E-Commerce Sales During a Price War in the Medical Product Industry," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019. doi: 10.1007/978-3-030-22335-9_1.

[11] C. G. Homer, G. Xian, C. L. Aldridge, D. K. Meyer, T. R. Loveland, and M. S. O'Donnell, "Forecasting sagebrush ecosystem components and greater sage-grouse habitat for 2050: Learning from past climate patterns and Landsat imagery to predict the future," *Ecol. Indic.*, 2015, doi: 10.1016/j.ecolind.2015.03.002.

CHAPTER 8

UNLOCKING THE POWER OF MULTIPLE REGRESSION: A COMPREHENSIVE APPROACH TO FORECASTING COMPLEX RELATIONSHIPS AND TRENDS

Dr Jaykumar Padmanabhan, Assistant Professor
Department of Decision Sciences, CMS Business School, JAIN Deemed to-be University, Bangalore, India
Email Id- p.jaykumar@cms.ac.in

ABSTRACT:

Multiple Regression is a statistical method that enables the identification of relationships between a dependent variable and multiple independent variables. In the context of forecasting, Multiple Regression can be used to create a predictive model that can estimate the value of a dependent variable based on the values of several independent variables. Multiple Regression is an effective tool for forecasting because it can account for multiple variables and identify significant predictors. This technique can provide organizations with valuable insights into future trends, enabling them to make informed decisions and plan accordingly.

KEYWORDS:

Dependent Variable, Forecasting, Independent Variables, Multiple Regression, Predictive Model, Significant Predictors.

INTRODUCTION

The use of PDLs is usually quite sensitive to the choice of variables in the equation. That makes testing the equation more complex. In general, you should include what are expected to be the relevant variables in the equation, rather It is possible that these outliers are "harmless," which means the parameter estimates of the dependent variables will be about the same whether the outliers receive special treatment or not. In that case, adding a dummy variable simply improves the sample period fit without improving forecast accuracy. In such cases, the treatment of these outliers is irrelevant, and they can be ignored. However, that is usually not the case. It is more likely that the disturbances causing these outlying values also affect the independent variables; several examples are provided next.

Before turning to economic relationships, we look at the statistical distortion that can occur from a purely random observation by choosing an artificial example where bad data are introduced into the sample observations. To illustrate how outliers and dummy variables can affect an equation, consider the equation in which the percentage change in consumption is a function of the percentage change in income and the yield spread with an average lag of half a year. The sample data are then altered by introducing one period of "bad" data for both consumption and income, and the regressions are recalculated with and without dummy variables to offset these outlying observations. The bad data used here are constructed by adding 100% to the actual change; e.g., if income rose 5% in that year, the bad data would show an increase of 105% [1].

Not only is the R 2 close to zero and the coefficients insignificant, but their values are far different and, in one case, the value of the yield spread variable switches signs. If the error occurs in the independent variable and a dummy variable is added, the equation in this case is unchanged. If the error occurs in the dependent variable and a dummy variable is added, the resulting R 2 is far overstated, although the parameter estimates do not change. With an error of this magnitude, though, it is clear that some adjustment must be made: either a dummy variable must be added or the erroneous observations must be removed from the sample. If dummy variables are used, it is usually a good idea to recalculate the regression without those data points as a cross-check to make sure the equation has not changed very much. That will also provide a better estimate of the underlying value of R 2 for this regression the outlier occurs in the dependent variable, but that value is uncorrelated with any of the independent variables, regressions.

One possibility is simply to omit all those observations from the sample period. However, that is not always advisable. Suppose the independent variable in question is used in a 20-quarter distributed lag; then 20 observations would have to be omitted for each missing data point. Sometimes data are available in slightly different form and can be combined spliced together. In other cases, quarterly data can be interpolated from annual series, or monthly data from quarterly series [2]. Naturally there is some risk in making up the data with a certain hypothesis already in mind, and then finding that the data support that hypothesis. All researchers would always prefer to have complete data sets. But when that is not possible, what are the “second best” alternatives?

The problem with omitting observations from the sample period is intensified when long lag structures are used. Suppose one of the independent variables enters the equation with a distributed lag of 20 quarters, and the entire sample period is only 80 quarters. The sample period has already been reduced from 80 to 60 observations to accommodate this lag. To lose another 20 observations just because of one single missing data point could reduce the sample size to the range where the results are inconclusive. Monthly and quarterly data are available for auto sales starting in 1959, domestic light truck sales starting in 1966, and foreign light truck sales starting in 1976. Before that, only annual data are available. One option is to start the equation in 1976; however, that reduces the sample period almost by half. Since light truck sales, and particularly foreign light truck sales, were not very important in the early years (presumably one of the reasons that monthly.

DISCUSSION

Simple interpolation is acceptable if the missing point appears in a series with a strong trend and a small variance around that trend. If there is no trend and the observations appear to be serially uncorrelated, one could simply use the mean value of the variable. That might be the case for percentage changes, for example. If one is taking deviations around the trend, the missing value would then be 0. Suppose one series (e.g., foreign truck sales) are available only on an annual basis for part of the sample period, while a closely related series (domestic truck sales) are available monthly or quarterly. Then one can interpolate monthly or quarterly series for foreign truck sales based on the annual data for that series and monthly and quarterly figures for domestic sales [3]. Calculate a regression relating the data series with the missing observation to other variables during the sub-sample period when all the data are available; then use the “predicted” value for missing observations. That is equivalent to estimating the equation by omitting sample points where data are missing if none of the variables is lagged; but where lagged values are used, the size of the sample period can be expanded.

In these examples it is assumed that monthly and quarterly data are seasonally adjusted. If they are not, the appropriate seasonal factor should be added to that month or quarter when estimating the missing data. To a certain extent, it is not known how well these methods work, because by definition the missing data do not exist (although experiments can be constructed where one "pretends" not to have some of the observations). In general, though, (3) and (4) usually work fairly well where the relationships fit well during the periods when all the data are available. Conversely, assuming the value is equal to some sample period average generally does not work well and should be tried only as a last resort [4].

It is a truism, yet one that can be repeated often, that the estimated model is only as good as the underlying data.

While examples of inaccurate data abound, it is useful to group them into the following general classifications:

a) Outright errors. These are input errors, or changes in definitions that are not properly reflected in the published data.
b) Data revisions. This is often a serious problem for macroeconomic data. The corrections are usually due either to (i) missing data in preliminary releases, or (ii) change in sample survey techniques. In some cases, such as the inclusion of business purchases of software in capital spending, the entire concept of the term is changed.
c) Restatement of profits or other company information. To a certain extent, some of this is due to mergers, acquisitions, or divestitures, but the main problem is retroactive write-downs.
d) Fraud. In the international arena, intentional fraud may occur when the government wants to make the country's output appear better than was actually the case (primarily, but far from exclusively, the case in former communist regimes).
e) Defects in survey method or obsolete surveys. For example, the CPI weights could be based on the market basket consumers bought almost a decade ago rather than what they buy today. If one is focusing primarily on the prices of apples and oranges, it probably doesn't matter. If the area of interest is CD-ROMs and Internet access, it probably does.
f) Lack of understanding of how to collect the underlying data. This is probably more often the case for LDCs, although occasionally a new series even for the US will have to be completely revised when the underlying process is understood more thoroughly.
g) Changes in growth patterns due to rebasing. In most cases, the growth rate will be reduced by moving to a more recent base year. However, in the case of computers, where the deflator declines, the opposite is true.
h) Seasonal data. Seasonal patterns do change over time, but no method of adjusting for seasonal data is perfect, and sometimes these methods distort the underlying data. Also, when seasonal factors are revised, monthly or quarterly changes are often far different in the revised data.

i) Reclassification of companies from one industry to another often distorts industry data. This is particularly severe in the case of conglomerates, where small percentage change may shift the entire company from industry A to industry B.
j) Consumer misreporting. For individual consumer data, individual income, assets, spending patterns, or saving might be misstated or misreported. In particular, people might understate their income because some of it was not reported to the Inland Revenue.
k) There are no "textbook" answers about how to determine whether the desired data series are reliable, and no exhaustive list that will include all possible data deficiencies. The above list, however, covers most of the areas where data problems occur.

In general, a dummy variable takes a value of 1 during designated periods and 0 elsewhere. For a series with constant seasonal factors, the seasonal dummies for quarterly data are 1 in the ith month or quarter and 0 otherwise. Sometimes, however, dummy variables either take on a variety of values, or they are combined with other terms. Thus in addition to seasonal dummies, one can distinguish among the following principal types of dummy variables. Examples are provided for each of these cases single or isolated event changes: wars, energy crises, strikes, weather aberrations changes in institutional structure: floating of the dollar, deregulation of the banking sector changes in slope coefficients: variable becomes more or less important over time • nonlinear effects: big changes are proportionately more important than small changes

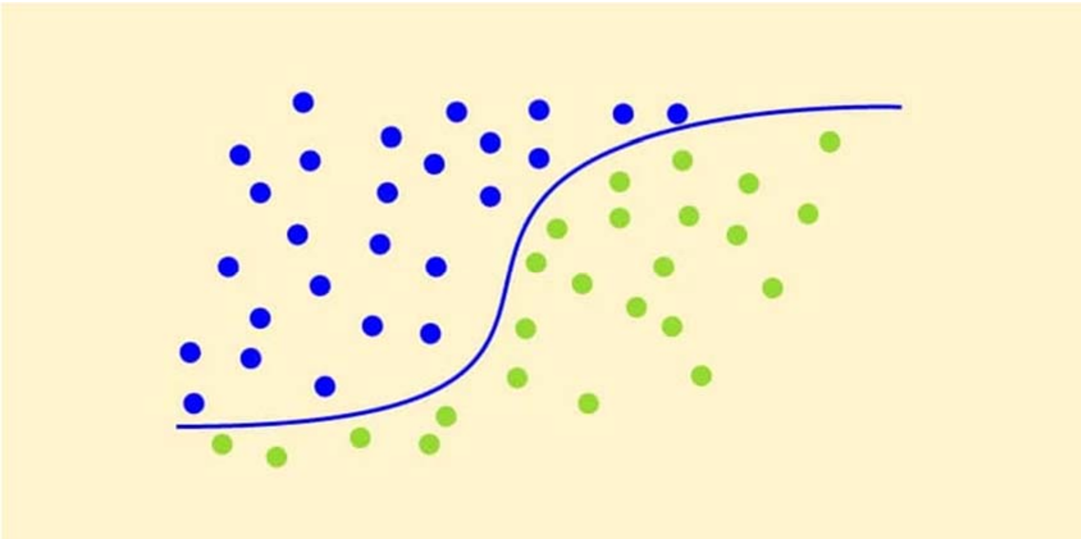


Figure 1: Illustrate the Multiple Regression.

As already noted, it is a simple matter to boost the sample-period goodnessof-fit statistics with dummy variables without improving forecasting accuracy. Dummy variables are often properly introduced to reflect institutional changes: deregulation of a particular sector or industry. Compare, for example, the behaviour of the airline, trucking, banking, or stock market sectors before and after deregulation [5].

Another type of far-reaching change could occur because of changed expectations. For example, when the Fed did not have a credible monetary policy, declines in the unemployment rate were widely thought to presage higher inflation. However, once credibility was reestablished, the tradeoff between unemployment and inflation disappeared. Company data for sales, orders, profits, etc., would obviously change if the company acquired another entity, or divested some of its divisions. For accounting and investment purposes, earnings per share can be restated so they are comparable, but any time-series data for total sales would show major shifts. Figure 1 illustrate the Multiple Regression.

In terms of econometric application, one of the primary issues is whether the dummy variable should be applied to the constant term of the equation, to some of the slope coefficients, or possibly to the entire equation. In the latter case, the same functional form can be estimated for two or more sub periods. It was already shown in the previous section that erroneous data can, under extreme circumstances, seriously distort the parameter estimates. That case was exaggerated to emphasize the point, which is that the significant criterion is whether the dummy variable is correlated with the other independent variables. If the outlier is due to a truly random event, then omitting a dummy variable will reduce the sample period fit but leave the parameter estimates unchanged. However, if the dummy variable is correlated with the other independent variables, then omitting it will bias the other coefficients. Consider the case of an auto strike. Consumers buy fewer cars because they realize there will be fewer choices in the showroom, so they may not be able to find their preferred make and model, and may also receive a smaller discount. If the strike is lengthy, that will not only affect the auto industry but the economy in general: disposable income will fall and, although striking workers are not counted as unemployed, the unemployment rate will rise as other workers are laid off. Without the use of a dummy variable for strikes, the coefficients for income and the unemployment rate would probably be overstated [6].

However, that is not the end of the story. The auto strike delayed sales, but it probably did not cancel them. The loss of sales during the strike period is generally made up in the following period. Thus the appropriate value of the dummy variable would be -1 during the strike and +1 in the next period. In other industries, such as the steel industry, if the strike were anticipated, the dummy variable might be +1 in the period before the strike and -1 during the strike. If the strike lasted longer than expected, the dummy variable might have the values +1, -2, and +1. Only if there were a permanent loss of sales would the values of the dummy variable sum to less than zero.

The same argument can be made for dock strikes: exports and imports both rise before the strike occurs, decline during that period, surge the next period, and then return to trend levels. In that case, the values of the dummy variable would also sum to zero over the period before, during, and after the strike. In some cases, the sum of the values of the dummy variable might be greater than zero. Suppose a hurricane devastates coastal areas. In the period before the hurricane, construction is at normal levels. After the storm ends, construction expenditures rise sharply for a while. If the rebuilding phase lasts several periods but gradually tapers off, the dummy variable might take the values 4, 3, 2, and 1 in the four periods following the storm. On balance, though, construction activity over the entire period will be higher than if no storm had occurred.

Sometimes the interaction is more complicated. The Nixon Administration imposed a wage and price freeze on August 15, 1971, that lasted for 90 days. That was followed by Phase II of controls, lasting through the end of 1972, during which wages and prices could rise by only a certain percentage determined by the government. During Phase III, which started on January 1, 1973, prices could be raised only by the amount that costs increased. Controls were ended on May 1, 1974, at which point prices briefly rose by record amounts [7].

It might seem clear that a dummy variable that was negative during controls and then positive for a while would be appropriate. In some commodity price equations the dummy variable is important, as will be shown later. However, when the macroeconomic inflation rate is correlated as a function of labour costs, money supply, and oil prices, the residuals do not show any such pattern. A wide variety of dummy variables this author tried are not significant. In this case the reason is not obvious. The inflation rate is, and should be, negatively correlated with productivity growth. During the period of controls, many firms had an incentive to understate the rise in prices and hence overstate the rise in output, since the current-dollar numbers could not be easily fudged. As a result, reported productivity growth soared to 3.5% during the period of controls and then declined to -1.6% when they were removed. It is quite unlikely that such a pattern actually occurred.

Yet both theory and empirical evidence suggest that productivity rises faster when inflation is lower, because capital goods are then purchased in order to earn their highest real rate of return, rather than being purchased as a hedge against inflation. Hence the strong negative correlation between productivity and inflation is theoretically as well as empirically robust. However, because of faulty data during that period, the correlation may be overstated. The use of a dummy variable should reduce that parameter estimate – if we had accurate data; but it is not available. In this particular example, a dummy variable is theoretically reasonable, but is not empirically significant [8].

This is perhaps an extreme example, but it illustrates how the use of dummy variables depends in large part on the correlation between the dummy variable and other independent variables. When they are correlated, it is good statistical practise to include a dummy variable: in that case, when used within reason, it is not just merely curve fitting or ad hoc adjustment. Over the past 25 years or so the US economy has undergone many structural changes involving deregulation. One of the most important was the deregulation of the banking sector in the early 1980s. Before then, growth in the money supply (M2) closely followed changes in the monetary base required reserves plus currency which could be closely controlled by the Fed. Since then, there has been no correlation between percentage changes in the monetary base and M2. Thus in estimating an equation for the percentage changes in M2, it is best to multiply the percentage changes in the monetary base by DBR – a dummy variable for changes in banking regulations which is 0 through 1980. 3 and 0 afterward. A comparison of these two series shows that, starting in late 1980, money supply growth accelerated at the same time that the growth rate in the monetary base decreased. After deregulation, changes in the money supply were more closely correlated to loan demand than to the monetary base, so the changes in business loans are multiplied by (1 - DBR). Also, the spread between the Federal funds and the discount rate, while still a significant determinant of changes in the money supply, is much less important after 1980 than before. In a previous section of this chapter, I went over how to estimate the bivariate relationship between consumption and income using levels, logs, first differences, and percentage changes. This case study illustrates the results of extending the consumption/saving function to take into account monetary variables, expectational variables, such as inflation and unemployment, and other significant economic factors.

Consumption is dependent on some metric of average or expected income, not just current income, according to the modern theory of the consumption function. A weighted average of past income could be used to calculate expected income, but other variables that reflect consumer wealth, such as stock and home prices, can also calculate expected income. Interest rates are crucial. Lower interest rates also mean greater access to credit and the ability to refinance home mortgages at lower rates, making them a more significant determinant of consumption than the cost of borrowing. These are sometimes included in an index of consumer attitudes, which is then entered as a separate variable. Attitudinal variables, such as the rate of unemployment and inflation, are also significant. This strategy is not employed in this case because, as will be discussed further in the text, changes in consumer attitudes that are unrelated to inflation, unemployment, and stock prices do not seem to be correlated with consumer spending.

In light of this, the theoretical function states that consumption is a function of current and lag income, the corporate bond rate at Aaa, the stock market index at S&P 500, the unemployment rate, the change in oil prices, and the relative price of homes. We are now going to think about the lag structure for each of these variables given that function. Those lags should be brief because changes in inflation and unemployment are expectational variables that affect timing rather than overall level of purchases. The lags for income, bond yield, stock prices, and home prices, on the other hand, could be significant based on theoretical considerations. Therefore, the first pass employs 12 quadrant PDLs, cubic polynomials, and a constraint at the far end; 12,3,2 in the EViews formula [9].

These findings are useless, though. All the signs, with the exception of income, quickly change as the lags lengthen. Further testing, which the user is encouraged to do, reveals that the signs flip-flop even as the lag structure is compressed. In the end, there is only a one-quarter lag in the ideal structure for the bond yield, stock price index, and relative price of homes. Even the income term's coefficients experience a sharp decline, followed by a longer lag before they recover. The lag structures for durable goods and services are thus different.

After determining that there is a long lag for income and short lags for all other variables, we then analyse these results using a variety of alternative formulations developed in this chapter, including OLS, WLS, logs, percentage change, ratio, and deviations from logarithmic trends. Only the summary t-statistics are presented here because the methodology has already been covered in some detail earlier in this chapter. The AR(1) transformation is used to make adjustments to all equations other than the one that uses percentage changes. Without that adjustment, OLS and WLS were identical in every way, so WLS is not listed because it is not applicable with the AR(1) transformation. The dependent variable in the ratio equation is the ratio of consumption to disposable income; in that equation, the percentage change in income correctly has a negative sign, reflecting the fact that consumption adjusts to income with a lag [10].

These findings are fairly illuminating. Starting with the levels function, it is simple to conclude that the relative home price variable is significantly less significant than the stock market variable.

The ratio function is no different in this regard. Relative home prices barely matter in the logarithmic function. Home prices are also not significant because random changes in the percentage change equation overwhelm the underlying function. Given that this is a longer-term effect, that might be anticipated. The fit is not improved by lengthening the lag in the percentage change equation, though, and the t-ratio for the change in stock prices becomes negative [11].

CONCLUSION

Multiple Regression is a powerful statistical technique that enables the identification of relationships between a dependent variable and multiple independent variables. This makes it a valuable tool for forecasting, as it allows for the consideration of multiple factors that may impact the outcome variable. By using Multiple Regression to create a predictive model, organizations can gain insights into future trends and make informed decisions based on those insights. The technique can identify significant predictors, quantify their effects, and help identify any interactions between variables.

REFERENCES

[1] D. Pavlyuk, "Feature selection and extraction in spatiotemporal traffic forecasting: a systematic literature review," *Eur. Transp. Res. Rev.*, 2019, doi: 10.1186/s12544-019-0345-9.

[2] R. Budhiraja, M. Kumar, M. K. Das, A. S. Bafila, and S. Singh, "A reservoir computing approach for forecasting and regenerating both dynamical and time-delay controlled financial system behavior," *PLoS One*, 2021, doi: 10.1371/journal.pone.0246737.

[3] W. Cao, W. Zhu, and Y. Demazeau, "Multi-Layer Coupled Hidden Markov Model for Cross-Market Behavior Analysis and Trend Forecasting," *IEEE Access*, 2019, doi: 10.1109/ACCESS.2019.2950437.

[4] Y. Wang, Z. Duan, Y. Huang, H. Xu, J. Feng, and A. Ren, "MTHetGNN: A heterogeneous graph embedding framework for multivariate time series forecasting," *Pattern Recognit. Lett.*, 2022, doi: 10.1016/j.patrec.2021.12.008.

[5] J. Jia, A. Zhao, and S. Guan, "Forecasting based on high-order fuzzy-fluctuation trends and particle swarm optimization machine learning," *Symmetry (Basel).*, 2017, doi: 10.3390/sym9070124.

[6] D. Maaß, M. Spruit, and P. de Waal, "Improving short-term demand forecasting for short-lifecycle consumer products with data mining techniques," *Decis. Anal.*, 2014, doi: 10.1186/2193-8636-1-4.

[7] A. K. Ozcanli, F. Yaprakdal, and M. Baysal, "Deep learning methods and applications for electrical power systems: A comprehensive review," *International Journal of Energy Research*. 2020. doi: 10.1002/er.5331.

[8] G. Marsden and N. C. McDonald, "Institutional issues in planning for more uncertain futures," *Transportation (Amst).*, 2019, doi: 10.1007/s11116-017-9805-z.

[9] W. Sulandari, Subanar, Suhartono, H. Utami, M. H. Lee, and P. C. Rodrigues, "Ssa-based

hybrid forecasting models and applications," *Bull. Electr. Eng. Informatics*, 2020, doi: 10.11591/eei.v9i5.1950.

[10] B. Chen and W. Li, "Multitime Resolution Hierarchical Attention-Based Recurrent Highway Networks for Taxi Demand Prediction," *Math. Probl. Eng.*, 2020, doi: 10.1155/2020/4173094.

[11] C. Lee, O. Kwon, M. Kim, and D. Kwon, "Early identification of emerging technologies: A machine learning approach using multiple patent indicators," *Technol. Forecast. Soc. Change*, 2018, doi: 10.1016/j.techfore.2017.10.002.

CHAPTER 9

EXPLORING ADVANCED METHODOLOGIES FOR ACCURATE AND RELIABLE FORECASTING

Dr Navaneetha Kumar, Professor
Department of Decision Sciences, CMS Business School, JAIN Deemed to-be University, Bangalore, India
Email Id- dr.navaneethakumar@cms.ac.in

ABSTRACT:

Advanced regression methodologies have gained significant attention in forecasting due to their ability to handle complex relationships between variables and capture non-linear patterns in the data. These methodologies include but are not limited to machine learning techniques such as random forests, neural networks, and support vector machines, as well as Bayesian regression models and quintile regression.

KEYWORDS:

Advanced Regression Methodologies, Forecasting, Machine Learning, Random Forests, Neural Networks, Support Vector Machines.

INTRODUCTION

The t-ratio falls sharply, but stock prices continue to be significant, and relative home prices become significantly more significant. The deviation-from-trend equation fits the 1980s and 1990s much better than the levels equation does, according to an analysis of the residuals from the levels and deviations-from-trend equations, even though both fail to track the detrimental effect of the first oil shock on consumption. The deviations-from-trend equation seems to be the most accurate forecasting equation in the sense that future oil shocks are much less likely to interfere with consumer purchasing patterns.

This illustration demonstrates that while it is plausible to assume that consumer spending patterns are based on long lags, this does not prove to be the case empirically. That finding by itself does not disprove the hypothesis that consumption is more strongly influenced by long-run average or expected income than by current income. It instead emphasizes a number of other outcomes. The availability of credit is the first factor that influences when many consumer purchases are made [1].

Second, lag changes in real disposable income are less closely related to expected income than they are to factors like current stock prices and relative home prices. Long distributed lags therefore do not perform well in this equation. Later in this text, we will reevaluate how well different consumption functions can predict future behavior, both with and without attitude variables. Additionally, this case study makes the suggestion that even though polynomial distributed lags are frequently helpful, you shouldn't use them if they don't work. Long lags are significant determinants for functions relating to bond yields, capital spending, and the Federal funds rate. Although the lags are much shorter, using PDLs to create accurate forecasting equations for consumer spending is not advisable, at least for discretionary purchases. According to conventional microeconomic theory, in an equilibrium, the marginal product of capital equals the cost of capital and the marginal product of labor equals the wage rate.

While the wage rate is clearly defined in terms of dollars per hour, the cost of capital has a time component because capital goods typically have a long lifespan. The cost of equity capital (stock prices), the rate of depreciation, and tax laws intended to affect capital spending (such as the accounting rate of depreciation and the rate of investment tax credit) must all be taken into account in addition to the cost of the capital good. The corporate income tax rate is also a factor to be taken into account. Additionally, the value of the marginal product of capital is influenced by how much the product costs in relation to the cost of the capital good. The average product of capital, which is output divided by capital stock, is proportional to the marginal product of capital if the firm is operating on the constant portion of its cost curve. According to this supposition, the optimal capital stock would be inversely related to the various components of the cost of capital and positively correlated with output [2].

For a number of reasons, this theory needs to be modified. First off, capital is "lumpy" in comparison to labor; one cannot buy a machine in half or a building in thirds. Second, the construction of a new machine or structure requires time. Third, there's the possibility that some businesses have excess capacity. In this case, even if output rises or capital costs fall, businesses might decide against investing in new machinery and equipment because the current stock is sufficient. Fourth, all the discussion so far has focused on net investment; even if output hasn't increased, businesses may replace worn-out machinery and other assets as it ages.

Empirical estimation of investment functions is challenging as a result of these modifications. The possibility of different lag structures for each of the variables adds another challenge, which is directly relevant to the discussion in this chapter. To ascertain the lag structure, PDLs are employed. Strong common trends may also cause multicollinearity issues, which recommends employing one of the techniques for trend eradication.

Different lag structures might exist for each of the independent variables, including GDP minus capital equipment, stock prices, the real bond rate, the impact of tax regulations on investment, the cost of capital equipment relative to the GDP deflator, and the relative price of oil. The latter term is included separately because the energy sector requires more capital than the rest of the economy, increasing the ratio of capital spending to GDP as relative oil prices rise. Both in the late 1970s and the first half of the 1980s, this occurred; when oil prices then dropped, the ratio of capital spending to GDP decreased for a number of years [3].

Given that firms factor in both output and the cost of capital when determining their capital spending plans, one might initially anticipate that the lag structures for all the terms would be similar. The situation is different, though. In comparison to the cost of capital, the lags for output are much shorter. There are numerous justifications for this that are outside the purview of this discussion, but the main point can be summed up as follows. Think about two capital goods: one has a lifespan of three years (personal computers, cars), and the other has a lifespan of twenty years (electrical generating equipment, jet aircraft). The cost of capital becomes more significant as a product's economic life lengthens.

DISCUSSION

Consumers can apply the same analogy: the interest rate is much more critical when purchasing a home than when purchasing a computer. Since a computer's lifespan is relatively brief, the decision to buy one will be primarily based on the output level in the most recent period. As a result, the demand for both long-lasting equipment, where the cost of capital with a long lag is important, and short-lasting equipment, where output with a short lag is important, can be seen as part of the capital equipment function. Indeed, that is what we discover, but we still need to make a change. The investment tax credit was introduced in 1962, increased in 1964, suspended in 1966, resumed in 1967, suspended in 1969, resumed in 1971, increased in 1975, expanded in 1981, and modified in 1982 before being discontinued in 1986. Throughout its existence, it was frequently used as a short-term policy variable to stimulate or deter the purchase of capital equipment. Nearly as frequently, adjustments for depreciation were also made. This user cost term element has a shorter lag because it tended to influence the timing rather than the amount of capital expenditure.

Last but not least, keep in mind that the stock market variable has two functions. Because it acts as a stand-in for expected output, it is significant with a short lag. With a longer lag, it is also significant because it measures the equity cost of capital. As a result, the weights begin quite high, drop almost to zero, and then increase once more. The l PDL usage is more of an art than a science. However, I provide the following fundamental principles for figuring out the lag structure [4]. Make an educated guess at the lag structure's approximate length first; if the guess is off, it probably won't affect the outcome, but it will take more time. Since we are aware that decisions regarding capital expenditures are based on data from several years, a reasonable starting point could be 3–4 years.

A cubic polynomial constrained at the far end is typically the default choice for PDLs, which means that the coefficients are assumed to be zero beyond the maximum length of lag specified. Although it is also possible to constrain the value of the coefficients to zero at the close end of the lag structure, this method is rarely employed as a first estimate. The code for variable R (let's say) with this lag structure in EViews would be written as PDL. R could represent a level or percentage change and it could also have a lag at the beginning. The EViews code would be to calculate the percentage change of R starting with a three-quarter lag. Typically, the initial outcome won't be satisfactory. There will be terms with the incorrect signs and insignificant terms in the lag distribution. A quadratic equation would be sufficient in some circumstances where the cubic term may not be significant.

Generally, it is preferable to shorten the lag structure until all remaining terms meet this requirement if the PDL equation's individual terms have a t-ratio of less than 2. It makes sense to lengthen the lag structure if the terms at the tail of the lag distribution seem to be highly significant as long as they continue to be so. At the start of the distribution, terms can occasionally be close to zero. This suggests a longer lag when the variable is first introduced. A zero value constraint for the coefficients at the close end of the distribution is occasionally also recommended [5]. Figure 1 illustrate the Comparison of basic and advanced regression forecasting methods.

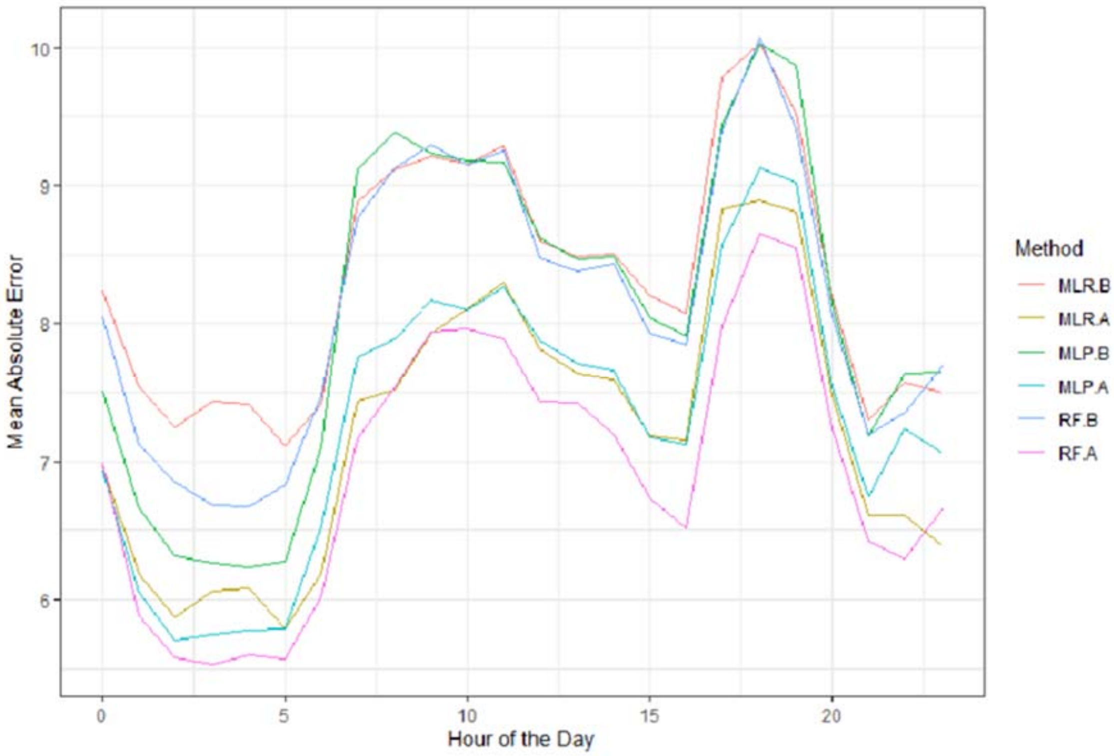


Figure 1: Illustrate the Comparison of basic and advanced regression forecasting methods.

They represent the "simple" cases. Sometimes a variable will start out being significant, then lose its significance or change its sign, before becoming significant once more. Other times, the variable will switch signs, with both significant positive and negative values. Should we keep or throw away these findings? But occasionally they are economically sensible. Each type is represented by one example in the investment function. The dual purpose of the stock market variable causes its coefficients to be highly significant, then decline, then rise once more. The structure is preserved even when the coefficients become negative because the negative terms't-ratios are never less than 1.0, indicating that they are not at all significant.

The other peculiar case is the value of the relative oil price coefficients, which change from negative to positive. In this case, economic theory can support the lag distribution. For example, when oil prices increase, the initial effect is to reduce real growth and profitability for most firms, which leads to investment reduction. The positive impact eventually outweighs the negative impact, however, as higher oil prices encourage investment in the energy sector [6].

For example, if the weights were assumed to decline monotonically, that assumption could be entered into the regression without using a PDL. PDLs are not advised unless the lag structure is assumed to be fairly long and complex. Without using PDLs, it is challenging to estimate an equation that yields reliable predictions for the investment function. Several additional observations on this equation seem pertinent at this time. In particular, we note that the t-ratio of the RCCE (-3) term is only 1.66, and the DW is an uncomfortably low 0.92, despite the apparent high R 2, which is fairly typical for time series with strong trends. The following addresses both of these points.

It would be tedious to repeat all of this information for each of the six forms of the equation listed above; instead, readers who are interested can estimate these equations independently. An overview of the findings is provided. The RCCE term is much more significant (t-ratio of 6.1) in the logarithm equation than the SP500 term, which disappears entirely. The capital stock, RCCE, RAAA, and RPOIL terms all disappear from the percentage change equation because random fluctuations dominate it, making it of relatively little use. All of the variables' significance is restored when using the capital spending to GDP ratio as the dependent variable, but the SP500 term is now only significant with a brief lag. In the equation for deviations from the trend, every term is significant, and the RPOIL term almost disappears. The results compared to the levels equation hardly change when using weighted least squares [7].

The Durbin-Watson statistic reveals significant autocorrelation in all equations except the percentage change equation, where nearly all the terms are dropped. The majority of the terms don't change much when the AR (1) adjustment is applied, with the exception of the relative price of oil, which becomes negligible. The issue of whether to remove it from the equation's final form is brought up by this. Although WLS barely makes a difference, it is not further considered, it is not immediately clear from these comments which of the five forms of the equation will produce the best forecast. Due to the variance's apparent high proportion of random fluctuations, the percentage change form can be discounted.

The logarithm equation has too large a coefficient for capital stock and too small a coefficient for stock prices relative to a-priori expectations. The user cost of capital term is not significant in the levels equation, which does not seem to be a reasonable result. Hence the choice is between the ratio and the deviations-from-trend. Further experimentation reveals that when the ratio equation is "tidied up" to remove PDL coefficients that are significant or have the wrong lag, several other terms become insignificant, so the final choice is the deviation-from-trend equation. Since DW is quite low checking the structural stability of individual equations is one of the key steps in model building. Unfortunately, it is sometimes ignored, because the results are not so easy to fix. Autocorrelation, heteroscedasticity, or multicollinearity can generally be reduced by straightforward and relatively simple adjustments. However, if tests show that the equation is unstable, that often means starting all over again with a different specification.

It is often difficult to find a reasonable equation that meets all the statistical tests, including the stability of coefficients. Of course, even accomplishing this goal provides no guarantee that the forecasts will be accurate. However, if the equation is unstable, mis-specified, or omits relevant variables, that almost guarantees the forecasts will not be accurate. Thus, satisfying the tests discussed in this chapter is a necessary condition for successful forecasting, even though it ais not sufficient. We first discuss further tests for the residuals of the equation, checking for normality, autocorrelation, and heteroscedasticity, then turn to some of the methods that allow the researcher to check for the stability of the equation.

These tests are then applied to a quarterly equation for new motor vehicles. An equation for housing starts is used to illustrate various alternative methods for adjusting the constant term. Finally, an equation for the dollar yen cross rate is used to illustrate some of the pitfalls of multi-period forecasting condition holds, the R2 and t-statistics are likely to overstate the actual goodness of fit.

One of the useful characteristics of the normal distribution is that it is completely described by the mean and variance. However, that is not true for other distributions. Hence if the residuals are not normally distributed, higher-order moments, particularly scenes and kurtosis, might be significantly different from the normal distribution. That is the basis of the Jarque–Bera test, previously mentioned when histograms were discussed. To review briefly, that test statistic is given as: where S is skewness, K is kurtosis, T is the number of observations, and k is the number of variables [8].

The probability and significance levels of JB are included in EViews, but it ais usually obvious whether or not the residuals are normally distributed by looking at the histogram. If more than one outlier are three or more standard deviations from the mean, those observations are presumably not drawn from the same population as the rest of the sample period data. You have to decide whether that unusual situation is a one-time event that will not recur, in which case it probably will not affect the forecast, or whether it should be treated by adding another variable, using dummy variables, omitting the outliers, or including some nonlinear transformation.

In particular, the residuals could still be serially correlated, as is often the case for economic time series. Since this is one of the most common results for regressions estimated with time-series data, further tests are often warranted. In some cases, the residuals of the estimated equations will be normally distributed, with no autocorrelation and no heteroscedasticity, yet the equations themselves will generate poor forecasts. In many cases, this occurs because there has been a shift in the parameters of the underlying structural equation. In the worst possible case, the structure remained constant during the entire sample period but then radically shifted just as the forecast period started. There is no secure for that disease, and when it does happen, the forecasts will be inaccurate. However, that is a fairly unusual circumstance. Usually, any shift in the structure can be detected during the sample period. Under those circumstances, a battery of standard tests can be used to uncover this shift, permitting the model builder to adjust the forecasts accordingly.

The Chow test6 was one of the first regression diagnostic tests to be developed, and is still one of the most important. The idea is quite straightforward. Divide the sample period into two (or more) sub-periods. Calculate the regression during all of these periods, and then calculate the regression separately for individual sub-periods. The Chow test then compares the sum of the squared residuals obtained by fitting a single equation over the entire sample period with the residuals obtained by estimating separate equations over each subsample period. If the residuals are significantly different, the coefficients have probably shifted, which increases the probability the equation is unstable and the co-efficients will shift again in the forecast period, generating poor forecasts. Econometricians realise that time-series regressions should be tested by omitting some of the data points.

The Chow test provides just such a test. However, since the choice of break point chosen by the researcher is somewhat arbitrary, it would be better to examine the equation over time as sample points are added one by one. In the past, that used to represent a tremendous amount of regression time on the computer, and was seldom done. However, EViews provides the algorithms to combine these results and shows them on a single graph for each test, thus making the comparison and analysis much easier to absorb and analyze.

The EViews programme performs six separate tests in this area; while some are a more important than others, taken together they provide a comprehensive analysis of how the parameters change over time and whether the equation can reasonably be expected to remain stable over a reasonably long forecast period. The programme calculates values of the residuals in the period immediately after the end of the truncated sample period. For example, the equation would be estimated through 1975.4 and those coefficients would then be used to predict a1976.1. If that estimate falls outside two standard errors as also calculated by the programmed), the test suggests the coefficients are unstable. This process is then repeated until the end of the full sample period [9].

The acronym stands for cumulative SUM of the residuals. In this test, the cumulative sum of the residuals are plotted; hence the bands widen over time. If the equation generates one-period errors that are random, will show the cumulative residuals remaining within their error bands. On the other hand, if the errors are cumulative (as would likely be the case if one were to use the lagged dependent variable on the right-hand side of the equation), they would increasingly lie outside the 2 as bands as time progresses have been a structural shift in recent periods that is noticeable in the residuals but has not occurred for a long enough time to warrant the inclusion of an additional variable.

This point is now illustrated with a quarterly equation for housing starts. It will be shown that I while the equation adequately tracks all the major cycles in housing starts, the residuals are highly autocorrelated; (ii) the residuals in recent years have all been positive; (iii) adjusting the forecasts by the average size of the recent residual materially improves the forecast: (iv) adding additional variables to try and explain the recent residuals worsens forecast accuracy, and (v) adjusting for autocorrelation using the AR(1) method also materially worsens forecast accuracy. While these results are based on only one equation, this author can state after many years of actual forecasting that results aof this sort are the rule rather than the exception [10], [11].

CONCLUSION

Advanced Regression Methodologies have emerged as a powerful tool for forecasting, particularly in cases where traditional linear regression models fail to capture the complex relationships between variables. Machine learning techniques such as random forests, neural networks, and support vector machines are effective in handling high-dimensional datasets with numerous variables and capturing non-linear patterns in the data. Bayesian regression models allow for the incorporation of expert knowledge into the model, improving forecasting accuracy. Quintile regression provides a more complete picture of the relationship between variables, particularly in cases where the data is skewed or outliers are present.

REFERENCES

[1] M. Wang, M. Rezaie-Balf, S. R. Naganna, and Z. M. Yaseen, "Sourcing CHIRPS precipitation data for streamflow forecasting using intrinsic time-scale decomposition based machine learning models," *Hydrol. Sci. J.*, 2021, doi: 10.1080/02626667.2021.1928138.

[2] S. Cenci, G. Sugihara, and S. Saavedra, "Regularized S-map for inference and forecasting with noisy ecological time series," *Methods Ecol. Evol.*, 2019, doi: 10.1111/2041-210X.13150.

[3] I. E. Livieris, S. Stavroyiannis, E. Pintelas, and P. Pintelas, "A novel validation framework to enhance deep learning models in time-series forecasting," *Neural Comput. Appl.*, 2020, doi: 10.1007/s00521-020-05169-y.

[4] A. Barrientos-Orellana, P. Ballesteros-Pérez, D. Mora-Melia, M. C. González-Cruz, and M. Vanhoucke, "Stability and accuracy of deterministic project duration forecasting methods in earned value management," *Eng. Constr. Archit. Manag.*, 2022, doi: 10.1108/ECAM-12-2020-1045.

[5] E. Meira, F. L. Cyrino Oliveira, and L. M. de Menezes, "Point and interval forecasting of electricity supply via pruned ensembles," *Energy*, 2021, doi: 10.1016/j.energy.2021.121009.

[6] D. S. Kumar, G. M. Yagli, M. Kashyap, and D. Srinivasan, "Solar irradiance resource and forecasting: a comprehensive review," *IET Renewable Power Generation*. 2020. doi: 10.1049/iet-rpg.2019.1227.

[7] H. Alkabbani, A. Ahmadian, Q. Zhu, and A. Elkamel, "Machine Learning and Metaheuristic Methods for Renewable Power Forecasting: A Recent Review," *Front. Chem. Eng.*, 2021, doi: 10.3389/fceng.2021.665415.

[8] M. C. Bueso, J. M. Paredes-Parra, A. Mateo-Aroca, and A. Molina-García, "Sensitive Parameter Analysis for Solar Irradiance Short-Term Forecasting: Application to LoRa-Based Monitoring Technology," *Sensors*, 2022, doi: 10.3390/s22041499.

[9] O. Timur, K. Zor, Ö. Çelik, A. Teke, and T. İbrikçi, "Application of statistical and artificial intelligence techniques for medium-term electrical energy forecasting: A case study for a regional hospital," *J. Sustain. Dev. Energy, Water Environ. Syst.*, 2020, doi: 10.13044/j.sdewes.d7.0306.

[10] N. Mohan, K. P. Soman, and S. Sachin Kumar, "A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model," *Appl. Energy*, 2018, doi: 10.1016/j.apenergy.2018.09.190.

[11] P. M. R. Bento, J. A. N. Pombo, M. R. A. Calado, and S. J. P. S. Mariano, "Stacking ensemble methodology using deep learning and ARIMA models for short-term load forecasting," *Energies*, 2021, doi: 10.3390/en14217378.

CHAPTER 10

BOX-JENKINS METHOD: A POWERFUL TOOL FOR TIME SERIES FORECASTING

Dr L.Sudershan Reddy, Professor
Department of Decision Sciences, CMS Business School, JAIN Deemed to-be University, Bangalore, India
Email Id- sudershan.reddy@cms.ac.in

ABSTRACT:

The Box-Jenkins Method is a time-series analysis technique that is widely used for forecasting future values of a time series. It involves identifying the order of autoregressive integrated moving average (ARIMA) models, estimating their parameters, and checking the adequacy of the model through diagnostic tests. The method is particularly useful when historical data is available, and the time series has a pattern that can be modeled using ARIMA. The Box-Jenkins Method has been successfully applied in industries such as finance, economics, and engineering. However, it has its limitations and assumptions that need to be carefully considered when applying it to real-world data. Overall, the Box-Jenkins Method is a powerful tool for forecasting time series data, and its widespread use is a testament to its effectiveness.

KEYWORDS:

ARIMA Models, Autoregressive Integrated, Box-Jenkins Method, Forecasting, Time-Series Analysis.

INTRODUCTION

The RMS may not be appropriate in all circumstances. In particular, individuals or firms might face an asymmetric loss function: if the predicted values are below the actual values, the firms may lose some sales, whereas if the predicted values are higher than the actual values, the firm may go bankrupt. It is also possible that in certain applications, such as financial markets, predicting the direction in which the particular market will move is just as important as predicting the magnitude of that move.10 In other cases, trend forecasts may be less important than accurate predictions of turning points There are many equations which, if correctly estimated and thoroughly tested, will generate forecasts that are well within the estimates indicated by the SE. In those cases, the econometricians have presumably done their work well, and no further comment is needed. However, as this author can attest, there are often times when an equation that appears to be robust by all standard statistical tests generates very poor forecasts [1].

The practical question is what to do in such situations. It is, of course, possible that the equation has been mis-specified, so the only reasonable solution is to start all over again. In many cases, however, forecast accuracy can be improved by using some of the following tools. First, some econometricians suggest recalculating the equation with an AR (1) transformation. Second, it is often advantageous to adjust the constant terms based on recent residuals.

Third, forecast error is often increased by faulty predictions of the independent variables; in some cases, using consensus forecasts may help. The advantages aand disadvantages of these methods, together with several examples, comprise the remainder of this chapter. It should also be noted that forecast error should be evaluated not only in comparison to the sample period error, but relative to errors generated from so-called naive models, which assume that the level or percentage change in a given variable is the same as last period, or in a somewhat more sophisticated version – that the variable is a function only of its own lagged values and a time trend. Hence the forecasting record of naive models is also considered in these examples.

Case study 1 on page 90 presented an equation for the annual percentage changes in constant-dollar retail sales at hardware and building materials stores for the period from 1967 through 1998. The data series itself has a standard error of 6.6%, so if one assumed that hardware sales would rise at the average amount every year, the average forecast error would be 6.6%. A regression with current levels of the change in disposable income, housing starts, and the unemployment rate, and lagged changes in the Aaa corporate bond rate, explains 89% of the variance and reduces the standard error to 2.2%. It would appear this equation does indeed predict most of the change in the hardware sales [2].

Yet equations of this sort are always subject to multiple sources of possible forecast error. As noted above, the first is the random generation process that ais reflected in the SE of the equation itself. The second test is the possibility that the structure will shift outside the sample period. The third is that the structure may remain unchanged, but forecasts for the unlagged values of the independent variables may be inaccurate.

Two major tests can be applied. The first one is to estimate the equation through a truncated sample period, ending in (say) 1993 and then forecast ahead, using the equation outside the sample period but inserting actual values of the independent variables. The second test is to use the consensus forecasts made each year for the percentage change in income, actual change in housing start, and level of unemployment. Since the bond rate is lagged, it is known with certainty when the forecasts are made and hence does not contribute to any error for one-year forecasts.

SE for the 1994–8 period calculated with an equation estimated through 1993 is only 2.1%, virtually the same as the SE of the fitted residuals if the equation is estimated through 1998. This indicates stability of the equation. Of course this is never a perfect test, as it could be claimed the equation fits so awell because we know what happened in the 1994–8 period and adjusted the sample period equation accordingly. In this particular case, however, the author actually used such an equation to predict hardware sales and can therefore warrant that the structural form of the equation did not change over that period.

DISCUSSION

The next test substitute's consensus forecasts for the actual values of income, housing starts, and the unemployment rate, and then recalculates SE for these five years. As shown in table 5.1, using predicted values for the unlogged independent variables boosted SE from 2.1% to 3.7%. These forecast errors are still smaller than the 5.2% error which would be generated by assuming the percentage change this year is the same as last year, but they are substantially higher than the ex post simulation errors. The error from the naive model is generated by assuming the change this year is the same as the change last year.

As shown in table 5.1, the difference in the RMS between the values when the sample period includes 1994–8 and excludes it is virtually nil. On the other Econometric modelling started in the 1920s and 1930s. Early attempts to build forecasting models centered on the estimation of equations to predict agricultural prices and quantities and stock market prices. The discipline then spread to building models for macroeconomic forecasting purposes. While early models included only 20 to 30 equations, the size rapidly increased, and by the 1970s some macro models contained thousands of individual equations.

While initial expectations for these macroeconomic models were quite optimistic, the quality of forecasts they delivered was disappointing. The nadir was reached in the 1970s and early 1980s, when models were unable to predict the simultaneous occurrence of double-digit inflation and unemployment. In particular, these models were also unable to predict any of the four recessions occurring between 1970 and 1982 [3]. These well-publicized failures, coupled with the inability of econometric models to predict financial markets accurately, led to a reexamination of the sources of error in structural models and a movement away from this approach toward non-structural models. As already noted, error from structural models occurred not only because of random elements but because the underlying data generation function often shifted, and it was difficult to predict the independent variables accurately.

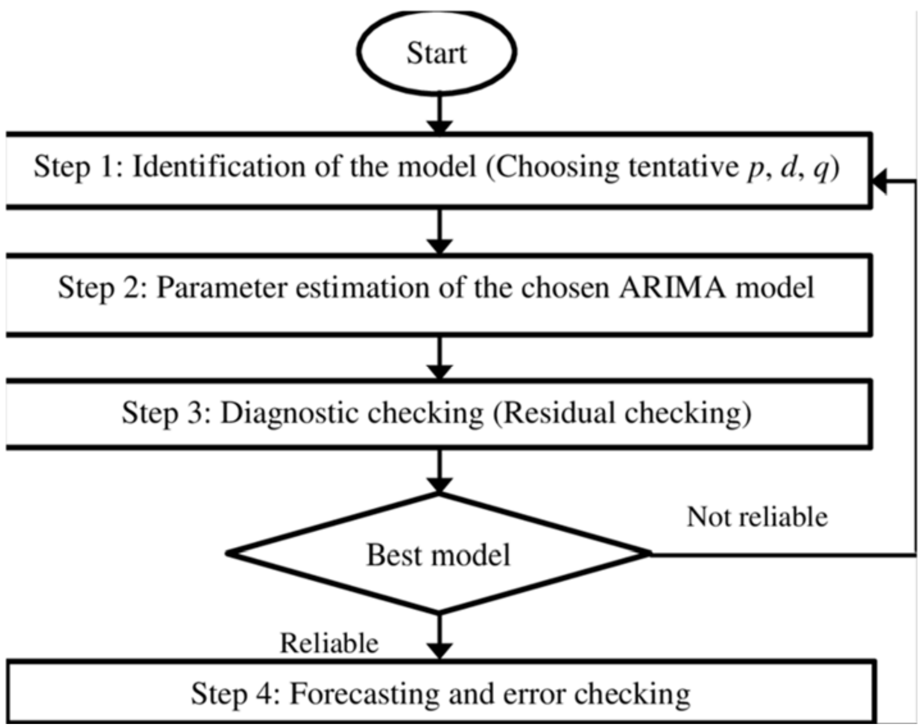


Figure 1: Illustrate the Forecasting procedure using Box-Jenkins approach

For these reasons, econometric analysis and development of forecasting models has shifted in the past 30 years towards placing a greater emphasis on non-structural models. This part of the book analyses single-equation models; multi-equation models are considered in chapters 11 and 12. In these nonstructural models, a given variable is a function only of its own lagged variables, lagged error terms, and truly exogenous variables such as time trends.

This author has been using econometric models to generate macroeconomic aand industry forecasts since 1963. Even initially, it was obvious that forecast accuracy was improved by adjusting the constant terms of the regression equations. There was always a certain amount of independent judgement used ain that procedure, but the changes often tended to be based on ad hoc rather than systematic methods. Sometimes, judgment played a greater role in determining the forecast than the underlying structure of the equations. Thus it became logical to raise the issue that forecast accuracy might be improved by systematizing the adjustment procedure or, in some cases, jettisoning the structural approach completely. Figure 1 illustrate the Forecasting procedure using Box-Jenkins approach.

The position taken in this text is that the choice of model is an empirical rather than a theoretical question. It is quite likely that some combination of methods structural models, time series models, and exogenous information in the form of field surveys or indexes of sentiment can provide more accurate forecasts than any single method, and this possibility is examined later in athe text. That approach is not a new development; the original Wharton econometric forecasting model, developed in the early 1960s, contained equations with terms for consumer and business sentiment as well as econometrically determined variables.

Even large sophisticated structural models can be reduced to the form where each variable is a function only of its own lagged values and truly exogenous variables. Admittedly, such models contain no estimates of elasticity's or multipliers, and cannot be used for policy prescriptions. Nonetheless, if the two approaches can be shown to be logically equivalent, it is worth testing these alternative statistical methods to see whether in fact they can reduce forecast error. Sometimes forecasting models are developed to track a large number of individual product lines, inventories, or production schedules. It may not be feasible to generate structural models for each individual item, so time-series regression equations represent a more realistic approach for forecasting and tracking these individual items [4].

Hence there are several reasons why time-series models are considered part of practical business forecasting. In most cases, these models perform better when the trend and seasonal factors have been removed from the original data we first turn to a discussion of those issues before focusing on the estimation of autoregressive and moving average models then discusses the possible benefits of combining different forecasting methods to reduce error. Theoretically, any time series can be decomposed into trend, seasonal, cyclical, and irregular factors. Under certain conditions, the irregular factor will be randomly distributed. However, this decomposition is often not straightforward because these factors interact. Also, the cyclical factor does not follow some preassigned function such as a sine curve, but contains random elements of timing and magnitude. Thus most of the discussion in this chapter will focus on removing the trend and seasonal factors. In some cases, the Hodrick–Prescott filter can be used to remove cyclical factors.

We first consider monthly data for general merchandise sales and discuss various methods for extracting the trend and seasonal factors. General merchandise sales, which comprise SIC 53, are similar to department store sales, but include discount stores, variety stores, and other establishments that sell a wide variety of goods. Sales in stores that sell a specific class of items – hardware, apparel, furniture, home appliances, jewelry, etc. are reported in other two-digit SIC codes in the retail sales sector. Later in this chapter the data for apparel sales (SIC 56) are used to illustrate the concept of shifting seasonal factors because of moving holiday dates. The seasonally unadjusted data for general merchandise sales.

Sales rise every December and decline every January the seasonal swings are much greater in more recent years, although one cannot tell from this graph whether the percentage changes are greater. The December peaks were smaller in 1990 and 1991, when there was a recession. As shown below, this is also true for earlier recessions, although it is not as obvious from the graph [5]. These points reveal substantial interaction between the trend, seasonal, and cyclical component. Furthermore, if one were to correlate retail sales with the usual economic variable's income, cost and availability of credit, expectations, etc. the residuals would be serially correlated, so the "irregular" component would not be random.

General merchandise sales exhibit a strong upward trend. However, it is not obvious whether that trend is linear. Indeed, regressing sales against a linear time trend quickly reveals that the residuals are large and positive near the end of the sample period. This is not surprising; it is more likely that the rate of growth of sales, rather than the actual change, would be constant. This equation can thus be refitted using the logarithm of sales as a function of a linear time trend. However, that doesn't work so well either, because the residuals in the middle of the sample period are larger, suggesting the growth rate of sales is smaller now than was the case earlier. Since these data are in current dollars, that is presumably due to lower inflation. The trend is not removed from the residuals until a quadratic function is used, in which the logarithm of sales is a function of time and time-squared.

Another possibility, considered later, is to re-estimate the function using constant-dollar sales, which may be useful for econometric analysis but would not answer the question of how much actual sales are likely to change in the near future without also predicting the change in prices. In some cases the researcher or model builder will want to identify and estimate the seasonal factors. This would be the case, for example, if someone wanted to know how much additional stock to order before the holiday shopping season, or how much to cut back on inventories before plant closings summer. In other cases, model builders want to use seasonally adjusted data to avoid spurious correlations. As already shown, retail sales rise sharply each December. Because people earn and spend more money, income and credit outstanding also rise sharply each December. Thus a simple correlation between these two series without seasonal adjustment would overstate the correlation between consumption and income, and consumption and credit. The methods for seasonal adjustment discussed in this section can be used for either of two purposes. In some cases, the desired objective is to prepare seasonally adjusted data. In other cases, the seasonal adjustment factors themselves are desired for forecasting purposes [6].

The underlying principal behind seasonal adjustment is quite simple if the seasonal factors remain constant. The problem is that, in most economic time series, they do not. The causes of changing seasonal factors can be divided into three categories: random changes, such as those associated with the weather; shifting patterns associated with variations in the calendar (such as the movable date of Easter, or different number of weekends per month); and shifts in factors related to economic decisions. The latter case is represented by such phenomena as the lengthening of the Christmas holiday shopping season, or fewer plant shutdowns during the summer.

One well-known example of the first type of change in seasonal factors is the pattern of housing starts in the Northeast and Midwest regions of the US, which decline sharply every January and February because of the cold weather. While a similar pattern occurs every year, the severity of winter varies considerably from one year to the next. The seasonal factor for housing starts in January and February for these regions is about, which means actual housing starts are multiplied by 2 to obtain the seasonally adjusted. Suppose actual housing starts in these two regions of the country usually average about 20,000 per month in January and February, or about 480,000 on a seasonally adjusted annual rate basis. Because of unusually mild winter weather one year, that figure rises to 27,000 per month. The seasonally adjusted annual rate figures would increase to 648,000. Assuming no offsetting weather effects in the South and West, seasonally adjusted housing starts at annual rates would increase 168,000, which might lead some economists to conclude that housing activity was accelerating [7], [8].

Other, wiser hands would point out that the surge was due to mild weather, but even that is not the end of the story. If in fact the underlying demand for housing has not changed, then housing starts in the Northeast and Midwest will actually total 480,000 by year-end. However, the seasonally adjusted data will tell another story. Suppose houses were started ahead of schedule in January and February, resulting in fewer starts during the peak months of June and July, when the seasonal factor is about 0.8. The switch of 7,000 starts from Jun/Jul to Jan/Feb will add 168,000 at seasonally adjusted annual rates in the first two months of the year, but subtract only 67,000 from the Jun/Jul figures. Hence the seasonally adjusted total for the year will be 101,000 more than the actual number [9]–[11].

CONCLUSION

The Box-Jenkins Method of forecasting is a widely used and effective technique for analyzing and forecasting time series data. The method involves identifying, estimating, and diagnosing ARIMA models based on statistical tests and analysis of historical data. The accuracy and reliability of the method have made it a valuable tool in industries such as finance, economics, and engineering, where accurate predictions are essential. However, the method has its limitations and assumptions, and its application to real-world data requires careful consideration of these factors. Overall, the Box-Jenkins Method provides a powerful framework for forecasting time series data and has contributed significantly to the field of time-series analysis.

REFERENCES

[1] N. Golyandina and A. Korobeynikov, "Basic Singular Spectrum Analysis and forecasting with R," *Comput. Stat. Data Anal.*, 2014, doi: 10.1016/j.csda.2013.04.009.

[2] S. A. Sannasiraj, V. Babovic, and E. S. Chan, "Local model approximation in the real time wave forecasting," *Coast. Eng.*, 2005, doi: 10.1016/j.coastaleng.2004.12.004.

[3] S. C. Huang and T. K. Wu, "Combining wavelet-based feature extractions with relevance vector machines for stock index forecasting," *Expert Syst.*, 2008, doi: 10.1111/j.1468-0394.2008.00443.x.

[4] I. Malone, "Recurrent neural networks for conflict forecasting," *Int. Interact.*, 2022, doi: 10.1080/03050629.2022.2016736.

[5] S. Maldonado, A. González, and S. Crone, "Automatic time series analysis for electric load forecasting via support vector regression," *Appl. Soft Comput. J.*, 2019, doi: 10.1016/j.asoc.2019.105616.

[6] H. Boraine and V. S. S. Yadavalli, "Electricity Load Forecasting With Artificial Neural Networks," *South African J. Ind. Eng.*, 2012, doi: 10.7166/14-2-263.

[7] H. T. V. Gouveia, R. R. B. De Aquino, and A. A. Ferreira, "Enhancing short-term wind power forecasting through multiresolution analysis and Echo State Networks," *Energies*, 2018, doi: 10.3390/en11040824.

[8] S.-C. Huang, "Combining wavelet-based feature extractions with SVMs for financial time series forecasting," *J. Stat. Manag. Syst.*, 2008, doi: 10.1080/09720510.2008.10701297.

[9] Q. Teng, J. Wang, Z. Ding, L. Zhang, and Z. Wang, "Data driven governing equations approximations using attention based multistep neural networks," *AIP Adv.*, 2020, doi: 10.1063/5.0015600.

[10] N. Sun, S. Zhang, T. Peng, J. Zhou, and X. Sun, "A Composite Uncertainty Forecasting Model for Unstable Time Series: Application of Wind Speed and Streamflow Forecasting," *IEEE Access*, 2020, doi: 10.1109/ACCESS.2020.3034127.

[11] S. G. Hassan, S. Iqbal, H. Garg, M. Hassan, L. Shuangyin, and T. T. Kieuvan, "Designing Intuitionistic Fuzzy Forecasting Model Combined with Information Granules and Weighted Association Reasoning," *IEEE Access*, 2020, doi: 10.1109/ACCESS.2020.3012280.

CHAPTER 11

EFFECTIVE COMMUNICATION OF FORECASTS TO MANAGEMENT: BEST PRACTICES AND STRATEGIES FOR SUCCESS

Dr.R.Satish Kumar, Professor
Department of Marketing, CMS Business School, JAIN Deemed to-be University, Bangalore, India
Email Id- dr.satishkumar@cms.ac.in

ABSTRACT:

Communicating forecasts to management is a crucial aspect of effective decision-making in any organization. Forecasts help managers anticipate future trends, identify potential risks, and make informed strategic decisions. However, communicating forecasts can be challenging, as managers may have varying levels of understanding and different information needs. To effectively communicate forecasts to management, it is important to tailor the message to the audience. This means using clear and concise language, avoiding technical jargon, and presenting the information in a format that is easy to understand. It is also important to provide context and explain the assumptions and limitations of the forecast, as well as any potential sources of error or uncertainty.

KEYWORDS:

Communication, Clear language, Decision-making, Forecasts, Management, Tailoring.

INTRODUCTION

Obviously, this discrepancy cannot remain on the books. The Bureau of the Census, which prepares these figures, goes back and adjusts the numbers so that the seasonally adjusted and actual totals are identical. Yet in the meantime, the figures reported on a monthly basis are misleading. Furthermore, the equilibration of the two series means the seasonal factors for January and February will be adjusted because of the mild winter, so if there is a severe winter the following year, the seasonally adjusted housing start data will understate the true level of starts.

No method will entirely compensate for this anomaly, and as a result, "seasonally adjusted" data from the government often contains unintended inaccuracies. This is an unavoidable error, not a criticism of the highly sophisticated methods used at Census, yet it does lead to forecasting errors. In recent years, the seasonal adjustment programmes have been refined to isolate or reduce the importance of outliers due to weather or other exogenous causes [1].

In other cases, changes in seasonal adjustment factors are due to predictable patterns. Since these changes are not random, the problems are usually less severe. In the case of changes in the calendar, the adjustments are straightforward. Some months may have more trading days than others, or more weekends; the seasonal adjustment algorithms can easily be adjusted to take this into account. The other major source of seasonal variation in this category revolves around changing dates for holidays, notably Easter. We show below how changes in the date of Easter Sunday affects the seasonal factor for apparel sales in March and April.

The more complicated issue arises when the seasonals shift due to economic factors. For example, the Christmas shopping season now extends further in both directions. Another example is that many manufacturing plants do not shut down as much in summer as they used to. Demand for power used to peak on cold days in the winter; now it peaks on hot days in the summer. More people vacation in Florida in August than used to be the case.

Many firms, such as hospitals and educational institutions, raise their prices only once a year. During years of rapid inflation, hospital costs rose an average of 10% per year. Since these changes usually occurred in January, the seasonal adjustment factors gradually incorporated this change. During the 1990s, the annual increase in hospital costs fell to about 3% per year. Thus, for a while, the seasonally adjusted figures for hospital costs in January would show a 7% decline. Eventually this problem was fixed, but when the rate of inflation slowed down in the US in the 1980s and 1990s, the seasonally adjusted data for the CPI initially showed an actual decline in January the current revised data have fixed this error, so you won't find it now.

Seasonal factors can also be affected by outlying values. Suppose some exogenous event an oil shock for the PPI – caused a change in one given month that was 10 times as large as an ordinary seasonal factor. If that one point is treated the same as all other observations, it will dominate the seasonals in other years. One possibility is to omit it entirely; another way is to dampen it [2]. However, it cannot be ignored entirely without distorting the seasonal factors in subsequent years. Census Bureau programmes have been designed to treat all these and other factors, and the methods they use will be briefly outlined. First, however, it is useful to understand how seasonal adjustment factors are calculated when the weights do not change and random fluctuations are relatively small. To illustrate these points, consider a typical monthly series before seasonal adjustment. The first step is to calculate a centred moving average, which for monthly data. The seasonal factors for each month can then be calculated as $y - \tilde{y}$. All the seasonals are collected for the first, second, twelfth month and averaged for each month.

Sometimes the sum of the seasonally adjusted data for any given year will not be equal to the actual data, as already discussed for housing starts. When that occurs, the seasonal factors must be re-scaled to eradicate that anomaly. One simple way would be to adjust each year separately: if the sum of the seasonally adjusted data were 102% of the actual data, all observations would simply be divided by 1.02. There is a problem with this method. Suppose in the following year the seasonally adjusted data were 98% of the actual data; then the seasonally adjusted series would have an unwarranted 4% rise from December to January. Hence the seasonal factors themselves must be smoothed by a moving-average method to eliminate the jumps that would occur when the seasonally adjusted data for any given year do not sum to the actual data.

DISCUSSION

If one wants to obtain the seasonal factors themselves, rather than the seasonally adjusted data, and if it is assumed the factors remain constant, a regression can be calculated in which the actual data are a function of seasonal dummy variables that are 1 in the appropriate month and zero elsewhere. Recall the earlier warning that you cannot include all 12 monthly seasonal factors plus the constant term, or the matrix will not invert. It is necessary to omit one of the seasonals, or the constant term itself [3].

The major problem with this method is that it usually does not work if the time series has a strong trend; as we have often seen, most of them do. In these cases, the methodology is similar except that the seasonal factors are calculated as y/y˜ instead of y - y˜. The error is 0.57% on average. Yet, there are a few times where the actual percentage shift lasts for more than 3 seconds. They seem to be selected from a separate population. These times frame the erasure of pricing restrictions. The energy shock in January 1980, the spike in oil prices in February and March 1986, the Iraqi invasion of Kuwait in August through October 1990, and the return to normalcy following the US and UN victory in the Persian Gulf War in February and March 1991 are some examples of events that occurred between these dates noteworthy is the fact that each of the two largest drops happened in February and March.

Hence, the seasonal adjustment algorithm would presume that prices always dropped significantly in those two months if no further changes were done. Hence, the seasonally corrected statistics would show significant increases in typical years. The user may set a specific s in the Census programmes, after which the observations are replaced by a weighted moving average of neighbouring values. Most often, s is set between 1.5 and 2.5, and when s rises over this range, the weights assigned to the outlying values steadily decrease until they are zero [4].

In the Census X-11 programme 3, the raw data are often corrected using the above methods for I trade day fluctuations where acceptable, (ii) moveable holidays when applicable, and (iii) outliers. This software then constructs a new set of seasonal components by recalculating the moving average using the updated data. For monthly data, the seasonal variables are scaled to add up to 12.000, while for quarterly data, they add up to 4.000.

The procedure is then improved a number of times more until the variations between iterations are negligible. The length of the moving average might vary (for monthly data) from 9 to 23 months in different versions of these programmes depending on the level of unpredictability in the series. Consequently, a sophisticated seasonal adjustment programme modifies the data based on the various trading days or the presence of moving holidays, reduces or eliminates the excessive influence of outliers, and modifies the length of the moving average based on randomness, further refining this process until a convergence process is reached. There are no statistically significant issues with any of these simple, albeit time-consuming, procedures.

These variables are subject to vary over time when using the above-described approach, which determines the seasonal components using changed data split by a smoothed trend. It is totally reasonable since doing otherwise would often lead to unsatisfactory outcomes. Yet, both in terms of statistical methodology and economic reasoning, such technique creates more concerns. Good projections are beneficial to a variety of service sector sectors, including financial institutions, airlines, hotels, hospitals, sports and other entertainment groups. Figure 1 illustrate the manual finance forecasting. Forecasting is used by the accounting and finance departments in many different contexts. Financial forecasting enables the financial management to foresee events before they happen, especially the need for obtaining money from outside sources. The creation of many pro forma, or projected, financial statements is the most thorough method of financial forecasting. The business can predict its future levels of receivables, inventories, payables, and other corporate accounts, as well as its expected earnings and borrowing needs, based on the projected statements [5].

Figure 1: Illustrate the manual finance forecasting.

Making company choices requires careful consideration of cash flow and estimates of revenue and cost rates. Moreover, accurate forecasting is necessary for asset market speculation. Both big and small airlines may profit from accurate forecasts of the load factor, fleet management, fuel, and other costs. Accurate hotel occupancy rate projections, for instance, have an impact on all other guest services provided in the hospitality and entertainment sectors. To anticipate costs and usage of emergency department staff, hospitals have traditionally employed forecasting systems. Forecasts are used in the sports business to determine how many tickets will be sold for each event. Revenue estimates are derived on a team's performance over a given year or years.

When deciding how many workers a company will ultimately require, human resource departments must utilise predictions heavily. This has an impact on the firm's resources and the need for personnel training. Resource planning and management choices might benefit from estimates of the number of employees in functional areas, the makeup of the workforce (e.g., part-time vs full-time), trends in absenteeism and tardiness, and productivity. In the public sector, forecasts are used to inform macroeconomic choices. Forecasts of significant economic indicators serve as a foundation for economic policy. Good predictions are essential for estimating the gross national product (GNP), employment, inflation rate, industrial output, and estimated income tax receipts for both individuals and corporations. These projections are used by the government to direct the nation's monetary and fiscal policies. Among the various applications for predictions, planning government spending on infrastructure, social insurance, and health care is one that heavily relies on population (or demographic) estimates.

The aforementioned examples of predictions' applications in different commercial and economic activities are by no means all-inclusive. This only illustrates forecasting's importance and range in decision-making. In recent years, forecasting has drawn a lot of attention as a planning tool. The necessity to succeed in a competitive, dynamic, and ever-changing global market contributes to this increasing focus. A second point is that 4 Forecasting for Management Decisions [6]. With the development of computers and easy access to data produced by businesses, companies are seeking for methods to enhance their decision-making procedures. Moreover, methodological advancements in forecasting have increased managers' capacity to utilise these tools successfully in making timely business and economic choices in both the public and private sectors. It takes both art and science to figure out how to include these advancements into the firm's actions.

Companies now have access to a broad variety of forecasting techniques. These vary from simple forecasting to quite complex quantitative techniques. Each of these approaches has advantages and disadvantages. Using them properly requires art. A manager must choose a certain technique based on both personal experience and professional judgement. The skill of predicting is in knowing when it is necessary and knowing how to combine qualitative and quantitative data. This paper addresses forecasting methods that may support management choices made with common sense.

The scientific foundations of model construction include the science of predicting. As in any other branch of science, scientists start by offering the most straightforward explanation for a phenomena. The model is often regarded as a suitable instrument for future prediction if it accurately captures the circumstances of the actual world and its outcomes match observable events. On the other side, scientists adopt more sophisticated models if the basic model is unable to adequately capture or explain the observed occurrence. The number of assumptions that must be made in a model increases with its complexity.

Simple models have been employed by economists to identify data patterns, which have subsequently been utilised to make future predictions. An economist may use an economic theory or model, which is a collection of definitions and presumptions, to explain certain kinds of occurrences. An economic theory explains the way in which certain economic variables interact, often in the form of a set of equations. The theory of consumer choice, for instance, contends that customer preferences, income levels, the cost of the product in issue, as well as the cost of comparable products and services, all influence how much of a given thing people will purchase. According to this notion, when a good's price increases, fewer people will normally buy it. There are theories in macroeconomics that suggest the overall amount of investment is influenced by the interest rate. These theories specifically suggest that higher interest rates would deter investment in real capital development (investment). We must ascertain the accuracy of these theories' predictions (forecasts) of economic events in order to assess their utility. To capture the effect of the many factors on the model under these circumstances, multivariate models are utilised.

An company must have a systematic forecasting process that can be swiftly implemented and changed in order to consistently provide accurate projections forecasting as necessary for Management Decisions 5. The forecaster benefits by following a procedure that is based on the scientific method, just as in any other scientific activity. The instructions that make up the language of the scientific method are descriptions of sequences of acts or procedures that are precise enough for anybody to follow. These guidelines are referred to as "operational definition." An operational definition ought to list all the steps required to repeatedly produce the same outcomes.

Forecasting procedures might be simple or complex. It starts when an organization's management needs a justification for a management choice. For instance, they can inquire as to whether a product's enhancements would result in a material rise in demand. In other circumstances, the management could request a projection if they must devote a significant amount of resources to a project or if a circumstance in the business environment indicates that a choice must be made [7].

The question "Forecasting for Management Decisions" serves as the foundation for a prediction. The forecaster is in charge of making sure that the forecast's goals are spelled out in detail. It is crucial to understand why the prediction is required and how the outcomes will be utilized. Depending on the kind of issue and the forecast's duration, organizations create unique predictions. For instance, opposed to production projections that are weekly or monthly, capital budget estimates are long-term in nature.

After the choice is taken to create a prediction, a theoretical model that addresses management concerns must be created. The link between the model's many variables may be elaborated on using the theoretical framework of the model. It also enables the division of impacts into internal and external elements. The term "internal factors" refers to those variables that the company may influence. Price, product quality, product features, marketing and advertising costs, and logistics are a few examples.

The company has no influence over factors that are external to it. The interest rate, the rate of inflation, income, employment, and the exchange rate in global trade are examples of exogenous influences. In addition, we may choose a trend or regression model for this assignment if management is interested in a long-term trend projection. The moving average, exponential smoothing, or Box-Jenkins models, for instance, may be used in the analysis if management is interested in a short-term prediction, i.e., weekly or monthly forecasts.

The forecaster is now prepared to acquire data that supports the analysis's conceptual framework after determining the theoretical model. The information may originate from company records or from other sources. The kind and quality of the data being obtained should be carefully considered. Unless it is essential to their everyday operations, businesses often do not gather disaggregated data. The company may need to obtain such data when highly disaggregated data are required to produce a projection. It could take more time and money to do so, which would make forecasting more difficult. When presented with a situation like this, the forecaster must weigh all of the potential outcomes before making a prediction [8], [9].

Data analysis should not be seen as only a mechanical procedure. The forecaster should get quite familiar with the characteristics of the company and the sector it represents before calculating the forecasting model. The market structure in which the business works, the sources of competition, the company's position within the industry, etc., are all examples of information that aids in forecasting and should be thoroughly examined. As a result, the forecaster may see the predicted model in a dynamic way. Changes may be made at this point if a need to reassess the findings arises. Models should be examined for validity and reliability. To evaluate the model's accuracy, forecast and actual outcomes should be compared. The procedure of evaluation at this point acts as a control. The validity of the model and forecasting for management decisions 7 both increase with the addition of new variables, factors, or changes to the time frame or periodicity of the data.

Presenting the findings to management is the last step in the forecasting process. The management of the company seeking the prediction must realise that, even if the duty of giving a forecast may be finished, the process of fine-tuning the forecast has not. To be clear, a good prediction is dynamic rather than static. The management would benefit from developing a process for regularly assessing its projections and from being aware that unforeseen market changes can affect the predicted model's underlying assumptions, necessitating a new estimate [10].

Forecasters in the information age have access to a broad variety of datasets that may be quickly accessed for use in predicting. Commercial and public asector institutions rely on these databases as their source of information for athe aggregate and disaggregate data. This paragraph offers a few instances of online resources. We must comprehend the numerous aspects that influence the choice of a forecasting approach in order to be able to assess its usefulness in a specific circumstance.

We shall emphasise those that concern data in this chapter. In particular, we will focus on the data pattern, the kind of historical connection in the data, and the degree of subjectivity involved in forecasting. We categorise the forecasting methodologies based on all three of these elements. Whereas multivariate forecasting makes use of previous associations to predict the future, univariate forecasting approaches rely on historical patterns. Qualitative predictions rely less on the manipulation of previous data and more on the forecaster's subjectivity and intuition. We'll talk about data patterns including horizontal, trend, seasonal, cyclical, autocorrelated, and a mixture of these patterns in the section after this one [11].

CONCLUSION

Communicating forecasts to management is a critical task that requires clear and effective communication skills. Forecasts can be valuable tools for helping management make informed decisions and plan for the future. However, presenting forecasts requires careful consideration of the audience, the data, and the methods used to create the forecast.

When communicating forecasts to management, it is important to use clear and concise language, avoid technical jargon, and provide context to help managers understand the implications of the forecast. Visual aids such as charts and graphs can also be helpful in conveying complex information.

REFERENCES

[1] C. C. Verschoor, "Building a World-Class Compliance Program: Best Practices and Strategies for Success," *Intern. Audit.*, 2008.

[2] T. M. Kiles, N. Borja-Hart, B. R. Wofford, and C. P. Renfro, "Screening for social determinants of health in community pharmacy: Identifying best practices, barriers, and strategies for success," *Journal of the American Pharmacists Association*. 2021. doi: 10.1016/j.japh.2021.05.004.

[3] L. Koharchik and P. M. Flavin, "Teaching Students to Administer Medications Safely," *Am. J. Nurs.*, 2017, doi: 10.1097/01.NAJ.0000511573.73435.72.

[4] L. Bruno, "Best Practices for University T ransit Bus Programs: Identifying Strategies for Success," *J. Chem. Inf. Model.*, 2019.

[5] A. Gupta, K. Gupta, and V. Vasudevan, "Excellence in the Accreditation Council for Graduate Medical Education Core Competencies: Strengthening the Mentor-Mentee Relationship," *Cureus*, 2020, doi: 10.7759/cureus.8564.

[6] J. A. Vilchez, J. Kruse, M. Puffer, and R. N. Dudovitz, "Teachers and School Health Leaders' Perspectives on Distance Learning Physical Education During the COVID-19 Pandemic," *J. Sch. Health*, 2021, doi: 10.1111/josh.13030.

[7] R. S. Authement and S. L. Dormire, "Introduction to the Online Nursing Education Best Practices Guide," *SAGE Open Nurs.*, 2020, doi: 10.1177/2377960820937290.

[8] R. G. Cooper, "The drivers of success in new-product development," *Ind. Mark. Manag.*, 2019, doi: 10.1016/j.indmarman.2018.07.005.

[9] S. Kraus, C. Richter, A. Brem, C. F. Cheng, and M. L. Chang, "Strategies for reward-based crowdfunding campaigns," *J. Innov. Knowl.*, 2016, doi: 10.1016/j.jik.2016.01.010.

[10] P. Meani, R. Lorusso, and F. Pappalardo, "ECPella: Concept, Physiology and Clinical Applications," *Journal of Cardiothoracic and Vascular Anesthesia*. 2022. doi: 10.1053/j.jvca.2021.01.056.

[11] E. T. Game, E. Meijaard, D. Sheil, and E. Mcdonald-Madden, "Conservation in a wicked complex world; challenges and solutions," *Conserv. Lett.*, 2014, doi: 10.1111/conl.12050.

CHAPTER 12

FORECASTING UNDER UNCERTAINTY: APPROACHES AND TECHNIQUES FOR DEALING WITH THE UNKNOWN

Dr. Anantha Subramanya Iyer, Associate Professor
Department of Marketing, CMS Business School, JAIN Deemed to-be University, Bangalore, India
Email Id- asi@cms.ac.in

ABSTRACT:

Forecasting under uncertainty refers to the process of predicting future events or trends in situations where the outcome is uncertain or unpredictable. This can be particularly challenging when dealing with complex systems or environments that are subject to a range of external factors that can impact the accuracy of the forecast. To address this challenge, researchers and practitioners have developed a range of methods and approaches for forecasting under uncertainty, including probabilistic forecasting, scenario planning, and simulation modeling. These methods seek to incorporate a range of potential outcomes and uncertainties into the forecasting process, enabling decision-makers to make more informed decisions in the face of uncertainty.

KEYWORDS:

Forecasting, Uncertainty, Prediction, Probabilistic forecasting, Scenario planning, Simulation modeling.

INTRODUCTION

The employment of a forecasting technique is determined by specific criteria, as we previously discussed. The selection of a forecasting technique is also influenced by patterns in the data, in addition to these qualities. For instance, certain forecasting techniques are better at identifying steady patterns of development, while others are more suited to identifying economic turning moments. The fundamental tenet of forecasting is that there is a pattern in the data and that it will probably persist into the future. Consequently, matching the pattern with a method that can manage it is a key factor in determining a given approach's capacity to provide a decent prediction in a particular context [1].

The observer may first see how the data have changed over time by making a straightforward observation of the data. When making predictions about the pattern of the data in the future, one might use the nature of this behaviour as a guide. We shall demonstrate how to utilise data patterns of this kind to construct projections in the later chapters of this book. Second, the pattern of the data can point to a correlation between two or more factors. In this situation, historical data on a single variable by itself cannot reveal the underlying trend.

The correlation between a country's population's income level and its economy's consumption of products is an illustration of this pattern of data. The level of consumption of goods and services increases along with income. The forecaster often encounters one of the horizontal, trend, seasonal, or cyclical patterns in the data while collecting information for a forecast. The forecaster is working with a horizontal data pattern when there is no trend in the data pattern. This suggests that there is no consistent rise or reduction in the observations. This is what is known as a stationary pattern in statistics. In this case, the forecaster has an equal chance of seeing that the subsequent value in the series will be higher or lower than the stationary value.

When determining if a horizontal pattern is visible in the data, the length of time is the most important consideration. The shorter the time frame, the amore probable it is to notice a horizontal pattern in the data. This pattern of data is often associated with two types of data: (1) items with consistent sales, and (2) faulty components originating from stable manufacturing processes. The trend might be stable, rising, or declining. A smooth linear or nonlinear graph may be used to represent it. In business and economics, trend patterns are linked to changing consumer preferences, technical advancements, the gross domestic product, income fluctuations, stock market fluctuations, industrial growth, and governmental tax policy changes. The pattern ranges from linear, to various nonlinear forms, as shown in Figure 2.2b-d, depending on the nature of the variable. In a pattern of linear trends, the variable of interest increases by a fixed absolute amount throughout each interval of time (t). The linear trend is denoted by the formula: $Yt = a + t > 0$ [2].

The number of time periods that have passed since the base year is expressed in equation by the positive constant increased to the power of t. This kind of tendency is often seen in sectors that flourish quickly at first before gradually declining. This kind of trend line is shown by the Gompertz curve. In a paper titled "On the Nature of the Function Expressive of the Law of Human Mortality," Gompertz demonstrated that "if the average exhaustions of a man's power to avoid death were such that at the end of equal infinitely small intervals of time, he lost equal portions of his remaining power to oppose destruction," then "he lost equal portions of his remaining power to oppose death." This is the basis for the Gompertz growth curve. Since the publication of this work, actuaries and economists have both been interested in the Gompertz curve. This curve has been used by economists to depict growth rates in a number of circumstances.

When a predictable and recurring movement is seen around a trend line for a period of one year or less, a seasonal pattern in the data is present. This implies that we need data that is provided on a regular basis, such as weekly, monthly, quarterly, etc., in order to be able to examine seasonal fluctuations. There are many causes for seasonal trends. We refer to this as an internally-induced seasonal pattern when a company choose to publish its profit and loss on a quarterly basis, as opposed to an externally-induced seasonal pattern that is caused by things like the weather. Clothing, heating oil, and the quantity of new automobiles sold at a certain time of the year owing to model changes are a few examples of data that exhibit seasonal trends.

DISCUSSION

Business and economic booms and recession follow a cyclical pattern. While they share characteristics with seasonal patterns, cyclical movements often endure for more than a year. Cyclical patterns are connected with economic cycles in a country's economy and vary in severity or amplitude. Cyclical patterns are the most challenging to foresee since they don't repeat themselves at regular periods of time.

Hence, it is hard to tell a smoothly growing trend apart from cyclical activity. To forecast cycles more accurately, forecasters have invested a lot of time and effort. These efforts, however, have not yielded the desired results. More information may become available as a result of methodological advancements and ongoing study [3].

Time series data may also display another pattern known as an autocorrelated pattern in addition to the ones mentioned above. This pattern only demonstrates the relationship between the values of the data in one time period and those in earlier times. The forecaster is aware that there is an automatic correlation between observations in a series when presented with an autocorrelated data pattern. This indicates that the value in June is positively correlated with the value in May if there is a large positive autocorrelation. Consumer behaviour, trends, and other elements. Seasonality must be taken into account when detecting autocorrelated data patterns, as discussed of Data Patterns and Forecasting Methods, while addressing the Box-Jenkins approach, this data pattern will be examined in further detail.

Let's look at a straightforward example to help you better comprehend what we mean when we talk about data patterns. Imagine that a manager of an online store would want to know the sales estimate one week in advance so that he may schedule the arrival of fresh shipments and replenish his inventory. While producing this prediction, the management has a variety of options. He might use the naive technique to predicting, which is the most straightforward method. With this approach, the manager just makes the assumption that the sales information from prior week's serves as a reliable predictor of what the sales may be in the weeks to come.

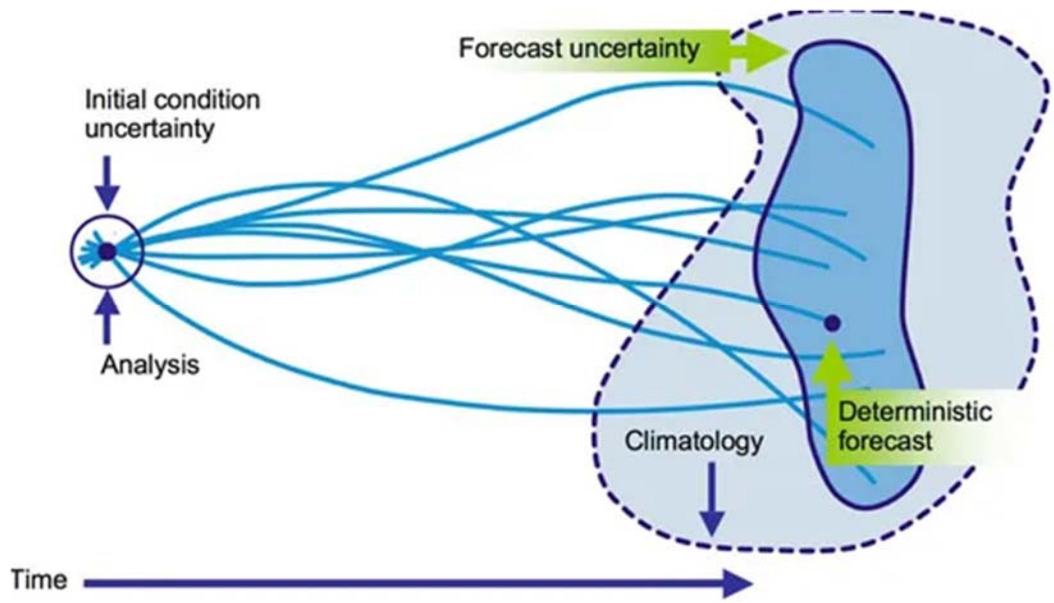


Figure 1: Illustrate the uncertainties in weather prediction.

The methodology's underlying assumption is that data patterns are consistent. The manager has acquired sales data for the previous 8 weeks in order to use this strategy contains the facts mentioned above. He bases his projection on the assumption that sales for the next week will not be smaller than those for the prior week. That is, he projects sales of \$41,000 for week 9. He anticipated the next week's sales based on the previous week's sales. We see that they have a horizontal pattern with sporadic swings around the average weekly sales amount. Instead of the manager's naive approach, a better method for predicting would be to utilise the average of the previous 8 weeks' sales. It would be simpler to avoid overreacting to data variations if the horizontal pattern's random fluctuations were eliminated by this averaging. We need some indication of the typical inaccuracy that may be anticipated over time in order to evaluate how accurate this straightforward prediction is every week e-Commerce Sales for a Naive Predictor Week: Actual and Projected Sales True Sales Estimated Sales. Figure 1 illustrate the uncertainties in weather prediction [4].

There are several accuracy metrics that may be used in a prediction. Finding the average mistake that may be anticipated over time is one technique to assess the accuracy of this prediction. The positive and negative values cancel each other out, as shown by the fact that if we merely took the mean of the mistakes in column 4, we would get a mean that is almost zero. This will create the appearance that our forecast's inaccuracy is little and might cause us to draw the erroneous conclusions. We may rely on the mean absolute deviation to solve this issue (MAD). The positive and negative signs associated with each error phrase are ignored when computing the MAD. The formula for calculating mean absolute deviation is MAD = n t = 1 |et | n.

The benefit of MAD is that it is simpler to understand and that each factor is given equal weight in the technique. Nevertheless, it's crucial to bear in mind that this measure of error misses the significance of systematic over- or underestimation since it only takes absolute numbers into account. The mean square error is a different indicator of a forecast's accuracy (MSE). The computation looks like this [5]:

MSE = n t = 1 (e 2 t) n

The MSE in this instance is 12.43. Note that this metric is helpful if management is less concerned with frequent little errors and more interested in preventing significant errors from happening. The penalty for a prediction from mean squared error is substantially greater for large deviations than for minor ones. For instance, while calculating the MAD, a 2 mistake counts just twice as much as a 1 error. Yet, the identical mistake counts four times as many times as a one error when MSE is calculated. As a result, the MSE as a measure of accuracy could be too sensitive to an error pattern that exhibits a few high numbers of mistakes. The MSE's interpretation is one of its main drawbacks.

Forecasters use mean absolute percentage error (MAPE) as another indicator of prediction accuracy. The only difference between this measurement and MAD is that it is given as a percentage. The measure has the benefit of accounting for the error term's magnitude in relation to the actual units of observation. Below is a computation of MAPE.

MAPE = n t = 1 |(et / Yt) 100| n

The calculated MAPE for this case is 7.45 percent, as indicated this number suggests that the average inaccuracy for this data set would be 7.45 percent. We may compare the accuracy of the same or different procedures on two completely distinct series using this measure of error. The mean absolute percentage error does not take systematic over- or underestimation into account, as was the case with MAD.

There are times when a forecasting technique routinely produces low or high predictions. Forecasters rely on the mean percentage error as a gauge of accuracy to identify whether such a bias exists in a strategy. The computation looks like this[6]: Trends in Data and Forecasting Methods 15 aMPE = a an at = 1 a(et / Yt a) an.

The MPE is predicted to be 0 if there is no bias found in the method. The forecasting approach is constantly underestimating if the calculated MPE is a high positive percentage. The forecasting approach, on the other hand, consistently overestimates if the MPE is a big negative number. When a model overestimates or underestimates, there are important management ramifications. For instance, an airline may incur more costs if it overestimates the sale of tickets. To meet such a rise of customers, the airline would erroneously buy more aircraft and recruit more staff. On the other side, there are financial repercussions if it overestimates seat sales. Underestimating reduces the negative cash flow associated with the acquisition of more aircraft and the employment of additional staff, which is expensive in terms of missed revenues.

The approach does not regularly overestimate or underestimate the sales of this e-commerce firm, according to the minor MPE of 2.29 percent. Evaluation of a forecast's accuracy in comparison to actual performance is another area of interest for management. The error between the ex post prediction and the actual observation determines the ex post assessment. In order to do this, the root mean square (RMS) and root percent mean square (R%MS) are utilised. Both equations and demonstrate how these mistakes are calculated.

RMS = n t = 1 e 2 t n

R%MS = n t = 1 (e 2 t / Yt) n

Similar to MAD, the root mean square is interpreted. The average error is calculated using the same unit as the original observation in RMS 16 Data Patterns and Forecasting Methods. The average error is expressed as a percentage and is known as the root percent mean square (R%MS). This error measurement is comparable to MAPE.

Today's forecasters have access to a wide range of forecasting methods. The technique of choosing is determined by the particulars of the circumstance as well as by the formal statistical standards that govern model selection within particular approaches. In general, accurate projections rely on the parsimony principle and a limited number of widely used tools and models. According to this rule, simple models win out over complicated ones when everything else is equal.

The models covered in this article are straightforward yet sophisticated enough to deal with the challenging commercial and economic situations seen in the real world. The goals of a forecast determine which approach is best to use. Most corporate managers want to make their judgements less ambiguous and be able to quantify the impact of different management initiatives. A good forecasting model reduces uncertainty for the decision-maker while simulating the effects of an uncertain future.

Before managers can utilise forecasting methods successfully, they need to be conversant with the main types of these approaches. Three types of forecasting techniques exist: quantitative models, qualitative models, and technology methods [7]. Statistical models and quantitative models are both unbiased methods of predicting. They rule the industry because they provide a methodical set of actions that are replicable and adaptable to a range of commercial and economic circumstances. Time series and regression algorithms are used in quantitative forecasting.

Time series models—more specifically, autoregressive models predict future values of a variable based only on previous observations of that variable. Another way to put it is that time is the independent variable in a time series model. For instance, if we believe that current vehicle sales are influenced by previous sales, we may formulate this time series model as:

Yt + 1 = Sales one time period in the future, where Yt + 1 = 0 + 1Yt + 1.

Yt = Revenue during the current time frame

Sales during the previous period equal Yt 1. Moving averages, exponential smoothing, time series decomposition, and Box-Jenkins are some of the autoregressive approaches. They are all covered in this book's later chapters. Economic theory provides the basis for regression or causal models. These models follow the theoretical connections between the variables as a reference. Regression models presuppose that one random variable's value (an "independent variable") may have an impact on another's value (a "dependent variable"). The model may be stated as follows in the context of our example on car sales:

Finding the precise shape of the connection between sales, pricing, and revenue is the objective of the regression model. To do this, estimations of the regression coefficients (0, 1, 2) that capture the structural or historical link between the dependent and independent variables are derived. On the other hand, nonstatistical or judgemental approaches to forecasting refer to qualitative ways of predicting. These methods mainly rely on professional judgement and the forecaster's intuition. When there is a lack of historical data, the qualitative methodologies are used. The Delphi process, the jury of executive opinion, sales force composite predictions, and focus group or panel consensus are the most popular of these methods.

The RAND Corporation created the Delphi method as a forecasting tool in 1963. A group decision-making procedure is used to determine the possibility that certain occurrences will occur. It is predicated on the idea that knowledgeable employees in firms have a better grasp of their industry and are better equipped to foresee future trends than theoretical techniques.

In the Delphi method, a group of experts responds to a series of anonymous questions. Before responding to the next series of questions, their replies are summarised. It is suggested that via this process of agreement, the group will converge on the "best" solution currently, it is also used for sales, marketing, and environmental forecasting.

The Delphi Method has the benefit of allowing the experts to live anywhere in the globe and not requiring them to physically come together. However, since the median represents the majority view, the procedure does not need unanimous agreement from all panellists. The anonymity of the comments helps eliminate issues with ego, dominant personalities, and the "bandwagon or halo effect" in responses. The method's drawback is that it might be challenging to maintain panellists for the several questionnaire rounds. Moreover, and probably more concerningly, the approach struggles to adapt to paradigm changes [8]. When skilled forecasters are used, the qualitative approaches, although missing the quantitative approach's rigour, do provide accurate predictions. The technological approach to forecasting includes the third group of forecasting methodologies. This group of techniques combines quantitative and qualitative methods to provide long-term forecasts. The model's goals are to react to sociological, political, economic, technical, and other developments in order to foresee them. As these models go beyond the purview of this book, we won't address them here.

After discussing the many forecasting models available, we'll move on to model selection and ways to gauge a model's accuracy. Business forecasters may use a variety of techniques to predict business and economic activity. These techniques include both qualitative and quantitative approaches. A management or researcher may choose from a number of different methods to generate a prediction, depending on the situation at hand, the time frame, the cost of doing so, and the forecast's degree of complexity. Methodologies that rely on a very subjective approach to predicting are referred to as qualitative approaches. While useful to forecasters, qualitative approaches cannot provide forecasts that can be standardised.

This strategy is said to have a flaw since it heavily relies on forecasters' judgement. It is not nearly as developed as its quantitative equivalent due to its reliance on the opinion of "experts" and the intuitive nature of qualitative procedures. As was already established, qualitative approaches include panel consensus, scenario analysis, Delphi, and market research. We shall exclusively discuss quantitative forecasting techniques in this work. The next chapters of this book will go into detail about these techniques. The forecaster must once again choose from a number of approaches for anticipating business and economic situations when using a quantitative methodology. There are no precise rules for matching a methodology with an economic or commercial issue since some of these approaches are so new. Yet, the forecaster must take into account both the features of the forecasting method and the characteristics of the decision-making scenario for which a forecast is to be created while choosing a methodology.

The time horizon, the application of the prediction for planning vs. control, the degree of detail needed, and the market's economic environment, i.e., stability or state of flux, may all be considered features of a decision-making scenario. The amount of time into the future for which the prediction is sought is referred to as the time horizon. Business interests often fall into four categories: immediate term (less than a month), short term (between one and six months), intermediate term (between six months and two years), and long term (more than two years). The modelling approach will probably change depending on the time horizon [9].

If the decision maker is interested in using the forecast for planning avs. Control, then there are implications for design, use, and evaluation of athe forecasting model. If the prediction is used for planning, such as when the decision-maker wants to know whether the present market situation will continue in the long run. If so, what trends are likely to recur, and how can the company benefit from knowing these trends? On the other hand, if management intends to use its prediction for control, they would prefer to rely on a model that can identify, at an early stage, if market circumstances have changed in a way that is unfavourable to the business. As a result, the model used should be able to spot pattern alterations, or more specifically, early turning points. Turning points are moments when a commercial or economic state changes course. For instance, we notice a turning point in the economy when it shifts from growth to decline. The amount of information needed in a prediction has a significant impact on the technique selection. In their decision-making processes, businesses use aggregate and highly disaggregated data. A production manager, for instance, would be concerned in the product lines offered by his or her organisation, but a corporate planning department would be primarily focused in corporate sales (very aggregate) (highly disaggregate) [10], [11].

CONCLUSION

Forecasting under uncertainty is a challenging but essential process for decision-makers and planners in today's rapidly changing world. Various methods and approaches, such as probabilistic forecasting, scenario planning, and simulation modeling, have been developed to incorporate uncertainty into the forecasting process. These methods aim to provide decision-makers with a range of potential outcomes and risks to inform better decision-making. However, forecasting under uncertainty remains an ongoing challenge that requires continued research and innovation to improve forecast accuracy and enhance decision-making. Organizations and individuals that can effectively manage uncertainty in their forecasting processes are better positioned to adapt to changing environments and make informed decisions that drive success.

REFERENCES

[1] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis-A survey," *IEEE Trans. Robot.*, 2014, doi: 10.1109/TRO.2013.2289018.

[2] R. V. Ramasesh and T. R. Browning, "A conceptual framework for tackling knowable unknown unknowns in project management," *J. Oper. Manag.*, 2014, doi: 10.1016/j.jom. 2014.03.003.

[3] M. Steinberger, M. Horn, and A. Ferrara, "Adaptive Control of Multivariable Networked Systems with Uncertain Time Delays," *IEEE Trans. Automat. Contr.*, 2022, doi: 10.1109/TAC.2021.3083563.

[4] B. Bouchaut and L. Asveld, "Responsible Learning About Risks Arising from Emerging Biotechnologies," *Sci. Eng. Ethics*, 2021, doi: 10.1007/s11948-021-00300-1.

[5] M. M. Alyannezhadi, A. A. Pouyan, and V. Abolghasemi, "An efficient algorithm for multisensory data fusion under uncertainty condition," *J. Electr. Syst. Inf. Technol.*, 2017, doi: 10.1016/j.jesit.2016.08.002.

[6] G. N. R. Cundill, C. Fabricius, and N. Marti, "Foghorns to the future: Using knowledge and transdisciplinarity to navigate complex systems," *Ecol. Soc.*, 2005, doi: 10.5751/ES-01444-100208.

[7] A. S. Dina and D. Manivannan, "Intrusion detection based on Machine Learning techniques in computer networks," *Internet of Things (Netherlands)*. 2021. doi: 10.1016/j.iot.2021. 100462.

[8] H. Su, C. Zhou, Y. Cao, S. H. Yang, and Z. Ji, "An intelligent approach of controlled variable selection for constrained process self-optimizing control," *Syst. Sci. Control Eng.*, 2022, doi: 10.1080/21642583.2021.2024916.

[9] M. Lahami and M. Krichen, "A survey on runtime testing of dynamically adaptable and distributed systems," *Softw. Qual. J.*, 2021, doi: 10.1007/s11219-021-09558-x.

[10] A. H. Tahoun and M. Arafa, "A new unmatched-disturbances compensation and fault-tolerant control for partially known nonlinear singular systems," *ISA Trans.*, 2020, doi: 10.1016/j.isatra.2020.05.007.

[11] C. Nemeth, P. Fearnhead, and L. Mihaylova, "Sequential monte carlo methods for state and parameter estimation in abruptly changing environments," *IEEE Trans. Signal Process.*, 2014, doi: 10.1109/TSP.2013.2296278.

CHAPTER 13

GENERAL LINEAR REGRESSION MODEL: A COMPREHENSIVE FRAMEWORK FOR MODELING RELATIONSHIPS AND FORECASTING OUTCOMES

Satyendra Arya, Associate Professor
Teerthanker Mahaveer Institute of Management and Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India
Email Id- satyendra_arya17@rediffmail.com

ABSTRACT:

The General Linear Regression Model is a statistical framework used to model the relationship between a continuous dependent variable and one or more independent variables. The model assumes a linear relationship between the dependent variable and the independent variables, with the coefficients of the independent variables representing the degree of influence they have on the dependent variable.

KEYWORDS:

Coefficients, Dependent Variable, General Linear Regression Model, Independent Variables, Linear Relationship.

INTRODUCTION

The approach used for predicting also much depends on the market's condition (whether it is stable or not). Forecasting investment decisions has become particularly challenging due to recent stock market volatility in the US (from October 2007 to the present). A system that can be continuously adjusted to the most current findings and information available is required under market situations like these. On the other hand, stable market circumstances demand for a forecasting technique that can be frequently evaluated to assess the model's usefulness.

When choosing a method, the aforementioned qualities are by no means all-inclusive. The amount of goods or product lines a firm may have, as well as its current planning processes, might have an influence on the forecasting approach that is chosen. A corporation must take into account the peculiarities of the different forecasting methodologies in addition to these aspects of a forecasting scenario. They include the length of the prediction, the distribution of the data, the kind of model, the price of the model, the degree of accuracy, and the simplicity of application [1].

The forecast horizon indicates to the forecaster the decision maker's preferred time range. Depending on whether a company is interested in a prediction for 1 month into the future, 1 year into the future, or 5 to 10 years into the future, several approaches might be utilised. For longer-term forecasts, forecasters often utilise qualitative approaches; for intermediate and shorter-term forecasts, they typically employ quantitative methods. When compared to other data, some may exhibit a seasonal trend or cyclical pattern. An average value and random variations around it make up data patterns and forecasting techniques. When one considers that trend patterns are long-term trends whereas seasonal fluctuations are periodic and repetitive throughout a year, the relationship between data pattern and time horizon becomes clear. It's crucial to keep in mind that certain strategies are more appropriate for a specific data pattern than others. Regression models, for instance, can handle almost all data patterns, but autoregressive techniques work better with time series that have few turning points.

A successful prediction depends on the kind of forecasting model utilised. It is the forecaster's obligation to pay close attention to the model's underlying assumptions since they differ amongst models. By using a model that can reliably foresee turning moments as opposed to projecting steady patterns of change, the prediction's result is quite different. Similar to this, some models are better at predicting short-term change than they are at predicting long-term conditions. These examples just highlight the significance of choosing a method whose underlying assumptions fit the properties of the data.

One must take into account the model's cost while choosing an acceptable one for predicting. While choosing a method, a number of expenses need to be taken into account. They include the price of internal data creation, the time needed to simulate different scenarios and update the model, the cost of conceiving and building a forecasting model (the labour and expert guidance needed to design a model), and lastly the time needed to compile the forecasting report [2].

The required degree of accuracy is another consideration when choosing a forecasting model. A general forecast of a market's future demand may be all that management is interested in. In some situations, a qualitative or judgmental approach may be sufficient. On the other hand, a more complex model that can appropriately address management's query is acceptable if management is interested in making investment choices that need a credible prediction of the long-term trend. The predicted financial worth of the choice is closely related to the forecast's accuracy.

Lastly, choosing a prediction model is influenced by how simple it is to use. Researchers and managers are interested in using methods that are simple to use while still being able to provide precise projections. The parsimony concept was used in this work to create forecasting models. As a result, we begin with the simplest model and technique and progress to more intricate ones. Simple models are acknowledged to be useful and frequently used in a variety of business and economic situations. The manager must also comprehend and be at ease with the methodology they select.

The reader is given access to a number of quantitative methodologies in this text, which they can then use to solve a variety of business and economic problems. The steps for creating accurate business and economic forecasts will be outlined in later chapters. Nations seek economic growth, full employment, and price-level stability as their major macroeconomic goals. Business decisions are made in the context of the macroeconomic of a nation. Firms take their cues from the economic environment in which they operate. When favorable macroeconomic conditions prevail, businesses expand operation and output. Contractions in the economy generally lead to slower growth patterns. Long-run economic growth has not always been steady as factors such as inflation, unemployment, recession, and depression have impacted it negatively. As was pointed out long-term trends, seasonal variations, cyclical movements, and irregular factors combine to generate widely divergent paths for businesses ain an economy. Given these conditions, how are businesses to predict future growth and contractions in the economy? In this chapter we will discuss specifically those cyclical linkages that bind the economy together during a typical business cycle, and how our knowledge of these linkages helps us in making good forecasts at the industry and firm level.

DISCUSSION

Economists use the term "business cycle" when they observe alternating increases and decreases in the level of economic activity, sometimes extending over several years. The duration and intensity of the cycle vary from cycle to cycle. Yet all business cycles display common phases, which are variously labeled by economists. A business cycle generally follows the four phases of peak, recession, trough, and recovery. Peak refers to the economic condition at which business activity has reached a temporary maximum and shows the onset of a recession or upper turning point. During this phase the economy is at full employment and the level of real output is at, or very close to, its capacity [3].

It is expected that the price level will rise during this phase of the business cycle. Recession follows a peak in the economy. In this phase a decline lasting more than 6 months in total output, income, employment, and trade will be noted. This downturn is marked by the widespread contraction of the business in many sectors of the economy. Because of inflexibility of downward price movement, one does not expect prices to fall unless a severe and prolonged recession or depression occurs. Trough refers to that phase of the business cycle where output and employment "bottom out" at their slowest levels. It is also called the lower turning point. The duration of the through may be short-lived or very long. In the recovery phase, output and employment surge towards full employment, and as recovery intensifies, the price level may begin to rise before full employment is reached.

Business cycles are products of unique series of historical events. Theories of business cycle vary widely in terms of detail and emphasis. The majority of them assume that the internal dynamics of a market economy lead to the observed regularity of fluctuations in aggregate economic activity. Several major studies have been conducted to determine the causes of business cycles and whether fluctuations in the macroeconomic have changed over time. If it is true that the fluctuations over time have been less volatile and more predictable than the past, then it gives the forecaster some hope that the industry or firm level forecasts would be more consistent. To measure economic volatility, Romer (1999) used the standard deviation of growth rates for the various macroeconomic indicators such as industrial production, GNP, commodity output, and unemployment rate.

What stands out from her findings is that there is extreme volatility in the interwar period. While she contributes most of this volatility to the Great depression (1929–1933), she notes also that there were also extreme movements in the early 1920s and the late 1930s. Second, there is some similarity of the volatility in the pre-World War I and post-World War II areas. Based on the four indicators that has been used, it appears that the volatility of ate U.S. macroeconomic has declined by 15 percent to 20 percent between the pre-1916 and the post-1948 eras. This basic similarity in volatility in the Standard Deviation of Percentage Changes Series 1886–1916 1920–1940 1948–1997 [4].

For the commodity output series, the interwar sample period stops in 1938 and the postwar sample period stops in 1996. For the unemployment series, the prewar sample period covers only the period 1900–1916 and consistent interwar data are not available. The standard deviation for the unemployment rate is for simple changes and so is expressed in percentage points rather than percent.

The Macro economy and Business Forecasts 37 prewar and postwar eras was also echoed in the findings of Sheffrin (1988) and Shapiro (1988). (1988). other studies in the 1970s and early 1980s reported declines in annual volatility of 50 percent to 75 percent (Baily, 1978; DeLong and summers, 1986). Forecasters can take comfort in knowing that the use of various monetary and fiscal stabilisers has contributed to less volatility in the macroeconomic.

Forecasters would also like to have more knowledge about the frequency and duration of recessions. Firm level expansions or contractions are costly investment choices. As a forecaster you must be able to provide top management in your organization with your prediction of a near-term recession or recovery in the economy. How does this foreseeable recession or recovery impact on your industry?

To analyse changes in the frequency and duration of recessions, Romer at (1999) has developed a new series of pre-World War II peaks and troughs. What her study shows is that recessions have not become noticeably shorter over time. She observes that the average length of recession is actually a 1 month longer in the post-World War II era than in the pre-World War I aera. Furthermore, there is no obvious change in the distribution of the length of recession between the prewar and postwar eras. It appears that most recessions lasted from 6 to 12 months during the time frame of the study [5].

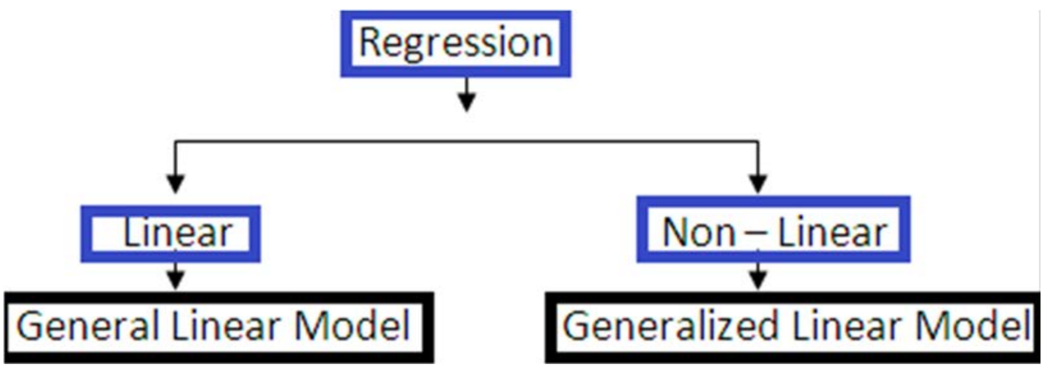


Figure 1: Illustrate the General Linear Model.

The Macroeconomy and Business Forecasts average during the volatile interwar period than in the postwar era. As late as December 1998, the U.S. economy had been expanding for 93 months. Adding this lengthy expansion raised the average postwar length to 56.1 months, aor 65 percent longer than the typical prewar expansion. These and other historical and cyclical insights provide invaluable background to the forecaster when the next recession looms on the horizon. As was pointed out earlier, many external factors contribute to each recession. However, there remains the basic continuity in the endogenous or internal patterns of economic activity within the business cycle that serves as a guideline. What is important to remember is that a forecaster must carry out detailed statistical analyses to segregate the unique factors affecting industries and individual firms from the general forces propelling commodity output, GNP, and employment. Figure 1 illustrate the General Linear Model.

Macroeconomic theories focus attention on the behavior of the economy as a whole. Given the real-world complexities and that many variables interact with each other in shaping the economy, it would be of value to forecasters to have the tools necessary to observe the interrelationships in the overall economy as well as the means to measure the impact of these variables on one another. It is in this sense that macroeconomic models could be looked upon as general equilibrium models in contrast to partial equilibrium models a(often used at the firm level) where single equation models are used to explain causality only in one direction.

In building macroeconomic models, the forecaster depends on economic theory and statistical techniques. Since these models portray the interrelationships of variables in the economy, the forecasts obtained from them are simultaneous. This means that the behavior of the variables is jointly determined [6]. Macroeconomic models have played an important role in serving as a tool in guiding economic policy. Economists such as Klein and Goldberger (1955) were the first to use a macroeconomic model for the American economy. In their model (known as the K/G model), they used Keynesian real aggregate demand to explain which components are determined endogenously and which ones are determined exogenously. The K/G model contained approximately two-dozen equations to link the variables of interest in the aggregate demand model.

The work of Klein and Goldberger guided other economic researchers in building macroeconomic models. Examples are the work of Suits (1962) and the models developed by the Office of Business Economics of the Department of Commerce. What is apparent from this first generation of macroeconomic models is that the interrelationships between the financial and real sectors were not emphasized. The second generation of these models not only included the financial sector in the model, but also expanded on the relationships of these models with input/output analysis. The second-generation models included the Brookings model, the FRB/ MIT model, and the Wharton Mark III model.

In addition to the models developed by the academic institutions and governmental agencies, commercial econometric forecasting services have flourished. Among them are the Chase Econometrics that provides services to major corporations and other paying clients. The models developed by these groups all seem to be complex and far reaching in their capability. They provide forecasts that link various sectors of the economy as well as deliver disaggregated forecasts on output and prices.

Available computer technologies and information systems (databases) allow today's forecasters to develop models that appropriately handle the interrelationships at the macroeconomy and link the firms to industry, and industry to the overall macroeconomic model. To see how these units of the economy are linked with one another, consider this simple diagram of the economy. In any economy, exogenous factors such as the monetary and fiscal policy of a nation as well as a wide range of other factor such as weather, wars, and labor strikes influence what happens in an economy. Endogenous variables also impact on the overall performance of the macroeconomy. These variables refer the Macroeconomy and Business Forecasts to those factors that are determined within the system. Sales and advertising expenditures within a firm are examples of the endogenous variable. In addition to these factors, the interaction between the firms within an industry and the role of the industry within the economy have to be incorporated into the macroeconomic model [7].

Given these interrelationships that exist between the macroeconomy and individual firms within the economy, forecasts that are generated by the macroeconomic models are quite useful to firms. Often the forecasted variables of the macro model are used as independent variables for making forecasts for the firm or industry. One cannot help but notice that, when building macroeconomic models, the forecaster faces a very large system of simultaneous nonlinear equations, as driven by economic theory. The task of collecting data, estimating the model, and solving the model to obtain forecasts requires skilled personnel and qualified judgment. Economists have pointed out that, however excellent the econometric model might be, the quality of any unconditional forecasts of endogenous variables will depend on the quality of forecasts of the exogenous variables used (Ashley, 1985, 1988). In this sense, the use of judgement in the Macroeconomy and Business Forecasts 41 prediction of future values of the exogenous variables is inevitable. Judgment is also used when the forecaster tries to specify the model. Other conditions that lead to the use of expert judgment are when an estimated model is inconsistent with strongly held prior beliefs.

Because of the changing economic environment, models have to be constantly revised so that appropriate forecasts are made. The complexity of the macroeconomic models calls for seasoned forecasters with the requisite skills. This, as well as the data collection and processing involved, are the reasons why firms depend on commercial forecasting entities. Economists have debated the value of macroeconomic models that are highly complex and yet have not produced forecasts that are useful. Most believe that, in building models, the principle of parsimony should be kept in mind. Proponents of the econometric approach would argue that, if models are appropriately built, they should be able to produce good forecasts. Given the differences of opinion among economists as to whether macroeconomic models or the simple ARIMA models provide better forecasts, it is wise to assess the quality of a forecast through comparison of competing forecasts. When asked to provide a forecast of macroeconomic variables such as gross national product, price inflation over the next 3 or 4 years, or the interest rates, it is advisable to consult with one or more of the model-building groups that have established a track record over the years [8].

In summary, individual firms or industries that use macroeconomic variables in making predictions of sales, revenues, inventories, etc., would have to consider how these macroeconomic variables were estimated. If forecasts are prepared in-house, a number of methodological approaches that will be discussed in subsequent chapters of this book could be adopted. Business decisions are made in the context of the overall economy. Firms base their decision on investment choices in production, distribution, and marketing on the economic outlook for the country. Long-term trends, seasonal patterns, cyclical movements, and irregular factors all play an important role in how an economy performs. Understanding how the various sectors of the economy are linked together is essential in making forecasts.

In discussing the cyclical linkages that bind the economy together during a business cycle, it was mentioned that firms take their cues from the economic environment in which they operate. When favourable macroeconomic conditions prevail, businesses expand operation and output. Contractions in the economy generally lead to slower growth patterns. Long-run economic growth has not always been steady as factors such as inflation, unemployment, recession, and depression can impact on it negatively. We mentioned that understanding the various phases of the business cycle is extremely helpful to forecasters. The four phases of the business cycle allow us to see how the economy will behave during a peak, recession, trough, or a recovery [9].

The historical picture of the business cycles in the U.S. presented in this chapter point to some very interesting conclusions fluctuations over time have been less volatile and more predictable than the past. To measure economic volatility, Romer (1999) used the standard deviation of growth rates for the various macroeconomic indicators such as industrial production, GNP, commodity output, and unemployment rate. Based on these indicators, it appears that the volatility of the U.S. macroeconomy has declined 15 percent to 20 percent between the pre-1916 and the post-1948 eras. Forecasters can take comfort in knowing that the use of various monetary and fiscal stabilisers has contributed to less volatility in the macroeconomy. Recessions over time have not become noticeably shorter. It was found that the average length of recession is actually 1 month longer in the post-World War II era than in the pre-World War I era. On the expansion side, Romer (1999) found that the length has expanded over time. It appears that the average time from a trough to the next peak is about 50 percent longer in the postwar period than in the prewar period [10], [11].

CONCLUSION

The General Linear Regression Model is a widely used statistical framework that helps to model the relationship between a continuous dependent variable and one or more independent variables. The model assumes a linear relationship between the variables, and its coefficients provide insight into the degree of influence each independent variable has on the dependent variable. The model is a powerful tool that can be used to analyze and interpret complex relationships between variables, make predictions and forecasts, and perform hypothesis testing. It can also be extended to include additional explanatory variables, such as categorical variables, interactions, and polynomial terms, making it a versatile tool in a wide range of fields, including economics, psychology, biology, and social sciences.

REFERENCES

[1] N. F. Che Nordin, N. S. Mohd, S. Koting, Z. Ismail, M. Sherif, and A. El-Shafie, "Groundwater quality forecasting modelling using artificial intelligence: A review," *Groundwater for Sustainable Development*. 2021. doi: 10.1016/j.gsd.2021.100643.

[2] C. Prentice, "Testing complexity theory in service research," *J. Serv. Mark.*, 2020, doi: 10.1108/JSM-09-2019-0353.

[3] K. Panuwatwanich, S. Al-Haadir, and R. A. Stewart, "Influence of safety motivation and climate on safety behaviour and outcomes: evidence from the Saudi Arabian construction industry," *Int. J. Occup. Saf. Ergon.*, 2017, doi: 10.1080/10803548.2016.1235424.

[4] S. Iyer and A. Mahajan, "Granger causality analysis of deviation in total electron content during geomagnetic storms in the equatorial region," *J. Eng. Appl. Sci.*, 2021, doi: 10.1186/s44147-021-00007-x.

[5] S. Ullah and C. F. Finch, "Applications of functional data analysis: A systematic review," *BMC Medical Research Methodology*. 2013. doi: 10.1186/1471-2288-13-43.

[6] S. W. Fleming, V. V. Vesselinov, and A. G. Goodbody, "Augmenting geophysical interpretation of data-driven operational water supply forecast modeling for a western US river using a hybrid machine learning approach," *J. Hydrol.*, 2021, doi: 10.1016/j.jhydrol.2021.126327.

[7] D. A. Shah, P. A. Paul, E. D. De Wolf, and L. V. Madden, "Predicting plant disease epidemics from functionally represented weather series," *Philos. Trans. R. Soc. B Biol. Sci.*, 2019, doi: 10.1098/rstb.2018.0273.

[8] A. Sahoo, S. Samantaray, and D. K. Ghose, "Prediction of Flood in Barak River using Hybrid Machine Learning Approaches: A Case Study," *J. Geol. Soc. India*, 2021, doi: 10.1007/s12594-021-1650-1.

[9] O. A. Awe, A. G. Woodside, S. Nerur, and E. Prater, "Capturing heterogeneities in orchestrating resources for accurately forecasting high (separately low) project management performance," *Int. J. Prod. Econ.*, 2020, doi: 10.1016/j.ijpe.2019.107556.

[10] C. M. Fernández Peruchena *et al.*, "Dynamic Paths: Towards high frequency direct normal irradiance forecasts," *Energy*, 2017, doi: 10.1016/j.energy.2017.05.101.

[11] L. T. Al-Bahrani, M. Seyedmahmoudian, B. Horan, and A. Stojcevski, "Distribution transformer load behavior, burden, and characteristics of residential consumers: A case study of Baghdad City," *Energy Build.*, 2020, doi: 10.1016/j.enbuild.2019.109693.

CHAPTER 14

A FRAMEWORK FOR MANAGING OUTLIERS AND ENSURING DATA ADEQUACY IN MULTIVARIATE DATA ANALYSIS

Chanchal Chawla, Professor
Teerthanker Mahaveer Institute of Management and Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India
Email Id- chanchalchawla0@gmail.com

ABSTRACT:

The presence of outliers in a dataset can have a significant impact on statistical analysis and modeling. Outliers are data points that differ significantly from other observations and may be the result of measurement errors, data entry mistakes, or rare events. Ignoring outliers or treating them as errors can lead to biased and inaccurate results. Therefore, outlier detection and treatment is an essential step in data analysis. Addressing outliers and issues of data adequacy is critical for obtaining meaningful results from statistical analysis. It requires a thorough understanding of the data, appropriate statistical techniques, and careful interpretation of findings.

KEYWORDS:

Data Adequacy, Imputation, Measurement Error, Outliers, Sample Size, Statistical Analysis, Statistical Power.

INTRODUCTION

It is interesting to note that expansions were somewhat shorter on average during the volatile interwar period than the postwar era. As late as December 1998, the U.S. economy had been expanding for 93 months. Adding this lengthy expansion raised the average postwar length to 56.1 months, or 65 percent longer than the typical prewar expansion. In making predictions of the GNP, inflation, interest rates, and other macroeconomic variables forecasters depend on macroeconomic models. These models serve as tools in describing the interrelationship that exists between variables of the economy as well as measure the impact of these variables on one another.

Macroeconomic models are powerful tools in making forecasts. However, because of their complexity and costliness, they are not recommended for use by individual firms. The verdict is still out as to how accurately these models predict future outcomes. Though driven by theory, the outcome of these models depends on the judgement and skills of the forecaster. The significant economic downturn in 2008 pointed to the complexity of the macroeconomic variables [1]. Economists Klein and Goldberger were the first to use a macroeconomic model for the American economy. As a policy tool, these models make significant contributions to the economy. Other researchers have expanded on the work of Klein/Goldberger by building models that show the relationship between input/output analysis and the macroeconomic models. Because of the complexity and cost associated with building these models, several groups including commercial firms have offered corporate clients forecasting models for the economy.

We'll look at techniques for evaluating data over time in this chapter. Planning for future developments is something that both researchers and management are interested in. Collecting the data is a necessary initial step in any study. In your role as a forecaster, you must remember that you would need to carefully evaluate the data to discover its source and the conditions under which it was collected unless you had gathered the data for a specified purpose. You would need to exercise caution in their usage depending on the type and operationalization of the data. For instance, because of how they have defined unemployment, the monthly unemployment statistics obtained by certain government agencies may not indicate any changes in the unemployment rate. In certain circumstances, a person must be let go for a certain number of weeks before being considered jobless [2].

It is crucial to bear in mind the data qualities that were emphasized in before using a strategy in predicting. Every management or researcher would need to pay close attention to these first phases in forecasting since we see data modifications and transformation as an essential component of forecasting. We will look at data transformation and modifications in then we'll look at the data patterns seen in a time series. The chapter is concluded by talking about how a time series is traditionally broken down. Researchers that study forecasting often use the aggregate data gathered by regional or federal government organizations. In circumstances when one wants to utilize these databases, the information could not immediately address the management's query.

Data must be modified and altered in order to be utilized for predicting under such circumstances. Time series data, for instance, may not be unique to management's requirements and could have been collected using the current dollar rather than a constant dollar. In place of information on aggregate output, for example, forecasters would want more detailed data. A management could be more concerned with the industrial consumer of a product than the end consumer. In situations like this, the forecaster may not benefit from the often provided aggregated data that just includes the ultimate customer. We'll go through the three business situations that lead to data modifications most often. These are the adjustments for the trading day, adjustments for price changes, and adjustments for population changes.

DISCUSSION

Companies rely on revenues, and sales depend on how many days the company is open each month. The number of business days is used by financial organizations to calculate the costs of borrowing and lending. Every businessperson will understand that there might be a 15% variation in the number of days between months. Examples of this change include the amount of days from January to February and from February to March. Although the actual quantity of sales or borrowing may not have changed, it is crucial to note that the recorded statistics will vary from month to month, whether it be the income from sales or the cost of borrowing [3].

We may change the statistics for the amount of working or trading days in a month once we keep this in mind. These two methods may be used by forecasters to change the working or trading days. The average daily number is the name of the first adjustment technique. It is calculated by dividing the monthly total by the number of working days, whether it be for production, sales income, expenses, etc.

For instance, we may change our data to represent the number of working or trading days in a month based on the information provided in Table 4.1. Assume that we have data on computer sales for each month for the years 2005, 2006, and 2007 as shown in Table 4.2. Let's look at the sales volume for February and March 2007 to better understand how the number of working days affects the amount of sales in each month. Computer sales increased from 5 million units in February to 6 million units in March, as can be seen. The sales have significantly increased by 20%.

Yet when we include in the amount of working days in each month, the gain in sales does not seem to be as big as it formerly was. February and March 2007 would have daily unit sales of 250,00 units (5 million/20 days) and 272,727 units (6 million/22 days), respectively. Although the sales volume is greater in March than it was in February, it is just 9.1%, which is much lower than the 20%. To develop short- and long-term business projections, today's decision-makers in business rely significantly on computers and databases.

The purpose of these business projections is to inform decision-makers about the possibility of profits and losses. In order to effectively do business in their respective industries, whether we are examining a little or big retail shop, an airline, a massive energy production, or a significant financial institution, short-term projections are used by all of them. The short-term forecasting technique's main benefit is its simplicity. It is crucial to remember that accuracy does not have to be sacrificed for simplicity. Moreover, we may assess the applicability, dependability, and need of the more complex models using these simple models as a standard. Yet, choosing a suitable model relies more on how well it can reproduce the pattern shown by the real historical data than it does on how sophisticated the model [4].

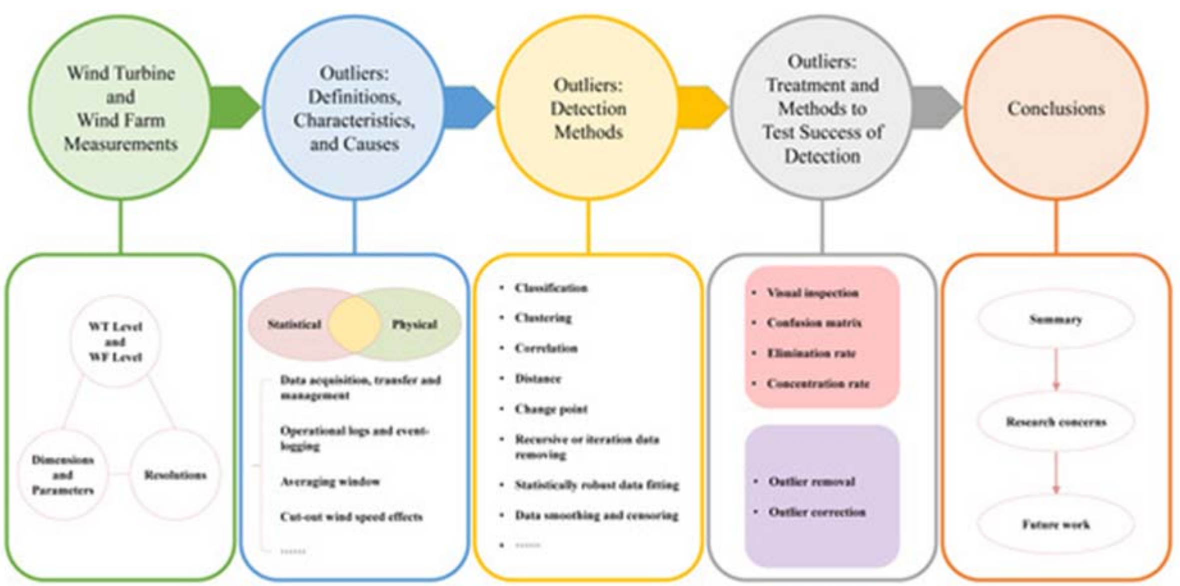


Figure 1: Illustrate the Detection and Treatment of Outliers in Processing Wind Turban.

This chapter will examine the short-term forecasting methods that may be used for projections that are weekly, monthly, and quarterly. The procedures are known as "smoothing approaches." These methods are easy to apply and straightforward, which makes them quite beneficial for managers.

By using these models, forecasting experts are able to differentiate between random fluctuations and the fundamental pattern in the data. Figure 1 illustrate the Detection and Treatment of Outliers in Processing Wind Turban.

The three different categories of forecasting techniques—naive, averaging, and smoothing will be covered. Each of these models has benefits and drawbacks of its own. If the pattern of the data is such that there is not much variation from one time period to another, the simplicity of the nave model makes it particularly effective for a rapid prediction. The model's drawback is that it can only predict the future based on the recent past, and data might include a lot of unpredictability.

Since they consider a full time series and its variations, averaging models are superior to the naive model. In the models that employ averaging, we just take a collection of observed values, calculate their average, and then use this average as a prediction. The analyst must have as many historical data points as the moving average requires for the moving average models. Hence, a 4-month average cannot be calculated until the end of period 4 to predict period 5, and a 5-month moving average cannot be computed until the end of period 5 to forecast period 6 [5].

The benefit of moving average models is that each new observation's projected value is based on a fresh average that is calculated. The moving average models also have the benefit of smoothing forecasts more effectively the more observations there are in the moving average. Later on, it is described how to decide whether to utilise more or less observations when calculating the moving average. These models are constrained by the need to keep the most recent N observed data in order to compute the moving average. That can take up a lot of room. The second drawback is that the moving average models give no weight to observations made before period and equal weight to each of the most recent N observations (t–N).

The exponential smoothing techniques provide a weighting system that would give the highest weight to the most recent observed values and diminishing weights to the recorded values of the distant past, overcoming the drawbacks of moving average models. Moreover, these models do away with the need for retaining previous values of the variable. We go into great depth on each of these models in the sections that follow, using examples to show how to utilise them. The accuracy of today's projections has a big impact on management choices. Fast market developments need prompt and adequate management response. The short-term forecasting techniques covered in this chapter serve as the foundation for projections that may be employed, for instance, in the production, finance, and inventory management divisions.

The simplest method of predicting, utilising the naive model, was used to start the chapter. When real historical data changes extremely gradually and with few turning points, this approach works well. The model's fundamental premise is that recent eras provide the most accurate forecasts of the future. The method's flaw is that it ignores everything that has happened since last year and fails to take into consideration any potential trends in the data. A suitable forecasting approach would be to employ an averaging method when managers are faced with economic and commercial situations that call for daily, weekly, or monthly forecasts of inventories. Simple, moving average, and double moving average are the three most used averaging techniques [6].

If the data are thought to be steady, then this approach is suitable. On the other hand, the moving average approach should be taken into consideration if the analyst is concerned with the most recent data observations. This method gives every observation the same weight as it becomes available. The method does not handle data series with trend and seasonality and is most effective with stationary data. When the data displays both randomness and a linear trend, the double moving average approach is advised. This approach uses the first moving average to construct a second moving average. To get a prediction for the next period, the difference between the simple moving average and the double moving average is added to the simple moving average along with the trend estimate. Compared to the basic and moving average approaches, this methodology is less affected by outliers since it considers significant random fluctuations. One of its drawbacks is that it does not take into account any potential seasonality in the series, and that there are computationally easier approaches like exponential smoothing.

With the help of more recent data, estimates may be continuously revised using exponential smoothing techniques. The more recent observations are taken into account over older ones. In addition to allowing for trend and being computationally more effective than the double moving average, exponential smoothing models also need less data than the double moving average does, which is 134 Forecasting with Smoothing Methods. Moreover, some flexibility is lost since the level and trend's smoothing constants could not be the same. Without initially deseasonalizing the data, this practise is not advised.

Brown's approach allows for the flexible use of two smoothing constants: one for the trend and one for the series level. When the data exhibits a linear trend, this approach is advised. The selection of the beginning values for the smoothed series and trend correction are issues with the Brown's double exponential smoothing method.

Last but not least, we spoke about winters' seasonal exponential smoothing. This model is an extension of Holt's model in which the model takes into consideration the trend and seasonality that a series may display in addition to the exponentially smoothed series. The model's seasonality component means that more data is needed than with the other approaches covered in this chapter. It can need more calculation time to simultaneously determine the three smoothing constants' ideal values than to use a regression model [7].

The European Monetary Union (EMU) was established with the goal of stabilising prices and minimising the uncertainty brought on by exchange rate swings for the benefit of investors, exporters, and importers. Theoretically, monetary unification has a variety of positive effects on a nation's economy. The first benefit is the removal of exchange rate risk with other members of the monetary union, which makes trading between them easier. Second, it strengthens competition by increasing the transparency of pricing discrepancies across member nations.

Thirdly, it may improve policy discipline; particularly, by giving control over monetary policy to a regional central bank, a national central bank may gain greater confidence in its dedication to maintaining price stability. The main expense of establishing a monetary union, however, is connected to this third gain. A country's central bank loses autonomous monetary policy control and, therefore, the capacity to stabilise the economy when it is struck by a shock when monetary policy power is transferred to a regional central bank. Depending on how high the cost is, the advantages of entering a monetary union can exceed the disadvantages.

The cost or need for independent monetary policy management is larger when member nations are subjected to various shocks and lessened when they are exposed to the same or comparable shocks, according to the optimal currency area theory (Mundell, 1961). High commercial integration among member nations is one element that lowers the chance of various shocks. A system of intraregional fiscal transfers and high labour mobility are two additional factors that reduce the cost.

In January 1999, the European Union officially launched its monetary union in the hopes that it would lead to price and exchange rate stability, convergence of interest rates, and fiscal restraint some people feared that the union would cause more issues for the 11 member countries, while others imagined more price stability and lower inflation for the member countries. By the year 2003, ait was apparent that the record of European Union was better than what the acritics had feared [8]. The European Monetary Union has played an important role in the monetary affairs of the member nations. According to Becker (2003) general government debt fell from a peak of 75 percent of GDP at the end of 1996 ato 69.5 percent in 2001. Additionally, EMU has improved conditions for growth. When compared to the U.S. in terms of size and other indicators, the EMU has given rise to a major economic area with significant potential for growth [9]–[11].

CONCLUSION

In the financial markets, it has served as an engine of integration and a catalyst for structural change. Many had expected the surrendering of national sovereignty in monetary, exchange rate and fiscal policy to trigger considerable increase in willingness to address structural reforms, but this has not been the case. The divergences in national inflation and growth rates which reflect lack of uniformed government policies have widened since the inception despite its single monetary and exchange rate policies. It appears that the integration of money and bond markets has progressed more quickly than that of the equity markets.

REFERENCES

[1] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, "Multivariate Data Analysis," *Vectors*. 2010. doi: 10.1016/j.ijpharm.2011.02.019.

[2] C. Beckett, L. Eriksson, E. Johansson, and C. Wikström, "Multivariate Data Analysis (MVDA)," in *Pharmaceutical Quality by Design: A Practical Approach*, 2017. doi: 10.1002/9781118895238.ch8.

[3] J. Thioulouse and S. Dray, "Interactive multivariate data analysis in R with the ade4 and ade4TkGUI packages," *J. Stat. Softw.*, 2007, doi: 10.18637/jss.v022.i05.

[4] J. F. Hair, W. C. Black, Jr, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*. 2019.

[5] C. Maione, F. Barbosa, and R. M. Barbosa, "Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: A review," *Computers and Electronics in Agriculture*. 2019. doi: 10.1016/j.compag.2019.01.020.

[6] C. Maione and R. M. Barbosa, “Recent applications of multivariate data analysis methods in the authentication of rice and the most analyzed parameters: A review,” *Critical Reviews in Food Science and Nutrition*. 2019. doi: 10.1080/10408398.2018.1431763.

[7] C. Buvé *et al.*, “Application of multivariate data analysis for food quality investigations: An example-based review,” *Food Research International*. 2022. doi: 10.1016/j.foodres.2021.110878.

[8] S. Goldrick *et al.*, “Multivariate Data Analysis Methodology to Solve Data Challenges Related to Scale-Up Model Validation and Missing Data on a Micro-Bioreactor System,” *Biotechnol. J.*, 2020, doi: 10.1002/biot.201800684.

[9] N. B. Pedersen, F. Zaefarian, A. C. Storm, V. Ravindran, and A. J. Cowieson, “Mathematical prediction of ileal energy and protein digestibility in broilers using multivariate data analysis,” *Poult. Sci.*, 2021, doi: 10.1016/j.psj.2021.101106.

[10] J. Hair, W. Black, B. Babin, and R. Anderson, “Multivariate Data Analysis: A Global Perspective,” in *Multivariate Data Analysis: A Global Perspective*, 2010.

[11] H. S. Afifi, I. Abu-Alrub, and S. Masry, “Discrimination of Samar and Talh honey produced in the Gulf Cooperation Council (GCC) region using multivariate data analysis,” *Ann. Agric. Sci.*, 2021, doi: 10.1016/j.aoas.2021.02.002.

CHAPTER 15

FORECASTING WITH A SINGLE-EQUATION REGRESSION MODEL

Puneet Sethi, Associate Professor
Teerthanker Mahaveer Institute of Management and Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India
Email Id- sethipuneet5@gmail.com

ABSTRACT:

Forecasting with a single-equation regression model is a statistical technique that uses a mathematical formula to predict the future behavior of a variable based on its past behavior and the behavior of other related variables. In this model, a dependent variable is regressed against one or more independent variables, and the resulting equation is used to predict the future values of the dependent variable. To improve the accuracy of the forecasts, researchers may use a variety of techniques, such as including lagged values of the dependent variable, adding more independent variables, or adjusting for seasonality or other temporal effects. They may also use diagnostic tests to assess the validity of the model and make adjustments as needed.

KEYWORDS:

Dependent Variable, Independent Variable, Lagged Variables, Forecasting, Single-Equation Regression Model, Seasonality, Temporal Effects.

INTRODUCTION

The growing importance of the euro as an international investment currency has made the market for euro-dominated issues more attractive for both investors and issuers. A key element behind these developments of the European bond market in this period (between 2001 and 2004) was the impetus for a better integrated and more liquid market and the increasing adversity of innovative products, such as index-linked bonds, real-time bond indices, fixed income exchange traded funds, credit derivatives and structured products (European Central Bank, 2004). (European Central Bank, 2004). In many business and economic situations, you will be faced with a problem where you are interested in the relationship that exists between two different random variables X and Y [1].

This type of a relationship is known as a bivariate relationship. In the bivariate model, we are interested in predicting the value of a dependent or response variable based upon the value of one independent or explanatory variable. For example, a marketing manager may be interested in the relationship between advertising and sales. A production manager may want to predict steel production as it relates to household appliance output. A financial analyst may be interested in the bivariate relationship between investment X and its future returns Y; or an economist may look at consumer expenditure as a function of personal disposable income.

Regression models are also called causal or explanatory models. In this case, forecasters use regression analysis to quantify the behavioral relationship that may exist between economic and business variables. They may use regression models to evaluate the impact of shifts in internal (company level) variables, such as discount prices and sales, and external economic factors, such as interest rates and income, on company sales.

To determine if one variable is a predictor of another variable, we use the bivariate modelling technique. The simplest model for relating a variable Y to a single variable X is a straight line. This is referred to as a linear relationship. Simple linear regression analysis is used as a technique to judge whether a relationship exists between Y and X. Furthermore, the technique is used to estimate the mean value of Y, and to predict (forecast) a future value of Y for a given value of X. In simple regression analysis, we are interested in describing the pattern of the functional nature of the relationship that exists between two variables [2]. This is accomplished by estimating an equation called the regression equation. The variable to be estimated in the regression equation is called the dependent variable and is plotted on the vertical (or Y) axis. The variable used as the predictor of Y, which exerts influence in explaining the variation in the dependent variable, is called the independent variable. This variable is allotted on the horizontal (or X) axis. When using regression analysis to make forecasts, it is important to determine the appropriate mathematical model that properly explains the relationship between the variables of interest. Economic theory of production, demand, finance, or trade, should be followed when specifying the model. For example, for quarter-to-quarter fluctuations in consumer expenditure on housing, cars, home appliances, etc., we turn to the theory of demand. On the other hand, when forecasting corporate profitability, financial theory will play a role in explaining model specification. Depending on the nature of the relationship between the dependent and independent variable, the forecaster may develop a linear, parabolic, logarithmic, or some other mathematical model. Once the mathematical model has been identified, the next step is to estimate the best-fitting model for the two variables of interest. In this chapter we elaborate on the use of the two-variable linear model. The linear relationship between the two variables Y and X is expressed by athe general equation for a straight line as: aY = a + bX [7-1] where aY = value of the dependent variable aa = regression constant, or the Y intercept ab = regression coefficient, or the slope of the regression line aX = given value of the independent variable

The appeal of using the linear regression for forecasting lies in its simplicity. This has given rise to the concerns that the linear model may be too restrictive and that the real-world business and economic conditions cannot be fully captured by the model. In this context the following two appoints should be kept in mind. First, although the linearity assumption of the model appears to be restrictive, it should be noted that the majority of business and economic data approximate linearity either directly or by some form of data transformation. Second, in the real world, we do not expect any simple relationship between variables to hold precisely. This means that the actual value observed for the dependent variable will inevitably differ somewhat from its expectation. In any particular case, for a number of reasons that we shall discuss shortly, we would expect the dependent variable at vary from its actual observed value in either a positive or negative direction. This suggests that our simple linear equation stated should be written as [3]:

$$Yt = a + bXt + \varepsilon t$$

The term εt a or error is simply representing all those factors other than the value of the independent variable that influences the value of the dependent variable. Since at a may take any value in any time period, we regard this as a random variable with a mean value of zero. This implies that the mean value a Forecasting with Simple Regression 165 of the dependent term corresponding to a given value of the independent term is thus (a + bX).

DISCUSSION

In regression analysis our objective is to estimate the value of the parameters a, and b expressed in Equation. Before we estimate these parameters, it was important to understand how we interpret each, and the assumptions associated with the error term. Parameter is the expected value of a dependent variable when the independent variable is zero. In business and economic scenarios, the interpretation of an irrelevant and sometimes misleading. For example, the expected demand for any goods when its price is zero does not mean a great deal, even if a linear relationship between expected demand and price appears to be reasonable in the range of prices observed. It does not make sense to attach any value to this relationship when the price is zero, as at his price falls outside the range of observed prices. Parameter b, on the other hand, has an important interpretation as the expected value of the dependent variable resulting from a one-unit change in the value of the independent variable [4].

The assumptions associated with the error term expressed in Equation are: a

(1) It is not correlated with the independent variable.
(2) It has a mean of zero.
(3) It has a variance of $\sigma 2$ a.

In the following sections of this chapter, you will be introduced to the techniques for estimating a regression equation, the standard error, and coefficients of determination and correlation. The concepts developed in this chapter can be applied to more than two variables, as will be discussed in the next chapter on forecasting with multiple regression.

As was pointed out in the introduction, simple linear regression analysis is concerned with the relationship that exists between two variables. Business and economic forecasters use theoretical knowledge and past behavior of the business and economic variables as a basis for selecting the independent variables that are helpful in predicting the values of the dependent variable.

Once we have determined that there is a logical relationship between two variables, we can portray the relationship between the variables through scatter diagram. A scatter diagram is a graph of the plotted points, each of which represents an observed pair of values of the dependent and independent variables. The scatter diagram serves two purposes: it provides for a visual presentation of the relationship between two variables, and it aids in choosing the appropriate type of model for estimation [5].

A manager of a clothing chain in Italy has postulated that there is a relationship between the sales volume and the amount spent on advertising. She has collected monthly data on these variables for the last 18 months and has asked a forecasting analyst to determine what type of a relationship athere might be between sales and advertising. Figure 1 illustrate the dependent and Independent Variables.

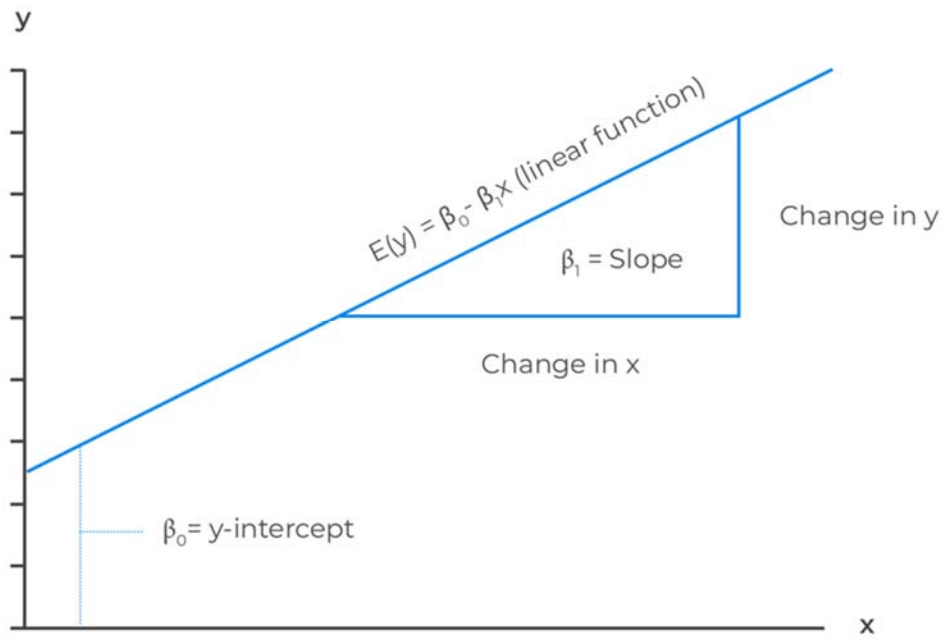


Figure 1: Illustrate the dependent and Independent Variables.

Following the standard convention of plotting the dependent variable along the Y-axis and the independent variable along the X-axis, the analyst has the monthly sales plotted along the Y-axis and the advertising expenditures along the X-axis. The scatter diagram is also used to determine if there is a linear or curvilinear relationship between variables. If a straight line can be used to describe the relationship between variables X and Y, there exists a linear relationship. If the observed points in the scatter diagram fall along a curved line, there exists a curvilinear relationship between variables. Forecasting with Simple Regression 167 a and b illustrate, respectively, a positive and negative linear relationship between two variables. Diagrams c and d show the positive and negative curvilinear relationships between variables X and Y.

Another curvilinear relationship is illustrated in e where X and Y rise at first, and then as X increases, Y decreases. Such a relationship is observed in many businesses. For instance, in product life cycle we note that sales tend to arise as the product grows to maturity and then decline as other competitor's enter the market [6]. The mathematical equation of a line such as the one in the scatter diagram that describes the relationship between two variables is called the regression or estimating equation. The regression equation has its origins in the pioneering work of Sir Frances Galton (1877), who fitted lines to scatter diagrams of data on the heights of fathers and sons. Galton found that the heights of children of tall parents tended to regress towards the average height of the population. Galton referred to his equation as the regression equation.

The regression equation is determined by the use of a mathematical method referred to as the least squares. This method simply minimizes assume of the squares of the vertical deviations about the line. Thus, the least squares method is a best fit in the sense that the Σ (Y − Yˆ) a2 a is less than it would be for any other possible straight line. Additionally, the least squares regression line has the following property:

$$\Sigma (Y - Y\hat{}) = 0$$

This characteristic makes the total of positive and negative deviations equal ato zero. You should note that the linear regression equation, is just an estimate of the relationship between the two variables in the population.

$$\mu y.x = A + BX$$

Forecasting with Simple Regression 169 where aμy.x = the mean of the Y variable for a given X value aA and B = population parameters that must be estimated from sample data. To understand the basic logic of the least square method, we have provided, as an example, the computation of the regression parameters. You can use Excel or any other software package to estimate the regression equation. The regression equation can be calculated by two methods. The first involves absolving simultaneously two equations called the normal equations. They are [7]:

$$\Sigma Y = na + b \Sigma X$$

$$\Sigma XY = a \Sigma X + b \Sigma X2$$

We use Equations to solve for a and b, and obtain the estimating or regression equation.

The second method of arriving at a least squares regression equation is ausing a computationally more convenient equation.

$$b = n (\Sigma XY) - (\Sigma X) (\Sigma Y) \text{ an } (\Sigma X2) - (\Sigma X)$$

$$a = \Sigma Y \text{ an} - b$$

$$\Sigma X \text{ an} = Y - bX[9]$$

You must keep in mind that the simple regression equation should not be used for prediction outside the range of values of the independent variable given in a sample. In order to graph the regression line, we need two points. Since we have determined only one (X = 10, Y = 65.22), we need one other point for graphing the regression line. The second point (X = 12, Y = 75.08) is along with the original data. The estimated sales revenue values of a65.22 and 75.08 should be treated as average values. This means that, in the future, the sales revenues will vary from sample to sample due to the response of the customers to advertising, price of the competing stores, the income levels of the customers, and a host of other factors[10], [11].

CONCLUSION

Forecasting with a single-equation regression model is a powerful statistical technique that can be used to predict the future behavior of a variable based on its past behavior and the behavior of related variables. This method is widely used in economic and financial forecasting, but its accuracy depends on the quality of the data used and the assumptions made about the relationships between the variables.

REFERENCES

[1] P. Hewage, M. Trovati, E. Pereira, and A. Behera, "Deep learning-based effective fine-grained weather forecasting model," *Pattern Anal. Appl.*, 2021, doi: 10.1007/s10044-020-00898-1.

[2] W. J. Corlett and D. J. Aigner, “Basic Econometrics.,” *Econ. J.*, 1972, doi: 10.2307/2230043.

[3] K. P. Amber *et al.*, “Energy consumption forecasting for university sector buildings,” *Energies*, 2017, doi: 10.3390/en10101579.

[4] M. Yousefi, M. Yousefi, M. Fathi, and F. S. Fogliatto, “Patient visit forecasting in an emergency department using a deep neural network approach,” *Kybernetes*, 2020, doi: 10.1108/K-10-2018-0520.

[5] G. Pang, J. He, Y. Huang, and L. Zhang, “A Binary Logistic Regression Model for Severe Convective Weather with Numerical Model Data,” *Adv. Meteorol.*, 2019, doi: 10.1155/2019/6127281.

[6] K. Dladla and H. Khobai, “The impact of Taxation on Economic Growth in South Africa,” *MPRA Pap.*, 2018.

[7] J. Bai, E. Ghysels, and J. H. Wright, “State Space Models and MIDAS Regressions,” *Econom. Rev.*, 2013, doi: 10.1080/07474938.2012.690675.

[8] V. I. Galkin and A. N. Koltyrin, “Investigation of probabilistic models for forecasting the efficiency of proppant hydraulic fracturing technology,” *J. Min. Inst.*, 2021, doi: 10.31897/PMI.2020.6.7.

[9] K. Drachal, “Forecasting crude oil real prices with averaging time-varying VAR models,” *Resour. Policy*, 2021, doi: 10.1016/j.resourpol.2021.102244.

[10] J. Xu, Y. Chen, L. Bai, and Y. Xu, “A hybrid model to simulate the annual runoff of the Kaidu River in northwest China,” *Hydrol. Earth Syst. Sci.*, 2016, doi: 10.5194/hess-20-1447-2016.

[11] J. Ahn, S. Hong, J. Cho, Y. W. Lee, and H. Lee, “Statistical modeling of sea ice concentration using satellite imagery and climate reanalysis data in the barents and Kara seas, 1979-2012,” *Remote Sens.*, 2014, doi: 10.3390/rs6065520.

CHAPTER 16

EXPLORING THE BASIC TIME-SERIES DECOMPOSITION MODEL: A COMPREHENSIVE ANALYSIS AND EVALUATION OF ITS COMPONENTS AND APPLICATIONS

Anushi Singh, Assistant Professor
Teerthanker Mahaveer Institute of Management and Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India
Email Id- anushigaur@rediffmail.com

ABSTRACT:

The basic time-series decomposition model is a statistical technique used to break down a time series into its underlying components, namely trend, seasonality, cyclist, and irregularity. The model assumes that a time series can be expressed as the sum of these four components, where the trend represents the long-term behavior of the series, seasonality refers to the recurring patterns that occur at fixed intervals, cyclist captures the medium-term fluctuations, and irregularity represents the random fluctuations in the series.

KEYWORDS:

Decomposition Model, Irregularity, Seasonality, Trend, Time Series, Statistical Technique.

INTRODUCTION

The regression equation is primarily used for estimation of the dependent variable, given values of the independent variable. Once we have estimated a regression equation, it is important to determine whether the estimate is dependable or not. Dependability is measured by the closeness of the relationship between the variables. If in a scatter diagram the points are scattered close to the regression line, there exists a close relationship between the variables. If, on the other hand, there is a great deal of dispersion between the points and the regression line, the estimate made from the regression equation is less reliable.

The standard error of estimate is used as a measure of scatter or dispersion of the points about the regression line, just as one uses the standard deviation to measure the deviation of the individual observations about the mean of those values.

The smaller the standard error of estimate, the closer the estimate is likely to be to the ultimate value of the dependent variable. In the extreme case where every point falls on the regression line, the vertical deviations are all 0; that is, $s_{y.x} = 0$. In such a situation, the regression line provides perfect predictions. On the other hand, when the scatter is highly dispersed, making the vertical deviation large ($s_{y.x}$ is large); the predictions of Y made from the regression line are subject to sampling error [1].

The $n - 2$ value in the denominator represents the number of degrees of freedom around the fitted regression line. Generally, the denominator is Forecasting with Simple Regression 173 $n - (k + 1)$

where k represents the number of constants in the regression equation. In the case of a simple linear regression, we lose two degrees of freedom when a and b are used as estimates of the constants in the population regression line. Notice that Equation requires a value of Y for each value of X. We must therefore compute the difference between each $\hat{Y}$ and the observed value of Y.

The above computational method requires a great deal of arithmetic especially when large numbers of observations are involved. To minimize cumbersome arithmetic, the following shortcut formula is used in computing the standard error of estimate can be effectively used to make forecasts in most business and economic environments. However, these models do not always capture the pattern that may be exhibited by many time series. For example, methods such as the exponential smoothing are suitable for shortterm forecasting of time series.

However, when faced with businesses and economic conditions that exhibit more complicated data patterns such as a combination of a trend, seasonal factor, cyclical, and random fluctuations, we need a more comprehensive method. Similarly, in using a regression model, the analyst depends on visual (scatter diagrams) and statistical analyses to determine the "best" model for forecasting purposes. The iterative process of determining which model is the "best" is expensive and time consuming [2].

Furthermore, the utility of the regression models depends heavily on satisfying the assumptions of these models. Even when most of the assumptions are satisfied, we were not certain whether the model would provide a good forecast in the long run. Because of the above limitations, analysts use the Box–Jenkins (B/J) methodology for business and economic forecasting. There are several reasons for using the Box–Jenkins method for forecasting. First, the methodology is able to capture a myriad of data patterns. Second, given a set of data, the "best" model can be identified.

Third, this forecasting approach can handle complex data patterns, using relatively well-specified rules. Fourth, reliability of forecasts can be tested through statistical measurement. Fifth, a significant difference of the Box–Jenkins (B/J) methodology is that it does not make assumptions about the number of terms used in the models or the relative weight given to them. When using the B/J models, the analyst selects the appropriate model based on established criteria, including the number of terms; then the software programme (Minitab, SAS, or SPSS) calculates the coefficients using a nonlinear least squares method. Forecasts for future periods are made with the calculated coefficients. Finally, confidence intervals for the estimates are easily constructed.

The limitation associated with this forecasting method has to do with the complex nature of the methodology and the associated cost. Because the methodology deals with much more general conditions, it is much more difficult to understand the fundamentals of the approach and how the method can be applied. The higher cost associated with using this approach is outweighed by the greater accuracy of the forecast [3].

DISCUSSION

Jenkins methodology can produce forecasts based on a synthesis of historical patterns in data without initially assuming a fixed pattern. This is not the case with the other methodologies. For example, when using the exponential smoothing technique, it is assumed that the data follow a horizontal pattern. In regression analysis the analyst assumes a pattern and then proceeds with the application of the technique.

The Box–Jenkins method is an iterative process that begins by assuming a tentative pattern that is fitted to the data so that the error will be minimized. The approach provides the analyst with explicit information, on theoretical grounds, to determine whether the assumed pattern is appropriate or not. If the correct pattern is not found, the Box–Jenkins method provides additional clues to the analyst so that a correct pattern is identified. Once a correct pattern is selected, the analyst can use it to make forecasts [4].

The major assumption of the B/J model is that the data series of interest is stationary. A data set is considered to be stationary when it fluctuates randomly around the mean of the series. Another way of stating this is that the average value of the time series is not changing over time. It should be noted that, while the stationary assumption appears restrictive, the data series that are nonstationary can be transformed easily to meet this assumption. Our interest in this data set is whether the pattern of these observations is due to some systematic relationship, and if so, how are they generated? From our discussion of regression analysis, we know that any observation on Y has two components. The first part is what is explainable by the model and is identified as (bt), and the second part is the random error (et).

In a similar fashion, the Box–Jenkins models assume that a time series is a linear function of past actual values, and random shocks or error terms. The expectation is that the error terms are distributed as white noise. By definition, white noise is normally and independently distributed (NID), having no patterns, a mean of zero, and an error variance that is lower than the variance of Yt . With these assumptions in mind, the Box–Jenkins models are classified as the autoregressive models (AR), the moving average models (MA), or a combination of the two, labelled as the autoregressive integrated moving average models (ARIMA). Before we discuss each of these models in detail, let us clarify the notations used in building them. The standard notation identifies the orders of autoregression by p, integration or differencing by d, and moving average by q. So, an ARIMA model could be looked.

Throughout our discussion of these models, you will become more familiar with these notations and recognise their versatility. An autoregressive model (AR) is one in which the current value of the variable is a function of its previous values and an error term. The reason this is called an autoregressive model is because Yt is being regressed on itself [5].

$$Yt = 0 + 1Yt - 1 + 2Yt - 2 + ... + pYt - p + \varepsilon t$$

Where Yt = Dependent variable

The Box–Jenkins Method of Forecasting Yt−1, Yt−2, Yt−p = Independent variables based on the dependent variable lagged (p) specific time periods

0, 1, 2, p = Computed regression coefficients

εt = Random error term measured in time t

An autoregressive (AR) model with a mean or a constant term of zero can have an order of one, two, or p, or it could exclude some of the lower-order terms respectively.

Order of 1: $Yt = 1Yt - 1 + \varepsilon t$ [10-5]

Order of 2: $Yt = 1Yt - 1 + 2Yt - 2 + \varepsilon t$ [10-6]

Order of p: $Yt = 1Yt - 1 + 2Yt - 2 + ... + pYt - p + \varepsilon t$

Orders excluded: $Yt = 2Yt - 2 + 6Yt - 6 + \varepsilon t$ [10-8]

The second type of Box–Jenkins model is called the moving average (MA) model. These models link the current value of the time series to random errors that have occurred in previous time periods. Equation specifies a moving average model.

$$Yt = \theta 0 - \theta 1\varepsilon t - 1 - \theta 2\ \varepsilon t - 2 - ... - \theta q\varepsilon t - q + \varepsilon t$$

Where Yt = Dependent variable

$\theta 0$ = The mean about which the series fluctuates $\theta 1$, $\theta 2$, θq = Moving average parameters to be estimated

$\varepsilon t{-}q$ = Error terms

εt = Random error term measured in time t

The highest order of the model is called q and refers to the number of lagged time periods in the model. Similar to the AR model, the MA could have different orders. For example, the MA model with one term is written as: $Yt = -\theta 1\varepsilon t - 1 + \varepsilon t$.

Note that, in the MA model, it is assumed that the current value of the series is a direct and predictable result of past random errors. The third model of Box–Jenkins is a combination of the AR and MA models. Thus, it is called the autoregressive integrated moving average or ARIMA. The model is written as: $Yt = 0 + 1Yt - 1 + 2Yt - 2 + ... + pYt - p + \varepsilon t - \theta 1\varepsilon t - 1 - \theta 2\varepsilon t - 2 - ... - \theta q\varepsilon t - q$

When using the ARIMA model, the analyst is able to use a combination of past values and past errors. As was mentioned earlier, the order of the model is The Box–Jenkins Method of Forecasting 273 commonly written as ARIMA (p, d, q). For example, when we say that the model is an ARIMA (1,0,0), it implies that we are dealing with an AR (1) type model. Similarly, an ARIMA (0,0,1) refers to an MA (1) model. To select the appropriate model for forecasting, we depend on the autocorrelation (AC) and partial autocorrelation (PAC) statistics of the time series [6].

The ACs and PACs provide us with the numerical measure of the relationship of specific values of a time series to other values in the time series. That is, they identify the type of model that will most closely capture the variation in the time series of interest. Software packages that handle the Box–Jenkins methodology will generate a table of ACs and PACs as well as the correlogram associated with each. The range of values of ACs and PACs is between −1 and +1. Once the computer software calculates the ACs and PACs for various time lags (depending on the nature and the periodicity of the data—daily, weekly, monthly, quarterly, etc.), a comparison is made with the theoretical distributions (patterns) of the ACs and PACs developed by the Box–Jenkins method.

The general guidelines to follow in determining what data pattern fits which model are:

a) When the autocorrelation coefficients gradually fall to zero, and the partial correlation has spikes, an AR model is appropriate. The order of the model depends on the number of spikes.
b) When the partial correlation coefficients gradually fall to zero, and the autocorrelation coefficients have spikes, an MA model is appropriate. As in the AR model, the order of the model equals the number of significant spikes.
c) When both the autocorrelation and the partial autocorrelation correlograms show irregular patterns, then an ARIMA model best represents the data pattern. Once again, the number of significant spikes equals the order of the model.

The theoretical distributions (patterns) for the various models of Box–Jenkins are provided below. Figure 10.2 (a, b, c, and d) shows the theoretical autocorrelation and partial correlation functions for the AR (1) and the AR (2) models, respectively. Note the behaviour of the correlation coefficients. A comparison of the autocorrelation coefficients with the partial correlation coefficients in the AR (1) and AR (2) models show that the autocorrelation coefficients gradually drop to zero, whereas the partial correlation coefficient drop to zero after the first time lag.

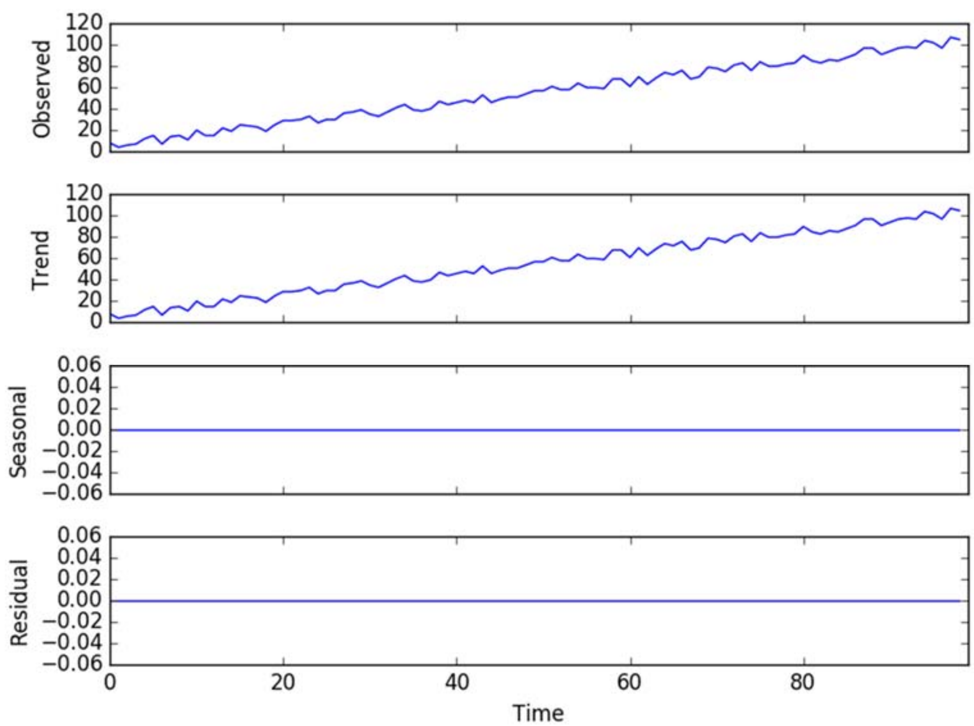


Figure 1: Illustrate the Time Series Data into Trend and Seasonality

The autocorrelation and partial autocorrelation coefficients for the moving average models (MA) are shown in Figure 10.3 (e–h). Again, note the behavior of these correlation coefficients. In the

AR (2) models, the autocorrelation coefficients gradually drop to zero, while the partial coefficients drop to zero after the second time lag. The theoretical distribution for the most common ARIMA model [7].

When selecting a model, you must keep in mind that the correlation coefficients are theoretical distribution patterns and that your actual data set may not conform with them identically. Nonetheless, the patterns presented in these figures are so common to most Figure 10.2 Autocorrelation and Partial Autocorrelation Coefficients for AR (1) and AR (2) Models. The Box–Jenkins Method of Forecasting 275 time series data that you should not have difficulty identifying a pattern. Figure 1 illustrate the Time Series Data into Trend and Seasonality.

Having analysed the AC and PAC for a time series and determined which model may serve our need, we are ready to estimate the model and verify its validity. Model validity is accomplished through statistical analysis of the residuals. Specifically, we examine the autocorrelation of the residuals to determine the randomness in error terms. Furthermore, the Q statistics provide us with a measure that tells us whether the autocorrelation of the entire residual series differs significantly from zero.

In the following sections of this chapter, each of the Box–Jenkins models are discussed more fully. The autoregressive model is an extension of the regression model. The only difference between the two is that, in the autoregressive model, the independent variables are simply lagged values of the dependent variable. The Box–Jenkins Method of Forecasting 277 on what the analyst is hoping to achieve. The general AR model is designated as ARIMA (p, 0, 0). The AR (1) model is specified as: Yt = 0 + 1Yt − 1 + εt [10-12]

where Yt = the dependent variable

0, 1 = coefficients chosen to minimise the sum of squared errors

Yt−1 = independent variable

εt = random error

The following example uses a very simple data set so that we can show how a forecast is made when an AR model is identified. For all practical purposes the sample size would have to be large enough to accommodate an accurate computation of the ACs and PACs. Accurate estimates of AC functions are made with a minimum sample size of 50 observations where k lags should not be larger than approximately n/4 [8].

A manufacturer of digital cameras in Japan wishes to make a forecast of sales in future years. The marketing manager has gathered the data shown in Table 10.2 for the last 19 years. Suppose that the ACs and PACs show that an ARIMA (1,0,0) model is appropriate for the data set. Furthermore, assume that 0 = 0.6 and 1 = 0.9 were fitted to data. Test the validity of the model by forecasting sales for the years 1991 to 2008 where the fitted model is.

(1) Single significant spike at lag 1 ARIMA (1,0,0) 1 > 0 Exponential decline with significant spikes at first, second, or more lags
(2) Single significant spike at lag 1 ARIMA (1,0,0)

(3) 1 < 0 Alternating exponential decline with a negative
(4) Single significant spike at lag 1 ARIMA (0,0,1) $\theta 1 > 0$ Single significant negative spike at lag 1
(5) Exponential decline of negative values, with more than one significant spike

ARIMA (0,0,1) $\theta 1 < 0$ Single significant positive spike at lag 1 Alternating exponential decline starting with a positive 278 The Box–Jenkins Method of Forecasting Since the actual value of sales is unknown in the year 2008, the forecast for the year 2007 is the forecasted value for the previous year. That is, the forecast for the year 2008 would be: Yˆ 08 = 0.6 + 0.9(17.286) = 16.157

Keep in mind that for the ARIMA (1,0,0) model, the absolute value of the coefficient 1 is normally constrained to be less than 1. This constraint is referred to as the bound of stationarity that states: | 1|<1. If the absolute value of 1 is greater than 1, then the series is not autoregressive, in which case transformation of the data through differencing should be used to conform to the stationarity assumption.

In our earlier discussion we mentioned that the MA models provide forecasts based on a linear combination of past errors. The general moving average (MA) model is specified as ARIMA (0,0,q). The assumption of stationarity also holds true for this model. Note that the mean of an MA (q) series represented by 0 is the constant term in the model as we assume the expected value of the error term to be zero. That is, E (εt) = 0.

The Box–Jenkins Method of Forecasting 279 t. In the MA model it is customary to show the coefficients (θ1, θ2, . . ., θq) with negative signs, even though these coefficients may be positive or negative. The simplest MA model with one term is given as[9]:

Similar to the autoregressive model, the absolute value of the coefficient θ1 is constrained to be less than 1. This constraint, which is related to the stationarity, is referred to as the bound of invertibility. If the absolute value of θ1 is greater than 1, the model is not stationary. To remedy the problem of nonstationary data, transformation of the data is necessary. As was highlighted in Equation [10-11], past values and past errors are used in the model to make future forecasts. Theoretically speaking, this integrated model can fit any pattern of data. However, the values of p (0, 1, 2, . . .) and q (0, 1, 2, . . .) must be specified before the method can be applied. You may recall that when p = 1 and q = 0 we were dealing with an AR model. Similarly, when p = 0 and q = 1, the model was an MA model. Finally, when d, p, and q can be different from zero, we are dealing with a mixed model. For example, when p = 1 and q = 1, the model is ARIMA (1,1) and is written as: Yt = 1 Yt − 1 + εt − θ1 εt – 1.

Note that the error term (εt) is influenced by Yt − 1 and at the same time by θ1εt − 1 which makes this equation nonlinear and highly effective in describing a wide range of data patterns. The theoretical distribution (pattern) of the ACs and PACs for this group of models was presented in Figure 10.4. The analyst is able to take a look at the computed autocorrelation coefficients and the partial autocorrelation coefficients of the data set and determine if the model follows an ARIMA pattern. Once it is determined that an ARIMA model should be used, then the forecasting process is similar to that of the earlier models [10].

In the beginning of this chapter we mentioned that one of the assumptions of the B/J methodology dealt with the stationary of the data. Given that data patterns may include trends, seasonal, and other factors that cause no stationary patterns, we must therefore eliminate them from the data series. This means that before applying the B/J models for forecasting, data must be free of trend or seasonal factors. This can be accomplished through a data. The Box–Jenkins Method of Forecasting 281 transformation procedure referred to as "differencing." The concept of differencing was addressed. Trends Trend patterns appear as an upward or downward movement that characterizes all economic activities in a dynamic economy.

Trend patterns may be of a deterministic or stochastic nature. A deterministic trend is a systematic period-to-period increase or decrease observed over a number of time periods. A stochastic trend, on the other hand, is a random variation of the data pattern. Whether we are faced with a deterministic or stochastic trend, these trend patterns must be made stationary using a transformation method illustrate how we transform a no stationary linear trend into a stationary data pattern, suppose we have a simple data set. We note that the data set has a deterministic linear trend, and violates the assumption of stationary. To remedy the problem, we could simply transform the series by taking the first difference as shown in column 2.

CONCLUSION

The basic time-series decomposition model is a powerful statistical technique that allows analysts to break down a time series into its underlying components, namely trend, seasonality, cyclist, and irregularity. By decomposing a time series, analysts can gain insights into the underlying patterns and relationships within the data, which can be useful for forecasting and decision-making in a wide range of fields, such as finance, economics, and engineering.

REFERENCES

[1] S. Xu *et al.*, "The fuzzy comprehensive evaluation (FCE) and the principal component analysis (PCA) model simulation and its applications in water quality assessment of Nansi lake basin, China," *Environ. Eng. Res.*, 2021, doi: 10.4491/eer.2020.022.

[2] D. Nagarajan, S. Varjani, D. J. Lee, and J. S. Chang, "Sustainable aquaculture and animal feed from microalgae – Nutritive value and techno-functional components," *Renew. Sustain. Energy Rev.*, 2021, doi: 10.1016/j.rser.2021.111549.

[3] H. Han and S. Trimi, "Social commerce design: A framework and application," *J. Theor. Appl. Electron. Commer. Res.*, 2017, doi: 10.4067/S0718-18762017000300005.

[4] T. Balachandran, A. K. Manoharan, V. Aravinthan, and C. Singh, "Component-Level Reliability Evaluation Model for Cyber Power Devices," *IEEE Trans. Smart Grid*, 2021, doi: 10.1109/TSG.2020.3016524.

[5] X. Geng, M. Qi, J. Liu, C. He, and Y. Li, "Application of principal component analysis on water flooding effect evaluation in natural edge-bottom water reservoir," *J. Pet. Explor. Prod. Technol.*, 2020, doi: 10.1007/s13202-020-01055-4.

[6] P. Saragiotis, "Business process management in the port sector: a literature review," *Maritime Business Review*. 2019. doi: 10.1108/MABR-10-2018-0042.

[7] J. G. Pitter *et al.*, “Planning, implementation and operation of a personalized patient management system for subjects with first suspect of cancer (OnkoNetwork): System description based on a qualitative study,” *BMC Health Serv. Res.*, 2019, doi: 10.1186/s12913-019-3957-9.

[8] A. Russo *et al.*, “Liquid biopsy tracking of lung tumor evolutions over time,” *Expert Review of Molecular Diagnostics*. 2019. doi: 10.1080/14737159.2020.1680287.

[9] D. Tewari *et al.*, “Withania somnifera (L.) Dunal: Phytochemistry, structure-activity relationship, and anticancer potential,” *Phytomedicine*. 2022. doi: 10.1016/j.phymed.2022.153949.

[10] M. M. A. Khan, R. Saidur, and F. A. Al-Sulaiman, “A review for phase change materials (PCMs) in solar absorption refrigeration systems,” *Renewable and Sustainable Energy Reviews*. 2017. doi: 10.1016/j.rser.2017.03.070.

CHAPTER 17

UNIVARIATE TIME-SERIES MODELING AND FORECASTING: A COMPARATIVE STUDY OF TRADITIONAL AND MODERN TECHNIQUES

Garima Rawat, Assistant Professor
Teerthanker Mahaveer Institute of Management and Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India
Email Id- grawat436@gmail.com

ABSTRACT:

Univar ate time-series modeling and forecasting is a statistical approach used to analyze and predict the behavior of a single time-dependent variable. The technique involves analyzing the past values of the variable to identify any patterns or trends and then using this information to make predictions about its future behavior. The modeling process usually involves identifying the appropriate mathematical model that best fits the data, such as autoregressive integrated moving average (ARIMA) or exponential smoothing. Once the model is selected, it is calibrated using the available historical data and then used to generate forecasts for future time periods.

KEYWORDS:

Forecasting, Pattern Recognition, Time-Dependent Variable, Statistical Analysis, Univariate Time Series.

INTRODUCTION

When faced with a nonlinear trend such as a parabolic condition, the time series is made stationary by taking the second differences. In the case of a growth curve, the original data series needs to be transformed into logarithms one of the indicators of economic activity in a nation is its index of productivity in the various sectors. Data in Table 10.5 represent the productivity index for the business sector in the U.S. between 1986 and 2005. Determine, θ0 using the data from 1986 to 2000, and then forecast years 2001 to 2005, using an ARIMA (0,1,0) model. From the plot of the data shown it appears that there exists a trend in the data as the series drifts upward, showing non-stationarity in the data. We can verify this nonstationary mean further by plotting the AC for the series.

The trend is removed from the data by taking the first difference of the data. We also observe that the PAC shown in Figure 10.7 drops off to zero after the first lag. Both of these autocorrelation functions suggest the use of an ARIMA (1,1,0) model with this set of data. Now we are ready to estimate the model. Based on our AC and PAC we Figure 10.5 Daily Stock Prices. The Box–Jenkins Method of Forecasting determine that an ARIMA (1,1,0) model would be appropriate. Our objective is to make sure that the model has a "good" fit. The "goodness-of-fit" is determined by the condition that the sum of squared residuals is minimized [1].

The estimated parameter 0.2148 is statistically significant as supported by the t-statistics. The p-value of 0.043 has the same interpretive value as in the case of regression models. You will note that we did not include a constant term in this model. Whenever we use differencing in the model, the constant term is omitted. However, when making forecasts, the computer programmed will take this into account and appropriately make future forecasts based on the model.

Faced with seasonal data, the model specified in Equation might not suffice and should be supplemented with seasonal parameters. The B/J models with seasonal factors are generalized as ARIMA (p, d, q) s. Seasonal models can be AR, MA, or ARIMA as shown in Equations to, respectively. To estimate the seasonal values of p and q (usually denoted as P and Q), we follow the same processes as the no seasonal data. That is, the AC and PAC for only the seasonal component are examined. If we are interested in the seasonality of the quarterly data, the values of 4, 8, 12, 16, 20, and so forth observations are of interest to us. On the other hand, for the monthly data the values of 12, 24, 36, 48, and so on will be examined for similar patterns.

The analyst should be aware that when seasonal patterns are overwhelmed by a dominant trend pattern you will see nonzero spikes for almost all time lags in the AC correlogram. To determine if there is a seasonal pattern in the data, it is best to examine the AC and PAC correlograms for the "differenced" series of the data. If seasonal patterns exist, then large spikes for time lags of 12 and 24 months capture the monthly seasonal pattern [2]. As an analyst for a large department store you are given the responsibility for forecasting sales. You have been given monthly sales data for the last 10 years as shown in Table 10.10. Use your professional expertise with the Box– Jenkins method to make a forecast for the department store. More than ever, today's company managers rely on timely and reliable information when making choices. Not just in home markets but also on a global scale, business has become very competitive. Economic and business analysts are trusted to provide guidance business executives. A forecast analyst's job is to provide several predictions to the decision-maker, including estimates for sales, cash flow, inventories, and costs. It can also include examining the consequences of tighter monetary policy on the corporate sector and the policy ramifications of tax increases.

There is a higher dependence on company estimates due to two causes that have been significant. First, the systematic use and efficacy of forecasting approaches have significantly increased. As a result, managers now rely more on statisticians, operations analysts, and decision scientists to help them with their everyday tasks. Second, easier data collection and rapid data analysis for decision-making are now feasible because too increased information accessibility (databases) and affordable, portable, and powerful computers. The skills and knowledge of people who create predictions as well as the comprehension of those who utilize them are necessary for the efficient use of forecasting technique [3].

DISCUSSION

Forecasters must assist management in understanding the usefulness and constraints of the predictions as organizations get increasingly interested in adopting quantitative forecasts in their decision-making. Also, they must present their results to management in a straightforward way that is clear of technical jargon. In this chapter, we'll go through the crucial phases in explaining a forecast's conclusions to management as well as the problems forecasters should take into account when writing their reports. We discussed the role that forecasting may play in creating market strategies of this book. Managers are becoming more aware of how important forecasting and prediction are in making business choices. Even though this position is crucial, some businesses are too tiny to establish a dedicated forecasting department. We advise managers at these enterprises to take some of the straightforward models covered in this article into account when creating projections for their own companies. These models' detailed step-by-step instructions make it feasible to create accurate predictions. We acknowledge that some familiarity with statistics and economics is necessary in order to use these forecasting models [4].

On the other hand, major firms who see forecasting as a crucial component of their business operations may rely on forecasting experts to provide various projections for their company. In most cases, a department or group is chosen inside an organisation to handle these tasks and report the results to management. The process of informing management of projected outcomes is time-consuming. For successful communication, it starts far earlier than when a projection is made and its outcomes are presented to management. A successful projection requires managerial involvement at this early stage of the process. Most managers are involved in other tasks at the company, therefore they may not completely contribute to forecast development. Holding discussion sessions with small groups of managers is one way to spark interest in forecasting among managers. In these workshops, managers learn about the advantages and constraints of forecasting and are prompted to consider how forecasting could support their decision-making. A one-on-one meeting with managers to explore potential applications to their area of interest might come after these sessions. Forecasting experts must comprehend the issues and worries that management has. They must enquire specifically about the appropriate time range for their projections as well as the amount of detail or aggregation that is necessary. If the findings are broader than what the management anticipates, a sophisticated forecasting model and its forecasts may not be useful [5].

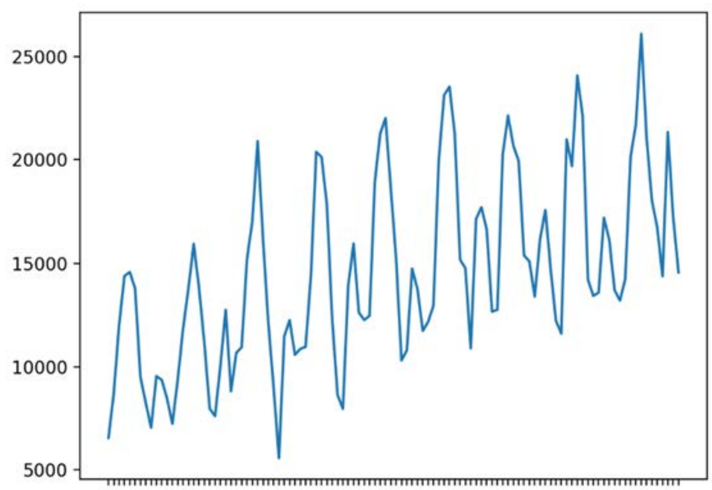


Figure 1: Illustrate the Deep Learning Models for Univar ate Time Series Forecasting.

The temporal horizon is equally important in management choices. If one were to produce a short-term vs a long-term projection, it would depend on whether they used facts from the recent past or those from a distant past. The forecaster must take this knowledge into account throughout the planning process if management, for instance, needs a forecast by the end of the next month. Mid-level managers lose interest in adopting forecasts when they take too long to develop and are not accessible when they are required, according to research. Figure 1 illustrate the Deep Learning Models for Univar ate Time Series Forecasting.

The forecasters must be highly skilled not just in forecasting methods but also in conveying to management the value and suitability of these approaches in various contexts. It is preferable to provide more than one prediction scenario when providing information to management. The model's assumptions should be made transparent so that management has the necessary data to make a decision. Bear in mind that emphasising the aspects that are essential in making a choice may be more helpful to the management than providing the manager with an abundance of information.

Management feedback is incorporated into a solid projection. It is vital to assess the forecast's efficiency when one has been produced and the findings have been compared to actual occurrences. This is closely related to what we discussed in Chapter 1. (See Figure 1.1). The forecasting expert must examine the trend in any potential forecasting errors. This has to be explained to management, and their opinion on how to make the prediction better should be sought for. While reformulating the forecasting model, management's opinions should be taken into consideration [6].

A system that takes adjustments into account and offers a feedback mechanism is essential to a successful forecasting process. It is advisable for managers and forecasters to be aware of several typical errors that reduce the efficiency of the forecasting process and, as a result, the prediction itself. Inadequate or subpar data systems used by the company that do not increase the forecasting process' viability. This implies that the outcome won't change regardless of how accurate the forecasting model and technical team are. Good forecasting will start with a commitment to assembling and maintaining disaggregate datasets. Before any forecasting is done, this has to be told to management as soon as possible. Middle management's lack of dedication to the forecasting process. The forecasting method is often approved by top management, but intermediate management does not recognise its worth or usefulness. In this circumstance, it is crucial that the forecaster explain to the middle management the value and efficacy of forecasting and how they may gain from it.

The decision-maker can discover that the facts or the prognosis are not readily accessible when needed. In order to prevent this issue, management and the forecaster must have a mutual agreement of how quickly the data is required and if the assignment can be finished in the allotted period. On occasion, the study uses the incorrect forecasting model, and the forecasting personnel may not want to see it put into practise. They could make things worse for the forecasting unit and themselves if they are unwilling to share their results out of concern that they will be accused of technical ineptitude. To prevent issues of this sort, a framework between management and the forecasting team must be developed.

Rather than functioning as a separate unit that only functions in response to sporadic requests from other corporate units, a forecasting unit must be incorporated into the current corporate structure in order to be successful. For example, the forecasting team is better able to provide a far more substantial response to the forecasting demands of the whole company when the forecasting unit is an important part of the budget and planning process. Forecasters may help management in its responsibilities as a controller and planner [7].

The forecaster must acknowledge that interpersonal and political savvy are just as crucial to effectively communicate conclusions to management as technical know-how. Control over the forecast means, to a considerable degree, control over the budgeting concerns and resource allocation for those companies that take the task of forecasting seriously. Forecasters may get embroiled in sensitive power conflicts inside the company in this situation. It is crucial to build interpersonal skills that safeguard the forecasting unit's objectivity and integrity inside the company. Listed below are some recommendations made by Bails and Peppers in 1993:

a) Get to know those who consult the prediction.
b) Discover the user's worries.
c) Address issues when creating a forecast for a specific unit before they become overwhelming.
d) You should stand up for your prognosis when faced with simply political criticism.

In conclusion, the forecaster's job is to provide insight into a wide range of decision options in order to have an influence on management choices. Management must clearly define forecasters' roles and responsibilities inside the company in order for them to be successful. The forecaster must also use caution while creating and presenting the forecast to management. This entails the time-consuming process of interrogating and nagging the user to determine what he or she really needs.

Successful company managers are aware of how crucial forecasting is to their choices. Within every business, forecasting has a wide range of applications. Predictions are created for planning and budgeting as well as for a variety of objectives, including those related to production, inventory management, marketing, advertising, sales estimates, and investment.

Businesses cannot afford to rely simply on the qualitative judgement of their management in today's competitive market, when business decisions are made at a much quicker pace and with a higher dependence on quantitative analysis. There are, however, drawbacks to relying only on the forecasting unit or department while business is booming. There is a need for improved forecasting processes and tools in order to make business choices, as shown by the recent collapse in the housing industry and the general weak economic performance in the U.S. Even though many people mistakenly believed that the market would steer the economy in the right direction in the latter part of the 1990s due to psychological factors like "feeling good" about the economy, Chairman Greenspan of the Federal Reserve Bank claimed that there is a "irrational exuberance" in the economy.

Considering that demand for forecasting department services is income elastic, forecasting staffing levels are often reduced in response to a dramatic decline in sales, prices, and profits. Companies started to slash costs in the 1980s as a result of falling prices in the oil, real estate, and financial industries. They did this by eliminating their research departments and lessening the need for forecasting. Businesses must rely on quantitative research and accurate forecasts at times like these [8].

The necessity for strategic planning as managers guide their firms into the future is highlighted by recent economic downturns in the United States and several other developed countries. Such planning tactics may efficiently use forecasting technologies. Nobel Prize winner Gary Becker (2001) notes that the September 11 attacks may have pushed the U.S. economy into a recession, which was already in decline. He notes that the sharp decline in stock prices during the first week after markets reopened is a reflection of pessimism and uncertainty about the future. Forecasters may significantly contribute to assisting business entities lessen the market uncertainty in this situation.

Any instability brought on by a crisis of major proportions on a national or worldwide scale would also have substantial ramifications for the forecasting function in company plans. Throughout the second half of the 1990s, the globalisation of trade, which allowed for a freer movement of people, money, and products, increased yearly U.S. growth by three-quarters of a percentage point. This growth rate may not continue in the foreseeable future as globalisation could halt and become more expensive. Businesses would likely have to spend more on security and insurance for employees who work abroad.

Providing Management with Predictions 313 property. More border checks can cause cargo movements to be slowed down, requiring businesses to stock up on additional merchandise. Tighter immigration regulations may have a substantial impact on the influx of skilled and blue-collar workers that has enabled businesses to grow while controlling salaries. In the meanwhile, a growing preoccupation with political risk has forced businesses to drastically reduce the scope of their new investments.

Forecasters must assist management in understanding the value and constraints of their predictions as company managers grow increasingly interested in adopting quantitative forecasts in their decision-making. Also, they must present their results to management in a straightforward way that is clear of technical jargon. The process of informing management of the findings is time-consuming and starts far before a prediction is created. It is advised that management be included in the forecasting process to make it more controllable [9].

This will provide the forecaster a clear grasp of management's requirements and how to address them. Alternative scenarios and the underlying assumptions should be communicated to management along with the findings. It is important to emphasise the error's meaning and its causes. Ultimately, the model should contain management input, and fresh findings should be compared to real occurrences. Due to upcoming national and international events as well as economic uncertainty, forecasting will be necessary. The accessibility of datasets and more complex quantitative studies improves forecasting methods.

Offered in this chapter since no clear long-term trend or cyclical pattern can be found. Compared to the 1950s, gas costs have increased significantly, although only in absolute terms. Additionally, even in nominal terms, there has been no rising trend since 1980. So, in this series, neither a log-linear transformation nor a Hodrick-Prescott filter will be able to detect any significant patterns or cycles [10], [11]. It is more beneficial to consider the cost of gasoline in relation to the CPI as a whole in order to spot any trends. The relative price of gasoline has a long-term tendency to stay constant, despite significant rises from 1973 to 1980. For trend analysis, it is perhaps the most logical working assumption. In instance, despite regular reports of a "shortage" of petroleum, there is little reason to anticipate increasing relative prices to return.

CONCLUSION

Unilabiate time-series modeling and forecasting is a powerful technique that allows analysts to analyze and predict trends in time-series data. By modeling the data using mathematical and statistical techniques, analysts can extract important insights and forecast future values with varying degrees of accuracy. Unilabiate time-series modeling and forecasting is a valuable tool for understanding and predicting trends in various fields, including finance, economics, and engineering. With the advent of powerful computing technologies and machine learning algorithms, the potential for accurate and reliable forecasting has only continued to increase in recent years.

REFERENCES

[1] A. Farouk Mohamed, “Comparative study of traditional and modern building techniques in Siwa Oasis, Egypt: Case study: Affordable residential building using appropriate building technique,” *Case Stud. Constr. Mater.*, 2020, doi: 10.1016/j.cscm.2019.e00311.

[2] A. Farouk Mohamed, “Comparative study of traditional and modern building techniques in Siwa Oasis, Egypt,” *Case Stud. Constr. Mater.*, 2020, doi: 10.1016/j.cscm.2019.e00311.

[3] B. . Oguntunde, S. . Arekete, M. O. Odim, and O. . Olakanmi, “A Comparative Study of Some Traditional and Modern Cryptographic Techniques,” *Int. J. Eng. Manag. Res.*, 2017.

[4] R. Kumar, “Supply Chain Management Executives Satisfaction: A Comparative Study of Traditional and Modern SCM Techniques,” *Anvesha*, 2016.

[5] G. Yadav and K. Paul, “Architecture and security of SCADA systems: A review,” *International Journal of Critical Infrastructure Protection*. 2021. doi: 10.1016/j.ijcip.2021.100433.

[6] M. A. Kamal, “Assessment of traditional architecture of lucknow with reference to climatic responsiveness,” *Archit. Eng.*, 2021, doi: 10.23968/2500-0055-2021-6-1-19-31.

[7] S. Bhattacharjee *et al.*, “Cluster analysis of cell nuclei in h&e-stained histological sections of prostate cancer and classification based on traditional and modern artificial intelligence techniques,” *Diagnostics*, 2022, doi: 10.3390/diagnostics12010015.

[8] L. H. Marfuah, G. Yudana, and W. Astuti, “Studi Perkembangan Pasar Tradisional Dan Toko Modern Ditinjau Dari Jarak, Aksesbilitas, Dan Perilaku Konsumen Di Kota Surakarta,” *Arsitektura*, 2017, doi: 10.20961/arst.v14i2.9126.

[9] E. Csákvári, M. Halassy, A. Enyedi, F. Gyulai, and J. Berke, “Is einkorn wheat (Triticum

monococcum l.) a better choice than winter wheat (triticum aestivum l.)? wheat quality estimation for sustainable agriculture using vision-based digital image analysis," *Sustain.*, 2021, doi: 10.3390/su132112005.

[10] Z. Pranskuniene, R. Dauliute, A. Pranskunas, and J. Bernatoniene, "Ethnopharmaceutical knowledge in Samogitia region of Lithuania: Where old traditions overlap with modern medicine," *J. Ethnobiol. Ethnomed.*, 2018, doi: 10.1186/s13002-018-0268-x.

[11] X. Zhou, Y. Peng, and B. Liu, "Text mining for traditional Chinese medical knowledge discovery: A survey," *Journal of Biomedical Informatics*. 2010. doi: 10.1016/j.jbi.2010.01.002.

CHAPTER 18

COMBINING FORECASTS: A COMPARATIVE ANALYSIS OF ENSEMBLE METHODS FOR IMPROVED TIME-SERIES FORECASTING ACCURACY

Shipra Agarwal, Assistant Professor
Teerthanker Mahaveer Institute of Management and Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India
Email Id- shipra21aug@yahoo.co.in

ABSTRACT:

Combining forecasts is a common practice in the field of forecasting, which involves combining the predictions from multiple models or experts to create a more accurate and reliable forecast. The idea behind combining forecasts is that different models or experts may capture different aspects of the underlying data generating process, and by combining their predictions, one can improve the overall forecast accuracy and reduce the risk of relying on a single model or expert.

KEYWORDS:

Forecasting, Forecast Accuracy, Model-Based Approaches, Model Combination, Simple Averaging, Weighted Averaging.

INTRODUCTION

As gasoline prices do not increase during booms and decrease during recessions, there is no discernible cyclical pattern. If anything, a recession occurs the year after a significant spike in petrol prices. The most logical presumption for the future is that there won't be any cyclical patterns. The percentage change data are shown in figure 6.26. Moving to the seasonal pattern, regression analysis of the percentage change in gasoline prices as a function of seasonal dummy factors indicates substantial positive seasonals in April, May, and June. According to this regression, the price of gasoline increases by an average of 1.2%, 1.2%, and 0.9% during those three months, but does not change overall for the remainder of the year. Yet, given the many changes in the energy markets over the last 50 years, it makes sense to use X-11 to test for varying seasonal weights. Figure 6.27 compares the X-11 seasonally adjusted data to the unadjusted data. There, it is shown that the irregular component outweighs the seasonals [1].

The X-11 application may be used by readers to analyse the real and seasonally adjusted statistics. The video demonstrates how seasonally adjusted statistics for the next several years indicated larger rises in those months than really happened after the significant reduction in gasoline prices in February and March 1986. Also take notice that the largest monthly percentage increase in gasoline prices happened in April 1999 at 15%, not during any of the energy crises. These anomalies will eventually be eliminated, although in prior years, the updated seasonally corrected statistics revealed a significant reduction in Building structural models was the strategy used previously in this work, on the theory that the best predictions would likely result from the best effort to estimate the underlying population parameters. Nonetheless, we have made an effort to illustrate the potential drawbacks of that strategy using both instances that worked and those that did not. Several of these cases also demonstrated how predicting accuracy may be increased by continuous modifications, although up to this point, this has mostly entailed an ad hoc methodology. The best elements from both structural and nonstructural projections may likely be combined to increase predicting accuracy; this subject is covered in chapter 8.

The structural approach may not be the best option for a number of reasons. The fundamental structure, to start, could change over time. The Federal funds rate is one prominent illustration of this; the fundamental factors alter each time a new Fed Chairman is installed. Second, for thousands of distinct product lines, a non-structural approach could be more practical. If someone had 10,000 components or SKUs; it would be very time-consuming and labor-intensive to create a structural econometric model for each individual item and maintain all of those models with the most recent information. A nonstructural approach would be much more effective in such circumstances [2].

The effectiveness of a structural equation's forecasting can be seriously questioned if it cannot explain a greater percentage of the variation than a model that solely employs the lagged dependent variable and past residuals. Non-structural models may also be used as a yardstick. Also, it was previously shown that, even when all of the independent variables' values are known, a single equation model may still provide correct predictions if mistakes in estimating those variables are made. The non-structural method merits careful study due to all of these factors.

Moving-average (MA) and autoregressive (AR) models have been used since the beginning of econometrics. G. Udny Yule's seminal papers on moving-average and autoregressive models were released in 1926 and 1927, and his article "Why Do We Sometimes Get Nonsense Correlations Between Time Series" was the first to make the distinction between the usual tests of statistical significance and serial correlation of residuals. Herman Wold and M. S. Bartlett conducted the first studies on so-called "mixed models," which combine AR and MA models, in 1938 and 1946, respectively.

Nevertheless, rather than using AR and MA models for forecasting, the main emphasis of these publications was on investigating the statistical characteristics of these models. It was still widely believed at the time that structural models would provide more accurate projections. George E. P. Box and Gwilym Jenkins are usually credited with making the first significant effort to demonstrate how AR and MA models may be used for predicting. Instead of inventing novel statistical methods, Box and Jenkins' main contribution was to demonstrate how AR and MA models could be combined to create "an integrated and well-defined approach to time series forecasting via model building stimulating a good deal of practical application over a wide range of actual time series," according to Granger and Newbold.

DISCUSSION

Identification, estimate, and diagnostic testing make up the iterative three-step technique that Box and Jenkins established. Identification is covered; at this point, we can summaries by saying that it entails determining the AR and MA models' length of lag using the parsimony principle the fewer terms the model has, the more likely it is to produce accurate forecasts, ceteris paribus and checking the residuals for serial correlation. Standard least squares approaches are used in estimation along with a number of trend removal techniques. The data should often be adjusted for seasonality. As multiple distinct models are likely to provide results that seem to be similar, diagnostic testing entails determining how well the equation predicts beyond the sample period and then accordingly changing the model. The idea of stationary, or the presumption that the series has no trends, served as the foundation for the statistical formulae and tests that were first created for AR and MA models. As most economic time series do in fact include patterns, it is necessary to start by removing such trends. To evaluate if the trend in any given series is significant, a number of tests have been established. Nevertheless, these tests often provide misleading findings and should be used with care [3].

When in doubt, remove it; while estimating these models, deleting the trend is often the best course of action. In order to create the conventional techniques for estimating ARMA models, the methodology used here first makes the assumption that the series do not exhibit any trend. The removal of the trend when it does exist and testing for stationarity are covered identification, estimate, and diagnostic checking the three phases of the Box-Jenkins procedure.

Differencing is only suitable in the unit-root situation, and incorrect differencing might be damaging, even asymptotically, since then we will have an accurate approximation to the dynamics in the data (emphasis added) (emphasis added). Contrarily, Plosser and Schwert assert that working with differenced data is always preferable to working with data in levels15 and Maddala asserts that "it is preferable to use differencing and regressions in first differences, rather than regressions in levels with time as an additional explanatory variable."

Several unit-root tests have been suggested more subsequently to address the issues with serial correlation in the residuals, 17 the most well-known of which is the Phillips-Perron (PP) test. The results are often comparable to those obtained with the ADF test, however. There are a number of places where these and other tests are compared. 18 Yet, these tests don't actually provide us with much information when it comes to actual business forecasting. The foundation of ARMA models is the idea of stationarity, and if that assumption is broken, the results are likely to provide predictions that are less accurate. Moreover, the results of these exams are often inaccurate or incorrect.

Before responding, it should be noted that ADF findings and related tests are often not particularly reliable. There is often little difference between $r = 1$, which denotes a unit root and necessitates differencing, and (let's say) $r = 0.95$, which denotes a stationary process and does not need differencing. The various methods often provide contradictory results when determining whether a trend in a particular series is caused by a unit root with positive drift or a correlation with time. Rudebusch evaluated real GNP19 (as it was then) for a unit root and found that it did not only fail to reject the unit-root hypothesis but also a stationary hypothesis, according to a widely cited finding. The outcomes for many macroeconomic time series are typical.

Even Maddala, who generally supports differencing, admits that one of the major drawbacks of this method is the loss of important long-term information in the data; in fact, it is this loss that prompts Diebold to come to the conclusion that levels should typically be preferred for estimating equations. As a consequence, the answer to the question above largely relies on the kind of outcome the researcher is after.

The levels technique is often used if one is mainly concerned with identifying and forecasting the long-term trend for a particular variable, even at the risk of any misleading connection with temporal trends; previously it was thought that utilising ratios may at least partly solve this issue. Nevertheless, many ARIMA models are designed to capture short-term variations, either for forecasting reasons or as monitoring models to identify when inventory, shipments, output, or other aspects of a certain company or sector are deviating from expected levels. In these instances, the author has discovered that the preponderance of the data strongly supports the use of differencing if a trend may be present. Regardless of whether the ADF or PP tests reveal the existence of a unit root that is often accurate [4].

Significant trends may be seen in retail sales, the consumer price index, and the S&P 500 stock price index, but not in the initial differences of these series. Of course, there are instances when it is unclear if series like interest rates or inventory/sales ratios contain important patterns. The real rate of interest, theoretically speaking, is devoid of any long-term trend. The series may, however, have a strong correlation with a temporal trend over the previous 50 years, which may be an example of a random walk with a positive drift. Most textbooks state that the recommended technique is to take initial differences while eliminating the trend. Nonetheless, it is often preferable to use percentage changes, which are comparable to first logarithm differences. Think about the graphs of monthly changes for percentage changes and initial differences in stock prices.

The results of the ADF and PP tests categorically disprove the hypothesis that the initial difference in stock prices has a unit root. Yet, the graph reveals that the fluctuations are significantly bigger towards the conclusion of the sample period, such that the premise of a constant variance is broken. Naturally, the test only applies to the residuals and not to the original series, but if we compute an ARMA(1,1) equation for d(sp500) and look at the residuals, which are shown in figure 7.4, almost the same pattern appears. As a result, the ADF and PP tests, which are based on the assumption that variance is constant, are unsuitable; the heteroscedasticity issue may be readily resolved in this case by using percentage changes.

The time series for unadjusted and seasonally adjusted garment sales that were examined in the previous chapter are now being reviewed demonstrates that both series exhibit the same rising trend, despite the fact that the variation for the unadjusted data is substantially higher. Yet, the results of the unit-root tests are inconsistent. The critical values for rejecting a unit root for this series, as calculated by McKinnon and included in the EViews software, are -4.0 at the 1% level, -3.4 at the 5% level, and -3.1 at the 10% level. For the ADF test and the PP test, the values for the seasonally adjusted series are -1.9 and -2.2, respectively. The idea that there is no unit root is categorically rejected, as it should be given the strong trend in garment sales.

More than ever, today's company managers rely on timely and reliable information when making choices. Not just in home markets but also on a global scale, business has become very competitive. Economic and business analysts are trusted to provide guidance business executives.

A forecast analyst's job is to provide several predictions to the decision-maker, including estimates for sales, cash flow, inventories, and costs. It can also include examining the consequences of tighter monetary policy on the corporate sector and the policy ramifications of tax increases.

There is a higher dependence on company estimates due to two causes that have been significant. First, the systematic use and efficacy of forecasting approaches have significantly increased. As a result, managers now rely more on statisticians, operations analysts, and decision scientists to help them with their everyday tasks. Second, easier data collection and rapid data analysis for decision-making are now feasible because too increased information accessibility (databases) and affordable, portable, and powerful computers. The skills and knowledge of people who create predictions as well as the comprehension of those who utilise them are necessary for the efficient use of forecasting technique. Forecasters must assist management in understanding the usefulness and constraints of the predictions as organizations get increasingly interested in adopting quantitative forecasts in their decision-making. Also, they must present their results to management in a straightforward way that is clear of technical jargon. In this chapter, we'll go through the crucial phases in explaining a forecast's conclusions to management as well as the problems forecasters should take into account when writing their reports [5].

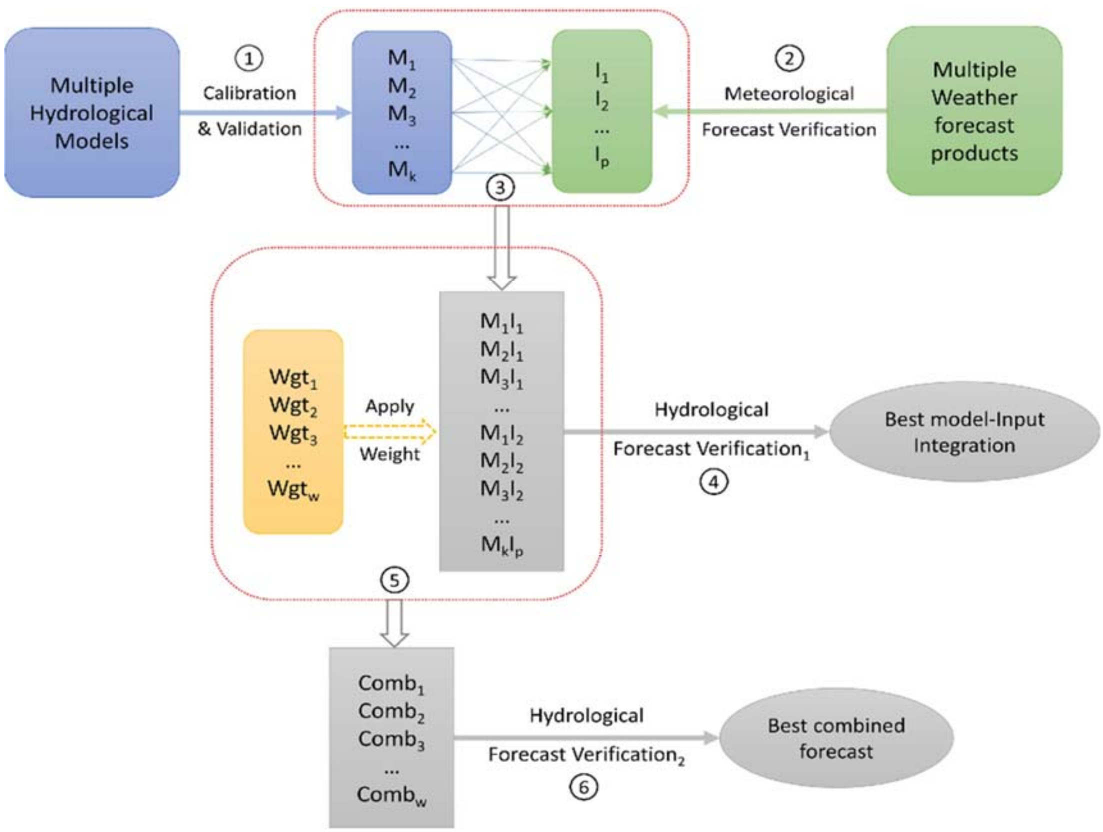


Figure 1: Illustrate the Identification of Combined Hydrological Models and Numerical Weather Predictions.

We discussed the role that forecasting may play in creating market strategies in Chapter 1 of this book. Managers are becoming more aware of how important forecasting and prediction are in making business choices. Even though this position is crucial, some businesses are too tiny to establish a dedicated forecasting department. We advise managers at these enterprises to take some of the straightforward models covered in this article into account when creating projections for their own companies. These models' detailed step-by-step instructions make it feasible to create accurate predictions. We acknowledge that some familiarity with statistics and economics is necessary in order to use these forecasting models.

On the other hand, major firms who see forecasting as a crucial component of their business operations may rely on forecasting experts to provide various projections for their company. In most cases, a department or group is chosen inside an organisation to handle these tasks and report the results to management. The process of informing management of projected outcomes is time-consuming. For successful communication, it starts far earlier than when a projection is made and its outcomes are presented to management. These are important inquiries that will direct and aid the forecaster in determining management's needs for a forecast. The forecaster's job is to provide a forecast that aids in decision-making when management foresees a prospective requirement. Figure 1 illustrate the Identification of Combined Hydrological Models and Numerical Weather Predictions [6].

A successful projection requires managerial involvement at this early stage of the process. Most managers are involved in other tasks at the company, therefore they may not completely contribute to forecast development. Holding discussion sessions with small groups of managers is one way to spark interest in forecasting among managers. In these workshops, managers learn about the advantages and constraints of forecasting and are prompted to consider how forecasting could support their decision-making. A one-on-one meeting with managers to explore potential applications to their area of interest might come after these sessions.

Forecasting experts must comprehend the issues and worries that management has. They must enquire specifically about the appropriate time range for their projections as well as the amount of detail or aggregation that is necessary. If the findings are broader than what the management anticipates, a sophisticated forecasting model and its forecasts may not be useful. The temporal horizon is equally important in management choices. If one were to produce a short-term vs a long-term projection, it would depend on whether they used facts from the recent past or those from a distant past [7].

The forecaster must take this knowledge into account throughout the planning process if management, for instance, needs a forecast by the end of the next month. Mid-level managers lose interest in adopting forecasts when they take too long to develop and are not accessible when they are required, according to research. The forecasters must be highly skilled not just in forecasting methods but also in conveying to management the value and suitability of these approaches in various contexts. It is preferable to provide more than one prediction scenario when providing information to management. The model's assumptions should be made transparent so that management has the necessary data to make a decision. Bear in mind that emphasising the aspects that are essential in making a choice may be more helpful to the management than providing the manager with an abundance of information.

Management feedback is incorporated into a solid projection. It is vital to assess the forecast's efficiency when one has been produced and the findings have been compared to actual occurrences. This is closely related to what we discussed in Chapter 1. (See Figure 1.1). The forecasting expert must examine the trend in any potential forecasting errors. This has to be explained to management, and their opinion on how to make the prediction better should be sought for. While reformulating the forecasting model, management's opinions should be taken into consideration. A system that takes adjustments into account and offers a feedback mechanism is essential to a successful forecasting process. It is advisable for managers and forecasters to be aware of several typical errors that reduce the efficiency of the forecasting process and, as a result, the prediction itself.

1. Inadequate or subpar data systems used by the company that do not increase the forecasting process' viability. This implies that the outcome won't change regardless of how accurate the forecasting model and technical team are. Good forecasting will start with a commitment to assembling and maintaining disaggregate datasets. Before any forecasting is done, this has to be told to management as soon as possible [8].
2. Middle management's lack of dedication to the forecasting process. The forecasting method is often approved by top management, but intermediate management does not recognise its worth or usefulness. In this circumstance, it is crucial that the forecaster explain to the middle management the value and efficacy of forecasting and how they may gain from it.
3. The decision-maker can discover that the facts or the prognosis are not readily accessible when needed. In order to prevent this issue, management and the forecaster must have a mutual agreement of how quickly the data is required and if the assignment can be finished in the allotted period.
4. On occasion, the study uses the incorrect forecasting model, and the forecasting personnel may not want to see it put into practise. They could make things worse for the forecasting unit and themselves if they are unwilling to share their results out of concern that they will be accused of technical ineptitude. To prevent issues of this sort, a framework between management and the forecasting team must be developed.
5. Rather than functioning as a separate unit that only functions in response to sporadic requests from other corporate units, a forecasting unit must be incorporated into the current corporate structure in order to be successful.

For example, the forecasting team is better able to provide a far more substantial response to the forecasting demands of the whole company when the forecasting unit is an important part of the budget and planning process. Forecasters may help management in its responsibilities as a controller and planner.

The forecaster must acknowledge that interpersonal and political savvy are just as crucial to effectively communicate conclusions to management as technical know-how. Control over the forecast means, to a considerable degree, control over the budgeting concerns and resource allocation for those companies that take the task of forecasting seriously. Forecasters may get embroiled in sensitive power conflicts inside the company in this situation. It is crucial to build interpersonal skills that safeguard the forecasting unit's objectivity and integrity inside the company. Listed below are some recommendations made by Bails and Peppers in 1993:

a) Get to know those who consult the prediction.
b) Discover the user's worries.
c) Address issues when creating a forecast for a specific unit before they become overwhelming.
d) You should stand up for your prognosis when faced with simply political criticism.

The forecaster's job is to provide insight into a wide range of decision options in order to have an influence on management choices. Management must clearly define forecasters' roles and responsibilities inside the company in order for them to be successful. The forecaster must also use caution while creating and presenting the forecast to management. This entails the time-consuming process of interrogating and nagging the user to determine what he or she really needs [9].

Successful company managers are aware of how crucial forecasting is to their choices. Within every business, forecasting has a wide range of applications. Predictions are created for planning and budgeting as well as for a variety of objectives, including those related to production, inventory management, marketing, advertising, sales estimates, and investment [10], [11].

CONCLUSION

Combining forecasts is a powerful technique that can significantly improve the accuracy and reliability of predictions. By integrating the predictions from multiple models or experts, we can reduce the risk of relying on a single forecast and gain a more comprehensive understanding of the underlying data generating process.

The choice of the combination method and the selection of individual models or experts are critical factors that affect the accuracy of the final forecast. As such, it is essential to carefully evaluate the performance of different combination methods and individual models or experts before making a final decision. When used appropriately, combining forecasts can enhance decision-making processes and lead to better outcomes in a variety of applications, including economic, weather, and financial forecasting.

REFERENCES

[1] S. G. Bandeira, T. M. G. de A. Barbosa, S. G. S. Alcalá, and R. O. Vita, “Comparison of selection and combination strategies for demand forecasting methods,” *Production*, 2020, doi: 10.1590/0103-6513.20200009.

[2] S. Xian and Y. Cheng, “Pythagorean fuzzy time series model based on Pythagorean fuzzy c-means and improved Markov weighted in the prediction of the new COVID-19 cases,” *Soft Comput.*, 2021, doi: 10.1007/s00500-021-06259-2.

[3] E. E. Elattar, S. K. Elsayed, and T. A. Farrag, “Hybrid Local General Regression Neural Network and Harmony Search Algorithm for Electricity Price Forecasting,” *IEEE Access*, 2021, doi: 10.1109/ACCESS.2020.3048519.

[4] Y. Liu *et al.*, "Wind power short-term prediction based on LSTM and discrete wavelet transform," *Appl. Sci.*, 2019, doi: 10.3390/app9061108.

[5] H. Li *et al.*, "Multivariate Financial Time-Series Prediction with Certified Robustness," *IEEE Access*, 2020, doi: 10.1109/ACCESS.2020.3001287.

[6] R. Singh and S. Srivastava, "Stock prediction using deep learning," *Multimed. Tools Appl.*, 2017, doi: 10.1007/s11042-016-4159-7.

[7] H. D. Chacon, E. Kesici, and P. Najafirad, "Improving Financial Time Series Prediction Accuracy Using Ensemble Empirical Mode Decomposition and Recurrent Neural Networks," *IEEE Access*, 2020, doi: 10.1109/ACCESS.2020.2996981.

[8] N. A. Hitam and A. R. Ismail, "Comparative performance of machine learning algorithms for cryptocurrency forecasting," *Indones. J. Electr. Eng. Comput. Sci.*, 2018, doi: 10.11591/ijeecs.v11.i3.pp1121-1128.

[9] T. Regina and P. Jodiawan, "Proposed Improvement of Forecasting Using Time Series Forecasting of Fast Moving Consumer Goods," *JIEMS (Journal Ind. Eng. Manag. Syst.*, 2021, doi: 10.30813/jiems.v14i1.2418.

[10] S. Zheng, Q. Zhong, L. Peng, and X. Chai, "Erratum: A simple method of residential electricity load forecasting by improved Bayesian neural networks," *Math. Probl. Eng.*, 2018, doi: 10.1155/2018/4276176.

[11] S. S. Jones, A. Thomas, R. S. Evans, S. J. Welch, P. J. Haug, and G. L. Snow, "Forecasting daily patient volumes in the emergency department," *Acad. Emerg. Med.*, 2008, doi: 10.1111/j.1553-2712.2007.00032.x.

CHAPTER 19

BUILDING AND PRESENTING SHORT-TERM SALES FORECASTING MODELS: A PRACTICAL GUIDE FOR BUSINESS DECISION MAKERS

Vivek Anand Singh, Assistant Professor
Teerthanker Mahaveer Institute of Management and Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India
Email Id- vivekanand.ima@gmail.com

ABSTRACT:

Short-term sales forecasting models are essential tools for businesses to effectively plan and manage their operations. These models help organizations predict future sales based on historical data and market trends, allowing them to make informed decisions on inventory management, marketing strategies, and resource allocation. Building a short-term sales forecasting model requires selecting appropriate variables, such as past sales, seasonality, economic indicators, and marketing activities, and applying appropriate statistical techniques to analyze the data. Machine learning algorithms can also be used to improve the accuracy of the model by identifying patterns and trends that may be difficult to detect with traditional statistical methods.

KEYWORDS:

Historical Data, Market Trends, Machine Learning, Short-Term Sales Forecasting, Variables, Statistical Techniques.

INRODUCTION

Businesses cannot afford to rely simply on the qualitative judgement of their management in today's competitive market, when business decisions are made at a much quicker pace and with a higher dependence on quantitative analysis. There are, however, drawbacks to relying only on the forecasting unit or department while business is booming. There is a need for improved forecasting processes and tools in order to make business choices, as shown by the recent collapse in the housing industry and the general weak economic performance in the U.S. Even though many people mistakenly believed that the market would steer the economy in the right direction in the latter part of the 1990s due to psychological factors like "feeling good" about the economy, Chairman Greenspan of the Federal Reserve Bank claimed that there is a "irrational exuberance" in the economy. Considering that demand for forecasting department services is income elastic, forecasting staffing levels are often reduced in response to a dramatic decline in sales, prices, and profits. Companies started to slash costs in the 1980s as a result of falling prices in the oil, real estate, and financial industries. They did this by eliminating their research departments and lessening the need for forecasting. Businesses must rely on quantitative research and accurate forecasts at times like these [1].

The necessity for strategic planning as managers guide their firms into the future is highlighted by recent economic downturns in the United States and several other developed countries. Such planning tactics may efficiently use forecasting technologies. Nobel Prize winner Gary Becker

(2001) notes that the September 11 attacks may have pushed the U.S. economy into a recession, which was already in decline. He notes that the sharp decline in stock prices during the first week after markets reopened is a reflection of pessimism and uncertainty about the future. Forecasters may significantly contribute to assisting business entities lessen the market uncertainty in this situation.

Any instability brought on by a crisis of major proportions on a national or worldwide scale would also have substantial ramifications for the forecasting function in company plans. Throughout the second half of the 1990s, the globalisation of trade, which allowed for a more free movement of people, money, and products, increased yearly U.S. growth by three-quarters of a percentage point (Engardio et al., 2001). This growth rate may not continue in the foreseeable future as globalisation could halt and become more expensive. Businesses would likely have to spend more on security and insurance for employees who work abroad. Providing Management with Predictions 313 property. More border checks can cause cargo movements to be slowed down, requiring businesses to stock up on additional merchandise. Tighter immigration regulations may have a substantial impact on the influx of skilled and blue-collar workers that has enabled businesses to grow while controlling salaries. In the meanwhile, a growing preoccupation with political risk has forced businesses to drastically reduce the scope of their new investments. Given the changing circumstances, forecasters may be of tremendous assistance in offering insightful projections [2].

Forecasters must assist management in understanding the value and constraints of their predictions as company managers grow increasingly interested in adopting quantitative forecasts in their decision-making. Also, they must present their results to management in a straightforward way that is clear of technical jargon. The process of informing management of the findings is time-consuming and starts far before a prediction is created. It is advised that management be included in the forecasting process to make it more controllable. This will provide the forecaster a clear grasp of management's requirements and how to address them. Alternative scenarios and the underlying assumptions should be communicated to management along with the findings. It is important to emphasise the error's meaning and its causes. Ultimately, the model should contain management input, and fresh findings should be compared to real occurrences. Due to upcoming national and international events as well as economic uncertainty, forecasting will be necessary. The accessibility of datasets and more complex quantitative studies improves forecasting methods. The breakdown of economic time series into their trend, cyclical, seasonal, and irregular components was discussed in the preceding chapter. Techniques for extracting the trend, long-term cyclical swings, and seasonal patterns were described, but the irregular component received very little attention. That's because it's supposed to be random in the traditional linear model [3].

DISCUSSION

The likelihood of such assumption being true has previously been shown. The strategies discussed in this chapter may be utilised to increase forecasting precision using data from the irregular component. The two main approaches are moving-average models, where predictions are modified by residual values from the past, and autoregressive models, which are crucial if the residuals are serially connected. As contrast to the notion in the previous chapter, where a moving average was some weighted combination of the current and lagged values of the variable itself, a moving-average model here refers to a regression in which the dependent variable is a function of prior residuals.

If the series has no discernible trend, the models are referred to as ARMA models. Integrated denotes the removal of the trend. These models, of course, lack any economic factors and are non-structural. The temporal trend has been eliminated in all instances where it is relevant, leaving just lagged values of the dependent variable and prior residuals as independent variables. Identification is the process of selecting the best (p, d, q) structure in an ARIMA model. 20 Because theory does not specify the ideal duration of lag or level of differencing, considerable guessing is involved in this procedure. The majority of the time, d is either 0 or 1, hence the following identification-related remarks are limited to figuring out what p or q should be set at [4].

There is often a compromise. If more words are included, the adjusted R2 will grow, but there is a greater chance that curvefitting may result in decreased prediction accuracy. In essence, a metric is required to balance the improvement in corrected R2 against the erroneous increase in correlation that results from the inclusion of more words. To account for the additional phrases, some kind of "penalty" term is required. The next three steps are given by Hannan and Rissanen (HR) establish the longest possible latency that an AR model may have first. Second, find the longest possible lag in an AR model using the Akaike Information Criterion (AIC), which has a "penalty" term. Finally, find the maximum delays for a mixed ARMA model using the Schwarz criteria (SC), which imposes a heavier penalty for extra coefficients. Here we discuss AIC and SC.

Identify the longest possible latency. The original HR technique advised applying a stationary series to a pure autoregressive equation until the corrected R 2 was maximised. Nevertheless, the author has shown that a correlogram analysis is a more effective method for determining the maximum length of lag—the point at which the partial autocorrelation coefficients (PAC) cease to be significant. The standard practise is to regard that lengthy lag as being caused by one-time external variables and dismiss that result in a situation where, for example, delays 1 through 4 were significant, 5 through 9 were not, but lag 10 was significant again [5].

The lone exception is when monthly data exhibit a strong PAC at lag 12, of the ARIMA model's guide to seasonal adjustment techniques. The lag estimated from the correlogram (or regression equation) will often be too lengthy. The word "penalty" is used in this context. The fundamental concept is to balance the lowering standard error as the number of AR terms grows with some penalty for adding terms to a pure autoregressive model. Figure 1 illustrate the Factors that affect the Impact Sale Forecasting.

The common test is called AIC,22 and it is a part of EViews and the majority of other regression software products. The highest degree of autocorrelation, k, is selected to minimise where Sei2 is the sum of the squared residuals, n is the number of observations, and k is the maximum degree. This estimate, according to Shibata23, is inconsistent and overstates the model's optimum degree. The AIC criteria often overestimates the ideal degree by a factor of 2 or 3, thus this discrepancy is not insignificant. We advise examining the correlogram rather than a pure autoregressive equation because of this. The goal of the research at this point is to discover the greatest plausible lag, not the ideal lag, thus this issue is not very significant. So, overstating that latency will need a few more calculations but shouldn't have an impact on the final model selection.

Figure 1: Illustrate the Factors that affect the Impact Sale Forecasting.

These procedures, like all of the formulae in this section, should be used with caution since they are recommendations rather than strict rules. In this instance, a potential weakness of SC is that it often produces inconsistent results depending on the length of the time series. Consider comparing a short and long time series; the small one, for example, has 60 monthly observations, while the long one, for example, has 600 monthly data, to understand how this may happen. When n = 60, the term (log n)/n is 0.068, whereas for n = 600, it is just 0.011 [6].

The Schwarz criteria will thus increase by 0.068 for the short series but only by 0.011 for the long series with the addition of a second lag term. While it is often close to the same, the change in s2 caused by adding another lag term might vary for short and long series. As a result, for longer series as opposed to shorter ones, the Schwarz criteria would tend to suggest a longer lag distribution. Because of this, it's not recommended to employ the Schwarz criteria entirely. In general, it is advisable to choose two or three distinct ARMA models, test each one's forecasting precision beyond the sample period, and then decide which model(s) to use more fully.

The monthly inflation data series, which was previously examined may be used to demonstrate these ideas. To recap, ocular examination shows a maximum AR latency of 5. Unfortunately, the AIC criteria suggests a 15-lag ideal value when we solve a pure autoregressive equation, which is much too lengthy. This anomaly was not caused only by the use of the monthly inflation data set. The reader can see that the highest lag for the first difference in apparel sales, which was previously covered in section 7.2, is 4, but the ideal latency, according to the AIC criteria. Now, using the Schwarz criteria, we estimate an ARMA model with a maximum latency of 5, discarding each term one at a time. First, observe that, for the long series, the SC value for the model is lowest, but for the short series, the SC value for the model is lowest. This supports what was said before. Second, there is a reversal of the monotonic trend that the SC values for the long series exhibit.

This clearly implies that is likely the longest model that should be taken into consideration, especially in light of the short series findings. Third, it appears that the shortest model should be taken into consideration because the gradient for the short series starts to decline after [7].

By this time, it should be clear that while AIC and SC may help with identification, they cannot offer firm conclusions. However, a brief diversion is taken to examine the estimation procedure, including seasonal adjustment, and description of the terms contained in Views, before turning to some helpful rules for assessing forecasting accuracy. So far we have ignored the issue of whether the data used in the ARMA process are seasonally adjusted or not, implicitly assuming they have already been adjusted by X-11 or similar processes. In the examples given above, seasonally adjusted data are used for inflation, and interest rates do not contain any significant seasonal factors. However, in many cases, it is useful to identify and measure the seasonal factors themselves; also, forecasting accuracy may be improved using data before seasonal adjustment in cases where the seasonal factors are not random but are related to the economic environment – such as retail sales and housing starts.

The discussion here centers on the use of monthly data; the same general analysis would apply for quarterly data, with the critical lag being four periods instead of 12. It is assumed throughout this discussion that the seasonal factors are multiplicative, as opposed to additive. Additive seasonal factors would be used only in the case where some of the observations were zero or negative; e.g., when the series is already a first difference or percentage change. We first consider the AR and MA processes separately. Suppose a seasonally unadjusted series has significant PACF for 1 and 2, then a long string of insignificant coefficients, then a significant coefficient again for a lag [8].

To understand the practical significance of these terms, assume for the moment that a1 is fairly close to unity and a2 is fairly small in the AR equation, so can be reasonably approximated as Dyt = h1Dyt-12. The pure MA case is somewhat more difficult to interpret intuitively, because there are no guidelines for the usual sizes of b1, b2, and h2. A fairly typical case for a series that is actually composed primarily of random elements would have ab1 and b2 fairly close to 0 and h2 fairly close to unity. The economic interpretation is that the value of y in the current period is highly correlated with the error term a year ago. However, if the series has not been seasonally adjusted, that error term would probably reflect the seasonal factor, and hence could also be measured by SAR.

This strongly suggests that, in a seasonal ARMA model, the SMA (12) term should not be included without at least trying the SAR (12) term as well; otherwise, the SMA(12) term itself serves as a suboptimal proxy variable for the seasonal factor. However, when both of these terms are tried together in the same equation, the results often look quite unusual [9].

These results are further evidence of a spurious negative correlation between SAR and SMA. On the one hand, the equation says the change in stock prices is positively correlated with the change a year ago, while the SMA term says the change in stock prices is negatively correlated with the error term a year ago. Essentially these two factors cancel each other, but simply glancing at the regression one might get the opinion they are both highly significant contributors to forecasts of the stock market.

To a certain extent this example is somewhat of a "ringer" because stock prices have no seasonal factor. It is included as a lesson to the unwary of what happens for those who blunder into an equation with no idea of the underlying structure. To avoid such nonsensical results, the following procedure is recommended. First, don't use either SAR or SMA terms unless the underlying series has a strong seasonal component. Usually this is apparent by eye, but if not, calculate a regression with seasonal dummy variables. If the seasonal pattern is strong, test for the significance of SAR. If that term is significant, add SMA, which is usually important if the values of the seasonal factors are changing. However, this procedure is best attempted one step at a time rather than adding all the variables at once and finding they are all significant.

The use of SAR and SMA terms often introduces spurious correlations. On the other hand, where seasonal patterns are strong, they cannot be ignored. Using seasonally adjusted data solves some but not all of the problems, because changing seasonal patterns are based on recent trends, which means that the forecaster must predict the seasonal factors as well as the underlying data. One alternative to these problems is to take first differences, or percentage changes, for the data over the past 12 months; i.e., the value of ay in any given month this year minus the value of ay in the same month last year. If the underlying variable has a significant trend, percentage changes are usually better, in which case the dependent variable would have the form (yt - ayt-12)/yt-12. If this formulation is used, the seasonal factors do not have to be estimated independently. Also, any shift that does occur can be estimated in the regression using either economic variables, truncated time trends, or dummy variables. In summary to this point, the following options can be chosen to minimize forecasting error using an ARIMA model with data that are not seasonally adjusted [10].

1. 1 Seasonally adjust the data using the X-11 programmed or a similar method. Assuming the variable has a significant trend, calculate first differences or percentage changes with the seasonally adjusted data, and then calculate the ARMA model.
2. 2 Do not seasonally adjust the data; take first differences or percentage changes, and then calculate the ARMA model with the original data using SAR(12) and possibly SMA(12) (12)
3. 3 Use seasonal differencing take first differences or percentage changes over the past year instead of the past month, and then calculate an ARMA model using those differences as the dependent variable.

To illustrate some results, reconsider three examples that have been used previously. The series for apparel sales has a strong trend and a strong seasonal pattern. The series for housing starts has no trend and a strong seasonal pattern. The monthly rate of inflation, this time as measured by the PPI for industrial commodities, has no trend and a weaker seasonal pattern; it is more influenced by random elements. For each series the following regressions are calculated: a(a) seasonally adjusted data, ARMA a(b) unadjusted data, ARMA, SAR(12) a(c) unadjusted data.

All these results are based on equations estimated from 1959.1 through a1997.12, with the period from 1998.1 through 1999.12 used for forecasting. In each case, the ARMA model includes the optimal lag structure based on the at-ratios of the individual terms, the overall adjusted aR a2, and the Schwarz criteria, although in general those did not have much influence on the optimal choice.

The results no consistent pattern, which may seem discouraging to the researcher. However, they do illustrate there is no one method that works best all the time. There are no cut and dried rules for choosing the ARMA model that will minimize forecast error. Also, of course, testing forecasting accuracy over a two-year period is hardly a comprehensive sample.

Nonetheless, the results are illustrative. In particular, note that the seasonally adjusted data are best for apparel sales and worst for housing starts, for which seasonal differencing gives the best results. The results for the monthly inflation rate for industrial commodities are about the same for seasonally adjusted data and seasonally differenced data when SAR(12) and SMA(12) are used. The big changes in the PPI are due primarily to the oil shocks, component, so the method of adjustment does not matter very much.

The main conclusion to be drawn from these experiments is that the optimal type of seasonal adjustment method depends critically on the nature of the seasonal factors. If they are always the same each year, the method does not matter at all. However, that is rarely the case for economic time series. The following rule of thumb can be used. If the seasonal process is predictable – e.g., the day on which Easter falls is obviously known in advance – the X-11 programmed will work better. If the fluctuations are due to random events, such as weather, oil arises, strikes, or other unforecastable interruptions, X-11 is likely to introduce spurious changes that will not reoccur. As a result, it is probably better to use seasonally differenced data in those cases [11].

CONCLUSION

Building and presenting short-term sales forecasting models is an important process for businesses to effectively plan and manage their operations. By analyzing historical data and market trends, businesses can predict future sales and make informed decisions on inventory management, marketing strategies, and resource allocation. The selection of appropriate variables and statistical techniques, as well as the application of machine learning algorithms, can improve the accuracy of the forecasting model. Presenting the results in a clear and concise manner, including highlighting the key drivers of sales and the assumptions underlying the forecast, is equally important.

REFERENCES

[1] N. M. P. Bocken, I. de Pauw, C. Bakker, and B. van der Grinten, "Product design and business model strategies for a circular economy," *J. Ind. Prod. Eng.*, 2016, doi: 10.1080/21681015.2016.1172124.

[2] C. M. Flath and N. Stein, "Towards a data science toolbox for industrial analytics applications," *Comput. Ind.*, 2018, doi: 10.1016/j.compind.2017.09.003.

[3] N. M. P. Bocken, I. De Pauw, C. Bakker, and B. Van Der Grinten, "Journal of Industrial and Production Engineering Product design and business model strategies for a circular economy Product design and business model strategies for a circular economy," *Taylor Fr.*, 2016.

[4] H. Anthem, M. Infield, and S. Morse-Jones, "Guidance for the rapid assessment of cultural ecosystem services," *Oryx*, 2016, doi: 10.1017/s0030605315001271.

[5] J. E. Perry, "Managing moral distress: A strategy for resolving ethical dilemmas," *Business Horizons*. 2011. doi: 10.1016/j.bushor.2011.05.005.

[6] Q. Hongdao, S. Bibi, A. Khan, L. Ardito, and M. Nurunnabi, "Does What Goes Around Really Comes Around? The Mediating Effect of CSR on the Relationship between Transformational Leadership and Employee's Job Performance in Law Firms," *Sustainability*, 2019, doi: 10.3390/su11123366.

[7] M. Asif, C. Searcy, A. Zutshi, and N. Ahmad, "An integrated management systems approach to corporate sustainability," *European Business Review*. 2011. doi: 10.1108/09555341111145744.

[8] J. Bleicher and H. Stanley, "Digitization as a catalyst for business model innovation a three-step approach to facilitating economic success," *J. Bus. Manag.*, 2016.

[9] J. McKenzie, C. van Winkelen, and S. Grewal, "Developing organisational decision-making capability: A knowledge manager's guide," *J. Knowl. Manag.*, 2011, doi: 10.1108/13673271111137402.

[10] R. Dixit and V. Sinha, "Investigating tools and techniques to promote workplace training transfer," *J. Work. Learn.*, 2022, doi: 10.1108/JWL-04-2021-0052.

[11] Q. Hongdao, S. Bibi, A. Khan, L. Ardito, and M. Nurunnabi, "Does what goes around really comes around? The mediating effect of CSR on the relationship between transformational leadership and employee's job performance in law firms," *Sustain.*, 2019, doi: 10.3390/su10023366.

CHAPTER 20

METHODS OF LONG-TERM FORECASTING: A COMPREHENSIVE REVIEW AND COMPARATIVE ANALYSIS OF STATISTICAL AND MACHINE LEARNING APPROACHES

Vipin Jain, Professor
Teerthanker Mahaveer Institute of Management and Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India
Email Id- vipin555@rediffmail.com

ABSTRACT:

Long-term forecasting is an essential aspect of many fields, including economics, finance, climate science, and social sciences. Methods of long-term forecasting involve the use of statistical and mathematical models to predict future trends and patterns over an extended period. The choice of long-term forecasting method depends on the nature of the problem being forecasted and the available data and resources. The use of multiple methods and a combination of qualitative and quantitative approaches can help improve the accuracy of long-term forecasts and enable decision-makers to plan for the future with greater confidence.

KEYWORDS:

Long-Term Forecasting, Mathematical Models, Scenario Planning, Simulation Modeling, Time Series Analysis.

INTRODUCTION

In using the common nonlinear algorithms, the answer that is obtained may differ depending on the starting point that is chosen. To understand this, think of a hilly terrain, with several peaks and valleys. One peak is the highest, so if were doing aerial scanning, there would be no difficulty in identifying that peak. However, suppose you are in one of the valleys all directions are up. It may be that the slope up to peak A is steeper than the slope to peak B but peak B is higher. Because of the steeper slope, you start up A. When you get to the top, you are unable to see from your vantage point that B is higher than A, so you think you have climbed the right hill when it was the wrong one. Alternatively, the valley may be so flat for a while that you have no idea which way to turn, so you remain stuck in the valley indefinitely.

Mathematically, then, any nonlinear algorithm could produce an incorrect answer for two reasons. First, it could reach a local rather than a global maximum. Second, it could fail to converge at all. Are these common enough to be of concern, or are they merely statistical lacunae of no practical significance? As it turns out, they happen quite often estimating ARMA models, it is not unusual for Views or other programs at print out a message that convergence was not achieved after 100 iterations [1].

Also, you will sometimes receive a message that the equation “failed to improve” after a certain number of iterations. In other cases, the results are obviously nonsensical; in that case, it is usually a good idea to simplify the model somewhat by dropping a term and see whether the problem persists. If it does, the almost likely answer is that the algorithm is starting from the wrong spot, in which case the researcher should specify estimated values to help get the estimation procedure pointing in the right direction. Of course, the values of the coefficients are not known in advance; if they awere, it would not be necessary to estimate the equation in the first place. However, when working with ARMA models that have been properly integrated, we are always dealing with stationary processes.

That means Saj < 1 and aSbj < 1. Thus reasonable starting points might be 0.5; other possibilities are a0.75 and 0.25. This will usually give you a better chance of reaching the right value, but if not, the remaining possibility is to simplify the model until convergence is achieved. EViews printouts for ARMA models contain two additional terms: back casts, and AR and MA roots. Back casting simply means the programmed fills in the residuals, so the sample period is not truncated by the number of period's equal at the longest lag. In practice that makes very little difference. The bottom part of the EViews printout for ARMA models contains information about the values of the inverted AR roots and inverted MA roots. They may be either real or imaginary; that doesn't matter. The key danger point occurs if the value of either of these roots is greater than unity, in which case the AR or MA process is explosive. If you have already taken first differences or percentage changes, that is unlikely to happen. If you are working with levels and alone or more of the roots is greater than unity, then the trend should be removed [2].

We now turn to testing different types of models. At this point it is assumed the trend has been removed, and either the data have been seasonally adjusted, aor seasonal factors have been incorporated in the estimation procedures (by using SAR(4) for quarterly data and SAR(12) for monthly data). It is usually more difficult to choose the optimal length of the AR and MA a processes, denoted as ap and q. In general, the overall goodness-of-fit statistics will look about the same for a large variety of different lag structures; however, the forecast accuracy of these equations may vary significantly. In general, the amore parameters that are estimated, the more likely that the equation simply represents an exercise in curve fitting that will not generate accurate predictions. The rule of parsimony is used here: use fewer rather than more parameters. However, this rule by itself explains very little. Does that mean ARMA(1,1) models are preferable to ARMA(2,2) models? Clearly more information is needed to make an informed choice. The parameters of ARMA processes, like other statistical models, are based on the assumption of a constant variance. If that is not the case, the equation may not generate accurate forecasts although the parameter estimates appear to be quite robust. When there appears to be any trend, it is usually better to remove the trend before proceeding further. If the trend is weak, the forecasts will be about the same whether the trend is removed or not, and no harm will be done by removing it [3].

DISCUSSION

Start by estimating ARMA models of low order, such as (1,1). Then expand the model as long as the terms remain significant. Occasionally the terms in a (2,2) or (3,3) model will not be significant, but the coefficients will then become significant again as the model is expanded further. That simply represents very high correlations among the terms and does not improve the model. That is often a warning sign that curve fitting is invading the equation. In general, the concept of parsimony should be observed.

Theoretically, the residuals should be randomly distributed (determined by looking at the correlogram). However, in practise that will not happen very often, and generally cannot be used as a criterion for determining the order of the ARMA model. Except for seasonal factors, the maximum structure will occur at the first PACF that becomes insignificant, but that sets the maximum rather than the optimal length of lag.

In order to indicate the practical difficulties associated with ARMA models, we look at several alternative equations for the Federal funds rate; the data series begins in 1955. First, we estimate a simple ARMA(1,1) model. Although the R 2 is high, that means nothing in this sort of model; the test comes when we examine the correlogram to see whether the residuals are randomly distributed [4].

This is not a very encouraging start, since one would hope to find that the significance of the residuals decreases as the length of lag increases. Also, there are no seasonal factors in the Federal funds rate, so significant coefficients a year ago are puzzling. Expanding the equation to ARMA(2,2) and then (3,3) does not help matters at all; the same PACF are still significant.

In that case, those PACF are no longer significant but we now find that PACF for lags 11 through 14 are all significant. If even more terms are added, the equation no longer converges, and it becomes clear this expansion process cannot continue indefinitely. What is causing the problem? In this case – and in many similar examples it is best to go back and look at the economics the residuals from a simple AR(1) model; it is clear that something very unusual happened in early 1980. More specifically, the standard error of the equation is 0.58, but the error for May 1980 is 6.44 more than 10 times the standard error. Clearly that observation is drawn from a different population. That error is so large that it affects the estimated parameters for the entire sample period would not have been predicted by any econometric model either, so it does not necessarily follow this is a failure of ARMA modeling. However, it illustrates what can happen when there are severe outliers and serves as yet another reminder that it is always a useful idea to plot the data before calculating any regressions.

By 1983, monetary policy had returned to normal, so an ARMA model is estimated starting in that year. Once again the process starts with a (1,1) model; the correlogram shows a very high PACF with lag 2, which strongly suggests trying a (2,2) model. For that model, the residuals are randomly distributed, but neither the MA(1) nor MA(2) terms are significant, raising the question of whether they are needed. First MA(2) is eliminated, and then MA(1); the resulting ARMA(2,0) model still has random residuals, so that is the one that is chosen. The economic interpretation of this equation says the Federal funds rate is highly influenced by its values one and two months ago, but is not as subject to exogenous shocks. This study analyses both seasonally adjusted and unadjusted data, since in some cases the seasonal factors themselves may be of interest. The procedure is as follows. First, several different methods for removing the trend are examined. Figure 1 illustrate the method of demand forecasting.

Figure 1: Illustrate the method of demand forecasting.

Second, an ARMA model is estimated with seasonally adjusted data, choosing the optimum p and q. Third, the model is estimated with seasonally unadjusted data, including AR(12) and MA(12) terms. In all cases, the sample period is truncated at the end of 1997, and the forecasting accuracy is assessed for the 1998–2000 period. The series used is in constant dollars; i.e., the dollar figures are deflated by the PPI for machine tools. The series for the volume of machine tool orders is shown in figure 7.9; the series has a significant trend, although it only explains about 5% of the variance. Both the ADF and PP tests indicate a unit root for this series, so the trend is removed[5].

While this series has a significant trend, the first differences are not materially larger at the end of the sample period, so they are used instead of percentage changes. The initial step, after plotting the data, is to estimate an ARMA(1,1) equation. The correlogram shows significant PACF at lag 6 and again at lag 12 even though the data are seasonally adjusted. The next step is to try an ARMA(2,2) equation; the additional terms are highly significant, but the same PACF correlations remain. When an ARMA(3,3) equation is tried, ridiculous t-ratios start showing up for some of the terms, and the equation almost does not converge. Another 258 UNIVARIATE TIME-SERIES MODELING/FORECASTING possibility is to try ARMA(2,2) plus AR(6), MA(6), AR(12), and MA(12), since that is where the significant correlations occurred. All these terms were significant but the correlogram revealed that the residuals remained serially correlated.

None of the ARMA models gives very useful predictions. There was a big increase in machine tool orders in January 1998, which this sort of model is not designed to capture. Note that when longer lags are added, the forecast error increases. This suggests that the correlations with longer lags are spurious or due to outliers rather than an integral part of the process. The seasonally unadjusted data series is considered next. In this case the ARMA(1,1) model with AR(12) and SA(12) gives results that are just about the same as when further terms are included; so using the principle of parsimony, the simpler form of the model is used. It can be seen that ARMA(1,1), (2,2), and (3,3) all give almost identical results when AR(12) and MA(12) terms are included [6].

However, just as was the case for the seasonally adjusted data, the correlogram shows that the residuals are serially correlated no matter how many lags are added (within reason). At least in this case, it is apparently no easier to obtain random residuals in time-series models than it was in structural econometric models.

Since there is no trend to be removed and both the ADF and PP tests verify this finding it is possible to proceed directly to estimating ARMA models. First, the ARMA (1,1) model is tried; the correlogram shows significant PACF for several lags. The ARMA (2,2) model is tried, but there is no improvement; similarly for the ARMA(3,3) model. Even an attempt to estimate an ARMA (6,6) model, which violates the principle of parsimony and generally gives inferior forecasts, still results in serially correlated residuals, so no useful purpose would be served by adding all those additional terms to the model.

None of them works very well; the big increase in the I/S ratio in 1998 and the decline in 1999 and 2000 are not tracked by any of the ARMA models. In this graph, the actual data for 1997 are shown in comparison with the actual and forecast values from 1998 through 2000. The previous two case studies have contained examples where the trend explains only a very small proportion of the sample period variance. We now turn to a case where a simple trend explains over 99% of that variance, and determine whether ARMA models can improve forecasting accuracy in this case. It is not initially clear whether trend removal using levels or logarithms will produce better results, so both methods are tried.

However, in this case the forecast results are instructive. While the ARMA models are all about the same, the forecasts using the detrended series for the logs are substantially better than those using the levels. That is not just a random result; the rate of growth in employment has been fairly steady over the sample period, and since the series has a strong time trend, that implies the actual changes in recent years will be greater than the sample period average, whereas that is not the case for the logarithmic approach. In this example, then, a clear difference emerges [7].

The superiority of the logarithmic equation still leaves the question unanswered of whether a simple equation log(employment) = a + b (time trend) would give essentially the same results; these results can also be compared with a naive model which says the percentage increase is equal to the average increase over the entire sample period 1947.01 to 1997.12. An equation in which log(employment) is a function of a time trend and an AR(1) adjustment provides forecasts that are very similar to the ARIMA(2,1,2) model; the AR(1) term provides virtually all the forecasting information, even though the other terms are also significant. That is often the case for government data with strong time trends where the data collection process often causes some artificial smoothness. The naive model falls below the actual level in 1998 and 1999 but, when slower growth ensues, is almost exactly equal to the level of employment by the end of 2000. Hence the naive model does just as well as ARIMA or trend models.

Both structural econometric predictions and non-structural time-series processes were covered in detail in the preceding seven chapters. These instances showcased both achievements and failures to highlight the fact that projections are never completely accurate. Economists discovered more than 30 years ago that forecasting mistakes might frequently be decreased by integrating several forecasting techniques; since then, many research have been conducted to support this claim. Yet, not everyone agrees with this viewpoint. Why not use the prediction that seems to be the best if there are many forecasts? Or, to put it another way, should projections based on painstakingly constructed and calculated models be altered by the observations of some wanderer off the streets?

Forecast inaccuracy is not usually decreased by combining predictions. When forecasts are based on econometric models that each have access to the same information set, as noted by Clements and Hendry, "combining the resulting forecasts will rarely be a good idea... when models do not draw from a common information pool, and are essentially of a different nature or type, or when models are differentially susceptible to structural breaks, then the case for combination is more persuasive [8].

The benefits of combining projections continue to be a contested topic, as is the case with the majority of forecasting-related topics. The combination of projections "may be superior than each of the elements," Clements and Hendry add. Such combination, however, goes against the idea of a progressive research plan. The latter would imply that when a model is determined to be deficient in certain aspects, it is preferable to improve the model. The objective is to create a model that "forecast the process better than its competitors."

It was shown many years ago that combining forecasts with weights inversely proportionate to the errors can minimise forecast error provided the predictions from two or more separate sources are unbiased, if the error variance stays constant, and if the underlying function has not changed.

The benefits of combining predictions have been the subject of a large body of literature, most of which Clemen has evaluated. Yet, in reality, these circumstances are seldom realised, despite the fact that combining projections would minimise forecast error. Standard regression analysis alone might be utilised to get the best projections if the underlying function and error variances didn't really change over time. Yet, when these prerequisites are not satisfied, the majority of the issues with practical business forecasting occur.

The following forecast combining techniques are discussed in this chapter:

a) Techniques that are not structural, such as ARIMA, Box-Jenkins, and Holt-Winters
b) The use of both structural and non-structural methodologies; the importance of judgement as shown by mood indices and consensus predictions; and
c) Modifying slope and constant terms in structural equations
d) Combining the aforementioned techniques.

The outcomes for these categories may be distilled into the next list. Secondly, integrating non-structural approaches does sometimes decrease prediction error, albeit not as often as may be anticipated based on conventional statistical metrics. Second, there is conflicting information about the effectiveness of combining structural equations with the ARIMA process; the solution will rely on the underlying characteristics of the specific time series and the degree of serial correlation. Finally, even if the projections from the consensus or mood index alone are no better than those from a naïve model, the involvement of judgement may often increase forecast accuracy. Fourth, a structural equation's constant and slope terms may always be adjusted wisely to increase prediction accuracy. Finally, some of these techniques are often used for modifying the initial structural equations in optimum predictions.

Nonetheless, it is helpful to quickly explain some of the ideas behind the combination of forecasting methodologies before moving on to these particular situations. To begin, think about two predictions and a regression equation where the actual values are estimated as a function of the predicted values produced by models (a) and (b) (b). If the first prediction's estimated regression coefficient is 1 and the second forecast's estimate is 0, then the first forecast has all the necessary information and combining forecasts adds no value. In this scenario, model (a) is said to forecast-encompass model (b), and merging predictions will not increase forecast accuracy. This test can only be used to evaluate a new model if it has a history of accurate predictions, which is obviously a need. Otherwise, it is a straightforward test to do [9].

The variance-covariance forecast combining approach was invented in this field and may be traced back to the foundational publication by Bates and Granger4. Nevertheless, Diebold5 subsequently demonstrated that this was a unique instance of the regression-based prediction combining technique, so our remarks are limited to that approach. Where ya and yb are predictions from models (a) and 6 yt+k = b0 + b1 yat+k + b2 ybt+k + et (b). As previously said, provided sufficient historical prediction data is available and the prerequisites outlined at the beginning of this section - impartial, constant variables, and an unchanging data production function - are met, that computation is fairly straightforward. They are not, for the most part. Hence Diebold suggests four different forms of changes to accommodate for changing circumstances. They fall into the following categories: I time-varying combination weights; (ii) serial correlation corrections; (iii) shrinkage of combining weights towards equality; and (iv) nonlinear regressions.

In situations when one forecasting technique gains prominence over time, such as when monetary policy changes become more relevant as a result of increased transparency, the first scenario may be applicable. In order to account for the most current residuals from the structural equation, this strategy fundamentally consists of combining structural or judgmental predictions. This chapter will demonstrate how often this strategy minimises forecast inaccuracy. The residuals are estimated statistically using the same equation, but they do so using an AR (1) technique. This approach may also be expanded to include an ARMA (p,q) process, although this is less likely to increase prediction accuracy.

The b coefficients change with time in the third instance, which has certain characteristics with the first scenario. Yet in this scenario, the user believes that the likelihood of weights approaching 1 /2 increases with the length of the time horizon. Model (a) may thus be given a significantly higher weight in the short term, but as the prediction horizon lengthens, model (aaccuracy)'s declines in comparison (b). This approach isn't often used in real life. The fourth scenario is when one approach seems to provide the most accurate projections during times of average growth but another way becomes more valuable at times of abnormally substantial shifts. Consequently, this alternate approach would provide a stronger warning signal without always forecasting "doom and gloom," for instance, when the chance of business cycle recessions, energy shocks, or stock market collapses is greater [10].

In general, the forecaster's estimation of the relative accuracy of these two methodologies determines a substantial portion of whether or not to combine predictions. The logical conclusion could be to choose the model that has historically produced the lesser error, but the explanation above shows why it might not be the best course of action. First, over time, the weights could change. Second, considering serial correlation of the residuals may often minimise prediction error. Finally, the relative benefit of one strategy may wane as the time horizon becomes longer. Fourth, one approach may be more effective at predicting trends than another is at predicting turning moments. Nevertheless, before moving on to some real-world instances of combining predictions, it is important to highlight the main causes of mistake in order to underline that when a previously unanticipated circumstance occurs, combining forecasting techniques may not even enhance accuracy. Sometimes the residuals throughout the sample period aren't even regularly distributed. Nevertheless, it is possible to lessen that issue by adjusting structural equations for autocorrelation and heteroscedasticity and using ARIMA techniques for time-series analysis. Yet, prediction mistakes are almost always greater than sample period data would suggest. Together with the random variance brought on by the sampling process, the primary causes of inaccuracy may be summed up as follows.

1. 1 The equation is incorrectly defined since some of the variables are either missing or entered incorrectly (e.g., the underlying relationship is really nonlinear).
2. 2 Inaccurate values for external variables are assumed over the forecast period, or endogenous variable forecasts are made.
3. 3 Data series are updated, resulting in estimations of equations based on data that subsequently prove to be erroneous. However, based on updated data, what first seemed to be a good projection now has a considerably higher inaccuracy.
4. 4 As is usually the case with AR processes, error accumulation occurs in multi-period predictions when the lagged dependent variable is present on the right side of the equation. This finding suggests that multi-period forecasting should not be conducted using equations having AR terms. In many instances, it is preferable to account for autocorrelation by changing the constant term.
5. The equation's structure has changed, either during the sample period or after it ended. Hence, none of the conventional statistical tests are relevant, and the prognosis is always enhanced by subjective considerations.

Forecast error will typically be greatly reduced by a mix of structural equations, time-series methods, and wise judgement. Combining predictions will be investigated with the potential that reason, in particular, often explains the majority of forecasting mistake. We stress in advance that depending on the properties of the underlying data series, various combinations will result in the best projections. When examining the topic of optimal forecasts, it is customary to use the relative mean square error (RMSE) as a metric of forecast accuracy. Although the absolute average forecast error (AAFE) is occasionally used, RMSE is typically preferred because, if the residuals are normally distributed, RMSE is the best indicator of forecast accuracy. Even if the assumption of normality is incorrect, it is very rare, if ever, that a different distribution is known with sufficient precision to replace a different measure. Data revisions that take place decades after the fact can sometimes obscure the question of how accurate a forecast was at the time it was made in the case of macroeconomic forecasts, which have been the most thoroughly documented and are significant because they serve as the starting point for many industry and company forecasts. There are several examples that may be used to demonstrate this idea.

Robert J. Eggert found that the consensus error for forecasting the real growth rate for the preceding three years had only been 0.2%, 0.1%, and 0.1% in the January 10, 1980, edition of the Blue Chip Economic Indicators. The statistics were later corrected many times, and the National Income and Product Accounts (NIPA) adjustments released in late 1999 revealed that these mistakes were far larger: 0.6%, 1.4%, and 1.3%. As a naïve model mistake for real GDP creates an error of around 1%, there is a significant difference between an average error of 0.1% and 1.1% when calculating predicting accuracy. So, despite no fault of the individual forecasters, data updates made a considerable time after the predictions were published may often diminish predicted accuracy. Which analyst was correct: the one who anticipated a significant increase in profits, which is what the business originally disclosed for the next quarter, or the one who anticipated a significant decrease in earnings, which really occurs when the SEC discovers accounting irregularities? Now that we've established this, we'll admit that data revisions don't always result in incorrect macroeconomic model forecasts; in most circumstances, changes only contribute a minor amount to forecast error. But, one must decide whether data is more crucial in the situations when it does make a difference: the data that will shortly be provided or the final, amended data?

Since it involves a decision, this question cannot be answered using conventional statistical methods; however, the majority of realistic business forecasters would prefer to make accurate predictions for the following year based on data that will be released during that year rather than data that will be revised five, ten, or even twenty years later. In general, US macroeconomic forecasters were unable to foresee the changing patterns of the 1970s and 1980s, which included four recessions in 12 years after almost a decade without one, two episodes of double-digit inflation, and historically high increases in interest rates. Several explanations were put out for this failure later on. The two energy shocks could not have been expected, according to those who had a more favourable opinion of econometric models. The previous models were criticised by several members of the "Chicago School" because they were Keynesian rather than monetary, and Bob Lucas noted that the older models did not take into account reasonable expectations. Some academics, like Christopher Sims, felt that structural models were subpar and recommended switching to vector autoregressive models, or VARs. The 1980s saw much discussion of all these different macroeconomic forecasting techniques with the hope that combining them would provide macroeconomic projections that would be more precise. Unfortunately, none of these techniques were successful in foreseeing the recession of 1990–1991; a thorough analysis of these shortcomings may be found in a 1993 NBER report titled Business Cycles, Indicators, and Forecasting. Keynesian econometric model creators made inaccurate predictions. Both monetarist econometric model builders' and rational-expectations theorists' predictions were inaccurate.

CONCLUSION

Long-term forecasting is an important aspect of many fields, and there are various methods available to make predictions about future trends and patterns. These methods include time series analysis, scenario planning, and simulation modeling. While each method has its strengths and weaknesses, the most effective long-term forecasting often involves using a combination of methods and a range of data sources to generate accurate and reliable predictions. It is important to note that no forecasting method is perfect, and there is always some degree of uncertainty involved in long-term predictions. Therefore, it is important for decision-makers to consider the potential risks and limitations associated with long-term forecasts and to use them as a tool to inform decision-making rather than as a definitive prediction.

REFERENCES

[1] Z. Z. Wu *et al.*, "Research Note: Information Technology, Contract Completeness, and Buyer-Supplier Relationships," *J. Oper. Manag.*, 2012.

[2] W. Nugroho, *Komunikasi dalam keperawatan gerontik*. 2006.

[3] S. Wongsawat, "Predicting factors for quality of life of elderly in the rural area," *Int. J. Arts Sci.*, 2017.

[4] S. Nugraha and Y. Ohara-Hirano, "Mental Health Predictor of the Sixth Batch Indonesian Nurse and Certified Care Worker Candidates Migrate to Japan under the Japan–Indonesia Economic Partnership Agreement in Pre-migration Stage," *J. Heal. Sci.*, 2014.

[5] 2018 Kesehatan, "Provinsi Jawa Barat Tahun 2012," *Dinas Kesehat. Jawa Barat*, 2012.

[6] Profil Kesehatan Provinsi Jawa Barat, "Resume Profil Kesehatan Provinsi Jawa Barat Tahun 2012," *Dinas Kesehat. Jawa Barat*, 2012.

[7] M. Caballero-Anthony, A. D. B. Cook, G. G. H. Amul, and A. Sharma, "Health Governance and Dengue in Malaysia," *Heal. Gov. Dengue Southeast Asia*, 2015.

[8] S. Idaiani and S. Suhardi, "Validity and Reliability of the General Health Questionnaire for Psychological Distress and Social Dysfunction Screening in Community," *Bull. Heal. Res.*, 2006.

[9] J. B. DinKes, "Profil Kesehatan Provinsi Jawa Barat Tahun 2012," 2012.

[10] M. Caballero-Anthony, A. D. B. Cook, G. G. H. Amul, and A. Sharma, "Health Governance and Dengue in Indonesia," *Heal. Gov. Dengue Southeast Asia*, 2015.

CHAPTER 21

UNRAVELING THE COMPLEXITIES OF SIMULTANEOUS-EQUATION MODELS: AN EMPIRICAL ANALYSIS OF ENDOGENEITY AND IDENTIFICATION ISSUES

Pankhuri Agarwal, Associate Professor
Teerthanker Mahaveer Institute of Management and Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India
Email Id- dreams.pankhuri@gmail.com

ABSTRACT:

Simultaneous-equation models are a type of statistical model used to analyze interdependent relationships among multiple variables. These models are commonly used in economics, social sciences, and other fields where causality is a key concern. Simultaneous-equation models involve a system of equations that describe the relationships among multiple variables, where each equation represents one variable as a function of the others. The equations are solved simultaneously to estimate the values of the variables, taking into account the interdependence among them.

KEYWORDS:

Interdependent Relationships, Multiple Variables, Simultaneous-Equation Models, Statistical Model.

INTRODUCTION

The index of leading indicators and numerous gauges of consumer and corporate mood did not accurately predict the economy. No one forecast a recession out of the 52 participants in the Blue Chip Economic Indicators study, according to the July 1990 report. Yet, the National Bureau claims that was the month the real recession started. So, despite improvements in forecasting methods during the preceding ten years, no one was able to foresee the slump. We may make use of trying to ascertain, with the benefit of hindsight, what really caused the slump before offering any speculative solutions. Undoubtedly, Saddam Hussein's decision to invade Kuwait and the subsequent increase in oil prices were contributing factors. Oil prices did, however, recover to their prior levels when Hussein was overthrown by UN troops, but the American economy remained exceptionally sluggish for the next two and a half years. There were undoubtedly more forces at play [1].

In hindsight, the lending restrictions that followed the collapse of the savings and loans sector turned out to be the fundamental cause of the recession. With the relaxation of banking regulations as part of deregulation in 1982, a large number of lending institutions issued loans with, at least in hindsight, a very low likelihood of repayment. Whether this was an error in judgement or blatant fraud is immaterial; when the regulations were altered in 1989, many bank lending officers tightened credit requirements significantly out of fear of going to prison for making risky loans.

Therefore, constant-dollar consumer credit outstanding decreased by an average of 3.9% per year in 1990, 1991, and 1992, while constant-dollar business loans decreased by an even greater 6.2% per year. Over the previous three decades, both of these measures of credit availability had increased by an average of about 4% per year in real terms, roughly in line with overall real growth. Despite the attention paid to the change in law and the extraordinary reduction in credit availability, analysts were unable to precisely predict how it would affect housing, capital investment, and consumer and consumer spending. Hence, using several forecasting techniques in this instance would not have decreased prediction error.

This example effectively illustrates our main argument that a change in the fundamental data production function's structure is the primary cause of prediction mistake. Every forecasting technique that is dependent on past events, which includes almost all techniques, may provide errors that are far bigger than would be predicted from prior experience when the underlying structural connections change. It's logical to believe that no one anticipated the full effects of the credit crunch, the Iraqi-Kuwaiti conflict, or the momentary doubling of oil prices. Nevertheless, imagine if in early 1991, when all of these things were known, you were asked to forecast the state of the economy. Which course of action would be the best?

Real growth always returned to above average rates during the first full year of recovery, according to historical business cycle data. But, this time, actual growth stayed below average for the next two years. How was it possible to predict such development? The increases in consumption and investment were exaggerated by historically calculated functions, as this author can confirm. Because of this, it would have been appropriate to make continuous modifications based on current experience [2]. It would have also been reasonable to adjust the slope terms in the formulae for consumer credit and commercial loans. Forecast accuracy would have also been enhanced by using discretion over how long the credit limits would last. These three techniques are among the most important ones for lowering prediction error when utilising structural models. The rest of this chapter concentrates on these components in relation to structural equations after a quick examination of the comparison of different non-structural approaches.

Many research have merged various non-structural forecasting techniques and shown that the resultant forecasting errors were decreased. The study by Makradakis and eight other authors is one of the most well-known ones. They looked at 1,001 series projections with an 18-period time horizon. They compared using naïve models, several exponential smoothing techniques, including Holt-Winters, and regression models. The average predicting error by combining several of these approaches is around 8% less than the single best method, which is Holt-Winters, according to this research, without going into all the specifics. For a comparable test, Newbold and Granger claim an improvement of 6%.

DISCUSSION

These statistics are likely close to the utmost that can be anticipated from combining techniques of non-structural estimate, according to the author's experience, some of which is detailed below. Granger and Newbold note that "the instances of combination offered so far are not actually tailored to display the process in its best possible light" after detailing their findings. Following all, it would be natural to anticipate that a combination would be most successful when the nature of the individual projections is quite different. 10 It is the same point that was made previously. [3] It will be shown that using several non-structural approaches in conjunction seldom increases prediction accuracy, especially for multi-period forecasts, when dealing with real economic time-series data.

We look at six economic time series that were previously utilised in this article to demonstrate this idea. These are all monthly statistics. Apparel sales, single-family home starts, and the PPI for industrial commodities are the three that are not seasonally adjusted. The real Federal funds rate, the S&P 500 index of stock prices, and the monthly rate of inflation as determined by the CPI are the other three variables, which are either seasonally adjusted or lack seasonal influences. The best ARIMA model, Holt-Winters (HW) exponential smoothing, and a naïve model that states the percentage change this period is the same as the previous period are all compared in every scenario. For seasonally unadjusted data, the percentage change over the preceding 12 months is utilised. While it may not seem to be a formidable rival, the naïve model is often taken into account in combinatorial predictions. There is nothing to be gained by integrating alternative smoothing techniques in our research since, as Makradakis et al. have already shown, the HW approach is often better than other exponential smoothing methods.

The equations were calculated using data from 1947.01, or the earliest month for which reliable data were available, through 1997.12, with the months of 1998.01 through 1999.12 set aside for the prediction period. Even when the ADF and PP tests did not reveal a unit root, trends were always eliminated using first-difference or percentage changes when levels were employed, the prediction errors were always greater [4]. The findings at first glance seem encouraging. The RMS error of the average of these three approaches for seasonally unadjusted clothing sales over the forecast period is much lower than any of the individual methods. Since the ARIMA model continuously overestimates real sales while the naïve model continually underestimates them, this is the cause. On the other hand, the combination forecast averages these mistakes, which results in a substantially reduced average prediction error.

The HW approach produces the least prediction errors for housing starts; for the industrial PPI, it is the ARIMA model. The naïve (no change) model produces the best estimate for the real Federal funds rate for the series devoid of seasonal influences. HW provides the greatest predictions for the stock market, while ARIMA with a full split provides the best predictions for the CPI. Given that the US economy was at full employment when the calculations were done, it was unclear which technique would be based on prior experience when these six data were picked. The actual Federal funds rate, which had previously varied dramatically, remained essentially stable. Housing starts and the stock market both increased more than would have been predicted based on previous data.

Evidently, the basic framework was not altered throughout the construction of another research. Realistic business forecasters are aware that this is seldom the case in the actual world. Nevertheless, the mistakes are highly connected, with the exception of clothing sales. The high and low numbers are often missed by all approaches. The benefit of merging projections decreases as the forecast series' correlation coefficient gets closer to unity [5].

At this point, we draw the conclusion that combining forecasting techniques will often only considerably decrease forecasting error if I the correlation between the mistakes is minor and (ii) the underlying structure, the data producing function, stays the same. In actuality, these circumstances are seldom present. As a result, combining naïve, exponential smoothing, and ARIMA models in practise is not likely to significantly increase forecasting accuracy. This does not necessarily imply that predicting accuracy is not increased by merging predictions. It does imply that combining different nonstructural, mechanical approaches does not provide the best outcomes. Instead, integrating structural, non-structural, and judgemental approaches that use various processes, have uncorrelated mistakes, and take structural alterations into consideration is more likely to provide superior results. The remainder of this chapter investigates a few of these pairings.

Indeed, it is feasible to integrate structural and non-structural forecasting techniques. Even though the major series exhibit substantial temporal patterns, many structural equations are approximated in levels rather than actual or percentage changes. This is true for many reasons than merely the fact that levels equations always provide superior goodness-of-fit statistics. The model creator can want to anticipate the trend itself or get long-term projections under equilibrium circumstances [6]. Long-term elasticities may be significant as well. A steel company would wish to calculate the short- and long-term price elasticity of replacement for aluminium, or an auto manufacturer might want to calculate the short- and long-term effect of a change in energy costs on sales of different kinds of new automobiles. Calculating how much alternative price reduction percentages might enhance their long-term sales growth rate can be of interest to computer makers.

Many structural equations are approximated in levels form for these and other comparable reasons. Yet, although ARMA processes are developed under the explicit premise that the variables do not have a significant trend, the processes are stationary since most economic time series contain substantial temporal trends. This poses a problem, one that is not limited to ARMA models. Consider a mixed set of independent variables, some of which show significant temporal patterns, like sales and income, and others which do not, such interest rates, relative prices, or percentage changes. Although interest rates and comparable prices would remain relatively stable over a long length of time, the value of the sales or income variables would increase 10-fold. Figure 1 illustrate the recursive simultaneous equation model.

Have a look at the common equation $Y = c + 1.5 X - 50r$, where Y and X are trend variables and r is a trendless variable (such as the real rate of interest). Let's say that at the start of the sample period, Y = 200, X = 200, and r = 5. Y would vary by 1.5% for every 1% change in X. Y would vary by 1.25% for every 1% change in r, or from 5.00 to 5.05. Assume that Y and X are both 2,000 at the conclusion of the sample period, and that r is still 5. A 1% change in X will still cause a 1.5% change in Y, whereas a 1% change in r will only cause a 0.125% change in Y. The importance of r decreases as X and Y grow in size. By using the percentage change of Y and X and leaving r in levels, this issue might be solved, but the parameter estimations would then become skewed as a result of the fluctuations outweighing the trend.

Using four-quarter % moving averages is a middle option that has often been employed in predicting macro models. Even so, such wouldn't be long-term projections; for them, equations in levels form could be chosen. Nevertheless, the equation will include both stationary and non-stationary variables if levels and ARMA processes are mixed, making the standard statistical tests incorrect. The projections might also include some significant mistakes [7]. There are several approaches that may be used to reduce this issue, some of which combine projections. On a structural level, the equation may be converted to ratio form, allowing one to estimate the relationship between Y/X as a function of r. Both variables are likely to be trendless in the scenario. Another option is to produce two sets of predictions: one that anticipates trends in both X and Y, and the other that links changes in Y to changes in r.

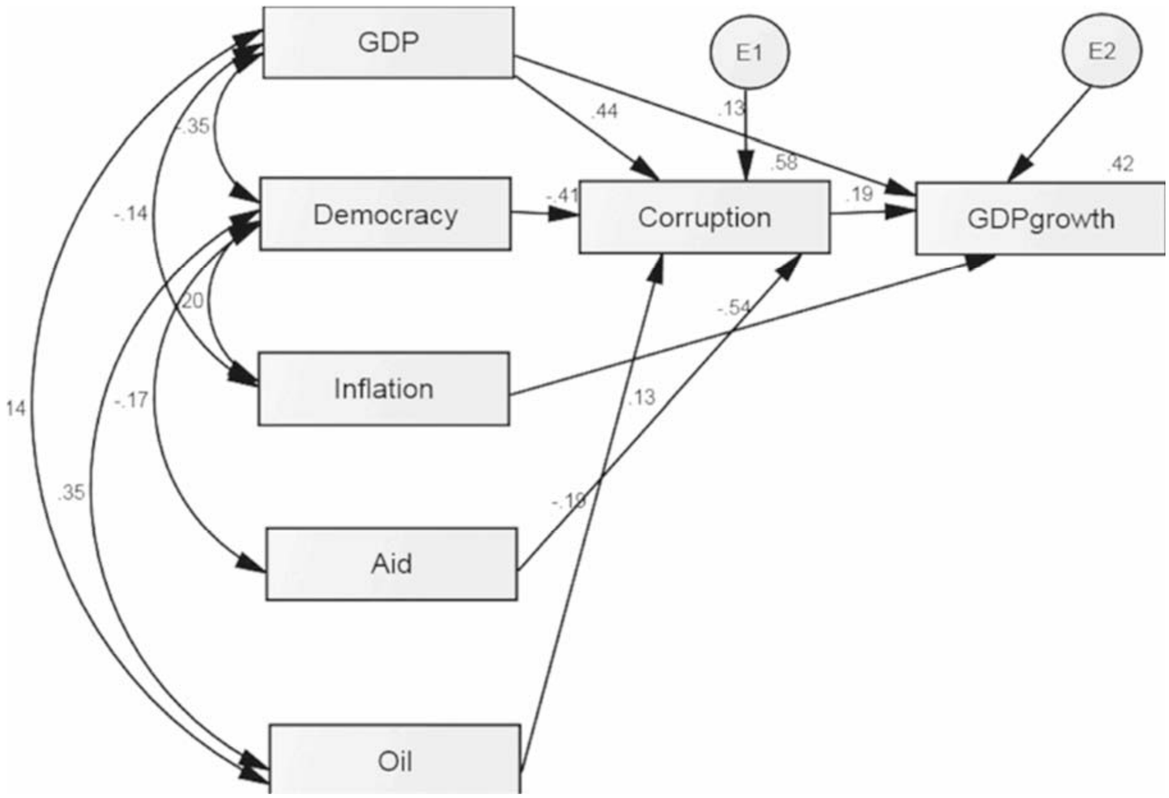


Figure 1: Illustrate the recursive simultaneous equation model.

The weights obtained by the techniques discussed in section 8.1 may then be used to combine these projections. When the Fed tightens, consumer and business confidence may deteriorate. As a result, a third alternative is to link cyclical fluctuations in the real rate of interest to some form of sentiment measure. Examples of this include the stock market's strong reactions to sudden changes in the federal funds rate, which may also have an impact on consumer and capital goods purchases. So, the major option is either a structural equation that weakens or removes the common trend in Y and X, or alternatively, a combined prediction that gives cyclical factors more weight.

Assume, for instance, that purchases of constant-dollar consumer durables (hereinafter, durables) depend on several attitude variables, such as stock prices, unemployment, inflation, relative oil prices, and the consumer sentiment index, as well as current and lagged disposable income, lagged housing starts, the yield spread, and these variables. None of the attitudinal factors, including the yield spread, exhibit any discernible patterns. On the other hand, there are notable trends in the prices of durables, income, and stocks. The challenge of how to this case study shows an equation that is used to anticipate durables on a quarterly basis. It places particular attention on the problems of detrending and simultaneous causality, as well as on how combining methods in this instance, with an indicator of consumer sentiment can improve forecasting accuracy [8].

The main factors that go into an equation for durables are disposable income, the price and accessibility of consumer credit, housing starts (since homeowners buy furniture and appliances when they move into a new home), and a number of psychological factors like stock prices, the unemployment rate, inflation rates, relative oil prices, and a consumer sentiment index. In an equation that is estimated in both levels and four-quarter percentage changes, all these factors were tested. The sample period for this equation is shortened in 1995 to evaluate forecast accuracy. 4. Also, the sample period does not begin until 1967.1 since we wish to examine the significance of the consumer sentiment index and before that time, reliable quarterly data were not available.

This chapter's main goal is to demonstrate how combining forecasting techniques may increase predicting accuracy. It is important to remember that if the underlying equation is not fundamentally robust, forecasting accuracy will suffer. This includes avoiding the use of independent variables that can be subject to simultaneity bias and keeping trends and trendless series separate unless the relevance of the latter will be effectively reflected in the forecast period. When such factors are taken into account, the prediction errors are often substantially bigger than what the sample period data suggest.

Insignificant coefficients are not given despite the fact that all of the aforementioned factors were tested. Equations were calculated both with and without the consumer confidence index, and they performed better on a quarterly basis with no lag. It should be noted that depending on whatever option is selected, the variables contained in the equations change significantly. The amount of variation that is explained by the consumer sentiment index across the sample period increases, and the serial correlation of the residuals is decreased. Given the strong correlation between the two, it is not unexpected that the unemployment rate disappears from the equation as consumer mood rises.

There are several peculiarities, however. It is unclear why relative oil prices matter so much in the levels equation but not in the percentage change equation; one can only assume that there is some erroneous association. It is also unexpected that the yield spread term vanishes when consumer sentiment is included. The lack of the stock market factor in the equation is also noticeable; this topic will be covered momentarily. The projections produced by these four equations for the years 1996.1 through 2000.4 are then compared; the results are shown in figure 8.3. Although if the actual values of all the independent variables are utilised, the errors are substantially bigger than would be anticipated from the sample period statistics, which implies the findings are better than would have been achieved on a true ex ante basis [9].

Although the levels equations likely suffer from erroneous trend correlation, the mistakes are a little bit reduced for the % change equations. The emotion variable, however, makes little to no contribution to either the levels equation or the % change equation. The optimal equation has an RMSE of 68.3 and a SEE of 13.4. After 1997, it would seem that the fundamental structural connection changed. In this aspect, the consensus projections are of little use. Chip Blue While there are no consensus predictions for consumer durables per se, data on total consumption and light vehicle sales are available new cars and light trucks. The fact that consumer mood hardly increased over the course of this five-year period, despite the rise in durable goods, did not help to enhance these estimates.

The index of consumer sentiment did not increase prediction accuracy throughout the 1990s; the very significant terms in the aforementioned regressions indicate its larger significance in earlier decades. As the residuals for 1996 and 1997 were almost negative before the actual gains far outperformed those projected by the equation or consensus projections, even making constant adjustments would not improve the forecasting performance. If the regressions are performed up to 2000.4, the stock market term appears as one of the most significant factors, and the predicting mistakes are randomly distributed, there is nearly always a solution. Nevertheless, the reader may confirm that the stock market factor is negligible and often has the erroneous sign when added to the regressions provided above or other equations of a similar kind for the sample period ending in 1995. In fact, running the same equation again for an additional five years without the consumer sentiment index results in a change in the t-ratio on the stock price term from -1.1 to +22.0, which is unmistakable proof of a structural shift.

Yet, merging projections and giving a stock market more weight does not completely resolve the problem. Notwithstanding the severe decline in stock prices at the beginning of 2001, the number of light motor vehicle sales increased significantly from a seasonally adjusted annual pace of 15.8 million in 2000 to 17.2 million in 2001. Also, it is not necessary to rely on lagging stock market values since motor vehicle sales improved later in the year and, when they declined after the terrorist events of September 11, 2001, they quickly recovered in October thanks to zero interest rate financing.

We shall soon provide an illustration of the crucial lesson that can be drawn from this situation. Let's briefly go through the available data, however. An approximated structural equation also produced good predictions for 1996 and 1997, however they were also somewhat below actual values. This equation seemed to closely match consumer durable sales until the end of 1995. Even with the inclusion of an emotion term, continuous modifications to reflect serial correlation of the residuals, and assigning more weight to trendless cyclical factors like the unemployment rate, inflation rate, and yield spread, the estimates for the years 1998 to 2000 were still significantly too low.

The ratios of spending on motor vehicle and component purchases to disposable income and spending on other consumer durables to disposable income. The numbers are in billions of dollars (chained) from 1996. The motor vehicle ratio shows almost no trend, and in fact, levels in the late 1990s are much lower than peak levels recorded in the 1970s and 1980s. The fraction of other durables, which is dominated by personal computers, has increased significantly since 1990, particularly in the second half of the decade. Personal computer purchases show relatively little link with housing starts, inflation, relative oil prices, or yield spread; yet, sales of this durable good category dominated in the second half of the 1990s [10].

We discover that the structural equation was constructed on a shifting basis even though it seemed to be sound based on traditional approaches. If motor cars, personal computers, and home durables had been calculated independently, the estimates would have been substantially more accurate. In that situation, we would have discovered that motor vehicle sales continued to be influenced by conventional cyclical factors and increased in early 2001 as a result of falling loan rates in prior quarters. On the other side, the booming stock market caused a surge in personal computer sales, which dropped 3.5% that quarter when the market crashed in 2001. It turns out that structural instability is the problem once again.

No matter how many additional projections are added to a bad equation, it will still provide poor predictions. The primary issue with the equation in Case Study 16 is that it essentially combines bicycles and apples. Durable consumer goods are not a uniform commodity. In particular, the group consists of two key elements with contrasting trends: "high tech" durables, mostly personal computers, which increase far faster than income, and "traditional" durables, such as cars and home appliances, which rise at about the same pace as income. These two groups should be handled differently. When a forecaster is asked to forecast "sales" for a certain business or sector, the same conundrum often arises. Even when used in conjunction with other techniques, projections are likely to be inadequate if the product is not homogenous, particularly if it includes certain components with quickly increasing sales and others with dropping ones. For reliable projections, a solid structural approach is always the best place to start [11].

CONCLUSION

Simultaneous-equation models provide a powerful framework for analyzing interdependent relationships among multiple variables. These models allow for the estimation of both direct and indirect effects, which is especially useful when analyzing complex causal relationships or predicting the effects of policy interventions. Simultaneous-equation models have been widely used in economics, social sciences, and other fields where causality is a key concern. However, they require careful specification and estimation, as well as careful consideration of the assumptions underlying the model.

REFERENCES

[1] J. Yan, X. Fu, T. H. Oum, and K. Wang, “Airline horizontal mergers and productivity: Empirical evidence from a quasi-natural experiment in China,” *Int. J. Ind. Organ.*, 2019, doi: 10.1016/j.ijindorg.2018.01.004.

[2] J. Peng, “Identification of Causal Mechanisms from Randomized Experiments: A Framework for Endogenous Mediation Analysis,” *Inf. Syst. Res.*, 2022, doi: 10.1287/isre.2022.1113.

[3] S. E. Wolfolds and J. Siegel, “Misaccounting for endogeneity: The peril of relying on the Heckman two-step method without a valid instrument,” *Strateg. Manag. J.*, 2019, doi: 10.1002/smj.2995.

[4] M. I. U. Husnain, A. Subramanian, and A. Haider, “Robustness of geography as an instrument to assess impact of climate change on agriculture,” *Int. J. Clim. Chang. Strateg. Manag.*, 2018, doi: 10.1108/IJCCSM-03-2017-0049.

[5] A. A. Bostian, M. B. Bostian, M. Laukkanen, and A. Simola, "Assessing the productivity consequences of agri-environmental practices when adoption is endogenous," *J. Product. Anal.*, 2020, doi: 10.1007/s11123-019-00564-7.

[6] Y. Zhou, "Have Free Trade Agreements Created Trade? Evidence from CEPA," *Pacific Econ. Rev.*, 2019, doi: 10.1111/1468-0106.12198.

[7] F. Ferreira, "You can take it with you: Proposition 13 tax benefits, residential mobility, and willingness to pay for housing amenities," *J. Public Econ.*, 2010, doi: 10.1016/j.jpubeco.2010.04.003.

[8] M. G. Tsionas and M. L. Polemis, "On the estimation of total factor productivity: A novel Bayesian non-parametric approach," *Eur. J. Oper. Res.*, 2019, doi: 10.1016/j.ejor.2019.03.035.

[9] A. S. Kaminaga, "An empirical analysis of social interactions: Implications for HIV and domestic violence prevention programs in rural Malawi.," *Diss. Abstr. Int. Sect. A Humanit. Soc. Sci.*, 2012.

[10] A. Advani and B. Malde, "Empirical Methods for Networks Data : Social Effects, Network Formation and Measurement Error," *IFS Work. Pap.*, 2014.

[11] F. O. AMADU, "PEER EFFECTS IN AGRICULTURAL EXTENSION: EVIDENCE OF ENDOGENOUS SOCIAL INTERACTION IN THE PERFORMANCE OF COMMUNITY KNOWLEDGE (EXTENSION) WORKERS IN UGANDA," *Appl. Microbiol. Biotechnol.*, 2014.

CHAPTER 22

COMPARING THE ACCURACY AND ROBUSTNESS OF ALTERNATIVE METHODS FOR MACROECONOMIC FORECASTING: A COMPREHENSIVE EVALUATION

Roma Khanna, Assistant Professor
Teerthanker Mahaveer Institute of Management and Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India
Email Id- romakhanna11@gmail.com

ABSTRACT:

Alternative methods of macroeconomic forecasting refer to approaches other than traditional econometric models that rely on statistical techniques and historical data to make predictions about future economic conditions. These methods are often used in situations where traditional models have limited predictive power, such as during periods of structural change, high uncertainty, or when the available data is limited or unreliable. One alternative approach to macroeconomic forecasting is based on the use of machine learning techniques, such as neural networks, decision trees, and support vector machines. These models are designed to learn patterns in data and identify relationships between variables without making explicit assumptions about the underlying economic structure.

KEYWORDS:

Alternative Methods, Decision Trees, Data Mining, Machine Learning, Neural Networks, Support Vector Machines.

INTRODUCTION

Economists have long noted that consumer spending patterns are somewhat influenced by past behaviors; hence, purchases made in the present are influenced by both past behavior and the current levels of income and financial resources. Many believe that the consumer mood index is a key influencer on purchases to the degree that it reflects these elements. Also, it's critical to consider anticipated future changes in economic and financial well-being. In fact, the consumer mood index is still one of the important factors that economic forecasters and financial experts regularly monitor. In that sense, it has successfully undergone a "market test," and one might therefore argue that it must provide meaningful information in order to maintain its popularity.

The first surveys created for macroeconomic forecasting were not mood indexes; rather, they were projections of consumer spending plans, corporate sales, and capital expenditures. Initially, it was believed that "hard" indicators of actual purchase intentions would be more important than "soft" indicators of consumer or corporate mood. Yet, it turned out that the opposite was true: while indices of mood continue to thrive and are highly anticipated by forecasters and market experts, indexes of consumer and company purchase intentions have been thrown in the trash [1].

Also, it was first believed that capital investment would be the area of the economy most conducive to the application of business indices of planning or emotion. Contrary to consumer spending, which often involves "impulse purchasing," the majority of capital goods purchases are undertaken by big corporations, who meticulously plan their capital expenditures well before they actually decide to make a purchase. Additionally, since most commodities and construction projects take a long time to complete, statistics on orders and appropriations should provide further, useful insight into the direction that capital investment in particular and the economy as a whole were taking.

According to this reasoning, immediately after the conclusion of World War II, both public and private organizations began to publish yearly reports on investment expectations; the Commerce-SEC statistics were also accessible on a quarterly basis. For the simple reason that they were not very helpful, these reports are no longer published. The "market test" problem rears its head here as well; the anticipations data were finally phased out since they did not increase prediction accuracy. Similar to this, the Conference Board no longer publishes information on capital appropriations by significant firms. Every month, information on new orders for capital goods is released, but only the correlation for the current quarter shows any real relationship with future capital expenditure. This news therefore seldom affects the financial markets.

Investment expectations, appropriations, and orders are not reliable indicators of actual capital expenditure for a number of reasons. Some have to do with the erratic delivery and building time frames. But more importantly, this lack of connection illustrates how corporations adjust their strategies in response to changing economic conditions. So, rather than representing elements that were anticipated to occur in the future, the anticipations indexes indicated factors that had already occurred. They don't really add anything new in that regard [2].

We now take a look at a potential house buyer sentiment index that the National Association of Home Builders has compiled and made available (NAHB). The sample period is less for this index since it only begins in 1985, yet the findings are very obvious estimation of the housing starts equation revealed that the demographic parameters, the percentage change in income, the vacancy rate, the unemployment rate, the percentage change in the real money supply, and the yield spread were the main independent variables. I warned that this problem will be brought up again later since it turned out that equation produced highly erroneous predictions unless they were modified by the constant factor.

It turns out that employing the NAHB index of homebuyer mood significantly lowers the prediction inaccuracy. The consumer confidence index does not increase the projection of single-family home starts, as the reader may confirm. Due to the shorter sample period and the introduction of the homebuyer attitude variable, the demographic and vacancy rate variables were removed from the housing start equation previously given. Financial journalist Joe Livingston has produced consensus predictions for many years, some of the regional Federal Reserve Banks also issue consensus forecasts, and able have been gathered and tallied jointly by NBER and ASA. 12

The projections created by Blue Chip Economic Indicators, in the author's view, are the most comprehensive and well-documented ones. These forecasts, which come from 50 or more contributors, have been available every month since January 1977. Many of the same panelists have taken part in the survey since its inception, and the series are internally consistent. The panelists are listed by name and organization as well, making it possible to compare the performance of consulting economists versus those employed by major corporations to see, for example, whether large New York City financial institutions are better at forecasting interest rates than economists in other parts of the nation.

DISCUSSION

Since the record of economists in predicting interest rates has been unusually poor and since interest rates are of great importance in determining both where the economy is heading and how financial markets will perform we next examine whether the consensus forecasts for interest rates can improve forecasting accuracy. This case study uses the Blue Chip Consensus forecasts for short-term interest rates, measured by the three-month Treasury bill rate on a discount basis, and long-term rates, measured by the ten-year Treasury note yield. For some years in the past, the Blue Chip consensus used the 4–6 month commercial paper rate to measure short-term rates, and the corporate bond yield and utility bond yield to measure long-term rates. All of the calculations here have been adjusted to reflect these shifts [3].

Since the Blue-Chip forecasts are issued monthly, it would be possible to calculate the impact of the consensus forecasts of a monthly basis. However, there usually very little change in the consensus outlook from one month to the next; fluctuations that do occur in that narrow a time frame are largely due to speculative trading patterns and short-term disturbances. Hence for the purposes of identifying the possible contribution of the consensus predictions to forecasting accuracy, we examine the forecasts on an annual basis. The consensus forecasts used are those provided to Blue Chip on the first working day of each January. These forecasts can be compared to I a naive model that says the level of interest rates this year will be the same as last year, and (ii) a structural model that relates interest rates to the inflation rate, the Federal budget surplus or deficit ratio, loan demand, and the unemployment rate. In addition, the bond rate forecast is also correlated with the Treasury bill rate lagged one year as well as these other variables.

The only unlogged variable that is significant is the change in the unemployment rate. However, using an unlogged variable with an assumed error of aero could bias the results. Thus instead we have used the difference between the consensus forecast of unemployment this year and the actual unemployment rate last year. This term is almost as significant as the actual change in unemployment, and removes any simultaneous-equation bias. The equations for the three-month Treasury bill rate and the corporate bond rate are estimated for the period from 1977 through 1999, using annual data. The independent variables are the difference between the consensus estimate of the unemployment rate this year and its actual value last year, and lagged values of the inflation rate, the percentage change in business loans, and the ratio of the Federal government surplus or deficit to GDP. Also, the lagged Treasury bill rate is used in the corporate bond equation [4].

The quarterly and monthly functions for interest rates discussed elsewhere in this text are more complex, but these annual functions capture the essence of those factors that have determined interest rates in the past. Also, as mentioned above, note that only lagged or consensus variables are used. The RMS So far we have seen how judgment can improve forecast accuracy by using survey results and consensus forecasts. The third major method of incorporating judgment on a statistically consistent basis as opposed to using qualitative expert opinion from the field is to adjust the constant term and slopes that have been estimated by least squares. Figure 1 illustrate the Statistical and Machine Learning forecasting methods.

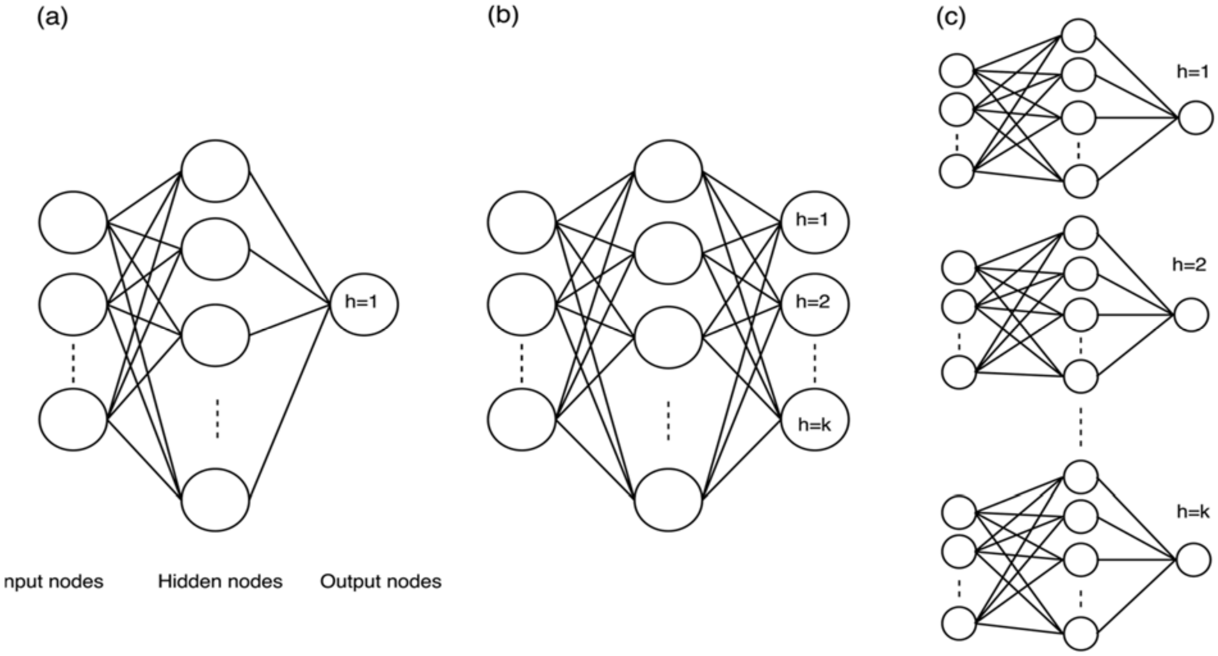


Figure 1: Illustrate the Statistical and Machine Learning forecasting methods.

The methodology for these two types of adjustment is quite different. In the case of the constant term, the appropriate procedure is to examine the residuals of recent periods and adjust the constant term in the forecast period based on those residuals; the other parameter estimates themselves are not changed. By comparison, changing the slope terms assumes the estimates of the slope terms vary over the sample period. Most forecasters adjust their predictions by including some information from the residuals of the equation over the past few periods. If the actual values start at drift off from those predicted by the equation, it would be foolish to ignore that evidence. Nonetheless, such adjustments invariably raise the question: what factors are causing this drift?

We have already examined the reasons why this might happen, but to review briefly, this type of forecast error usually occurs either because some unobservable variable is missing, or the underlying structure has changed. By definition, unobservable variables such as expectations cannot be measured precisely. In the equation for single-family housing starts, adding the index aof homebuyer sentiment materially reduced forecasting error for the 1998–9 period. However, it is not always possible to find useful indexes of sentiment or confidence. Besides, the shift might be due to other factors: the introduction of new product, or a new process for manufacturing or distributing that product [5].

Sales of personal computers jumped when Internet usage became popular, and the rate of inflation declined when the use of the Internet permitted people to purchase goods at lower prices. Changes of this sort would presumably be captured by the recent patterns of residuals of equations used to explain computer sales, or prices of goods that can be purchased "on the computer."

There are no hard and fast rules for estimating what proportion of the recent residuals should be incorporated into the forecast. Basically the forecaster is either the shift is permanent, in which case the full value of recent residuals should be included, or the shift is temporary, in which case the adjustment should gradually return to zero over the next few quarters. If the residuals of the estimated equation have a high degree of serial correlation, and all the recent residuals have approximately the same value, it is reasonable to apply that adjustment to the forecast. However, that raises the question of why the autocorrelation coefficient is so high: perhaps the equation specified, and not all of the variables are actually linear. Perhaps one or amore variables is missing. Other times, a shift has occurred outside the sample period, as was the case for the consumer durables equation after 1997. Usually the reason for a series of positive or negative residuals is not immediately obvious [6].

If the residuals are not highly auto correlated, the usual procedure is to assume that any adjustment in the constant term will return to zero. Yet the drawback of quickly moving back to zero is as follows. Suppose the error over the past year is an amount equal to 4% of the mean value, and the series usually changes by an average of 6% per year. If the last period is used as the starting point, the slope estimates indicate that the change the following year should be equal to the average change of 6%. But if the constant term returns to zero, the predicted change would be only 2%. In that case, the forecast would be based almost entirely on judgment, with the values predicted by the equation almost ignored.

This dilemma emphasizes there is no mechanical way to adjust the constant term in the forecast period unless the source of the error is known, in which case the obvious remedy is to improve the equation over the sample period. Forecasting, like politics, is the art of the possible. Often, the best equation that can be estimated over the sample period will have significant autocorrelation and residuals that bunch near the end of the sample period. Such cases suggest that combined methods of forecasting be used, perhaps by including expert judgment, surveys, or consensus forecasts. But suppose the item to be predicted is sales of an individual firm, for which none of these alternative methods exist.

If the forecaster thinks the recent residuals can be explained as a result of some development that has affected the dependent variable but cannot yet be quantified, the average value of the recent residuals should probably be continued into the forecast period. If, on the other hand, there does not seem to any available reason for the bunching of these residuals, and hence the effect appears to be random, the forecast should be generated without incorporating the values of the recent residuals.

This "hand-crafted" approach will work for someone who is forecasting just few variables, but is not efficient when thousands of variables are being predicted, such as a system designed to keep track of sales and inventories. That is one of the reasons why ARMA procedures are used to keep these equations on track. The examples in this text are primarily designed to show how a298 a forecasts can usefully be improved when the model builder has the time and resources to examine each equation individually [7].

There are no definitive answers about how to treat constant adjustments. In general, though, we suggest the following. If there are strong reasons for estimating an equation in levels form even though the residuals exhibit significant autocorrelation, future forecasts should probably be adjusted by recent levels of the residuals. Otherwise, it is better to remove the trend from the equation, are specify it so the residuals are not auto correlated, and then use constant adjustments only when clearly identifiable changes have occurred in the underlying relationship. We next consider the possibility of estimating an equation in which the values of the parameters change over the forecast period. Perhaps the importance of particular variable has increased over time, is more important in downturns than in upturns, or exhibits other forms of nonlinearity.

The sample period fit can almost always be improved by shifting the slope estimates, but the question is whether that will really improve the forecast accuracy, or just boost the goodness-of-fit statistics in the sample period – so-called a"curve fitting." The tendency, especially for those without considerable experience in actual forecasting, is to over-engineer the equation during the sample period [8]. Keeping that caveat in mind, the three most common methods of changing the slope coefficients over the sample period are I dummy variables for truncated periods and nonlinearities that make the equation piecewise linear, (ii) slope coefficients that are a function of a time trend, and (iii) changes that occur continuously over the sample period but are not tied to specific economic developments. The latter category involves the use of Kalman filters, which are not covered in this book. The use of dummy variables has already been discussed.

One key example of parameter shifts can be seen in the equation to predict net exports of the US economy. Since this series itself has a strong trend, some method should be used to reduce if not eliminate the trend, since trendless variables such as the value of the dollar are important independent variables. In this case, the best method is to take the ratio of net exports to total GDP, both in constant dollars; the graph is shown in figure 8.6. The data start in 1969 because all major foreign exchange rates were fixed before that date, so the underlying data generation function was clearly different.

Net exports are defined as exports minus imports, and are negatively correlated with the trade-weighted average of the dollar. Imports are positively related ato changes in US GDP, whereas exports are positively related to foreign GDP. Hence net exports would be positively correlated with various measures of aGDP in other countries and negatively related with changes in GDP in the US. As the US has integrated more fully into the global economy, the share of GDP accounted for by international commerce has grown. Thus, it would be logical to conclude that in more recent years, a specific change in US GDP or the value of would result in a bigger shift in the net export balance.

Weights have changed: trade with Canada, Mexico, and the developing countries in Asia has grown substantially more significant, whilst trade with Europe, Japan, and OPEC has become relatively less important. These modifications can also be reflected in the coefficients. The exceptional fall in the net export ratio from 1996 to 1999 makes this forecasting problem difficult. Ex post, it was brought on by the demise of the Asian economy and the strengthening of the American economy.

It is unclear whether the drop can be correctly anticipated using an equation with estimates only going as far back as 1995. An equation with constant parameters will do a very poor job of predicting this downturn using the standard variables changes in real GDP in the US and its major trading partners, plus the value of the dollar whereas an equation with changing parameter estimates related to a time trend generates much more accurate forecasts.

While the origin of imports has also changed, this has no impact on the equation since the rise of imports is correlated with changes in US GDP. Canada's GDP is not taken into account since it closely resembles the American economy. Hence, one may anticipate that the significance of GDP for The development of the statistical theory and its empirical application for estimating and evaluating business forecasting models were the main topics of the text's first eight chapters. This information also demonstrated how to assess for parameter robustness during the sample period and enhance prediction accuracy [9].

Actual examples of forecasting models will be shown in the subsequent four chapters. Although chapter 10 focuses on long-term models that may be used to predict trends and changes. The extra issues that arise when using multi-equation models for forecasting. In certain circumstances, the performance of each equation alone may be satisfactory, but the forecasting accuracy suffers when the equations are combined. A small macroeconomic forecasting model with a focus on the actual methodology used to select exogenous variables and modify constant terms in the model. The chapter also assesses how much macroeconomic forecasts can be enhanced by combining structural methods, constant adjustments, and various exogenous inputs, such as sentiment indexes or consensus forecasts.

The strategies for creating and presenting management with short-term sales predictions are discussed in this chapter. This information was created with the customer in mind, including how to convey details about the model's underlying theory and underlying assumptions, as well as the outlook and alternate projections. Due to this, only structural models are taken into account here, however the projections might be improved by it is crucial to establish clearly what the forecasting model is anticipated to do before beginning to estimate multiple regression equations. The forecaster should specifically decide the following.

Sometimes management requests a prediction of overall sales, sales by product category, or sales by geographical area. They may sometimes be curious about the expected market share for certain items. In other instances, the model will serve more as a monitoring tool to check on whether production, sales, and inventory continue to be in line with projections. To estimate real or percentage changes, assess the likelihood that sales or cash flow would fall below a certain threshold, identify the growth rate beyond which expansion is desired, or keep an eye on product lines to check whether they are still on track, management may seek predictions.

Annual forecasts may be preferred in certain circumstances, but they will be monitored on a monthly basis to see if objectives are being fulfilled. Airlines want to know how to best use their current aircraft in the near term to increase the proportion of seats that are full; in the long term, they want to know how many new aircraft to purchase. In the near term, utilities must distribute their current resources to address the peak-load issue; in the long term, they must decide how much capacity to add.

If the data series are not sufficiently long and varied to allow for substantial correlations with the independent variables, the modelling effort is unlikely to be robust. If, for instance, a company had data indicating that sales increased 10% annually for the previous 10 years, there would not be enough variety to provide accurate parameter estimates. The data have to span at least one substantial cyclical variation period. Unless monthly and quarterly data are utilised, robust models should be built on 40–50 observations, which is uncommon for data from particular product lines. The genuine degrees of freedom, however, are often far lower than represented by the number of observations minus the number of variables since these data are not always really independent [10].

Limit your promises. Expecting prediction period errors to be fewer than what sample period data show is often unreasonable. Even if the forecasting process is reliable, those who assert that it is possible face the danger of losing credibility. A forecast range is often helpful in addition to a point prediction. In certain circumstances, missing goals is significantly more severe than having sales increase more quickly than expected. In other instances, a corporation can be forced to put consumers on allocation and lose business to rivals if growth rates are not precisely predicted. When providing the projections, be aware of the dangers on both sides.

Management may sometimes want the best estimate notwithstanding all potential outcomes. Financial analysts, for instance, are paid to choose equities that beat the market as a whole, not stocks that will outperform the market in the event that the Fed eases, oil prices drop, or there is a revolution in Russia. In other situations, management needs to know how much sales would be impacted if real growth were to slow down by 2% the next year or if the dollar rose in value. Other scenarios are sought in these circumstances. These options may not necessarily depend on large-scale variables. Managers can be curious about the impact a new competitive threat will have on their company and the best course of action.

Economic activity may in certain situations influence overall growth, although one brand may benefit at the price of another. Macroeconomic considerations, for instance, have a significant role in determining overall vehicle sales, yet SUV market share may increase temporarily at the cost of "conventional" automobiles before beginning to drop once the market has reached saturation. When that turning moment is most likely to occur is something management may want to know. While the total number of airline miles may be calculated within a very small range, customers may travel more for work than for pleasure, or vice versa.

This is somewhat dependent on whether the forecast consists of thousands of distinct product lines or just a few somewhat aggregate things. Also, it depends on whether management feels more at ease using a prediction that is connected to important economic considerations or one that is made automatically and relies less on discretion. According to the author's experience, the majority of customers need an explanation of the underlying causes of your forecasts. Even if it turns out to be true, a prediction that states "Sales will climb 6% next year" based just on lagged figures and error terms without any further explanation often won't be warmly accepted.

The main independent variables in a structural model must "make sense" to the users of the predictions. If the words employed in the model have little to no apparent relationship to the variables being predicted, then there is little purpose in providing an ostensibly statistically robust model. Customers will be suspicious if phoney correlations are being used. In certain instances, management may want that you use judgement, sales predictions, consensus estimates, management objectives, and other in-house business data in the forecasting process; in other instances, they will have employed an econometrician to discourage you from using this approach. If data are available, one should strive to ascertain the extent to which prior judgement knowledge has enhanced structural model estimates.

While the prediction period inaccuracy will often be greater, you must demonstrate that the model has tracked properly in the past to create trust. Find out whether management accepts the sample period error or if they anticipated a lesser mistake. The majority of executives are aware that the sample period error will almost always be more than the forecast period error, therefore if the latter is greater than anticipated, a different approach may need to be explored. Examine the impact of integrating various forecasting techniques on sample period error reduction. Remember that these techniques must also be believable; management will not be satisfied by just adding another forecasting technique since doing so sometimes lowers prediction error. The presentation of different scenarios, including optimistic, consensus, and pessimistic ones, is often helpful. More precise solutions, nevertheless, are preferable in certain circumstances. Alternatives for the oil sector may be built around a scenario in which a new military conflict breaks out in the Middle East [11], [12].

An alteration in governmental laws, such as the choice to phase out internal combustion engines in favour of more energy-efficient modes of transportation, might be modelled using intervention analysis. In these circumstances, modifications to the particular variables would often outweigh the significance of modifications to the macroeconomic environment. How well do you foresee the drivers' behaviour? This is not a problem if the time being projected simply uses delayed data. Yet, if you discover a significant connection with an unlagged variable that offers a very small sample period error but that variable is not highly predictable, the findings may not be of much help. In the same spirit, confirm that the dependent variable has not accidentally wound up on the right-hand side of the equation in a disguised form.

CONCLUSION

Alternative methods of macroeconomic forecasting are becoming increasingly important in providing valuable insights and predictions about economic trends and behavior. These methods include machine learning, big data analysis, and qualitative forecasting techniques such as expert opinion surveys and scenario analysis. The use of these methods can help policymakers, investors, and researchers to better understand the complexities of the modern economy, particularly in situations where traditional econometric models are limited in their predictive power. By incorporating these alternative methods into their forecasting process, stakeholders can gain a more comprehensive and accurate view of economic conditions, ultimately leading to more informed decision-making and better outcomes.

REFERENCES

[1] Y. Guo, F. He, C. Liang, and F. Ma, "Oil price volatility predictability: New evidence from a scaled PCA approach," *Energy Econ.*, 2022, doi: 10.1016/j.eneco.2021.105714.

[2] Z. Zeng and M. Li, "Bayesian median autoregression for robust time series forecasting," *Int. J. Forecast.*, 2021, doi: 10.1016/j.ijforecast.2020.11.002.

[3] J. T. Kristensen, "Factor-based forecasting in the presence of outliers: Are factors better selected and estimated by the median than by the mean?," *Stud. Nonlinear Dyn. Econom.*, 2014, doi: 10.1515/snde-2012-0049.

[4] C. L. Dunis and J. Alexandros Triantafyllidis, "Alternative forecasting techniques for predicting company insolvencies: The UK example (1980-2001)," *Neural Network World*. 2003.

[5] D. Hendry, "Reformulation empirical macroeconomic modelling," *Oxford Rev. Econ. Policy*, 2000, doi: 10.1093/oxrep/16.4.138.

[6] M. Guidolin and A. G. Orlov, "How Good Can Heuristic-Based Forecasts Be? A Comparative Performance of Econometric and Heuristic Models for UK and US Asset Returns," *SSRN Electron. J.*, 2014, doi: 10.2139/ssrn.2469120.

[7] "Erratum: Loan growth, capitalization, and credit risk in Islamic banking (International Economics (2020) 161 (100–119), (S2110701719302884), (10.1016/j.inteco.2019.11.007))," *International Economics*. 2021. doi: 10.1016/j.inteco. 2020.12.002.

[8] G. Jungmeier, "The Biorefinery Fact Sheet," *Int. J. Life Cycle Assess.*, 2017.

[9] A. Pestova and M. Mamonov, "A survey of methods for macroeconomic forecasting: Looking for perspective directions in Russia," *Vopr. Ekon.*, 2016, doi: 10.32609/0042-8736-2016-6-45-75.

[10] A. Carriero, A. B. Galvão, and G. Kapetanios, "A comprehensive evaluation of macroeconomic forecasting methods," *Int. J. Forecast.*, 2019, doi: 10.1016/j.ijforecast. 2019.02.007.

[11] S. Smeekes and E. Wijler, "Macroeconomic forecasting using penalized regression methods," *Int. J. Forecast.*, 2018, doi: 10.1016/j.ijforecast.2018.01.001.

[12] C. Zhang, Y. X. Tian, Z. P. Fan, Y. Liu, and L. W. Fan, "Product sales forecasting using macroeconomic indicators and online reviews: a method combining prospect theory and sentiment analysis," *Soft Comput.*, 2020, doi: 10.1007/s00500-018-03742-1.

CHAPTER 23

EXPLORING THE PREDICTIVE POWER OF CONSUMER AND BUSINESS SENTIMENT INDEXES FOR MACROECONOMIC FORECASTING: A COMPARATIVE ANALYSIS

Charu Agarwal, Assistant Professor
Teerthanker Mahaveer Institute of Management and Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India
Email Id- charu.management@tmu.ac.in

ABSTRACT:

Consumer and business sentiment indexes have become popular tools for forecasting economic activity. These indexes are based on surveys that measure the attitudes and expectations of consumers and business leaders towards the economy. The consumer sentiment index reflects how optimistic or pessimistic consumers are about their current financial situation and the overall state of the economy. The business sentiment index reflects how optimistic or pessimistic business leaders are about the current and future economic conditions.

KEYWORDS:

Business Sentiment, Consumer Sentiment, Economic Activity, Indexes, Forecasting, Surveys.

INTRODUCTION

Even if the findings are excellent, the process has been a waste of time if you can't persuade the end user to purchase the projections. The presentation must be made in clear, reliable words. Several of these tips place equal emphasis on statistical robustness and presentation. Although it is important to keep that in mind, accurate forecasting is often a required but not sufficient requirement for success. The author is aware of several instances when economists were terminated despite making correct predictions. The technique was seen as being convoluted in certain instances, and the economists were unable to support their claims with sound methodology. In other instances, it came down to "shooting the messenger"; when the economy entered a recession and sales fell, the economics team was let off despite properly predicting the collapse.

When management urges employees to "think positively" despite a faltering economy, it may be prudent to outline both the circumstances in which sales will increase and those in which they will decrease. It is unlikely that making the blanket claim that "sales are going to fall next year and there is nothing you or anyone else can do about it" would result in positive performance reports or consultant retention. The car sector is a notable exception, since most businesses are not big enough for changes in their sales to have an impact on the economy as a whole. There shouldn't be any simultaneity bias in the parameter estimations as a result of incorporating variables like interest and inflation rates, industrial output and capacity utilisation, and credit availability. Therefore, these variables do not meet the requirement of being known at time t unless delayed values are provided [1]. Macroeconomic variable values that are not lag-adjusted are likely to result in more prediction error than would be predicted from sample period statistics.

Use delayed values of the independent variables wherever feasible if the prediction period is just a few months or quarters. Use of a distributed lag that includes both lagged and unlagged variables is sometimes advised. Key drivers may be predicted through consensus. If there are no significant economic shocks that option offers a credible approximation of what may be anticipated. Moreover, using consensus projections might act as a foundation for other scenarios. The management should be informed of the repercussions if the general expectation is for interest rates to fall but you believe they will increase.

If the drivers are different from the kinds of factors that consensus predictions often anticipate, the model should definitely be modified to include these variables as well. Unless they are essential to the model, in which case scenario analysis might be used, avoid using variables that cannot be properly predicted in general. For instance, stock prices are notoriously difficult to forecast, but if they have a significant impact on the prediction variable, it is sense to provide alternate estimates based on educated predictions about the stock market. Finding an exceedingly high correlation with a variable that cannot be properly predicted does not often help much.

Secondly, think about some of the key problems that might occur when predicting the kinds of factors that the company can influence. Companies sometimes mix up intended targets with real sales growth. Let's say a bank wishes to add 15% more loans to its portfolio each year. Making riskier loans allows them to achieve that aim temporarily, but if enough of them fail over time, it may be compelled to combine or even close its doors. But, with better risk management models, the bank may be able to increase its odds of success and increase loan activity without materially increasing risk [2].

The first stage in adopting growth objectives is to ascertain how much sales are anticipated to rise in light of global, local, and demographic trends. The company may then opt to expand, combine, or enter new markets in an effort to accelerate this growth rate. The only way the company can increase sales of its current product lines faster than overall sales is by buying rival companies or stealing market share from other businesses, whether local or foreign. The second outcome might be achieved by cutting costs via increased capital expenditures, improved product quality, greater promotion, or the establishment of enterprises in underserved areas. They are all effective strategies for increasing market share, but since they are all expensive, management must decide if the rate of return is acceptable. They must also be mindful that other businesses may launch a similar counteroffensive. Several high-tech companies learned the hard way in 2000 that mindless growth is often a prescription for catastrophe.

The author is aware of numerous situations where businesses operate under the assumption that their sales can grow faster than the rate indicated by economic conditions simply by urging the sales staff to "do better" without spending the additional funds necessary to achieve this goal. Perhaps all of this seems obvious when put so bluntly, but the author is aware of many instances where firms operate under this assumption. In this situation, econometric models provide a helpful roadmap for what to anticipate in terms of general national or global growth for a certain product or service [3].

DISCUSSION

The majority of industrial or retail businesses use a system that carefully tracks inventory levels. To compare inventory/sales ratios with earlier levels and longer-term trends, several tools are available. If previous patterns persist, these programmers often perform well; nevertheless, when sales alter unexpectedly, econometric modelling frequently offers an extra dimension that can assist optimize stocks and boost profit margins. The following issue is a typical one. An average of 6% has been added to sales each year. They rise 10% in one month above levels from the previous year. That could mean the economy is improving, the demand for the product has risen in spite of no change in macroeconomic conditions, the firm has gained market share relative to its competitors, the gain is a one-time fluke a caused by exogenous conditions (batteries and flashlights were purchased ahead aof a hurricane warning), or random fluctuations have produced an increase athat is not related to either economic conditions or exogenous variables. It is the task of the econometric model-builder to ascertain what really transpired [4].

Top management often cites a range of external factors when reporting to investors, including severe weather, increasing or declining commodity prices, the election, the World Series, the insurgency in Afghanistan, etc. While some of these circumstances may have really had an impact on sales, it is more probable that underlying demand slowed down or expenses increased more quickly than expected. Their positions may be in peril if senior management is unable to resolve these issues swiftly. Several businesses participate in groups that publish industry data gathered and processed by a third party in order to avoid antitrust issues in order to respond to queries of this kind. It will show if your gain was shared by or came at the cost of rivals. The monthly statistics released by the Census Bureau may be used to make comparisons and can provide insight into how quickly broad retail lines are expanding at the retail level.

Decisions on inventories are often made at the micro level. Clothing retailers will rapidly learn which trends are no longer in style this year and which are "hot" things. Especially for perishable goods, supermarkets will fill their shelves with products that will sell rapidly. Steel service centres will soon be able to tell whether demand for rods, beams, wire, and other goods is increasing or decreasing. Inventory management systems will inevitably use ARIMA models for tracking tens of thousands of items since it is difficult to monitor the economic aspects impacting thousands of various product categories or SKU items. Yet, the determination of total demand is greatly influenced by macro variables. The best modelling attempts will often combine these two categories of variables [5].

The quantity of capital expenditure, the scope of growth, and the amount spent on advertising, and the technique for managing and regulating inventories may all be determined by a corporation using econometric models. This data may be utilized as independent variables and elasticity's can be estimated when historical values are known. Even in the absence of this data, judgments may be made based on what growth rate can be anticipated given economic circumstances and how much internal factors can influence that growth rate.

The potential reaction of rivals should be taken into account as the third sort of variable. These factors are undoubtedly beyond the firm's control, at least if it is behaving lawfully, but they are often predictable. Competition will almost always equal price reductions, thus it would be better to raise prices if there were shortages. Even if the news that a rival wants to develop a shop nearby may come as a surprise at first, the current business will have time to respond while construction is underway. The underlying question is whether these aggressive man oeuvres can be econometrically predicted.

We investigate how these many characteristics may be included in a model that predicts how quickly loans are expected to increase for a certain bank. The public disclosure of personal credit information is prohibited under federal privacy laws. But, the broad strategy may be outlined here based on research conducted by this author and his colleagues [6]. Determine the overall increase that may be anticipated in the specific category of loans, such as consumer credit cards, auto loans, equipment leasing, house mortgages, commercial real estate, etc., depending on macroeconomic circumstances. After that, create projections using a particular macromodel or consensus estimates by estimating a regression equation that links the increase in loans to these factors plus interest rates. Let's say that this regression indicated that a certain loan category will expand by 8% annually over the course of the next two years.

Second, the bank must choose whether to focus on lending in its local area, where loan officers may meet prospective borrowers in person and evaluate their character, or whether to extend to other parts of the nation. If the latter choice is used, it may be used to estimate how much the demand for loans will increase in quickly developing regions. Finally, it calculates how much money should be allocated to marketing, hiring more loan officers, and using computerised credit scoring algorithms to increase demand. In certain circumstances, data will be available to experimentally estimate these parameters; in other circumstances, the values will be relied on senior management's opinion.

Fourth, the bank needs to assess the competitive response by other banks. For instance, if a large bank in a big city launches an aggressive marketing effort promoting the fact that "for a short time only, you may borrow money from us BELOW PRIME," they must evaluate how rivals will respond in order to maintain their market share. Reducing interest rates may, in part, boost demand for loans, but because it would also reduce profits, it could not be viable. The bank would want to know, at the very least, an estimate of the increase in loan demand that would result from a 1% reduction in interest rates [7].

Forecast numbers for macroeconomic factors are often calculated outside of the company. Nonetheless, senior management may feel that field judgment may improve forecasting accuracy for factors within the firm's control or those indicating rival reaction. This brings up the broader question of how judgment ought to be included in sales forecasting algorithms. Most business executives believe they have a deeper understanding of their industry than a staff economist or an outside consultant hired to develop a model. They consequently occasionally assert that statistical modelling cannot replace sound judgment. However, forecasts by management could be inaccurate for a variety of reasons. First, executives may prefer an optimistic forecast for reasons of morale.

Second, they may be poorly informed about factors affecting the overall domestic or world economy, even though they understand the dynamics of their own business very well.

It may be seen as a morale booster, or an incentive to the sales force to sell more product. Many firms believe it is worthwhile to set sales targets “too high” in order to discourage slack performance by sales personnel. Top management might “tweak” credit scoring models in order to accept a larger proportion of loans. Production schedules may be set above realistic levels in order to encourage the sales force to increase their effort. It is always possible that sales personnel will underestimate the likely sales growth, thus permitting them to exceed their quotas and earn additional bonuses. However, such attempts are generally seen as self-serving and are not likely to influence company forecasts. In this author’s experience, it is much more common to hear that this company doesn’t want any recession forecasts demoralizing their staff. Negative economic conditions ahead are seen as a “opportunities” and “challenges” rather than nightmares, and many corporations like to infuse their staff with a “can do” mentality [8].

The problem is particularly severe for new product development, where the manager in charge of the new product or service naturally tends to have an optimistic outlook for its future. Yet even with existing products, the usual puffery that “our” product is better than anyone else’s sometimes gets translated into an overly optimistic forecast. When these are not reached, excess inventory accumulates, too many people are hired, too many plants are built, and in extreme cases the company is forced to close. Mindless expansion of retail outlets is one key example of this sort of thinking, as is spending hundreds of millions of dollars on dot.com advertising without giving more than a passing thought to how much sales will rise.

Admittedly, the headline cases in the financial press generally tend to be the failures; when a company performs well because sales or product managers accurately predicted demand that is deemed far less newsworthy. And turning novice loose with an econometric model who has little understanding of athe industry will invariably generate poorer forecasts than the expert judgment aof an experienced product manager. Nonetheless, having said all this, there attends to be a “cheerleader” effect in which pessimistic forecasts are viewed as unwelcome.

The severe drawbacks to judgmental forecasts have long been noted. In an early work, P. E. Meehl around that judgmental forecasts were inferior to decision rules because people applied different factors at different times. See his book, Clinical Versus Statistical Predictions: A Theoretical Analysis and Review of the Literature (University of Minneapolis Press, Minneapolis), 1954. More recently, R. Hogarth and a seasoned business judgment certainly should not be ignored. If sales of a particular product have been rising at 15% per year over the past 10 years, exogenous factors that could cause growth to decline to 5% next year must be carefully considered. Senior management may be aware of changes in the product, or in its method of delivery, that will have a material impact on its sales growth in the future. Econometricians ignore such factors at their own peril [9].

One of the oldest but nonetheless useful clichés in the financial sector is a “never confuse genius with a bull market.” This aphorism also applies to situations where sales of a particular good, after stagnating or growing at modest rates for several quarters, suddenly surge. Often the tendency is to assume that there has been a change for the better, whereas in many cases this improvement are presents a temporary one-time shift. In these situations, management forecasts will often become too optimistic. Figure 1 illustrate the Forecasting with Business and Consumer Survey Data.

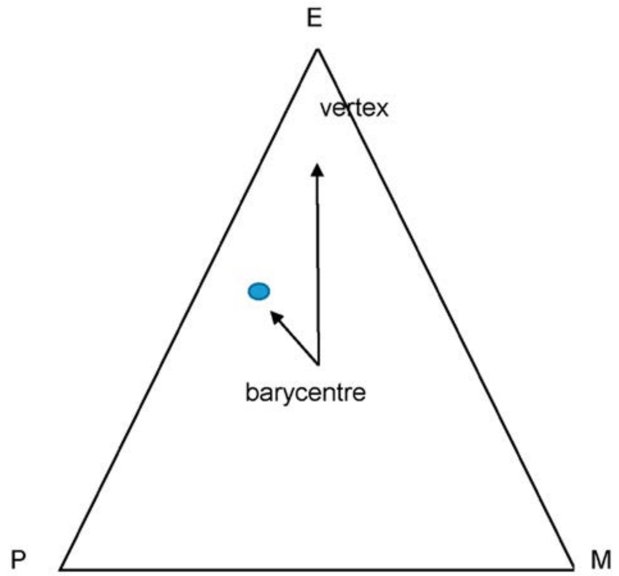


Figure 1: Illustrate the Forecasting with Business and Consumer Survey Data.

As these comments should make clear, no blanket statement can be issued about when judgment will or will not improve forecasting accuracy. However, in this author's experience, a number of useful rules of thumb have emerged/

1. In the long run, forecasts based primarily on technology invariably understate the actual growth that will occur; but in the short run, management tends to overstate the benefits from emerging technology. See: dot.com companies in a 2000.
2. Judgmental forecasts are generally weighted too heavily by events of the recent past. If sales have been above average, that superior performance is expected to continue indefinitely. If sales have been sluggish, it is often the case that subpar performance is expected to continue.
3. Most judgmental forecasts miss turning points. Of course that is also true for many econometric forecasts. However, if signs of a recession are brewing, management too often assumes their sales will not be negatively affected.

Regardless of their expertise at running individual businesses, most top managers possess no special knowledge about the general state of the economy. Hence forecasts of company sales that override macroeconomic concerns are adapt to be inaccurate. On the other hand, sales forecasts based on detailed knowledge of products or markets – precisely the sort of information. In this author's experience, forecasting errors that senior management generally make are not generally tied to individual product performance but reflect an inaccurate assessment of macroeconomic conditions. That is one of the reasons why the importance of macro forecasts is highlighted throughout this text. However, as also noted, macro forecasts have not had an outstanding track record in the past, which leads to the following suggestion.

In terms of communicating with top management, one of the best ways to overcome this problem is to prepare alternative scenarios for the macroeconomic drivers: optimistic, consensus, and pessimistic. In the past, it has been observed (say) that sales of a particular product line increase an average of 2% ape year faster than the underlying economic drivers. Thus, if the economy is expected to continue to advance at average rates, that relationship would continue to hold. If, however, the economy stumbles, sales would probably grow at a much slower rate. The same analysis could be applied to international developments – especially in view of the collapse of various economies around the world throughout the 1990s.

There may also be another reason for top management to phrase forecasts in conditional terms. Public companies are invariably requested to provide a "guidance" to investment analysts, although in mid-2000 the requirement of a "greater" public disclosure paradoxically led to less guidance to selected analysts. Sometimes, when earnings fall far short of expected targets, class action suits follow. If management were to state their expectations in terms of conditional forecasts, such lawsuits could usually be avoided. This will not solve all these problems all the time, but will often provide an additional safety net [10].

Even if lawsuits are not the issue, management performance may appear to be a more satisfactory in the eyes of the Board of Directors if the shortfall in sales reflects an unexpected decline in economic activity rather than poor business planning. These generalizations do not take us very far in determining the weight that should be given to judgment, field input, surveys, and consensus forecasts in generating sales forecasts. We need to be more specific. As a first pass, one should judge the usefulness of such input on what has happened previously. If a given method – whether econometric or judgmental a– has been used previously, one can determine whether these inputs have improved forecast accuracy. Even this method is not foolproof; if a bad forecast is made, it can be pointed out that Sales Manager X is no longer on the job, having been fired for his poor estimates, but Manager Y is much more adept at prediction. Conversely, Manager X did such a great job of forecasting.

Moved on to a better job, and his replacement Y isn't as good. Nonetheless, in spite of the cliché that past performance is no guarantee of future accuracy, the track record of how well these methods have worked in the past should provide valuable guidance about whether they will be helpful in the future.

Sometimes independent judgment reinforces factors that would be captured by the underlying economic relationships. For example, a rise in housing starts will boost sales of home furnishings over the next several months, a relationship that presumably would be reflected by the regression equation and also indicated by input from the sales force [11]. What happens when the input from the field is markedly different from what the modelling process is showing? In most cases, the input is more optimistic; it is unusual for informed opinion to be unduly pessimistic. It is important to examine the reasons for this optimism. Claims that the sales force is better motivated, a new management team is in place, the company lost market share last year but will gain it back this year, etc., are generally worthless pieces of information. The customers will not buy more product just because the management team has improved. On the other hand, significant changes in product quality or service, different market positioning, increased advertising, etc., may have a significant impact on sales that would not be captured in the modeling process.

The importance of judgmental input can also depend on the degree to which product identification is important. A company selling steel finds its demand will not magically increase if the underlying demand for autos, construction, and machinery remains unchanged. Possibly one firm can take market share away from another, but except for price-cutting which will invariably be matched by the competition – this is often wishful thinking. Conversely, 1 rating by J. D. Power & Associates for an automobile brand that had not previously been ranked near the top would presumably boost sales over the following year. However, no econometric model would have adequately predicted the sales of Firestone tires in late 2000.

In this author's experience, the attempt to impose a greater degree of optimism on the forecasts based on outside experience, sales force input, and judgment of top management is usually unwarranted. Surveys are a different matter: surveys of consumer and homebuilder sentiment can improve forecast accuracy. General guidelines in these matters are as follows.

1. One should be cautious about using input from the field if there is no prior attack record. Determine whether informed judgment or surveys improves forecast accuracy before incorporating it into the methodology [12].
2. Judgmental forecasts that are more optimistic than the econometric model approach should be supported by actual changes that have occurred, not just wishful thinking that this year will be better than last year.
3. Negative surprises, such as a competitor building a new plant or opening up across the street, a new international competitor, or the discovery of a product defect, should be taken into account even if they are not part of the previous modeling experience. If there has been a major change in government regulations, intervention analysis the use of dummy variables to capture a shift in the underlying data generation function may be warranted.

CONCLUSION

To the extent that judgment from the sales force and senior management reflects a change in product quality, advertising strategy, or expansion, the forecast should also be adjusted by those explicit factors. Conversely, claims that sales really ought to be better this year because they were disappointing last year are unlikely to have any correlation with sales performance. The vaguer the reasoning, the less likely the input will affect the actual change in sales. However, this doesn't cover all cases. Indeed, suppose that because of inferior forecasting performance in the past, a new team is assembled to try a new approach to forecasting. As part of this new approach, field surveys are commissioned, and input from the sales force, which had not previously been utilized, is also solicited. In some cases, these judgmental factors should be included. Suppose that the quality of a particular product had been subpar, but the manufacturer remedied this defect, and various consumer surveys now show that quality has improved and consumers are beginning to recognize this. Sales would then be expected to rise much faster than levels predicted from a regression equation based on past experience. Sales and orders may also be influenced by new product lines, major trade shows, or a new advertising campaign. In such cases, judgment factors can reinforce the trends and cycles that would otherwise be predicted by regression analysis.

REFERENCES

[1] A. T. Vouldis and D. P. Louzis, "Leading indicators of non-performing loans in Greece: the information content of macro-, micro- and bank-specific variables," *Empir. Econ.*, 2018, doi: 10.1007/s00181-017-1247-0.

[2] J. C. Fuhrer, "What role does consumer sentiment play in the U.S. macroeconomy?," *New Engl. Econ. Rev.*, 1993.

[3] Y. Qiu, "Forecasting the Consumer Confidence Index with tree-based MIDAS regressions," *Econ. Model.*, 2020, doi: 10.1016/j.econmod.2020.06.003.

[4] S. Stetsenko and A. Moholivets, "Methodical approaches to forecasting and early detection of economic cyclicity," *Ways to Improv. Constr. Effic.*, 2021, doi: 10.32347/2707-501x.2021.47(2).120-131.

[5] I. Lolić, P. Sorić, and M. Čižmešija, "Personal finances vs. The overall economic conditions: What drives the new eu member states' stock markets?," *Comp. Econ. Res.*, 2017, doi: 10.1515/cer-2017-0031.

[6] X. Wang and Y. H. Chen, "Is the consumer sentiment index capable of forecasting the transaction activeness of clients? Case study for a commercial bank in China," in *Emerging Markets Finance and Trade*, 2012. doi: 10.2753/REE1540-496X4801S102.

[7] J. A. Laub, "Assessing the servant organization; Development of the Organizational Leadership Assessment (OLA) model. Dissertation Abstracts International," *Procedia - Soc. Behav. Sci.*, 1999.

[8] J. A. Laub, "Assessing the servant organization; Development of the Organizational Leadership Assessment (OLA) model," *Diss. Abstr. Int.*, 1999.

[9] K. J. A˚ström *et al.*, "Abstracts," *J. Power Sources*, 2015.

[10] K. McAlinn, K. A. Aastveit, J. Nakajima, and M. West, "Multivariate Bayesian Predictive Synthesis in Macroeconomic Forecasting," *J. Am. Stat. Assoc.*, 2020, doi: 10.1080/01621459.2019.1660171.

[11] Y. Dendramis, G. Kapetanios, and M. Marcellino, "A similarity-based approach for macroeconomic forecasting," *J. R. Stat. Soc. Ser. A Stat. Soc.*, 2020, doi: 10.1111/rssa.12574.

[12] P. Richardson, "Nowcasting and the use of big data in short-term macroeconomic forecasting: A critical review," *Economie et Statistique*. 2018. doi: 10.24187/ecostat.2018.505d.1966.

CHAPTER 24

THE NEXUS BETWEEN POPULATION AND NATURAL RESOURCE TRENDS: A COMPREHENSIVE ANALYSIS OF IMPACTS AND FUTURE OUTLOOK

Bindoo Malviya, Professor
Teerthanker Mahaveer Institute of Management and Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India
Email Id- bindoomalviya@gmail.com

ABSTRACT:

Population and natural resource trends are two interconnected phenomena that have significant impacts on global sustainability and development. The world's population has been steadily increasing over time, and this trend is projected to continue into the future. As the population grows, so does the demand for natural resources such as water, food, energy, and land. The increasing demand for resources has led to the depletion of natural resources, environmental degradation, and climate change. These trends pose significant challenges to sustainable development, poverty reduction, and social equity.

KEYWORDS:

Climate Change, Environmental Degradation, Natural Resource Depletion, Population Growth, Resource Demand, Sustainability.

INTRODUCTION

In making a presentation, management invariably wants to know what is going to happen to sales and profits of their company, usually by major product line or division. A general discussion of where the overall economy is heading is only of limited interest. Having said that, however, a simple declarative statement that sales are going to rise 6.3% next year even if accurate will not impress very many managers. The thought process is at least as important as the ultimate result, although both are expected to be clearly presented and accurate.

Any presentation of the overall economic environment should be tailored to the type of product or service being forecast. If the product is a cyclical consumer or capital good, variables that determine overall consumer and capital spending are generally most important, including consumer and business sentiment. Indeed, where consumer goods are being predicted, a brief discussion of consumer confidence is usually appropriate. There is no comparable variable of business confidence for capital spending; the NAPM (National association of Purchasing Managers) index is generally viewed as an accurate snapshot of the current state of manufacturing, but has little predictive value [1].

Interest rates are almost always important; even if the industry is not cyclically sensitive, the cost of financing inventories, new capital expenditures, expansions, or mergers and acquisitions is also important. In many cases, the fluctuations in the stock market are an important determinant of sales, but since almost people think "you can't predict the market," alternative scenarios are generally considered useful. If the foreign arena is of interest and importance, the forecasts of key foreign currencies the yen, pound sterling, euro, Canadian dollar, and possibly the Mexican peso and the Australian dollar should be presented. A general scenario of growth in major regions of the world will usually suffice, although if the Company or industry in question has a major investment in some volatile apart of the world, that deserves further mention.

Early in the presentation, the consultant should present a list of the key drivers that affect their markets. The forecasts for these variables should be summarized, followed by a discussion of how variables affect the company or industry markets being predicted. After these have been presented, it is often useful to show the actual and simulated values for the underlying equations. In some cases it may be worthwhile to show what would happen if estimated instead of actual values were used for the independent variables (dynamic simulations).

Managers and clients generally will often want to see how well the equation fits over the sample period, on the logical ground that it certainly will not do any better when true ex ante forecasts are generated. The actual statistics should always be held available in case anyone questions the results, but most managers are not interested in t-ratios, Durbin–Watson statistics, or the Schwarz criterion. Instead, a simple graph showing actual and forecast values should be sufficient [2]. The drivers should usually be represented in elasticity's: for example, a 1% change in variable X causes a 0.4% change in the dependent variable. If lags are used for the independent variables, the lag structure should be briefly summarized, although few are interested in the intricacies of polynomial distributed lags.

Finally, managers generally expect to hear a discussion of the forecast risks, usually by including alternative scenarios with estimated percentage for each case. Especially after a recovery has been ongoing for several years, a discussion of when the next recession is likely to occur is usually relevant. If the industry has an international presence, a discussion of the economic trouble spots around the world is appropriate at this juncture. If the business of the client is directly related to commodity prices agriculture, energy, metals, or lumber a synopsis covering the alternative possibilities of major changes in prices in either direction is warranted.

The remainder of this chapter consists of abbreviated versions of actual presentations made by this author to various clients. The confidential company or industry data used cannot be reproduced here, so published industry data are substituted, but the systematic approach is the same. Two examples are presented: one for capital goods, where management is primarily interested in annual changes, and the other for consumer goods, where the key factor of interest to management is the level of sales this month relative to a year ago [3].

DISCUSSION

Both presentations are designed for short-term forecasting one to two years in advance; long-term trend projections are covered in the following chapter. In both cases, the general approach follows the same structure: graphical explanation of the key drivers affecting the variable to be predicted a• presentation of the actual equation used for forecasting a• discussion of the macroeconomic outlook a• discussion of other key factors in the equation (if relevant) a• presentation of standard forecast for the next two years a• discussion of other variables that could be relevant, including judgmental factors a• alternative scenarios and associated probabilities a• presentation of alternative forecasts for next two years. In general, these graphs and charts are extracted from PowerPoint presentations; in the interests of space, some of the intermediate text has been omitted. The message is to indicate what is relevant for presentations to clients or top management, and what should generally be omitted. While the researcher will presumably have undertaken a wide variety of structural tests to verify stability, they are not included in these presentations because the client is seldom interested in a barrage of statistical details.

Predicting changes in various components of capital spending is one of the areas in which econometric modelling is often thought to be useful because the wide cyclical swings are tied to clearly defined and measurable economic forces. In the presentation given below, NIPA a5 data are used; in practice, this author used confidential industry data collected by an association of firms in this industry. The macroeconomic outlook is discussed next. At the end of 2000, the main appoints of interest were the following. First, the economy is slowing down, and the Fed is expected to ease in early 2001. Because of the lags involved, though, that would not boost purchases of construction equipment very much until a2002. Second, housing starts are expected to improve substantially in 2001, although to a certain extent the decline in interest rates would be partially offset aby weaker consumer confidence and a sluggish stock market. Third, it was assumed oil prices would return to about $25/bbl, but an alternative scenario analysis was prepared since no one really knows what will happen to oil prices [4].

At this point some questions from management or the client can be anticipated. First, the availability of credit has been discussed, but what about the cost? Second, what about the F.W. Dodge index of construction contracts, since only residential construction appears to be included in this equation? Third, what about attitude variables as represented by the index of consumer sentiment and stock prices? In fact all of these variables and many others were tried in the regression equation and were found to be insignificant, but that is not usually the answer management wants to hear. Another approach is warranted. You should explain that while the cost of credit is also important, it is highly correlated with the availability of credit, and is also included indirectly through the importance of mortgage rates in the housing start equation. You can also note that since bond rates already declined in the second half of 2000 in anticipation of Fed easing, they are unlikely to decline much more over the next year or two.

The second issue needs to be finessed somewhat differently, since one would ordinarily expect that changes in non-residential construction would be a major variable in this equation. Initially we had also expected to find that. The actual result which illustrates the lack of correlation between the percentage change in construction contracts as measured by F. W. Dodge and the residuals from the equation on which the forecasts are based. However, that is not the best answer to management. It is better to base your answer on which indicates that the correlation between Dodge contracts and actual purchases of construction equipment has awakened in recent years. Explain that contracts for non-residential construction increasingly lag behind changes in economic activity because of the increased delays in obtaining permits and other legal documents, so they now aserve as a lagging rather than a leading indicator [5].

Finally, consumer confidence and the stock market can be discussed. In this case it is useful to point out that the surge in both these variables in 1999 was actually accompanied by a decline in purchases of construction equipment, largely because of the decline in oil prices and the aforementioned difficulty in obtaining permits. One can also concede that, although these variables are still important over the longer run, they have not recently been correlated with annual changes.

At this point, forecasts should be presented for the most likely, or "standard" economic forecast, an optimistic forecast in which the Fed eases soon and housing starts rebound rapidly, an alternative in which oil prices remain at a$35/bbl but interest rates are little changed, and another alternative in which high oil prices causes the Fed to tighten, hence reducing the availability of credit and housing starts. Ordinarily these would be termed "pessimistic," abut in the case of construction equipment, higher oil prices helps their business.

Note that the gain in construction equipment purchases in 2001 is likely to be robust in spite of slower growth in the economy because of the lagged impact of higher oil prices and previous increases in housing starts. In 2002, by comparison, the lagged effects of the tightening of credit conditions and slowdown in housing starts should result in much more modest gains in purchases of construction equipment. Also, oil prices probably will not rise above peak 2000 levels.

That concludes this presentation. However, the reader may legitimately raise the issue of whether this example is "too simple," in the sense that it does not seem to use many of the sophisticated statistical and econometric tools discussed elsewhere in this text. Also, annual data have been used instead of quarterly data, which greatly simplifies the lag structure. These points are addressed in the next example [6].

This example focuses on percentage changes from year-earlier levels at retail furniture stores. In this case the list of drivers is substantially larger and the lag structure is more complicated, although the specific details are not communicated to the client. The dependent variable is the percentage change in retail furniture store sales from year-earlier levels in constant dollars. All independent variables are also 12-month percentage changes except the yield spread, which airs in levels. Unlike the forecast for construction equipment, which depends largely on lagged variables, this graph shows a wide variety of options for furniture sales, arranging from 7% for the very optimistic forecast to an actual decline for the pessimistic forecast. Under the standard forecast, sales would rise only about a3% per year over the next two years, compared with about 9% in 2000. That reflects the decline in housing starts and stock prices that have already occurred.

Sometimes it is appropriate to assign weights to these various forecasts; for example, 50% to the standard forecast, 20% to the moderately optimistic outlook, 20% for the very optimistic outlook, and 10% for the pessimistic forecast. That would give a weighted average gain of about 4% for furniture sales in 2001 and about 5% in 2002. If management is pessimistic about the overall economy, however, they should be made aware of the fact that, at least based on historical evidence that would point to virtually no gain in furniture sales next year [7].

We now turn to the development of a sales forecasting model for bicycles. The data are based on monthly figures for total bicycle sales calculated by the US commerce Department, and start in 1959. Rather than simply presenting the final results, we will show how this model is derived The first step is to graph the ratio of bicycle sales to total disposable income, and compare this with the ratio of car sales to income to see whether there is any correlation. Car sales are chosen because I both are methods of transportation, and (ii) car sales is a macro variable for which consensus forecasts are already available.

Several factors are apparent from this graph. First, bicycle sales lag motor vehicle sales by an average of 1–2 years; the precise lag structure can be determined empirically. Second, a huge surge in bicycle sales occurred in the early a1970s that has never been repeated. Third, the overall correlation between these two variables is not very high; other variables will obviously be needed. The surge in bicycle sales from 1968 to 1973, and its disappearance at hereafter, deserves further comment, and indeed points out the limitations of econometric analysis. It would be possible, by judicious juggling of independent variables for population, income, credit, and other terms, to explain this bulge, but that would be a useless exercise for forecasting purposes. What really happened is that the social unrest associated with the Vietnam War generated a desire in some younger members of society to move away from "gas guzzlers" and other material goods and move back to a simpler age when bicycles were more important. The end of the Vietnam War, the resignation of anion, and the ensuing recession after the first energy shock brought that era at a quick end [8].

Is this just ex post rationalization, or was this known at the time? In fact, this example is based on an actual model the author and colleagues built many years ago for a leading bicycle manufacturer. In the late 1960s, there was a sharp upsurge in the number of bicycles purchased by "trendy New Yorkers," so it awes assumed that trend would spread to the rest of the country, which did indeed occur later. When those same New Yorkers stopped buying bicycles, the national sales curve dipped about a year later. The fact that income dropped, unemployment rose, interest rates increased, credit was tightened, and oil prices arose may all have been contributing factors, but could not explain most of the cyclical effect that actually occurred. In this case, exogenous judgment was required to generate accurate forecasts.

Now fast forward to 2000, and assume you are given the assignment of building – or updating – a model to predict bicycle sales. For practical purposes, the area of the late 1960s and early 1970s is irrelevant for predicting sales in the future; on a statistical basis, one can quickly verify that assumption by estimating the function for pre-1977 and post-1977 periods and noting that the parameter estimates are significantly different. Consider the data from 1977 through 1999. However, that equation does not work very well. The population ratio for those aged 20–24 is not significant, even though that should be one of the prime age groups. Second, the DW is very low. Third, that equation misses the declines in 1980, 1985, and 1991, and the surge in 1987.

Perhaps some index of consumer sentiment, such as the one published by the Conference Board, should be added. Also, while oil prices should not affect a bicycle sales negatively, high oil prices may be associated with negative sentiment that is not measured in the usual indexes, whereas a drop in oil prices boosts sentiment and also increases discretionary income. The unemployment rate and stock market might also serve as useful measures of consumer sentiment. The equation is estimated with those variables, but they are barely significant, and the DW remains very low. The next step is to try adding an AR (1) adjustment. However, that merely reveals that the series is essentially a random walk; the AR coefficient is 0.997, and none of the other variables is significant with the correct sign except for the unemployment rate. So far it might seem that nothing has been accomplished.

The trend has already been removed from the series for bicycle sales through dividing by disposable income; yet the residuals still exhibit a very high degree of serial correlation. At this point, we should quickly review the steps that have been taken. First, the dependent variable was divided by income and compared with the ratio of car sales to income. Second, various population variables were added. Third, various attitudinal variables were added. Fourth, the equation was tried with an aAR(1) transformation, but without success. Fifth, monthly percentage first differences were used, but also without success. Sixth, the variables were transformed into annual percentage changes, with much more satisfactory results. This is typical of the amount of experimentation that is necessary in building forecast model.

This equation is still far from perfect note the low DW statistic but at least all of the variables make sense. Both population terms are significant with the correct sign, gasoline prices are significant with a negative sign, motor vehicle sales are significant with an average lag of 18 months, and stock prices and the unemployment rate are also significant. Since the residuals are still highly correlated, the equation is once again estimated using the AR(1) transformation. Here again, the results are not very robust, and most of the variables are only marginally significant; in particular, the stock market variable drops out [9].

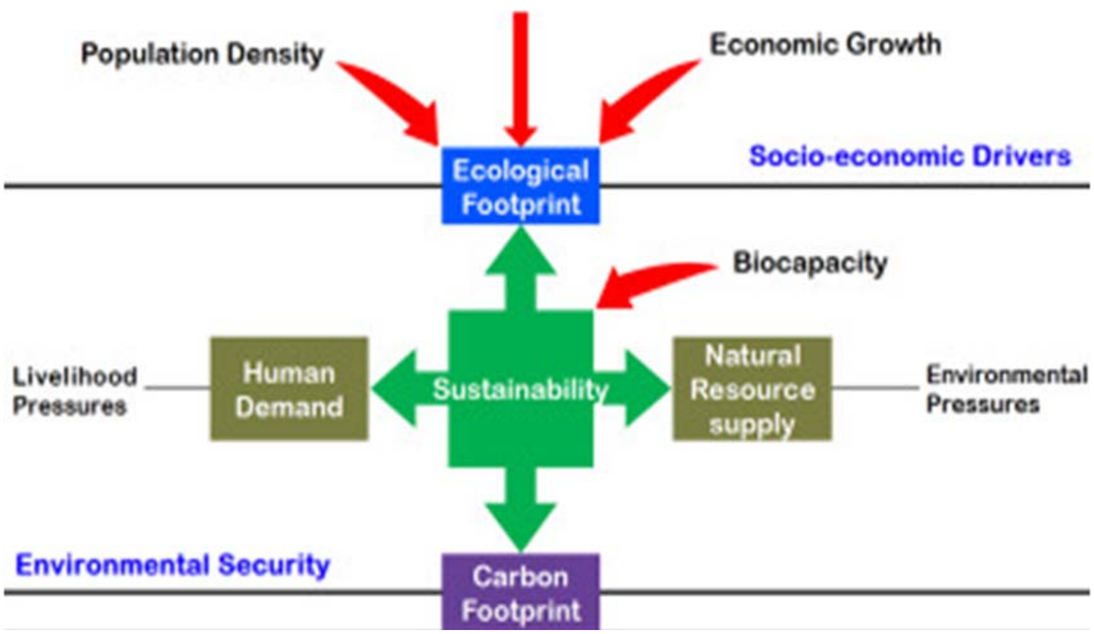


Figure 1: Illustrate the Environmental Performance.

The final equation without the AR(1) adjustment shows that the annual percentage change in the ratio of bicycle sales to disposable income is a function of the ratio of motor vehicle sales to disposable income, lagged 1 a1 a/2 years; changes in population for groups aged 16–19, 20–24, and 25–29; the change in the relative price of gasoline over the past two years; and consumer attitudes as represented by changes in the stock market and the unemployment rate. Figure 1 illustrate the Environmental Performance.

When these variables are included, the Conference Board index of consumer sentiment is not significant. There is still the bothersome detail of the low DW statistic. One standard test is to estimate the equation to the end of 1997 with and without the AR(1) term and then use those parameters to predict 1998 and 1999 sales. The results are revealing: the RMS error with the AR(1) term is 72.4 (million dollars), while it is only 53.2 without the AR(1) term. For comparative purposes, the average value of bicycle sales during that period was $3.04 billion. That result is fully comparable with results obtained for other equations in this text.

One final note on the low DW statistic. Recall that the data used here are on commerce Department statistics and hence are smoothed; the regressions that awe calculated with actual company data did not exhibit the same autocorrelation problem. In many cases, where data have been artificially smoothed by methods of data collection, a low DW does not necessarily indicate that the equation is specified, or that variables have been omitted. It often means the commerce Department has smoothed the data. Only a small proportion of the steps actually taken to estimate this model have been shown; the rest represented empirical experimentation with the lag structure. The variables included in this equation are what one would ordinarily expect in an equation for a discretionary purchase: income (as reflected by using the ratio form), population trends, and attitudinal variables. While it a342 would have been possible to improve the sample period fit by "squeezing" some of the larger outlying residuals, our experience suggests this would not improve the forecast accuracy.

The lessons that can be learned from this example that are generally applicable to building short-term sales forecasting models and using them for forecasting and are summarized next[10]:

1. If the series in question shows growth rates that are inconsistent with historical perspective, check to see whether one-time exogenous factors are at work. If they are, the econometric approach should be combined with judgment.
2. Determine the relationship between sales of this product and relevant consumer or capital spending variables, with particular attention to the lag structure, since short-term forecasting accuracy will be improved if lagged values of the independent variables can be used.
3. It may be necessary to experiment with alternative forms of the dependent variable: in this case, neither levels nor monthly percentage changes generated reasonable results, but annual percentage changes worked much better.
4. Adjusting for autocorrelation is likely to result in larger forecasting errors, even although the sample period statistics appear to be more robust. Hence one should always generate ex post forecasts with and without the AR(1) term to check this possibility.

The data for constant-dollar new orders for machine tools are graphed, and appear to have no significant trend. Hence the initial choice of independent variables would also be trendless. The usual variables in an equation for capital goods would include both variables that measure output or capacity utilization, and variables that reflect the cost and availability of credit, such as the real bond rate and the percentage change in real commercial loans. Both of these variables are lagged, although since the dependent variable is a series for orders rather than shipments, the lags are shorter than would be the case for an equation predicting shipments or purchases of capital goods.

Data for machine tools and the rate of capacity utilization for the manufacturing sector. Note that while neither series has a strong time trend, machine tool orders have tended to rise over time relative to the rate of capacity utilization. That suggests the introduction of an additional variable that does include a time trend, one logical choice being the level of industrial production. The cost of credit is measured as the corporate bond yield minus the five-year average rate of inflation, and the availability of credit is measured by the percentage change in real commercial and industrial loans. The unemployment rate represents the fact that when unemployment is low, firms are more likely to increase their capital spending, ceteris paribus. Most of the estimation work undertaken is used to determine the lag structure that appears to provide the best sample-period statistics and the best ex post forecasts [11].

The graph of the actual versus simulated values is given in figure 9.16. Most aof the cyclical swings are adequately tracked, as is the increase during the 1990s. The DW statistic is still too low by standard tests; however, if the sample period starts in 1977, it rises substantially; furthermore, all the variables remain significant and the coefficients do not change very much. One again the issue arises about whether this equation would predict better with an AR (1) adjustment. The sample period is truncated in 1997 and then used to predict 1998 and 1999 values with and without the AR(1) term. The RMS error is 2.01 without the AR (1) term and 2.18 with it. Here again, the autocorrelation adjustment increases the forecast error. The standard error for the 1997–9 period for the ARIMA (2,1,2) model is 4.30, so that formulation is not considered further.

Nonetheless, this equation does not predict very well. The equation can be are estimated through 1996 and then used to predict monthly values for 1997 through 1999, with the results shown in figure 9.17. Almost all of the actual values are above the simulated estimates. This suggests adjusting the constant term. It is often said that economists – like most other researchers – would much rather present their successes rather than their failures. The farm equipment equation provides an example of an equation that appeared to work well for years, but went off the track recently. Nonetheless some important lessons can be learned. The first step is to graph the pattern of constant-dollar purchases of farm equipment and farm income, as shown in figure 9.18. However, that is not a very promising start, since there appears to be no correlation between these two variables; even with a substantial lag farm income is an important determinant of farm investment, as would be expected, but only when other variables are included in the equation. In this case, a snap judgement cannot be made just by looking at simple correlations.

where INVFARM is the constant-dollar investment in farm equipment, ACREAGE is total planted acreage, INFL is the rate of inflation, CRED is constant-dollar consumer credit outstanding, FAAA is the Aaa corporate bond yield, YFH is constant-dollar net farm income, DITC is the rate of investment tax credit, DOLLAR is the trade-weighted average of the dollar, and LOAN is constant-dollar business loans. D67 is a dummy variable, 0 before 1967. 1 and 1 thereafter, because the value of the dollar was fixed before 1967. DFARM equals 1 in the fourth quarter of years when the investment tax credit was in force and farm income rose; otherwise 0 (see below). Purchases of farm equipment are closely related to the cost and availability of credit as well as farm sector conditions.

The relevant farm variables are net farm income in constant prices, the value of the dollar, and farm acreage. The value of the dollar is included because it impacts farm profitability: when the dollar is overvalued, the volume of farm exports may not drop very much if farmers have excess supplies, but profitability is sharply curtailed. Because the trade-weighted average value of the dollar did not change very much before 1967 – the international gold standard remained in place to 1971, but in the intervening four years the British pound devalued, the DM appreciated, and the US devalued the dollar by 15% – the value of the dollar is multiplied by a dummy variable that is zero before 1967. Because this would introduce a discontinuity, a dummy variable that starts in 1967 is also entered separately.

Credit conditions are also important. The real Aaa corporate bond rate – divided into the nominal rate and the inflation rate over the past two years – consumer credit, business loans, and the rate of investment tax credit are all significant. These terms have the signs that would be expected for an investment function in a commodity-based industry, and the elasticity's appear reasonable. When farm income rose during the year, farmers boosted their capital expenditures of farm equipment late in the year (generally the fourth quarter) in years when the investment tax credit was operative, thus reducing their tax burden. In years when income declined, so the tax burden was lower, the ITC (investment tax credit) was not used. After the ITC ended, this pattern disappeared [12].

CONCLUSION

Population growth and natural resource depletion are two critical trends that have significant impacts on global sustainability and development. The increasing demand for natural resources and the depletion of these resources pose significant challenges to sustainable development, poverty reduction, and social equity. Moreover, these trends have varying impacts on different regions of the world, with some experiencing rapid population growth and resource depletion, exacerbating poverty and inequality.

REFERENCES

[1] S. Tedesco, J. Barton, and B. O'Flynn, "A review of activity trackers for senior citizens: Research perspectives, commercial landscape and the role of the insurance industry," *Sensors (Switzerland)*. 2017. doi: 10.3390/s17061277.

[2] B. A. Bryan *et al.*, "Land-use and sustainability under intersecting global change and domestic policy scenarios: Trajectories for Australia to 2050," *Glob. Environ. Chang.*, 2016, doi: 10.1016/j.gloenvcha.2016.03.002.

[3] R. Deepa, A. Anandhi, and R. Alhashim, "Volumetric and Impact-Oriented Water Footprint of Agricultural Crops: A Review," *Ecological Indicators*. 2021. doi: 10.1016/j.ecolind.2021.108093.

[4] J. Paul, K. Kadirgama, M. Samykano, A. K. Pandey, and V. V. Tyagi, "A comprehensive review on thermophysical properties and solar thermal applications of organic nano

composite phase change materials," *J. Energy Storage*, 2022, doi: 10.1016/j.est.2021.103415.

[5] I. Marotta, F. Guarino, S. Longo, and M. Cellura, "Environmental sustainability approaches and positive energy districts: A literature review," *Sustainability (Switzerland)*. 2021. doi: 10.3390/su132313063.

[6] Y. K. Gautam, K. Sharma, S. Tyagi, A. K. Ambedkar, M. Chaudhary, and B. Pal Singh, "Nanostructured metal oxide semiconductor-based sensors for greenhouse gas detection: Progress and challenges," *Royal Society Open Science*. 2021. doi: 10.1098/rsos.201324.

[7] Z. T. Richards and J. C. Day, "Biodiversity of the Great Barrier Reef- how adequately is it protected?," *PeerJ*, 2018, doi: 10.7717/peerj.4747.

[8] T. Ahmad and D. Zhang, "A critical review of comparative global historical energy consumption and future demand: The story told so far," *Energy Reports*. 2020. doi: 10.1016/j.egyr.2020.07.020.

[9] Y. Wang *et al.*, "The above-ground strategies to approach the goal of geothermal power generation in China: State of art and future researches," *Renewable and Sustainable Energy Reviews*. 2021. doi: 10.1016/j.rser.2020.110557.

[10] E. Y. Menteşe and A. Tezer, "Impacts of Infrastructure Developments on Ecosystem Services Potential in Istanbul," *Front. Environ. Sci.*, 2021, doi: 10.3389/fenvs.2021.614752.

[11] M. Z. Hauschild *et al.*, "Identifying best existing practice for characterization modeling in life cycle impact assessment," *Int. J. Life Cycle Assess.*, 2013, doi: 10.1007/s11367-012-0489-5.

[12] M. Law, "Oxford Textbook of Public Health," *Postgrad. Med. J.*, 1985, doi: 10.1136/pgmj.61.719.848.