CHRIST
(DEEMED TO BE UNIVERSITY)
BANGALORE · INDIA

Case Study Analysis of Forecast Failures of Econometric Models due to Economic Shocks

SUBMITTED TO

DR RAJESHWARI U R

DEPARTMENT OF ECONOMICS

BY

AKSHATH RAJEEV, 2040809

ARNAV SINHA, 2040811

ATHARSH BALAJI R, 2040815

Table of Contents

Case Study Analysis of Forecast Failures of Econometric Models due to Economic Shocks

Abstract: The stock market crash during the Global Financial Crisis 2007-08 and the 2020 COVID-19 pandemic provides researchers and analysts with a lot of data to study, in the most unique and multifaceted ways possible. This paper uses the same data to add to the literature of forecast failures of econometric models during economic shocks by showcasing that a statistically significant model fails to accurately predict stock prices and indices. Specifically, the paper builds two ARIMA models for the two time periods, one from January 1998 to August 2008 and the other from January 1998 to January 2020, predicting S&P 500 values based on 6 other variables, and showcasing with the help of T-Test that there exists a statistically significant difference in the difference between the predicted and actual values during the normal economic circumstances and the difference between the predicted and actual values during economic shocks. This is done after a comprehensive Exploratory Data Analysis involving a thorough Descriptive Statistical Analysis as well as testing for various properties like Multicollinearity, Autocorrelation, and Heteroskedasticity and treating the data for the same. Further, Autocorrelation Plots and Augmented Dickey-Fuller Test is performed to check for Stationarity of the data. Finally, the paper concludes by talking about the policy implications and restating the real-life applicability and impact of the problem under consideration.

Keywords: 2007-08 Financial Crisis, COVID-19 Pandemic, Forecast Failure, ARIMA Model

Introduction

Our complex and highly intertwined economic and financial world is quintessentially dependent on models. Economic and financial models play a crucial role in understanding, analyzing, and predicting economic and financial outcomes. These models are built using mathematical and statistical methods and help in making informed decisions and formulating policies by governments, businesses, and individuals. The complexity and interdependence of the economic and financial world make it challenging to make predictions about future trends and outcomes. Econometric models help to break down this complexity by simplifying real-world situations into a set of mathematical equations. The models then use historical data and other relevant information to predict future outcomes based on these equations. Econometric models are used to make predictions about various economic variables such as GDP, inflation, interest rates, and stock prices. They are also used to evaluate the impact of government policies and regulations on the economy. Moreover, econometric models can help to identify patterns and trends that might not be immediately apparent from raw data. By providing a systematic approach to data analysis, these models can help to test hypotheses, evaluate the impact of different policies, and compare different scenarios.

Despite their usefulness, econometric models are not perfect and can sometimes fail to accurately predict economic outcomes, especially during times of economic shocks such as recessions, financial crises, or pandemics. Economic shocks are sudden and significant events that can have a major impact on the economy, causing unpredictable changes in economic variables such as GDP, inflation, and employment. During these events, econometric models may experience forecast failures due to their limitations in capturing the complex and dynamic relationships between variables in the economy.

Literature has identified many common reasons due to which econometric models fail. Firstly, economic shocks often occur in situations where there is limited data available, making it difficult for econometric models to accurately capture the relationships between variables. Secondly, models that are not well-specified for the economic situation at hand can lead to inaccurate predictions. For example, models that do not incorporate key variables or interactions between variables may not be able to fully capture the impact of a shock. Thirdly, economic shocks can cause instability in the relationships between variables, making it difficult for econometric models to accurately capture these changes. Furthermore, econometric models often rely on assumptions about the distribution of errors, the stability of relationships between variables, and the presence of outliers. When these assumptions are not met, models may produce inaccurate forecasts. Lastly, some econometric models can be overfitted to the historical data, leading to a lack of generalizability and poor performance during times of economic shocks. This is because economic shocks can cause sudden changes in the economic landscape that econometric models may not be able to capture, leading to forecast failures.

This paper thus focuses on studying these forecast failures of econometric models. Specifically, the paper attempts to showcase how even all-encompassing econometric models fail to accurately work and consequently predict the given variables amidst economic shocks. The paper particularly focuses on two economic shocks- The 2007-08 Global Financial Crisis and the COVID-19 Pandemic. Both these shocks are widely different in their characteristics; however, both had a profound, deep-rooted and permanent impact on each and every facet of our lives, changing the economic and financial landscape and the way world works and the way people operate in it. Thus, this paper attempts to present how econometric models fail to accurately forecast essential variables, in our case, specifically the S&P 500 in two different cases of economic shocks as mentioned previously.

The paper also concludes by attempting to justify that despite these limitations, econometric models continue to be widely used and relied upon because they provide valuable insights and information about the economy and financial markets and that they are a useful tool for economists, financial experts, and policymakers to make informed decisions about the future of economies and financial markets.

Literature Review

Macroeconomic Forecasting

The research paper "Central Bank Macroeconomic Forecasting During the Global Financial Crisis: The European Central Bank and Federal Reserve Bank of New York Experiences", focuses on the forecasting practices of central banks during the global financial crisis. It particularly looks at the performance of The Federal Reserve Bank of New York (FRBNY) and the European System of Central Banks (ESCB) in predicting key macroeconomic variables (Alessi et al., 2014). The forecasting tools and monetary policy processes have been tested by the events of the past 5 years and have shown to have worse performance during the Great Recession compared to prior to the crisis. The article identifies three main failures in the real-time forecasting of FRBNY: misunderstanding of the housing boom, insufficient analysis of new forms of mortgage finance, and not giving enough weight to the adverse feedback loops between the financial system and the real economy.

The global financial crisis presented challenges for central bank forecasters as their models based solely on historical estimates were insufficient. Forecast performance was worse during the crisis compared to prior to it, but was still comparable to that of outside forecasters. The article suggests that the use of high-frequency financial data and a scenario-driven approach could have improved the forecast performance. Overall, the article highlights the limitations of the forecasting tools and monetary policy processes used by central banks during the financial crisis. It suggests that central banks should give more consideration to the role of the financial sector in their macroeconomic models and use high-frequency financial data to enhance their forecasting abilities. (Alessi et al., 2014)

Furthermore, the paper “A Short History of Macro-Econometric Modelling” by David F Hendry traces the history of empirical macro-econometric model building, focusing on the early pioneers of the field and the difficulties they faced. The forecasting industry in the US, which emerged during 1910-1930, was wiped out by the Great Depression, which it failed to predict (Hendry, 2020). The era of big macro-econometric models began after the success in predicting the balanced budget effects of the Kennedy stimulus in the early 1960s. However, systematic forecast failures in the 1970s led to increased criticism of both the models and their Keynesian theoretical basis.

The paper explains the various reasons why macro-econometric models can fail, including incomplete specifications, simplistic dynamics, incorrect economic theory, restrictive models of expectations, mismeasured data, incorrect exogeneity conditions, naive model selection, assumption of stationarity, misunderstanding of identification, and linear approximations (Hendry, 2020). The misunderstandings about the causes of forecast failure are often used to further political agendas, even if the failure is unrelated to the quality of the theoretical basis and its empirical implementation. Lastly, the paper offers a summary of a complicated history with many successes and failures, but also undoubted increases in knowledge and understanding overall. Despite such advances, our understanding remains incomplete, given the non-stationarity of the global economy and environment, and may always be so as greater knowledge accrues.

Forecast Failures in Econometric Forecasting

In the past decade, many countries have adopted inflation targeting as a framework for their monetary policy which uses the forecasted rate of inflation as the target for policy decisions, instead of waiting for actual inflation to deviate from the target. However, inflation forecasting is inherently uncertain, and incorrect predictions can lead to incorrect policy decisions. Therefore, improving inflation forecasting is a key concern for those involved in implementing and evaluating monetary policy. It is most effective when the central bank's forecasting model accurately reflects the inflation process in the economy. In this situation, the uncertainty in the forecast can be expressed using conventional confidence intervals or fan charts commonly used by leading inflation targeting central banks. These probabilistic forecasts communicate to the public that the predicted inflation rate is only an average expectation, and that similar inflation rates are equally likely. Based on the representation of uncertainty in the forecasting model, certain inflation rates are deemed unlikely to occur in the future.

However, the notion of model accuracy is challenged by the frequent failures in economic forecasting. When a forecast fails, the errors are typically larger and more systematic than what would be expected if the model were correct. In other words, outcomes that the forecasts consider highly unlikely, such as those outside the confidence interval, occur more frequently than predicted. Therefore, the assumption of model accuracy is unrealistic and represents a weak basis for setting interest rates based on forecasts.

A forecast failure means that the claim of a "correct" forecasting mechanism is no longer valid. The question then becomes whether the error in the model was detectable at the time of the forecast. It is possible that even a thoroughly tested model can produce a forecast failure due to a shift in its parameters after the forecast was made. Regime shifts, which are common in economics, cannot be predicted, leading to repeated forecast failures. The challenge is to quickly identify the cause of the shift and prevent future forecast failures. This is particularly important in the context of inflation targeting, where poor forecasts over a period of years can occur due to unforeseen structural changes. However, sometimes forecast failures occur because of shocks and changes in parameters that occurred before the forecast was made but were not detected by the forecasters. This can be due to the limited statistical power of tests to detect parameter instability. There are also practical challenges that make it difficult to quickly identify shifts in regimes, such as limited time and resources and uncertainty about the reliability of recent data. These factors make it challenging for forecasters to assess the significance of changes in parameters. Forecast errors caused by unanticipated shocks in economics are unavoidable and can cause problems for forecasters. To combat this, it is important to have a flexible and strong forecasting process. Including historical shocks in uncertainty calculations can improve the robustness of forecasts, preventing them from appearing too precise. This could have been a problem in the past, with low and stable inflation rates potentially leading forecasters to underestimate uncertainty. To address this, all inflation explanatory variables are also forecasted, and our ex-post forecasts do not rely on variables that would have been unknown at the time of the original forecasts. Additionally, the forecasts are updated regularly as new information becomes available, with model parameters re-estimated and initial conditions adjusted. (Nymoen, 2004)

The study makes a distinction between two types of failures in forecasting: ex-ante forecast failure and ex-post predictive failure. The former refers to incorrect predictions about future events and can be caused by factors such as data errors or incorrect assumptions about unmodeled variables. The latter occurs when a model is not constant over the entire data set and is a well-known concept. The analysis highlights the challenges in using econometric models for forecasting. According to Hendry and Mizon (1996), a model must either correct itself dynamically or change its target as the economic situation changes. The issue is that important factors affecting forecasting may not be causal variables when the model is not correctly specified in a world with changes in constant elements (trends, etc.). This opens up the possibility of creating forecasting methods that are resilient to significant structural shifts, separating forecasting from modelling and making it impossible to claim that the best forecasting model should be chosen for policy purposes. An example of poor forecasting in the UK housing market during the period of 1973-75 can be seen as a result of changes in financial markets due to Competition and Credit Control regulations. Models created by Hendry in 1984 resulted in inaccurate predictions of the significant increase in house prices during one of the most rapid booms. Similar instances of incorrect forecasting have been seen in other areas, such as consumers' expenditure in the mid-1970s and early 1990s, and UK M1.

The conclusion is that the presence of structural changes in data does not clash with congruent modelling. On the contrary, it can provide opportunities to improve our understanding and improve the model. However, it is still possible to have poor forecasts due to unforeseen events such as wars. The main reason for forecast failure is the shift in the average value, as demonstrated by various simulations.

Model specification errors, collinearity, and complexity are not the main cause of forecast failure but can make the problem worse. Hence, modelling methods do not seem to be a solution to the problem of forecast failure. Modelling unexpected changes after they occur may lead to overfitting, making the forecast less accurate. The preferred approach is to include dummies for identified structural changes, not just outliers, in the model and possibly compute the forecast confidence intervals from a model that does not include these dummies.

There is a significant issue in forecasting when models are mis-specified and there are changes in constants and trends, as the causal variables may not always have the most impact. This can lead to the creation of non-congruent forecasting devices that are resilient to structural breaks. This creates a divide between forecasting and modelling and challenges the idea that the best forecasting model should be used for policy. However, analysing the reason for forecast failure can lead to progress in modelling. There have been instances where a model that initially produced poor forecasts eventually provided a constant and accurate explanation. This was due to incorrect measurement of some of the explanatory variables during the forecasting process, which was corrected with the use of more accurate measures. As a result, a larger, non-constant model was initially thought to be necessary, but it was later found that the original model could still provide an accurate explanation with a different set of parameters, confirming its constancy. Examples of this phenomenon can be found in Hendry (1996b) and Ericsson, Hendry, and Prestwich (1997). (Hendry & Doornik, 2002).

A prediction or estimation of future events is referred to as a forecast. This is done by taking current information and projecting it into the future using a set of systematic procedures. However, this process can be challenging because the future is uncertain for two reasons: one being the uncertainty where we have a grasp on the probabilities involved and can therefore factor these into our predictions; the other being the uncertainty that we don't yet understand.

Models that are based on past data and experiences, called empirical models, can consider the impact of unexpected events from the past and therefore provide an accurate representation of the past. However, new and unpredictable events are always likely to happen in the future, making it more uncertain compared to the past. Therefore, any practical method for forecasting economic events must consider and prepare for such unforeseen occurrences.

The persistent and repetitive changes are now modelled using stochastic trends. Structural breaks, which are sudden and significant changes that are almost always unanticipated, are a major cause of inaccurate forecasts. This leads to a significant decrease in the accuracy of predictions, typically compared to the historical performance of a model. Currently, there is no established method for modelling such breaks, though much effort is being invested in developing non-linear models, which mainly focus on rare events. The nine sources of forecast error are: (1) shifts in the coefficients of deterministic terms (2) shifts in the coefficients of stochastic terms (3) mis-specification of deterministic terms (4) mis-specification of stochastic terms (5) mis-estimation of the coefficients of deterministic terms, (6) mis-estimation of the coefficients of stochastic terms (7) mis-measurement of the data (8) changes in the variances of the errors, and (9) errors cumulating over the forecast horizon.

The success of a forecast should not be used as the sole criterion for selecting a model, especially when it comes to policy models. Similarly, the failure of a forecast should not result in the rejection of a model. Efforts to predict future uncommon occurrences that involve sudden changes must be taken into account. However, many rare events do not follow a pattern like economic cycles. Nevertheless, these rare events can still be partly predictable as they have underlying causes. This can be achieved by observing high-frequency data, which can reveal the sudden changes in real-time. An economic forecasting theory that takes into account structural breaks, incorrect models, and other factors, has vastly different consequences compared to a theory that assumes stability and accurately specified models. It has been proven that concepts that are easily proven under the assumption of stability and correct models do not apply to a more realistic scenario, where realism refers to consistency with the evidence of forecast failures and results from forecasting contests. (Hendry & Clements, 2001).

Concerns regarding Econometric Modelling

The paper titled "Problems and Issues in Evaluating Econometric Models" by James B Ramsey and Jan Kmenta focuses on several key themes related to the accuracy and reliability of econometric models. The first theme is the level of detail and accuracy in the specification of the model. There is a difference between formal and informal models, and a more precise distinction between parametric and nonparametric models. The second theme concerns the robustness of the model's inferences in the face of errors in the model's specification. It is often said that all models are approximate and therefore all models are in error. While this may be true to some extent, it is important to recognize that models may contain errors and to use inferential procedures that are less sensitive to the most likely errors. The third theme involves formal methods for comparing models. As econometric models have become more complex in structure, it has become necessary to develop new methods and criteria for choosing between different models of a given economic situation. The fourth theme is the role of time series analytical methods in econometric models. In the past, there was a dichotomy between purely statistical data analysis with no economic theory content and theoretically specified models that did not consider the complexity of time series structure in the stochastic elements. Recent efforts have aimed to reconcile the conflicts between these two approaches. The fifth and final theme involves the potential benefits and costs of using experimental data to test economic hypotheses. On one hand, the use of experimental data can provide valuable information not directly observable from historical data, regardless of the inferential methods used. On the other hand, experimental data can serve as a substitute for further refinement and improvement of methods for extracting information from historical data (Ramsey & Kmenta, 1980).

Adding to the literature in this topic, the paper titled "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left" by Jerry Hausman deals with mismeasured variables in regression models and approaches to deal with it. The problem of mismeasured variables in statistical and econometric analysis is a well-known issue dating back to the 1870s. In a linear model, if both the left-hand side and right-hand side variables are mismeasured, the estimate of the regression coefficient remains biased downward, with less accuracy and a decrease in R^2 (Hausman, 2001). To solve this problem in linear specifications, the use of instrumental variables is often relied upon. These are variables that are assumed to be correlated with the true value but uncorrelated with the error term. However, in non-linear models with measurement error in the right-hand side variables, two-stage least squares estimate no longer result in consistent estimates of coefficients, leading to the need for alternative solutions. One approach is to use a polynomial specification. The use of instrumental variables and polynomial specifications are ways to mitigate the impact of mismeasured variables in econometric analysis.

Another influential paper titled "Data Problems in Econometrics" by Zvi Griliches highlights the challenges and limitations posed by data in the field of econometrics. It points out that the quality of econometric results relies heavily on the quality of data used, and that even small inaccuracies in data can have significant impacts on the results (Griliches, 1984). It emphasizes the largely second-hand nature of economic data and the consequences that flow from the distance between econometricians as users of data and its producers. There are also serious definitional problems about the borders of the economic activity. While many macro series are subject to errors, most of these errors rarely fit into the framework that has been developed in Econometrics. The paper also focuses on the limitations of using aggregate data, such as the aggregation bias, which can result in incorrect inferences about underlying relationships. It also discusses the challenges in accurately measuring variables, such as measurement errors, which can lead to biased results. The author concludes that these limitations in data can lead to incorrect inferences about relationships between variables and can impact policy decisions based on these results. It also deals with the problems of missing data. Missing variables cause a loss of efficiency as well as the possibility of serious bias in estimated coefficients (Griliches, 1984). The article highlights the importance of careful consideration of the data used in econometric analysis and the need for more rigorous methods to improve data quality.

Unpredictability is defined as a random variation in a known distribution that remains unchanged even when considering available information, as demonstrated by Doob (1953). If a variable is inherently unpredictable, it cannot be modeled or forecast more accurately than its unconditional distribution. Unpredictability can also stem from unexpected changes in the distribution at unforeseen times, with shifts in the mean being the most damaging. Intrinsic unpredictability can be considered as "known unknowns" as the probabilities of outcomes can be estimated, while extrinsic unpredictability is similar to "unknown unknowns" as it is impossible to calculate the conditional or unconditional probabilities of outcomes beforehand. The different forms of unpredictability have significant impacts on economic analysis, modeling, and forecasting. If faced with extrinsic unpredictability, inter-temporal economic theory, forecasting, and policy analysis can be faulty, but the eventual outcomes can still be modeled afterwards.

The idea being presented is that neither economic agents nor professional economists are successful in accurately predicting changes or shifts in the economy. This has significant consequences for economic theories that depend on stable conditions and the use of conditional expectation for predictions. When the distribution of the economy changes over time, it becomes irrational to use the conditional expectation, and agents may need to use other methods. While it is not possible to predict the timing and extent of major independent events, it is possible to remove the problem from a model by using appropriate indicator variables after a shift has occurred. However, even if the causes are known and variables are added, it may not improve forecasting unless future shifts in these variables can also be predicted. For example, changes in oil prices can greatly affect inflation and are often included in models of price inflation, but they remain just as difficult to forecast as inflation itself.

There are several benefits of using a congruent encompassing model for forecasting despite the presence of unexpected shifts. The four potential advantages are: 1) such models produce the lowest variance for the innovation error, which is a key factor in forecast error; 2) modeling mean shifts in-sample eliminates one source of systematic mis-forecasting; 3) excluding irrelevant variables that may shift during the forecasting period reduces the chance of forecast failure; 4) in congruent models, decisions about model selection can be based on standard inference procedures, such as t-tests. The robust form of an econometric model that is based on economic theories can produce accurate forecasts for long periods of time even when the original model fails to do so, despite the presence of structural changes. (Hendry & Mizon, 2014)

A different approach to voicing concerns w.r.t econometric models and such was taken in the paper "Mathematical Models and Economic Forecasting: Some Uses and Mis-Uses of Mathematics in Economics" by David F. Hendry. Th paper examines three scenarios in which mathematics is used or misused in economics and econometrics. The first scenario focuses on the importance of mathematical analysis in understanding economic forecasts and why they can fail. The author argues that mathematical analysis is crucial in understanding the properties of forecasting models, regardless of the specific model and behavior being forecasted. The main reason for forecasting failure is identified as location shifts, while other factors such as model mis-specification and data mis-measurement play a secondary role. The second scenario concerns the issue of model selection when there are more candidate variables than observations. The author emphasizes that advanced mathematical analysis is necessary to understand the properties of extended general-to-specific procedures and their ability to correctly select the best model. Despite widespread belief that model selection is problematic, the author argues that there is no substantive mathematical proof to support these claims. The third scenario deals with the mathematics of inter-temporal optimization and expectations, particularly the concept of "rational expectations." The author argues that current mathematical approaches to this issue lead to misleading results when applied to realistic economies. The author highlights that conventional notation fails to address the different times relevant to expectations formation and that more powerful and general mathematical techniques are needed, with assumptions that better reflect economic reality (Hendry, 2012).

Traditional econometric models have been found to be less effective in forecasting compared to simple time series models. The issues with static regression models include unsteadiness in structure, poor forecasting accuracy, and false regression. In recent times, more advanced dynamic models such as the autoregressive distributed lag model (ADLM) and the error correction model (ECM) have been introduced in the field of tourism demand forecasting. Relying only on the dummy variable approach to capture exogenous shocks in the time series is inadequate, as it fails to correct for the effects of these shocks and can lead to incorrect estimation of model parameters due to uncorrected outliers. Given that tourism time series are often highly impacted by these factors, using a comprehensive outlier detection method is crucial for accurate forecasting.

In this study, it was found that the simplest model, known as the naive model, generated the most accurate long-term forecasts, followed by the TVP and SARIMA models. It was also shown that the TVP model, which incorporates seasonal and calendar effects, outliers, and parameter identification, consistently produced excellent results for short-term forecasts of one to three quarters ahead, aligning with the findings of Smeral and Wüger. On the other hand, the static regression model was found to produce relatively inaccurate forecasts for all forecasting horizons, as it lacks a dynamic structure and is unable to effectively capture fluctuations in quarterly tourism demand. Hence, when choosing the optimal forecasting model, it is important to consider the specific market conditions of the destination. The frequent success of the destination-specific models suggests that the TVP model is consistently effective across various market conditions. (Song et al, 2008).

Types of Prediction Methods

The article by (Demyanyk & Hasan, 2010) titled "Financial crises and bank failures: A review of prediction methods" is a review of the econometrics and operations research methods that are used in the empirical literature to study and predict financial crises and mortgage defaults. The authors argue that an interdisciplinary approach that combines these two methodologies is beneficial as isolated methods may not be capable of accurately predicting such events. The paper reviews the usage of Neural Networks (NN) in analyzing financial data, stating that as they are nonlinear and do not make any assumptions about the statistical distribution of the data, making them useful in practical situations where financial data does not meet the requirements of certain statistical models.The authors also provide an analysis of the circumstances surrounding the subprime mortgage crisis, which is believed to have triggered the global financial crisis. The paper mentions (Demyanyk & Hemert, 2008) and how they analyzed the subprime crisis empirically, utilizing a duration statistical model that allows estimating the so-called survival time of mortgage loans, i.e., how long a loan is expected to be current before the very first delinquency or default occurs, conditional on never having been delinquent or in default before. The authors also show that the above-mentioned monotonic deterioration of subprime mortgages was a market-wide phenomenon.

The paper also reviews many other papers analyzing the types of models used and their results such as the following:

In their study conducted between 1980 and 1994, Demirguc-Kunt and Detragiache used a multivariate Logit model to examine the factors that contribute to the likelihood of a banking crisis globally. They found that banking crises are more likely to occur in countries with low GDP growth, high real interest rates, high inflation rates, and the presence of an explicit deposit insurance system. In a subsequent study in 2002, Demirguc-Kunt and Detragiache specifically looked at the relationship between explicit deposit insurance and stability in the banking sector across countries. They confirmed and strengthened their previous findings that explicit deposit insurance can harm bank stability. This happens because banks may be encouraged to finance high-risk, high-return projects because of the insurance, which can lead to more losses and failures. The authors found that deposit insurance has a greater negative impact on the stability of banks in countries where the institutional environment is weak, where the coverage offered to depositors is more extensive, and where the scheme is run by the government instead of the private sector.

The prediction of a company's default risk has been a topic of hundreds of research articles since (Altman's, 1968) proposal of using the "Z score." This has been reviewed in the works of (Kumar & Ravi, 2007) and (Fethi & Pasiouras, 2009). Many studies have proven that the techniques used in operations research for intelligence modeling can be applied to predict bank failures and crises. For instance, (Celik & Karatepe, 2007) discovered that artificial neural network models can forecast non-performing loans relative to total loans, capital relative to assets, profit relative to assets, and equity relative to assets. Additionally, (Alam et al.,2000) found that fuzzy clustering and self-organizing neural networks are effective tools for identifying banks that are likely to fail.

The use of Early Warning Systems (EWS) by central banks to monitor the risk of banks has been widespread. Nevertheless, the persistence of banking crises in recent decades, such as the Asian crisis, the Russian bank crisis, and the Brazilian bank crisis, highlights the challenge of safeguarding the banking system. In the United States, regulators must conduct assessments of bank risk every 12-18 months, based on the Federal Deposit Insurance Corporation Improvement Act of 1991. A rating system, known as CAMELS, is used to indicate the safety and soundness of banks and consists of six components: capital adequacy, asset quality, management expertise, earnings strength, liquidity, and sensitivity to market risk.

In 2008, Davis and Karim conducted a study to evaluate the effectiveness of statistical and intelligence techniques in predicting banking crises. They compared the Logistic Regression and Signal Extraction EWS methods and found that the choice of estimation model impacts indicator performance and crisis prediction. The Logit model was found to be a better global EWS, while Signal Extraction was more suitable as a country-specific EWS. In a follow-up study, Davis and Karim tested whether the Logit and binomial tree approaches could have predicted the subprime crisis in the US and UK. The results showed that among the global EWS for the US and UK, the Logit model performed the best, although it only had a limited ability to predict the crisis.

(West, 1985) used a combination of Logit modeling and factor analysis to assess and describe the financial and operational characteristics of banks. He analyzed data from Call and Income Reports and Examination Reports of 1,900 commercial banks across various states in the US and found that the factors identified by the Logit model were similar to those used in CAMELS ratings. He concluded that his method of factor analysis and Logit estimation was useful for evaluating the conditions of banks.

Discriminant Analysis (DA) was a popular technique for analyzing and predicting bank failures for many years, with linear, multivariate, and quadratic subcategories (e.g., Karels & Prakash, 1987), (Haslem et al., 1992). However, DA requires a normal distribution of regressors, and when this requirement is not met, Logit models can be used instead. DA is suitable for analyzing cross-sectional data, while hazard or duration analysis models should be used for time series data on bank firms or loan defaults.

(Canbas et al., 2005) developed an Integrated Early Warning System (IEWS) that combined DA, Logit, Probit, and Principal Component Analysis (PCA) to help predict bank failure. They first used PCA to identify three significant financial components that explain changes in the financial condition of banks, then employed DA, Logit, and Probit regression models. By combining these methods, they created the IEWS. The authors tested the predictive power of the IEWS using data from 40 privately owned Turkish commercial banks and found that it had greater predictive ability than other models in the literature.

Neural Networks (NN) is the most popular intelligence technique used today. It has developed from the fields of artificial intelligence and brain modeling and has mathematical and algorithmic elements that mimic the human nervous system. NN processes information by connecting artificial neurons using a connectionist approach, with the structure of the model changing based on information flow during the learning phase. (Boyacioglu et al., 2008) compared NN, Support Vector Machines (SVM), and multivariate statistical methods for predicting bank failures in Turkey using financial ratios from the CAMELS ratings. MLP and LVQ were found to be the most successful NN models. Back-Propagation Neural Networks (BPNN) is a commonly used NN model for classification and prediction problems, consisting of input, hidden, and output units. BPNN outperforms other methods in most cases, as seen in studies by (Tam, 1991) and (Ravi & Pramodh, 2008). (Tam & Kiang, 1992) found BPNN to be superior to other methods for a one-year-prior sample, while linear discriminant analysis was better for a two-year-prior sample.

(Tsionas & Papadakis, 2009) developed a statistical framework that can be applied to stochastic DEA, using a Bayesian approach based on simulation techniques to make inferences about efficiency scores. The authors tested their method on the efficiency of Greek banks and found that most of them perform close to market best practices. (Cielen et. al., 2004) compared the performance of a DEA model (Minimized Sum of Deviations), a linear programming and DEA combination (MSD), and a rule induction model (C5.0) in predicting bankruptcy. Using data from the National Bank of Belgium, they found that MSD, DEA, and C5.0 achieved correct classification rates of 78.9%, 86.4%, and 85.5%, respectively. The authors concluded that DEA was more accurate than the C5.0 and MSD models. (Niemira & Saaty, 2004) used an Analytic Network Process (ANP) based multiple criteria decision-making model to predict the likelihood of financial crisis, testing it on the US bank crisis of the 1990s. They found that the ANP framework improved the reliability of information processing and reduced judgmental forecast error, and concluded that it is a more flexible and comprehensive methodology than traditional models. (Ng et. al., 2008) introduced the Fuzzy Cerebellar Model Articulation Controller (FCMAC) model, which uses a compositional rule of inference called FCMAC-CRI(S) to integrate fuzzy systems and neural networks (neural fuzzy networks) for localized learning. The new network analyzes patterns of financial distress from public financial information and provides insights into financial distress through fuzzy IF-THEN rules. The authors compared the accuracy of the FCMAC-CRI(S) model to Cox’s proportional hazard model and the GenSoFNN-CRI(S) network model and found that the new approach was better than the benchmark models.

In conclusion, the authors argue that the interdisciplinary approach of using econometrics and operations research methods is beneficial for future research on financial crises and mortgage defaults.

The use of Neural Networks, a type of artificial intelligence, has been found to be a useful tool in predicting such events, particularly when combined with statistical methods (Demyanyk & Hasan, 2010).

Research Gap

Even though the literature is quite comprehensive when considering the topic of forecast failures of Econometric Models, none of the authors carry out a comprehensive case study analysis specifically focusing both on the 2007-08 Global Financial Crisis and the COVID-19 pandemic. This paper is unique considering the objectives of the same is to present a detailed case study analysis showcasing the forecast failures of econometric models in predicting S&P 500 even when the model is statistically significant otherwise. We seek to compare the output or the forecasted value of the model from the original or actual values of S&P 500 and show that due to economic shocks, forecasts based on comprehensive and even statistically significant econometric models do not work and do not accurately capture the numerous unknown variables. This may seem basic and obvious on the first glance; however, the aim of the paper is to add to the literature of forecast failures of econometric models and present the urgency of relying on models that cannot be used during economic shocks to forecast or even to predict such crisis which continue to be a feature in the highly intertwined and complex economic and financial system.

Research Objectives

- To showcase how a statistically significant ARIMA model fails to accurately predict the S&P 500 in two different cases- the 2007-08 Global Financial Crisis and the COVID-19 Pandemic.
- To generalize through comprehensive review of literature and the case study analysis the issue of forecast failures of econometric models, its consequent implications and potential solutions

Research Methodology

To begin with, the paper follows a thematic style of literature review, concluding the same by finding the relevant research gap. Clearly, as the paper aims to predict and forecast S&P 500, it is crucial to establish the kind of models that are being used and the variables that are being used in these models.

After a comprehensive study of the literature (Alessi et. al., 2014; Demyanyk & Hasan, 2010; Hendry & Mizon, 2014; Hausman, 2001), we have concluded that the following variables would best encompass the movement of S&P 500: *a) The Dow Jones Industrial Average (DJIA), b) The Nasdaq Composite Index, c) The Russell 2000 Index d) The VIX Index (also known as the CBOE Volatility Index), e) The yield on 10-year Treasury bonds, along with the historical values of S&P 500.*

The dataset consists of Daily values of the above-mentioned variables with the time period for these variables ranging from January 1998 to January 2020. The reasoning behind this time period is that the paper aims to test the forecast of the econometric model for 2 different economic shocks as mentioned earlier, the 2007-08 Financial Crisis and the COVID-19 Pandemic. Specifically, the paper aims to use data from January 1998 to August 2008 for the first case and attempts to showcase that the model fails to accurately predict the market crash seen on September 29, 2008 and the following period. Similarly, for COVID-19 pandemic case the data being considered ranges from January 1998 to January 2020, attempting to showcase that the model fails to accurately predict the stock market crash as seen on February 24, 2020 and the following period.

For forecasting we are considering the *ARIMA (AutoRegressive Integrated Moving Average) model* which is time series model. This model is considered after a thorough study of the structure of the data. It was found out that the data was non-stationary. To check for stationarity of data we used *Autocorrelation Function Plots (ACF)* as well as the *Augmented Dickey–Fuller (ADF) test*. Given that

ARIMA model can be used for non-stationary data and that the model itself converts the non-stationary data to stationary data, this was selected. Further, given that daily values are being taken, research has shown that ARIMA fits well for such type of data (Alessi et al., 2014; Nymoen, 2004; Demyanyk & Hasan, 2010).

The data and all the variables are sourced primarily from www.finance.yahooo.com. The paper first aims to conduct an Exploratory Data Analysis (EDA) of the relevant variables. For EDA we perform a Descriptive Statistical Analysis, focusing on the statistical properties of the variables and the data in general. Further, we test the data for important assumptions like *Heteroskedasticity, Autocorrelation and Multicollinearity*. We used *Breusch-Pagan (BP) test* to check for Heteroskedasticity and further used the *Log-Log Transformation* to treat the data. Next, to check for Autocorrelation we used the *Durbin-Watson (DW) test*. Finally, for checking Multicollinearity we used the *Variance Inflation Factor (VIF) value*.

Further, the best ARIMA model was found out using the *order of Differencing* and by studying the *lags of ACF and PACF plots*. Lastly, the paper would predict S&P 500 values based on the model and compare them with actual values of S&P 500 to draw meaningful conclusions as based on the research objectives. All the analysis of the data, building of the model and prediction is done using the R software and RStudio.

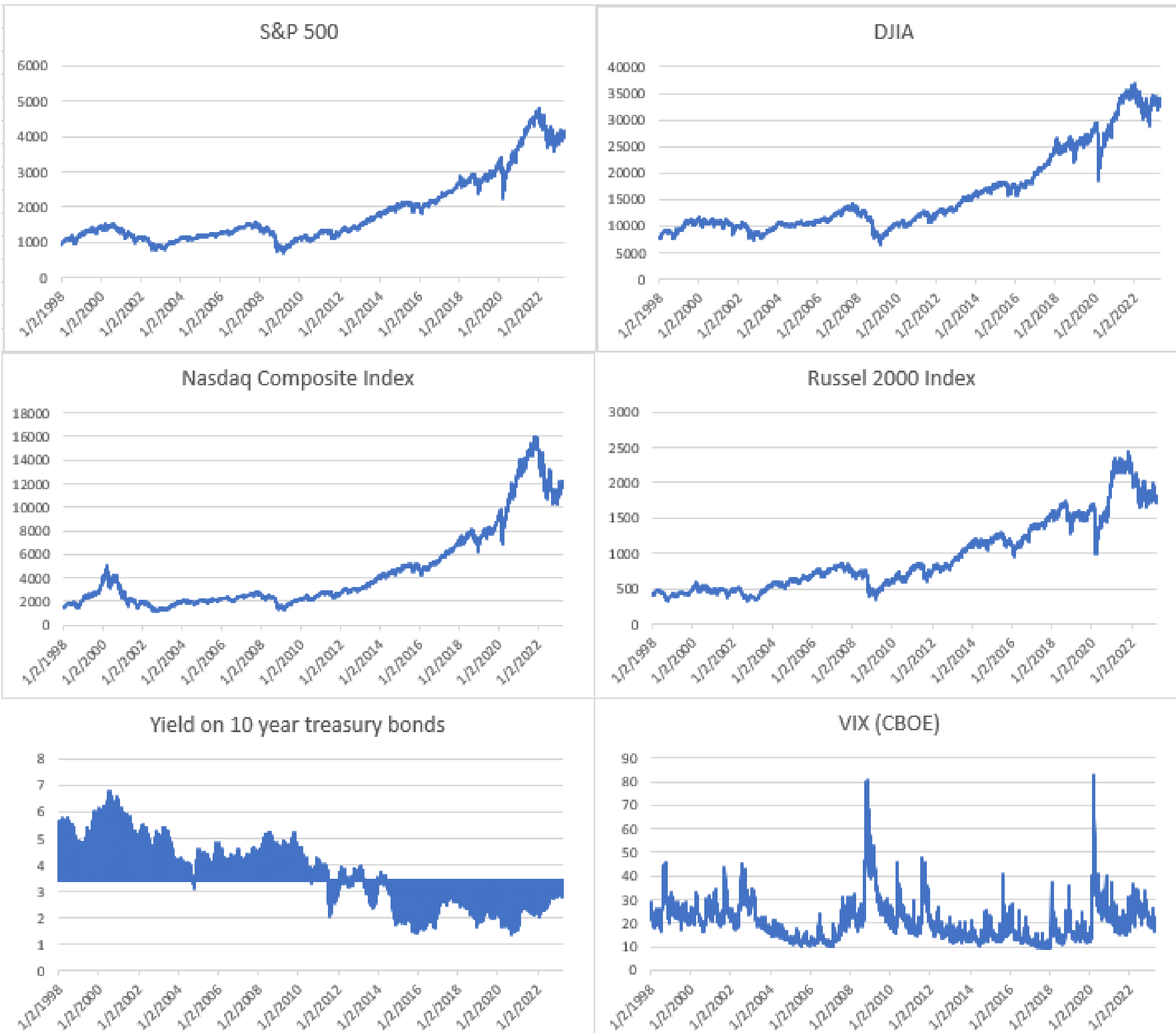
Significance of the Study

The significance of the study clearly lies on the quintessential nature of econometric models and their extensive usage. Showing that complex models fail and cannot predict the complex nature of our financial and economic world is crucial. In a peripheral level it may look like these economic shocks our uncommon, however historic literature would powerfully and strongly reject this point, we will be able to clearly conclude that financial failures are recurring events and occur more frequently than not and have wide ranging adverse impacts. Time and again, the world at a larger level and minutely speaking, economic models have failed to either accurately predict and forecast or to flag and regulate the highly convoluted economic system under which we operate. These failures have led to the ruin of livelihoods of millions, if not billions of people around the world. Thus, highlighting the importance of our study, which aims to underline and highlight the uncanny reliance on untrustworthy models trying to predict variables of the world riddled in uncountable unknowns and innumerable “shocks”, adding to the sea of literature showcasing the forecast failures of econometric models, however, uniquely presenting this problem as detailed further in the research gap section.

Analysis and Interpretation

Descriptive Statistical Analysis

Analyzing the trend



An upward trend is observed for stocks in general shown through the trend of S&P 500, DJIA, Nasdaq Composite Index and Russel 2000 Index. Further, the VIX (CBOE) representing market's expectations for volatility over the coming 30 days shows a spike during the 2007-08 Financial Crisis and COVID-19 pandemic periods.

Summary Statistics

S&P 500		DJIA		Nasdaq Composite Index	
Mean	1846.25	Mean	15959.1	Mean	4513.99
Standard Error	12.1379	Standard Error	97.6117	Standard Error	44.1377
Median	1401.11	Median	12408.6	Median	2801.61
Mode	1130.2	Mode	10655.2	Mode	1858.24
Standard Deviation	968.447	Standard Deviation	7788.17	Standard Deviation	3521.62
Sample Variance	937889	Sample Variance	6.1E+07	Sample Variance	1.2E+07
Kurtosis	0.75465	Kurtosis	0.05739	Kurtosis	1.40144
Skewness	1.32205	Skewness	1.13148	Skewness	1.52726
Range	4120.03	Range	30252.6	Range	14943.3
Minimum	676.53	Minimum	6547.05	Minimum	1114.11
Maximum	4796.56	Maximum	36799.7	Maximum	16057.4
Sum	1.2E+07	Sum	1E+08	Sum	2.9E+07
Count	6366	Count	6366	Count	6366

Russel 2000 Index		Yield on 10 year treasury bonds		VIX (CBOE)	
Mean	957.131	Mean	3.630534402	Mean	20.5688
Standard Error	6.39591	Standard Error	0.01498094	Standard Error	0.10688
Median	778.02	Median	3.4	Median	19
Mode	462.99	Mode	3.4	Mode	13.42
Standard Deviation	510.312	Standard Deviation	1.195287542	Standard Deviation	8.52778
Sample Variance	260418	Sample Variance	1.428712309	Sample Variance	72.7229
Kurtosis	-0.14047	Kurtosis	-0.632547885	Kurtosis	7.14056
Skewness	0.90614	Skewness	0.273421026	Skewness	2.00437
Range	2132.46	Range	5.415	Range	73.55
Minimum	310.28	Minimum	1.366	Minimum	9.14
Maximum	2442.74	Maximum	6.781	Maximum	82.69
Sum	6093094	Sum	23111.982	Sum	130941
Count	6366	Count	6366	Count	6366

Apart from Yield on 10-year treasure bonds, all the other variables are positively skewed as can be interpreted from the Skewness of each being greater than 0.5. Negative kurtosis shows that the distribution has lighter tails than the normal distribution which is true for Russel 2000 Index and Yield on 10-year treasury bond. Further, positive values of kurtosis indicate that distribution is peaked and possess thick tails, which is true in the case of VIX (CBOE), S&P 500, DJIA and Nasdaq Composite Index.

Analyzing the properties of the data

Heteroskedasticity

A regression model was first built for the data. Heteroskedasticity for this model was then checked using the Breusch-Pagan Test. The p-value obtained for the BP Test was less than 2.2e-08. Thus, we reject the Null Hypothesis of BP Test and conclude that heteroskedasticity is present in the model. To correct this, we transformed the data into logarithmic form. After the Log-Transformation, we ran the BP test again on the transformed data. Now, the p-value was 0.0047. Here the p value is still less than 0.05. Hence, we say that heteroscedasticity is still present in the model even after performing logarithmic transformation. But after the logarithmically transforming the data, we have minimized the heteroscedastic nature of the data and thus will have less impact on the model. Thus, we can proceed to our parameter estimation.

Multicollinearity

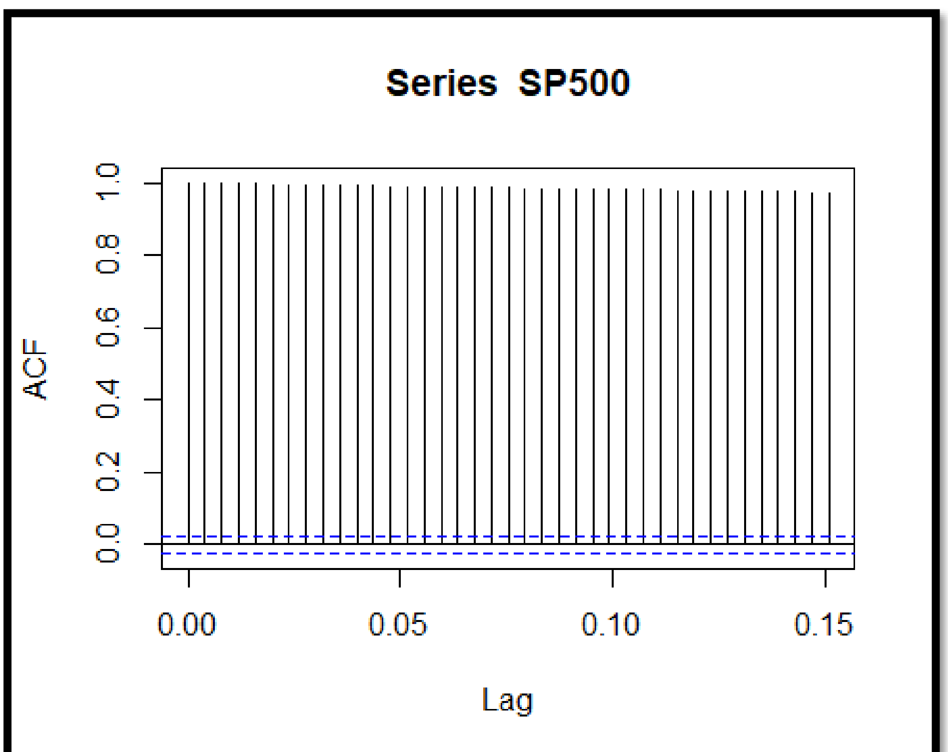
Now, for checking for multicollinearity in the data, we used the Variance Inflation Factor (VIF) value. VIF for the various variables of the transformed data was as follows: DJIA: 47.700, Russel 2000 Index: 24.053, Nasdaq Composite Index: 22.332, Yield on 10-year treasury bonds: 2.402095 and VIX (CBOE): 1.465. Using VIF, we can say that DJIA, Composite Index and Russel 2000 Index are the main sources of multicollinearity since VIF is above 10. However, given that the data are all stock market indices, and how the data are correlated to each other, the data is bound to have high multicollinearity. Moreover, in most cases than not, with economic data, it is not possible to drop variables. Further, literature on treating Multicollinearity, particularly for economic data shows that multicollinearity can be ignored. Thus, we go ahead with our analysis, ignoring multicollinearity.

Autocorrelation

Next, we check for autocorrelation present in the data. We use the Durbin-Watson test which gives a p-value of < 2.2e-16 which is less than 0.05 which is the significance level. Thus, we reject the null hypothesis and confirm that there exists autocorrelation. The DW statistic is 0.02 which tells us that there is positive autocorrelation. Given that autocorrelation implies non-stationarity of the data, we plan to use ARIMA model to estimate parameters, knowing that ARIMA handles non-stationary data and converts it into stationary data as required for the analysis.

Stationarity

Using ACF Plots and the ADF Test



The dotted lines are threshold values and the lags that lie within these lines are negligible. Since all lags lie above the threshold this means that all the lags are significant and hence there exists very high autocorrelation. Thus, the data is not stationary and needs to be converted to into a stationary format to make it forecastable.

ADF test

```
adf.test(SP500)

##
##  Augmented Dickey-Fuller Test
##
## data:  SP500
## Dickey-Fuller = -1.692, Lag order = 18, p-value = 0.7088
## alternative hypothesis: stationary
```

ADF test was used to verify the conclusions made from the ACF plot. Here since the p value is greater than 0.05, we accept the null hypothesis. Thus, it can be concluded that the data is not stationary.

Treating Non-Stationarity using the method of Differencing

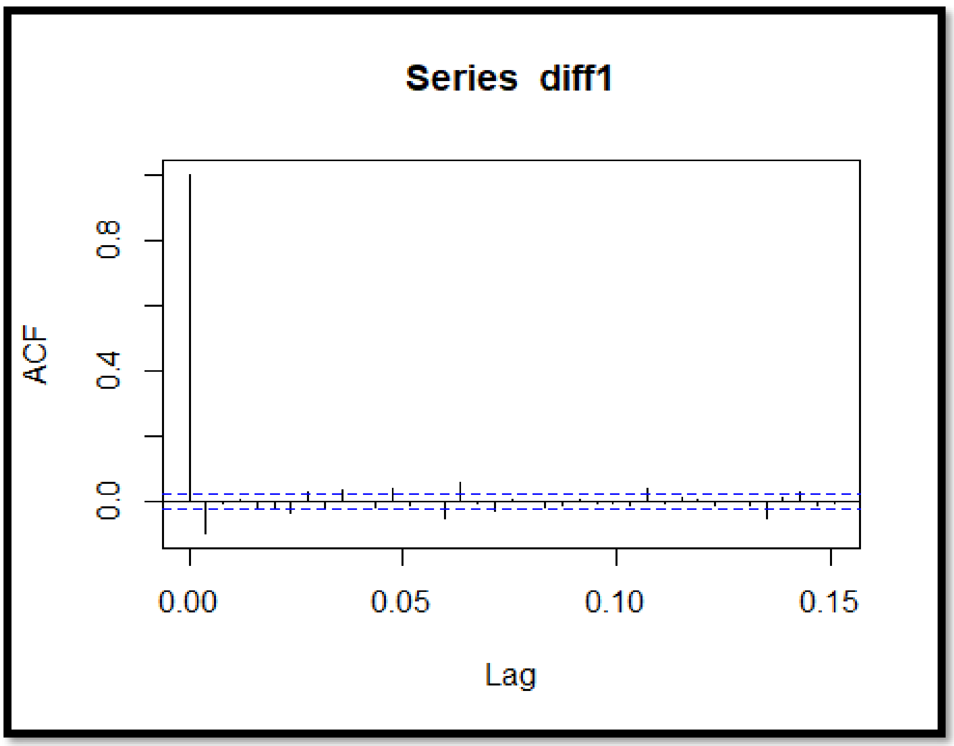
From the time series plot it was clear that the data had an increasing trend component. Thus, differencing was performed to eliminate the trend component and make the data stationary.

```
adf.test(diff1)

## Warning in adf.test(diff1): p-value smaller than printed p-value

##
##  Augmented Dickey-Fuller Test
```

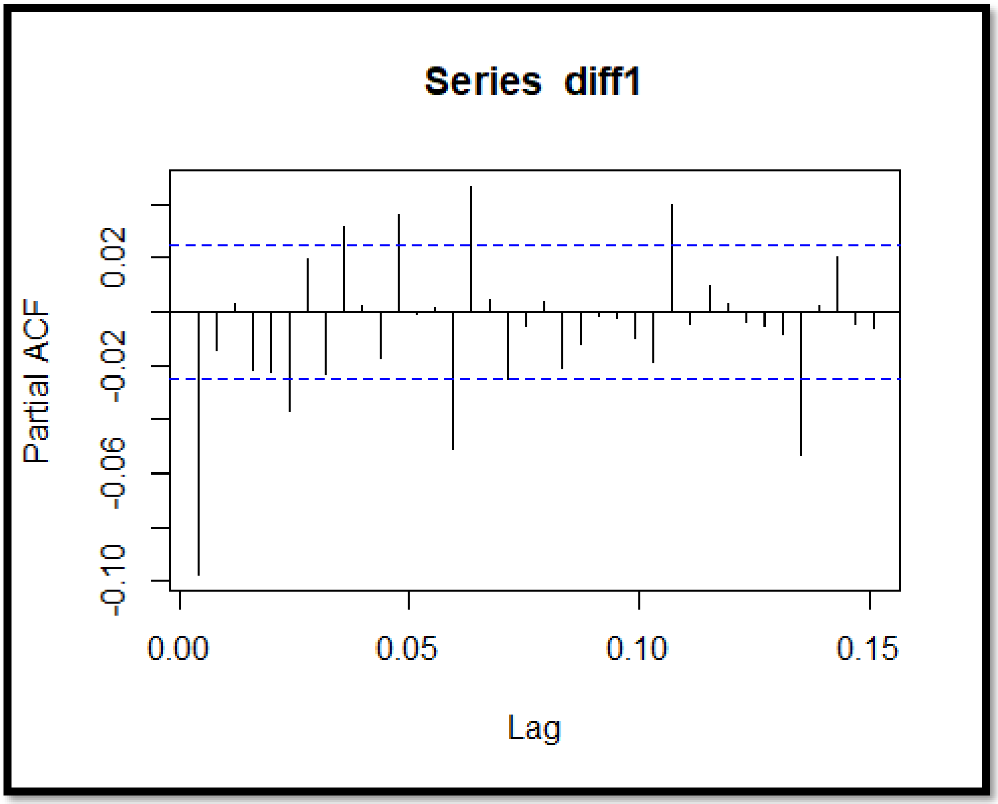
```
##
## data:  diff1
## Dickey-Fuller = -18.692, Lag order = 18, p-value = 0.01
## alternative hypothesis: stationary
```

The p value of the adf test after 1st differencing is less than 0.05 so we reject H0 and hence the data is stationary is now. Thus, it can be concluded that the data is stationary after 1st differencing. From this we can also conclude that d=1 for the ARIMA model.

Choosing the ARIMA Model



From the ACF plot constructed after differencing it can be observed that it cuts off after lag 1 which is significant. Thus, the order of MA is 1, ie MA(1) should be used. Now using Partial ACF plot, we have the following:



From the PACF plot constructed after differencing it can be observed that it cuts off after the 1st lag which is significant. So, the order of AR is 1, i.e., AR (1) should be used.

Thus, using the order of differencing and the number of significant lags in ACF and PACF plots respectively, the parameters of the ARIMA model to estimated are p = 1, d = 1 and q = 1, ie ARIMA(1,1,1) model should be constructed.

Building the ARIMA Models

CASE 1: The 2007-08 Global Financial Crisis

```
#MODEL 1: 1998 - 2008
##
## Call:
## arima(x = CIA3[1:2682, 8], order = c(1, 1, 1), xreg = CIA3[1:2682, 9:13
])
##
## Coefficients:
```

```
        ar1      ma1  log(DJIA)  log(Comp)  log(Russel)  log(Yield)  log(V
IX)
      0.3679  -0.3118     0.6606     0.2036       0.0243      -1e-04  -0.0
165
s.e.  0.2554   0.2609     0.0067     0.0048       0.0069       2e-04   0.0
011

 sigma^2 estimated as 9.433e-07:  log likelihood = 14793.79, aic = -29571.
59
```

The ARIMA model is fitted with the order (1, 1, 1) with the exogenous variables using the data between the time period from January 1998 to August 2008. The econometric model is generated using the coefficients obtained:

$$\mathbf{Y_t = 0.3679 * (Y_{t-1}) - 0.3118*(e_{t-1}) + 0.6606*(X1_t) + 0.2036*(X2_t) + 0.0243*(X3_t) - 0.0001*(X4_t) - 0.0165*(X5_t) + e_t}$$

Where,

Y_t: represents the log dependent variable at time t.

Y_{t-1}: represents the log dependent variable at time t-1.

X_t: represents the log independent variable at time t.

e_t: represents the error term at time t.

- 0.3679 = the AR coefficient of order 1.
- -0.3118 = the MA coefficient of order 1.
- 0.6606 = the slope of log (DJIA).
- 0.2036 = the slope of log (Composite Index).
- 0.0243 = the slope of log (Russel).
- 0.00 = the slope of log (Yield of Treasury Bonds).
- 0.0165 = the slope of log (VIX).

Forecast of S&P500 Index during Normal Economic Condition

```
forecast_d = predict(fitted_model, newxreg = CIA3[2683:2692,9:13])
forecast_d

## $pred
## Time Series:
## Start = 02-09-2008
## End = 15-09-2008
## Frequency = 1
##  [1] 3.106368 3.106426 3.093677 3.094689 3.102882 3.092234 3.094378 3.0
99785
##  [9] 3.099252 3.081085
```

The values for 10 days during normal economic conditions from 02/09/2008 to 15/09/2008 are forecasted using the fitted ARIMA model and the log(S&P500) values are obtained. These values are converted to the Index values and are compared to the actual values. The difference between the Actual values and the predicted values are calculated.

DATE	log(S&P500)	Predicted	Actual	Difference
02-09-2008	3.106368	1277.5209	1277.58	0.059140713
03-09-2008	3.106426	1277.6915	1274.98	-2.711483509
04-09-2008	3.093677	1240.7292	1236.83	-3.899190652
05-09-2008	3.094689	1243.6237	1242.31	-1.313728945
08-09-2008	3.102882	1267.3075	1267.79	0.482514757
09-09-2008	3.092234	1236.6135	1224.51	-12.10354743
10-09-2008	3.094378	1242.7335	1232.04	-10.69348386
11-09-2008	3.099785	1258.3023	1249.05	-9.252327717

12-09-2008	3.099252	1256.759	1251.7	-5.058988383
15-09-2008	3.081085	1205.2718	1192.7	-12.57181277

Forecast of S&P500 Index during Economic Shock (2007-08 Financial Crisis)

```
## $pred
## Time Series:
## Start = 29-09-2008
## End = 10-10-2008
## Frequency = 1
##  [1] 3.054600 3.073970 3.072296 3.057406 3.051454 3.035643 3.014438 3.0
07159
##  [9] 2.978626 2.974386
```

The values for 10 days during the Financial Crisis from 29/09/2008 to 10/10/2008 are forecasted using the fitted ARIMA model and the log(S&P500) values are obtained. These values are converted to the Index values and are compared to the actual values. The difference between the Actual values and the predicted values are calculated.

DATE	log(S&P500)	Predicted	Actual	Difference
29-09-2008	3.0546	1133.965913	1106.42	-27.54591335
30-09-2008	3.07397	1185.686841	1166.36	-19.32684099
01-10-2008	3.072296	1181.125375	1161.06	-20.06537532
02-10-2008	3.057406	1141.316248	1114.28	-27.03624797
03-10-2008	3.051454	1125.781221	1099.23	-26.55122106
06-10-2008	3.035643	1085.532924	1056.89	-28.64292386
07-10-2008	3.014438	1033.803504	996.23	-37.57350437

08-10-2008	3.007159	1016.620821	984.94	-31.68082084
09-10-2008	2.978626	951.9760008	909.92	-42.05600078
10-10-2008	2.974386	942.7271181	899.22	-43.50711811

T-Test

The paired T-test is used to check if there is a statistical significance in the means of the predictions in both the periods. The output of the t-test shows:

-Test: Paired Two Sample for Means

	Economic Shocks	*Normal Condition*
Mean	-30.39859666	-5.706290779
Variance	69.72654649	25.44403923
Observations	10	10
Pearson Correlation	0.576144727	
Hypothesized Mean Difference	0	
df	9	
t Stat	-11.43410649	
P(T<=t) one-tail	5.80346E-07	
t Critical one-tail	2.821437925	
P(T<=t) two-tail	1.16069E-06	
t Critical two-tail	3.249835542	

The p - value obtained is $5.803*10^{-07}$ which is less than 0.05. Hence, we reject the null hypothesis and conclude that there is a statistically significant difference in the means of the predictions. This showcases that the econometric model fails to provide the accurate predictions during economics shocks.

```
#MODEL 2: 1998 - 2020
##
## Call:
## arima(x = CIA3[1:5556, 8], order = c(1, 1, 1), xreg = CIA3[1:5556, 9:13
])
##
## Coefficients:
          ar1     ma1  log(djia)  log(comp)  log(russel)  log(yield)  logv
ix
      -0.2906  0.3213     0.6908     0.1930       0.0804       0e+00   -7e
-03
 s.e.   0.2192  0.2177     0.0048     0.0037       0.0042       1e-04   6e
-04

 sigma^2 estimated as 7.758e-07:  log likelihood = 31195.38,  aic = -62374
.76
```

The ARIMA model is fitted with the order (1, 1, 1) with the exogenous variables using the data between the time period from January 1998 to January 2020. The econometric model is generated using the coefficients obtained:

$$\mathbf{Y_t = -0.2906 * (Y_{t-1}) + 0.3213*(e_{t-1}) + 0.6908*(X1_t) + 0.930*(X2_t) + 0.0804*(X3_t) + 0.0*(X4_t) - 0.007*(X5_t) + e_t}$$

Where,

Y_t: represents the log dependent variable at time t.

Y_{t-1}: represents the log dependent variable at time t-1.

X_t: represents the log independent variable at time t.

e_t: represents the error term at time t.

- -0.2906 = the AR coefficient of order 1.
- 0.3213 = the MA coefficient of order 1.
- 0.6908 = the slope of log (DJIA).
- 0.930 = the slope of log (Composite Index).
- 0.0804 = the slope of log (Russel).
- = the slope of log (Yield of Treasury Bonds).
- 0.007 = the slope of log (VIX).

Forecast of S&P500 Index during Normal Economic Condition

```
## $pred
## Time Series:
## Start = 03-02-2020
## End = 14-02-2020
## Frequency = 1
##  [1] 3.511811 3.518693 3.524744 3.526156 3.522324 3.525373 3.525638 3.5
29745
##  [9] 3.528322 3.528205
```

The values for 10 days during normal economic conditions from 03/02/2020 to 14/02/2020 are forecasted using the fitted ARIMA model and the log(S&P500) values are obtained. These values are converted to the Index values and are compared to the actual values. The difference between the Actual values and the predicted values are calculated.

DATE	log(S&P500)	Predicted	Actual	Difference
03-02-2020	3.511811	3249.4585	3248.92	-0.538538848
04-02-2020	3.518693	3301.3609	3297.59	-3.77087438
05-02-2020	3.524744	3347.6805	3334.69	-12.99048016

06-02-2020	3.526156	3358.5823	3345.78	-12.80233956
07-02-2020	3.522324	3329.0782	3327.71	-1.368224164
10-02-2020	3.525373	3352.5325	3352.09	-0.442525737
11-02-2020	3.525638	3354.5788	3357.75	3.171184793
12-02-2020	3.529745	3386.4526	3379.45	-7.002590738
13-02-2020	3.528322	3375.3748	3373.94	-1.434771276
14-02-2020	3.528205	3374.4656	3380.16	5.694440499

Forecast of S&P500 Index during Economic Shock (COVID 19 Pandemic)

```
## $pred
## Time Series:
## Start = 24-02-2020
## End = 06-03-2020
## Frequency = 1
##  [1] 3.505574 3.492079 3.490450 3.470615 3.465859 3.485977 3.473419 3.4
91338
##  [9] 3.475901 3.470484
```

The values for 10 days during the start of COVID 19 Pandemic from 24/02/2020 to 06/03/2020 are forecasted using the fitted ARIMA model and the log(S&P500) values are obtained. These values are converted to the Index values and are compared to the actual values. The difference between the Actual values and the predicted values are calculated.

DATE	log(S&P500)	Predicted	Actual	Difference
24-02-2020	3.505574	3203.125833	3225.89	22.7641671
25-02-2020	3.492079	3105.124372	3128.21	23.08562801

26-02-2020	3.49045	3093.499143	3116.39	22.8908573
27-02-2020	3.470615	2955.391364	2978.76	23.36863551
28-02-2020	3.465859	2923.203163	2954.22	31.01683652
02-03-2020	3.485977	3061.801278	3090.23	28.42872194
03-03-2020	3.473419	2974.534429	3003.37	28.83557147
04-03-2020	3.491338	3099.830878	3130.12	30.2891223
05-03-2020	3.475901	2991.58261	3023.94	32.3573901
06-03-2020	3.470484	2954.500039	2972.37	17.86996133

The paired T-test is used to check if there is a statistical significance in the means of the predictions in both the periods. The output of the t-test shows:

t-Test: Paired Two Sample for Means

	Economic Shocks	*Normal Condition*
Mean	26.09068916	-3.148471956
Variance	22.16567702	38.18655683
Observations	10	10
Pearson Correlation	0.020108951	
Hypothesized Mean Difference	0	
df	9	
t Stat	12.01903282	
P(T<=t) one-tail	3.79831E-07	
t Critical one-tail	2.821437925	

P(T<=t) two-tail	7.59663E-07	
t Critical two-tail	3.249835542	

The p - value obtained is $3.983*10^{-07}$ which is less than 0.05. Hence, we reject the null hypothesis and conclude that there is a statistically significant difference in the means of the predictions. This showcases that the econometric model fails to provide the accurate predictions during economics shocks.

Policy Implication

This study has crucial real-world implications. In its truest sense, the paper showcases how unpredictable our complex financial world is. Forecast failures during economic shocks highlight the limitations of econometric models in capturing extreme events and uncertainties. Policymakers may need to enhance their risk management strategies by incorporating more conservative assumptions and stress-testing their policies against a wider range of scenarios, including worst-case scenarios. This may involve building more robustness into policy frameworks, such as maintaining larger policy buffers, establishing contingency plans, and being prepared to respond flexibly to changing economic conditions. Further forecast failures during economic shocks may indicate the need to improve the calibration of econometric models to better account for extreme events and tail risks. Policymakers may need to review and update the assumptions, parameters, and data used in their models to ensure that they are relevant and accurate, especially during periods of heightened uncertainty or crisis. This may involve incorporating more real-time data, improving model validation techniques, and conducting regular model reviews and updates.

Moreover, relying solely on econometric models for economic forecasting can be risky during periods of economic shocks. Policymakers may need to diversify their forecasting tools and approaches to include other sources of information and expertise, such as scenario analysis, qualitative assessments, and expert opinions. This can provide a more comprehensive and holistic view of the economic outlook, especially when econometric models may struggle to capture the complexity and dynamics of rapidly changing economic conditions. Also, to ensure that public confidence is not undermined in economic policymaking, policymakers would need to be transparent about the limitations of their models, acknowledge forecast errors, and communicate clearly and effectively with the public and other stakeholders about the uncertainties and risks associated with economic forecasts. This can help manage expectations, reduce uncertainty, and build trust in the policy process.

Lastly, forecast failures during economic shocks can provide valuable lessons for policymakers to learn from and improve their policy frameworks. Policymakers may need to conduct post-mortem analyses of forecast errors to understand the root causes and identify areas for improvement. This may involve revisiting the underlying assumptions, methodologies, and data used in econometric models, as well as evaluating the effectiveness of policy responses. Learning from forecast failures can help policymakers better prepare for future economic shocks and enhance the resilience of their policy frameworks.

Summary and Conclusion

In conclusion, this paper has examined the forecast failures of econometric models during economic shocks, focusing on the 2007-08 Global Financial Crisis and the COVID-19 pandemic as two unique case studies. Through the construction of ARIMA models and a comprehensive exploratory data analysis, the paper has demonstrated that statistically significant forecast failures occur when predicting S&P 500 values based on other variables during economic shocks compared to normal economic circumstances. To summarize, the paper constructed two ARIMA models for two different time periods, one from January 1998 to August 2008 and the other from January 1998 to January 2020. These models were used to predict S&P 500 values based on six other variables. The paper then presented statistical evidence, using T-Test, to demonstrate a significant difference between the predicted and actual values during normal economic conditions and economic shocks. This analysis was conducted after conducting a comprehensive Exploratory Data Analysis that included Descriptive Statistical Analysis and testing for properties such as Multicollinearity, Autocorrelation, and Heteroskedasticity, with appropriate data treatment. Additionally, Autocorrelation Plots and Augmented Dickey-Fuller Test were utilized to assess the Stationarity of the data.

The paper emphasizes the real-life applicability and impact of the problem under consideration and discusses the policy implications of forecast failures of econometric models during economic shocks. Overall, this research adds to the existing literature on forecast failures of econometric models and underscores the need for continued research and improvement in modeling techniques to enhance their accuracy during times of economic shocks.

References

Alessi, L., Ghysels, E., Onorante, L., Peach, R., & Potter, S. (2014). Central Bank Macroeconomic Forecasting During the Global Financial Crisis: The European Central Bank and Federal Reserve Bank of New York Experiences. Journal of Business &Amp; Economic Statistics, 32(4), 483–500.

Clements, M. P., & Hendry, D. F. (2001). Economic forecasting: Some lessons from recent research. *European Central Bank Working Paper Series*, (82).

Demyanyk, Y., & Hasan, I. (2010). Financial crises and bank failures: A review of prediction methods. Omega, 38(5), 315-324.

Griliches, Z. (1984). Data problems in econometrics.

Hausman, J. (2001). Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. Journal of Economic Perspectives, 15(4), 57–67.

Hendry, D. F. (2012). Mathematical models and economic forecasting: Some uses and mis-uses of mathematics in economics. In Probabilities, laws, and structures (pp. 319-335). Dordrecht: Springer Netherlands.

Hendry, D. F., & Doornik, J. A. (1997). The implications for econometric modelling of forecast failure. *Scottish Journal of Political Economy*, *44*(4), 437-461.

Hendry, D. F., & Mizon, G. E. (2014). Unpredictability in economic analysis, econometric modeling and forecasting. *Journal of Econometrics*, *182*(1), 186-195.

Hendry, D. F. (2020). A short history of macro-econometric modelling. Journal of Banking, Finance and Sustainable Development, 1(1), 1–32.

Nymoen, R. (2004). A recent forecast failure. *Notat tilgjengelig på http://folk. uio. no/rnymoen/index. htm OMMENT AR*.

RAMSEY, J. B., & KMENTA, J. (1980). Problems and Issues in Evaluating Econometric Models. Evaluation of Econometric Models, 1–11.

Song, H., Smeral, E., Gang, L., & Chen, J. L. (2008). Tourism forecasting: Accuracy of alternative econometric model revisited. In *30th International Symposium on Forecasting*.