*Journal of International Financial Trends*
**Vol. 2** (2026), No. 1, pp. 33-45
DOI: https://doi.org/10.55578/jift.2601.003

ARTICLE

# Data Science vs Econometrics: Testing Different Forecasting Approaches

Colin Ellis [1,*]

[1] *Hult International Business School, London, United Kingdom*

[*]Corresponding author. Email: colin.ellis@hult.faculty.edu

*Received: 22 December 2025, Accepted: 23 January 2026, Published: 26 January 2026*

**Abstract**

Forecasting is a critical activity for economists, financial analysts, and businesses engaged in budgeting and planning. Recent advances in Data Science promise greater forecasting sophistication and accessibility, often through automated or semi-automated tools. However, these techniques also carry risks, particularly when users are unaware of the underlying data-generating processes. This paper compares popular Data Science forecasting approaches with long-standing Econometric methods using simulated data under varying assumptions about noise and structural complexity. The results show that while Data Science methods can perform competitively in highly noisy environments, structural econometric models consistently outperform them when data noise is low or when underlying relationships are complex. This reinforces a long-standing insight often missed by automatic approaches: the source and structure of shocks matter for accurate forecasting.

**Keywords:** Data Science; Econometrics; Forecast comparisons

## 1. INTRODUCTION

In recent years, "Data Science" has become an increasingly popular and important field in financial analysis, academia, business and policymaking. At its heart, this field encompasses a variety of statistical techniques to interpret and uncover insights in often large quantities of data. One key application of this is in the area of forecasting, where statistical patterns are identified using models that then predict future values of data or time series. In a business context, this can include forecasts of revenue and profitability that guide management decisions on investment, hiring and other factors. The rise of such Data Science techniques - often largely automated and with attractive user interfaces (UIs) - has been notable as computing power has increased.

Recent work in financial machine learning demonstrates that modern data-driven methods - Large Language Models (LLMs), deep neural networks, and reinforcement learning - can materially improve prediction, allocation, and execution decisions across financial tasks. Fine-tuned LLMs applied to financial newsflow generate return forecasts that outperform sentiment-based baselines, with aggregated token representations and decoder-only models delivering the strongest portfolio enhancements [1]. In related work, deep portfolio-optimisation frameworks can achieve higher cumulative returns, improved Sharpe ratios, and lower drawdowns by jointly modelling temporal structure, cross-asset correlations, transaction costs, and risk [2]. In credit risk, workflow-based machine learning combining Weight-of-Evidence encoding, ensemble learning, and multi-objective optimisation improves both predictive accuracy and profitability, with deep ensembles consistently outperforming traditional statistical and tree-based models [3]. Reinforcement-learning approaches have provided insights into dynamic trading behaviour: for instance, modelling limit order management as a Markov decision process shows that queue position, surrounding depth, volatility, and tick-size constraints jointly determine order value, and that the option to cancel accounts for roughly a fifth of a limit order's expected value [4]. There are reasons to be cautious: a review of explainable AI in finance highlights the sector's dependence on opaque models and the dominance of SHAP (SHapeley Additive exPlanations),

https://doi.org/10.55578/jift.2601.003

feature-importance, and rule-based techniques, while underscoring persistent challenges around transparency, regulatory compliance, and methodological consistency [5]. But collectively, these studies show that new Data Science methods can substantially enhance financial decision-making, while also highlighting the growing importance of explainability and risk-aware optimisation.

These new techniques, and others, therefore offer greater sophistication for analysts. But forecasting share prices, revenues or some other financial data is nothing new. Economists have been forecasting developments at individual businesses - or economies as a whole - for decades. Econometrics is the particular branch of statistics that focuses on this activity, alongside similar attempts to unpick patterns in data and test how they correspond to economic theory.

The purpose of this paper is, in a deliberately limited context, to compare the forecasting performance of common Data Science and Econometric methods using simulated data. The key advantage of this approach is that it allows us as testers to know exactly what the underlying properties of the data are; however, the modelling approaches that we are testing are blind to the underlying assumptions in the simulations. As such, this allows us to test and compare the forecasting ability of different models and different approaches. In turn, this then offers guidance as to when Data Science may be more appropriate for forecasting purposes, and if and when Econometrics may be more accurate.

At its heart, this experiment addresses a very old question at the heart of economic analysis. Suppose an analyst is forecasting profitability for a company that has some dependence on oil - either a beneficial or costly link. A natural question they may ask could be: what is the impact of a $10 movement in the price of oil? To a pure data scientist, this question would be approached by examining data on oil prices, company profitability, and other variables: and while it could be expressed within some range of uncertainty, there would be a central estimate. To an economist, in order to answer this question we must first understand why the oil price moved by $10: the source of the shock matters. This is a critical point noted by both policymakers [6] and researchers [7].

The rest of this paper is set out as follows. The next section offers a brief discussion of Data Science and Econometrics and their differences, together with some common forecasting approaches that we will test. Section 3 discusses how simulations can be used to test different models and forecasts, and focuses on two sets of simulations: one where the data structure relatively simple or well-behaved; and one where it is more complex, and arguably realistic. Section 4 discusses results from these different sets of simulations, focusing on a simple metric of relative forecasting performance across the Data Science and Econometric approaches. Finally, Section 5 concludes.

## 2. Exploring different statistical approaches

Statistics is the specific branch of mathematics that deals with the collection, analysis, interpretation, presentation, and organization of data. It is a powerful tool used in various fields, notably finance and economics, but also medicine, engineering, and more. The study of statistics is essential for making decisions that are informed by data, and more broadly for understanding how the world around us works.

At the core of statistical analysis is the process of summarizing and analyzing often large amounts of data, and trying to identify significant patterns and trends. For instance, in public health statistics are used to track the spread of diseases, evaluate the effectiveness of treatments, and inform policy decisions – this was crucial during the recent COVID-19 pandemic. In business, companies use statistical analysis to understand market trends, customer behavior, and to make strategic decisions about investment, hiring and other long-term bets.

Broadly, a key distinction can be drawn between descriptive statistics and so-called inferential statistics. The former focuses on summarizing data so that it can be easily understood by non-experts. This often means focusing on central tendencies in statistical distributions (e.g. means or modes) alongside indicators of variation (such as range and standard deviation). The main function here is to present data effectively, and build understanding.

In contrast, inferential statistics focus on making predictions about as-yet unobserved outcomes. This can include extrapolating to ‘population’ outcomes based on data samples, but also forecasting movements in revenues, profitability and stock prices. Typically, such inferential analysis is presented alongside estimated measures of uncertainty, for instance indicating the degree of uncertainty around a forecast. It is this aspect in particular – forecasting future outcomes – that this paper will focus on. This helpfully allows us to abstract from very real challenges in other areas of statistics, such as data cleaning and collection.

Both Data Science and Econometrics offer a range of different approaches to forecasting future outcomes. In recent years, Data Science techniques have gained prominence and popularity, offering forecasters new techniques. The fundamental question that this paper seeks to address is simple: does Data Science or Econometrics produce better forecasts?

In order to address this, we start by exploring each broad field, and some common applications, in turn.

### 2.1. Data Science: two common applications

To practitioners, Data Science is known as an interdisciplinary field that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data. It involves using scientific methods, processes, algorithms, and systems to analyze structured and unstructured data. The goal of Data Science is to uncover hidden patterns, correlations, and trends that can inform decision-making and drive strategic planning.

At its core, Data Science is about transforming raw data into actionable insights. This process typically involves several stages, including data collection, data cleaning, data analysis, and data visualization. Data Scientists use a variety of tools and techniques to perform these tasks, including programming languages (like Python and R), statistical software, and machine learning algorithms.

In recent years, Data Science has experienced a meteoric rise in prominence. This can be attributed to several factors. Two that go hand in hand are the rapid increase in the quantum of data generated each day; and advancements in technology that allow for the highly efficient analysis of large data sets, including data processing frameworks. As part of this, advancements in machine learning and artificial intelligence are often cited. Neither of these is a new concept, in the sense of having existed for several decades; but the current power of these approaches has advanced to the point where computers and algorithms can now make predictions without human interaction.

There are many different Data Science techniques, including simple regressions such as Ordinary Least Squares (OLS), hierarchical clustering, natural language processing, and A/B testing. Ongoing advances in machine learning and artificial intelligence have already made it easier for humans to use these techniques, often with automatic selection of key parameters or other model choices. For the purposes of this paper, we will focus on two specific forecasting techniques that have gained prominence in recent years. These are the so-called LASSO approach, and General-to-Specific (or GETS) modelling.

#### *2.1.1. A primer on LASSO*

LASSO, which stands for Least Absolute Shrinkage and Selection Operator, is a form of regression analysis. The key difference, compared with OLS, is that LASSO performs both variable selection and regularization to enhance the prediction accuracy and interpretability of statistical models. The approach is based on initial work by Tibshirani [8], with further development by Zou *et al* [9] and others.

The core principle of LASSO is to adjust the normal (OLS) regression process by incorporating a “penalty” term that constrains the sum of the absolute values of the model parameters. This penalty is controlled by a so-called “regularization parameter”, which can be assigned or (semi)automated.

One of the key benefits of LASSO is its ability to automatically select variables. By shrinking some coefficients in the model to zero, LASSO effectively excludes those variables from the model. This enhances model simplicity, making it easier to interpret, and is particularly useful for large datasets that have a large number of potential predictors.

As a consequence, LASSO is also beneficial in addressing multicollinearity. Multicollinearity occurs when predictor variables are highly correlated – this can make it difficult for statistical approaches to accurately estimate the (relative) size of different coefficients. By penalizing the absolute size of the coefficients, LASSO reduces the variance of the estimates, which can lead to more stable and reliable models.

One downside of LASSO is that it can often lead to very small models, for example only selecting one variable from a group of correlated regressors. The choice of the regularization parameter is crucial, although algorithms can select that automatically if required.

In summary, LASSO is a powerful tool for creating parsimonious models that balance complexity and predictive performance. This makes it a useful forecasting tool.

#### *2.1.2. A primer on GETS*

The General-to-Specific (GETS) approach in statistical modeling is a systematic method used to simplify complex models while retaining their core characteristics.

The process starts with large, unrestricted models – an initial forecasting model that includes many forecasting variables that may or may not be relevant. The GETS process then automatically tests for any mis-specifications in the model, to check it is a good representation of the data; and then starts to automatically exclude variables that are found to be insignificant, step by step. The aim of GETS is to arrive at a parsimonious model that accurately corresponds to the data but is also small and tractable for forecasting and interpretation. Once a final model has been obtained, many GETS processes then also perform validation, for example by testing the stability of the estimated model parameters across sub-samples.

GETS processes are nothing new in and of themselves: Hendry [10] offers a useful history here. But, as with other Data Science techniques, recent advances in computing power and coding have shifted the process from a manual one, where human inspection was required to remove insignificant variables, to a now wholly-automated process where the final model – and its forecast outputs – are delivered to users without any human intervention. As such, they are an increasingly popular tool for forecasters.

In practice, LASSO and GETs are similar approaches, statistically speaking. They both involve a step-wise process, where in each step a variable is examined to ascertain whether it should be added to (or removed from) the list of regressors. A key difference is that LASSO shrinks estimated coefficients towards zero in a data-dependent way, while GETS does not. But depending on parameterization, it may be the case that LASSO and GETS yield similar results.

### 2.2. Econometrics: two common approaches

As noted earlier, there are many commonalities between Data Science and Econometrics, the statistical branch of Financial and Economic forecasting. Financial forecasters often use regression analysis, GETS and even LASSO models in their work. But the key difference is that, further to that analysis, econometrics also seeks to test or uncover underlying economic relationships from the data.

The key difference here is between what Econometricians often call “reduced form” and “structural” analysis. Reduced-form analysis imposes no structure on the data - in essence, it is very similar to Data Science in “letting the data speak”. The focus is on uncovering patterns in the data that can be used to build understanding and make predictions. One challenge here is that reduced form analysis does not in any way uncover genuinely causal relationships: and it is often easy to mistake correlation for causality, particularly if users of such reduced-form analysis are relatively inexperienced.

In contrast, structural econometric models aim to uncover the underlying economic mechanisms that sit beneath the data. This implies incorporating economic theory and assumptions to specify some aspects of how different economic agents interact, which makes these approaches more complex than reduced form models: they require a degree of insight or judgement from an economist. However, that structure - if not mis-specified - then allows for a deeper insight into the causal relationships between data series, and is more useful for estimating the impact of economic factors, such as demand shocks or policy changes.

For the purposes of this paper, we will focus on two different econometric approaches in order to compare and contrast against the Data Science approaches detailed earlier. The first is based on a so-called structural VAR, or SVAR; and the second is based on cointegrated factor demand equations.

### *2.2.1. Structural VARs*

At its heart, a vector autoregression or VAR is a simple, parsimonious model that summarizes statistical relationships between different variables. In finance and economics, VARs are often used to estimate relationships between key economic variables such as GDP growth, inflation and interest rates. We will refer to such a simple three-variable model in the illustration below.

The first output from a VAR analysis will be a reduced-form model – as with the Data Science approaches described earlier, there is no structural interpretation that lets us establish or infer causality. So economists then need to impose some form of structure on the model in order to allow such a causal interpretation. The popularization and insight of imposing structure owes much to the work of Christopher Sims, as Christiano [11] details.

There are many different ways to do this. A popular approach is the so-called Cholesky decomposition. This essentially assumes that, when a shock hits the economy, it affects different variables in the model at different times. For the simple three-variable model outlined above, a common assumption is that a monetary policy shock – for example a rise (or cut) in policy interest rates that was not expected – will affect interest rates first, then GDP growth, and finally inflation. This simple difference in the timing of the impacts is enough to transform the estimated (reduced-form) residuals from the model into structural economic shocks.

However, the Cholesky method has often been criticized, and there are other approaches to structural identification as well. One simple one is the imposition of sign restrictions, which can either be short- or long-term. In the three-variable model, we could impose that an unexpected rise in interest rates (a monetary tightening) would initially raise interest rates, decrease GDP growth, and decrease inflation. But crucially all three could happen at the same time. Several papers detail the process and outcomes from sign-restriction VARs, including Uhlig [12] and Ellis *et al* [13].

Sign restrictions also allow the reduced-form model residuals to be transformed into economic shocks – and then the different impact of those shocks can be examined. To return to the question asked at the start of this paper, if a demand shock leads to a $10 increase in the price of oil, its impact on oil prices – and the broader economy – will be different than if a supply shock drove the $10 increase. Understanding which structural shocks are playing out is therefore crucial not only for interpreting the data, but for forecasting as well.

### *2.2.2. Cointegrated structural relationships*

Before starting any form of modelling, it is important to first understand how the data behave. One crucial aspect here is stationarity. A data series is stationary if its mean and variance do not change over time; or put simply, that data are well behaved.

The reason stationarity matters so much is that many data series – including company revenues, profits and even employment – will typically be non-stationary, even if the company is relatively stagnant. As such, any analysis that is not designed to adjust for this statistical property has the potential to be misleading.

One particular technique in econometrics exploits non-stationarity in data to uncover structural economic relationships. A particular example that lends itself to this is the estimation of factor demand equations. These are long-term relationships that describe how wages, prices, output and employment evolve (for labour demand) or how prices, output, investment and the price of investment goods evolve (for capital demand). The focus here is on uncovering the equilibrium relationship will become evident once short-term deviations from equilibrium have dissipated – or in other words, the equilibrium that the economy will tend towards. Ellis & Price [14] describe this in a UK context for the capital side of corporate behaviour; and Ellis [15] extends it to cover both labour- and capital-demand relationships.

The standard modelling approach here – known as a Vector Error-Correction Model, or VECM – therefore incorporates not just the long-term economic relationships, but also information about how the economy moves back towards those equilibria, while being (in principle) relatively agnostic about the short-term dynamics that are present in the economy. It therefore offers a very powerful means of testing for underlying structural relationships in economic data, and the degree to which those influence economic dynamics.

A key element of cointegrated approaches is that they tie movements and differences in levels together over time. Arguably, this is both more realistic and applicable to real-world examples than simple dynamic models that focus only on changes – Chief Financial Officers care about the level of revenues, not just the growth rate – but it is also more complex, from a forecasting perspective. Very simple models of non-stationary data series will be mis-specified and mis-leading almost by construction: unless forecasting models account for the potentially explosive nature of the variance of non-stationary data series, they will fail. That makes this form of data and modeling another good one to test.

## 3. Generating data & testing forecasting approaches

In order to test the forecasting performance of the different Data Science and Econometric approaches described above, we need to find some data to use. One option here is to use economic or financial data, either from companies or economies, to construct models, use them to forecast and then compare outcomes.

The challenge with this approach is that, ultimately, we will never know which model is “right”. Without certainty about the underlying relationships that exist between different data, it would be possible for a model to look credible, and perform well over a short sample, but ultimately still be misleading.

The alternative is to generate simulated data against which we can test different forecasting approaches. Critically, the underlying data generating processes (DGPs) used here are based on assumptions; but those DGPs are also ‘invisible’ to the models that will be used for forecasting purposes. Instead, these models can only see data similar to those available to economists or business forecasters. There are several ways to do this in practice; but for this analysis, we first generated the underlying relationships, and then layered on ‘noise’ that distorted the signals that the models are able to see. This means that we can monitor both ‘true’ underlying relationships, and the noisy data that are available to uncover them; and, importantly, we can vary the degree of ‘noise’ and see what sort of impact it has on forecasting performance.

It is important to emphasize that, from the perspective of the hypothetical modeler, many ‘truths’ are unknown, including the underlying relationships in the data, and the degree of noise applied: the ‘observed’ data is ultimately a combination of these two elements. Essentially, the application of different statistical models is a mechanism for trying to uncover the differences between this noise and the true relationships.

Following the structure described earlier, we applied two separate DGPs. The first was based on a simple structure where two variables – assumed to be changes in selling price and changes in quantity sold – were driven by underlying structural shocks to demand and supply. In standard economic fashion, a positive demand shock raised both price and quantity; and a positive supply shock would raise quantity but lower price.

Formally, this first, simple DGP can be specified as:

$$\Delta p_t = \alpha D_t - \beta S_t + \mu_{p,t}$$

$$\Delta q_t = \gamma D_t + \delta S_t + \mu_{q,t}$$

Where $p$ denotes price, $q$ denotes quantity, $D$ denotes demand and $S$ denotes supply. All four coefficients ($\alpha, \beta, \gamma, \delta$) are positive and greater than zero, and $\mu_{i,t}$ denotes random noise (Gaussian shocks) in price and quantity changes. The demand and supply variables themselves can be thought of as innovations from period to period. Importantly, while this is the underlying DGP, the models we will estimate only observe changes in price and quantity.

This means that, when estimating models that would be used to create forecasts, the underlying demand and supply shocks were not visible. Instead, the models had to capture this as best they could given their nature and structure, where differences in approach across the different models reflected the differences between the Data Science and Econometric approaches described earlier.

The second DGP was more complex, relating the level of different variables to each other. In particular, instead of looking at changes in prices and changes in quantities, the underlying structural relationship was between the level of prices and the level of quantities. Furthermore, any deviations from this structural relationship could have an impact on future changes in prices and quantities, as the system moved back to equilibrium. Technically, this means that different stochastic processes are related to one another, and have a tendency to adjust after shocks. As with the simpler DGP described above, the underlying structure here again had 'noise' layered on top of it, so that the data that were visible for forecasting purposes were not the same as the underlying structural relationship.

Formally, this second and more complex DGP can be specified around the long-run relationship between the two key variables, *x* and *y*:

$$x_t = \rho + \vartheta y_t$$

The dynamic form of the DGP is then:

$$\Delta x_t = \sum_{1}^{j} \tau_j \, \Delta x_{t-j} + \sum_{1}^{j} \varphi_j \, \Delta y_{t-j} + \sum_{1}^{j} \omega_j \, \Delta z_{t-j} + \emptyset_x (x_{t-1} - \rho - \vartheta y_{t-1}) + \varepsilon_{x,t}$$

$$\Delta y_t = \sum_{1}^{k} \tau_k \, \Delta x_{t-k} + \sum_{1}^{k} \varphi_k \, \Delta y_{t-k} + \sum_{1}^{k} \omega_k \, \Delta z_{t-k} + \emptyset_y (x_{t-1} - \rho - \vartheta y_{t-1}) + \varepsilon_{y,t}$$

where z represents a vector of other variables affecting the dynamic path of the two key variables (but not the long-run relationship), and the $\varepsilon$ terms are again Gaussian noise. Typically, the estimated lag structure across the two equations would be the same ($j=k$), but that does not have to be the case.

Armed with these simulated data, we can then estimate models on *part* of the time series up to a cut off point: use those estimated models to construct forecasts for observations beyond that cut off point; and then compare those forecasts against the actual outcomes in the simulated data. This approach is a formal "out of sample" test of forecasts, which is critical for testing the value of predictions and predictive models. In-sample fit, which models are often designed to maximize by default, is no guarantee of out-of-sample forecast performance.

Given the slightly adversarial nature of this exercise – testing Data Science techniques against Econometric ones – it is also helpful to have a simple summary statistic that captures the key results. The key metric we will use here is the 'multiple' – how accurate Data Science forecasts are relative to Econometric ones. This approach has previously been used for comparing stress tests [16].

Formally, the "multiple" here is calculated as the standard deviation of the forecast error from a Data Science forecasting approach, divided by the standard deviation of the forecast error from an Econometric approach. If Data Science is more accurate than Econometrics, this multiple will be below 1; if it is less accurate, the number will be higher than one. While more detailed statistics and comparisons are also available, the key advantage of using this multiple is that it offers a simple and intuitive interpretation.

## 4. Simulation results from simple and complex data simulations

This section reports results from the two sets of simulations described earlier. The first used a simple DGP that linked changes in prices to changes in quantities. The second used a more complex DGP, arguably more prevalent in real-world applications, where the relationship was between levels of the variables, and deviations from equilibrium could affect future changes.

### 4.1. Results from the simple DGP

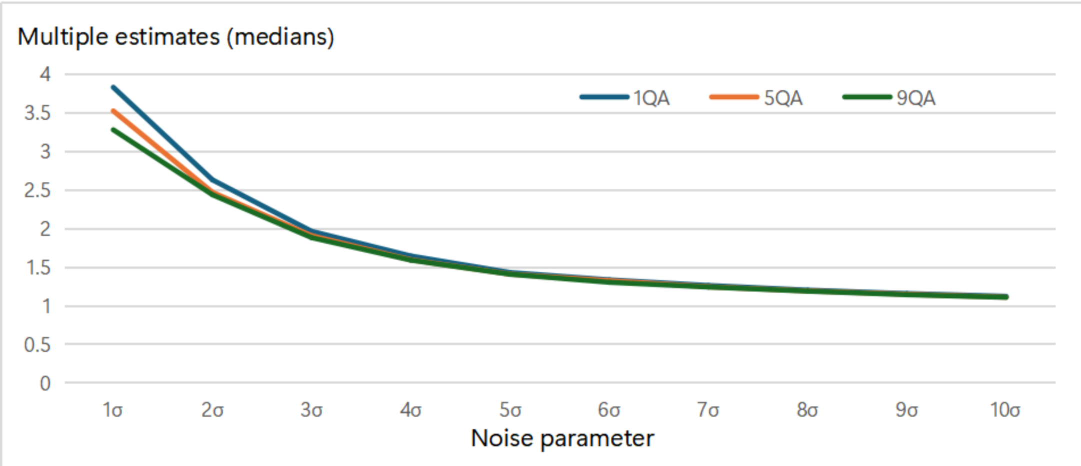
To start with, we focused on simulations that were based on the simple underlying data structure. This simple DGP assumed that changes in prices and quantities were related to unobserved demand shocks and supply shocks, and some 'noise' that we introduced, which could vary across simulations. For simplicity, the noise terms were assumed to be Gaussian – zero mean and with fixed variance within any given set of simulations (the variance could vary across simulation sets, specified in units of σ, the lowest calibration). All calibrations of DGPs in this paper were based on empirical results from quarterly data: so one period is equal to one quarter in the simulations. Based on the 'observed' price and quantity data made available to the hypothetical modeler, we then built Data Science and Econometric models to forecast future changes, and compared the results.

To do this, 5,000 simulations were run: creating data, estimating models, and testing forecasting performance each time for a new batch of data. These were run across an initial spread of 25 different parameterizations for the 'noise' terms, but this was cut to 10 after initial work indicated faster-than-expected convergence in results.

The forecasts were constructed and compared one quarter ahead (1QA) of the data available to the model; but also five periods ahead (5QA) and nine periods ahead(9QA), consistent with the quarterly data frequency that economists and financial analysts often use. The initial results compared the LASSO Data Science approach with the SVAR Econometric approach, where the structural parameterization was also unknown and had to be estimated in the model: only the functional form of the short-term restrictions was specified.

Exhibit A presents outcomes from these simulations, in particular the average (mean) estimate of the multiple comparing forecast performance from the Data Science and Econometric approaches. The vertical axis is the multiple, and the horizontal axis represents the degree of 'noise' in the data that was used to mask the true underlying relationships in the data.

**Exhibit A: Relative forecast performance for a simple DGP**




Source: Author's calculations.

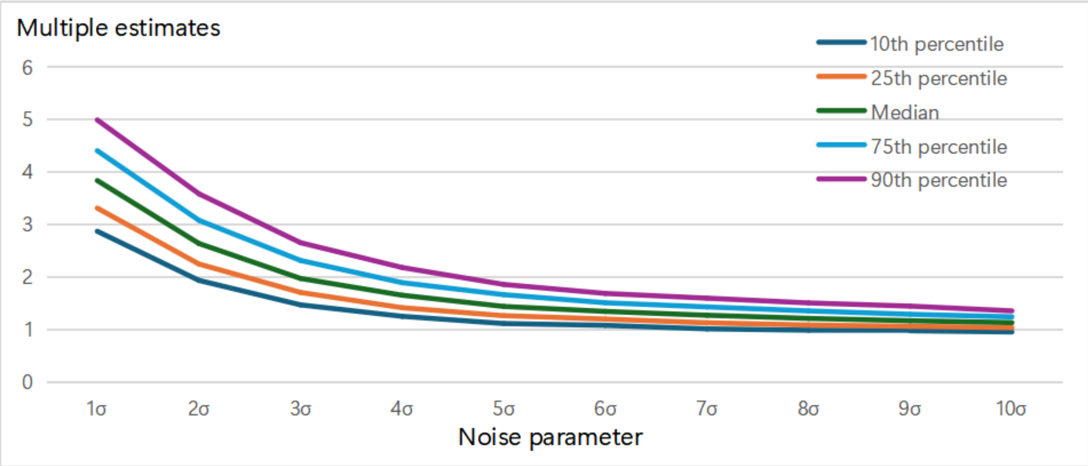
Two key results jump out very clearly. Firstly, although there are three lines on this chart, they all almost on top of one another. This means that the *relative* forecast performance – Data Science versus Econometrics – is similar across forecast horizons (one, five and nine quarters ahead). Unsurprisingly, the individual forecast performance of both approaches deteriorated the further into the future the forecast was – it is much easier to forecast next month than what will happen in two years' time. But interestingly, in this relatively simple framework, the relative performance of the two approaches deteriorated at the same pace.

The second key result is that, in this testing framework, Data Science forecast performance is at best only on a par with Econometrics; and then only when the 'noise' in the data is high. Where noise is low, the forecasting performance of the SVAR approach is significantly better than the LASSO approach – the multiple is 1.5 or even higher. And when noise is high, the central estimate of the multiple converges towards 1 – indicating similar forecasting performance – but does not fall below it.

To clarify, there were individual simulations where a LASSO forecast was better than the SVAR forecast; but that is not surprising given the randomness that we have injected into the simulated data. The starting point for judging the simulations results is the central tendencies; but we can also inspect the broader distribution of results.

Exhibit B presents a combination of these central and distributional outcomes from the 50,000 simulations, this time for the 1-period ahead forecasts. Exhibit B demonstrates that the central tendencies shown in Exhibit A are a decent guide to the overall distribution; the different percentiles either side of the median converge at a similar pace and in a similar manner. Interestingly, however, they are not symmetric even at high noise values. The 90th percentile at the highest noise calibration is 1.34, indicating DS forecasting performance was 34% worse than E; but the 10th percentile was 0.94, indicating it was only 6% better in that instance.

**Exhibit B: Distributional results for one-quarter-ahead multiples, simple DGP**




Source: Author's calculations.

Importantly, the broad results shown above were robust to various forms of sensitivity analysis. Most obviously, this included using a different Data Science framework – the GETS approach – instead of LASSO. Perhaps this is unsurprising, given the similarities between LASSO and GETS. But it is striking that, at best, the Data Science approaches only match the forecast performance of the Econometric approach at high noise levels; and are worse with less noise.

**4.2. Simulation results from a more complex (realistic) DGP**

The second group of simulations were based on a more complex underlying data structure, based on relationships between levels of non-stationary variables, akin to the VECM framework described in Section 2.2.2.

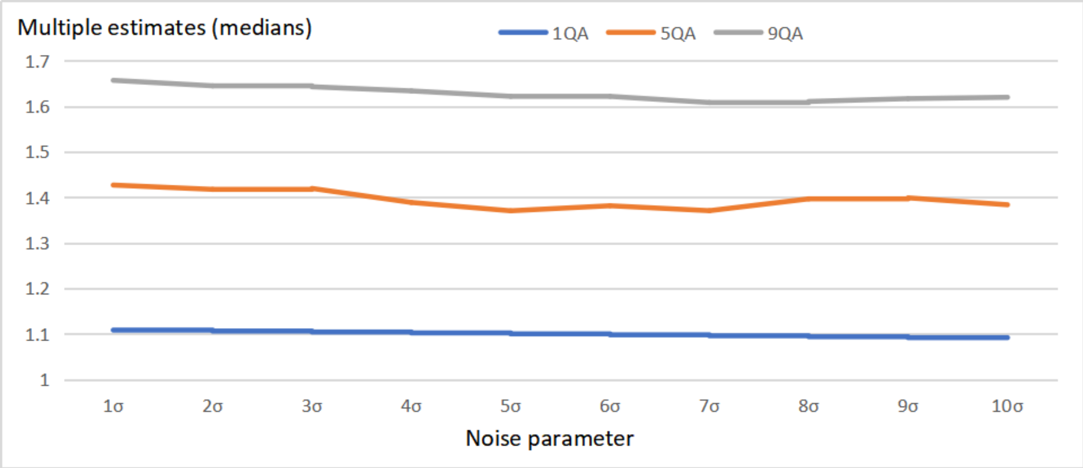
As in the previous section, two variables were constructed that were cointegrated in the underlying data, and with a degree of convergence when deviations from that cointegrating relationship were observed. Once again Gaussian noise was added to the underlying data, so that the 'observed' data made available to the assumed modeler conflated this relationship and the noise terms; and then Data Science and Econometric models were built to forecast future data, and results were compared.

One difference with this approach was that we gave both sets of models – Data Science and Econometrics – less information about the underlying structure of the relationship. In the simulations presented earlier, the complete observed time series were available, and the models were estimated based on the simulated data that directly related to the underlying relationship.

In this simulation, the approach was different, relating to the underlying DGP being specified in levels, rather than changes. For reduced-form approaches, the difference should be negligible, because – together with a constant – a time series (or history) of changes can always be accumulated to get the exact evolution of the level of different variables. But for the structural Econometric approach, the functional form of the long run relationship was not given to the hypothetical modeler – instead, it had to be estimated as part of the process. This essentially means that the Econometric model here is a "quasi-VECM" approach, as we are not specifying or estimating the structural relationship first; this is the normal approach, for instance as discussed in Ellis & Price (2004). Basically, we are making life harder for the structural Econometric approach – we are applying only the most limited form of analytical judgement – where the Data Science approaches were already agnostic.

Once again 5,000 simulations were run across ten different calibrations for the variance of the 'noise' terms, increasing in size. A plot of the resulting mean multiples is shown in Exhibit C.

**Exhibit C: Relative forecast performance for a more complex DGP**




Source: Author's calculations

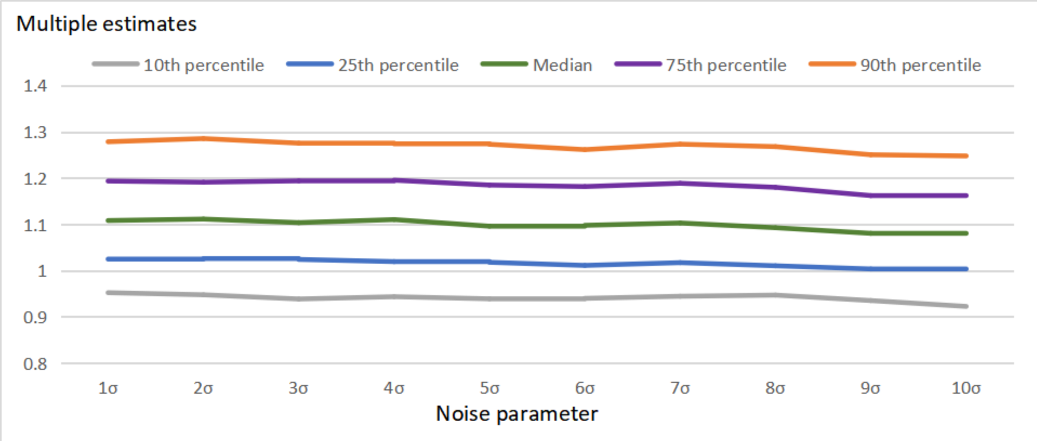
The results from this more complex DGP are significantly different from those in the simple approach. An initial glance could suggest that forecast accuracy is higher here, given the lower multiples in Exhibit C than in Exhibit A. But is it important to note that the multiples are ratios of different forecast errors – and actually, the absolute forecast errors (for the econometric approach) underpinning Exhibit A are smaller than those in Exhibit C. It is just that the ratio of errors (Data Science compared with Econometric) is lower.

There are two substantive points of note from the chart. First, there is a very obvious difference in relative forecast performance at different forecast horizons. The Lasso approach is closest to the quasi-VECM approach when forecasting one quarter ahead; but performs significantly worse five quarters ahead, and even worse again at nine quarters ahead. It is interesting that the deterioration in relative performance is bigger between one and five than between five and nine; this could be linked to the inherent uncertainty that all forecasting approaches face at longer horizons.

Second, there is also clearly different relationship between the degree of noise in the data, and the relative forecasting performance. Now, the degree of convergence in forecast accuracy between Lasso forecasts and the structural econometric (quasi-VECM) approach is very limited. Instead, the lines are broadly flat in Exhibit C as 'noise' in the data increases, compared with the significantly convex and being downward-sloping curves in Exhibit A. This suggests that greater complexity in the underlying relationships between data may be a significant hindrance to the LASSO forecasting method, particularly if it involves trying to interpret stationary noise against a backdrop of non-stationary data. Yet this is exactly what many businesses do when they try to forecast revenue and cost developments over time. Structural identification – an econometric approach, rather than a pure data-driven one – matters, even when noise is large.

One commonality is that the distributional results effects around the central tendencies shown in Exhibit C did correspond closely: Exhibit D presents results for one-quarter-ahead forecasts. Although the central tendencies behave differently, the distributional picture in Exhibit D is similar to that shown in Exhibit B.

**Exhibit D: Distributional results for one-quarter-ahead multiples, complex DGP**




Source: Author's calculations

As with the simpler simulations, once again the broad results here were robust to various forms of sensitivity analysis, including replacing the LASSO approach with a GETS modelling approach. As with the earlier results, that is likely to reflect the high degree of commonality between different Data Science forecasting methods.

**5. Discussion & conclusions**

Forecasting is important in many different fields, but particularly in finance and accounting. Being able to predict future revenues and profits with a degree of accuracy is critical input for business leaders to be able to manage, plan and direct resources.

Thanks to the continued expansion of computing power, and wider availability of different forecasting techniques, it is now relatively simple for non-technical specialists to build models and produce forecasts, including for complex financial metrics and processes. The range of options that fall under the "Data Science" description have particularly risen in recent years. This democratization of forecasting options and power is a welcome step.

But while Data Science techniques offer new tools to practitioners, many challenges, assumptions and even pitfalls can still lead to bad forecasting performance, under certain conditions. When those assumptions and risks are hidden behind a nice user interface, users may not even be aware they exist. One particular challenge is the issue of addressing causality, and uncovering structural relationships within the data – something that economists have long grappled with.

To illustrate these issues, this paper has used Data Science techniques to produce and test forecasts against approaches from the field of Econometrics, where analytical judgement is crucial to address these issues. Using simulated data – both from a simple process and from a more complex and realistic one – we test relative forecasting performance across these approaches.

It is important to remember that these are only two sets of simulations, albeit ones that try to replicate the real-world scenarios that financial and economic forecasters have to deal with. And the results are striking. In the simple simulations, the two approaches do yield similar forecasting power; but only when the 'noise' in the data is relatively high. When the underlying relationships are less uncertain, the structural model outperforms the reduced-form one. In the more complex simulations, the reduced-form approaches fare worse; even the limited analytical judgement applied to the structural approach yields significant and persistent gains in forecasting power, even when 'noise' in the data is high. At the moment, machine learning and artificial intelligence techniques are unable to apply judgement in this manner; and even the so-called 'causal AI' models are not typically focused on addressing the issue of identifying structural economic relationships.

For financial institutions, these findings have direct implications for the design and governance of forecasting frameworks. Automated or machine-learning-based approaches may be attractive because they scale easily and require limited analytical judgement. However, the simulations here suggest that such approaches may be materially less reliable when underlying relationships are complex or when the signal-to-noise ratio is relatively high. In applications such as stress testing, expected-loss modelling, and internal capital planning, structural econometric models may therefore offer more robust and interpretable forecasts, particularly when understanding the economic mechanism behind a projection is as important as the projection itself.

For regulators and supervisory authorities, the results reinforce the importance of model governance expectations that go beyond simple measures of in-sample fit or short-run predictive accuracy. If reduced-form or fully automated tools cannot reliably distinguish between shocks with different economic implications, then prudential oversight may be weakened. Requiring institutions to demonstrate an understanding of the structural drivers embedded in their models, to test robustness under alternative structural assumptions, and to articulate the economic narrative behind their forecasts is therefore a natural complement to technical model validation.

This is obviously good news for economists who are used to making such analytical judgment calls; they can help companies identify and understand the underlying factors driving changes in revenues and profits, and improve forecasts. But the broader conclusion from this analysis is that understanding *why* revenues or profits are changing is still important – or in other words, the source of the shock matters. Unfortunately, enthusiastic but uninformed users of Data Science forecasting techniques may be totally unaware of this proviso.

**Funding Statement:** This research and manuscript preparation were conducted without external funding or financial support.

**Author Contributions:** The author contributed to the research and writing of this article and has read/agreed to the published version of the manuscript.

**Data Availability Statement:** The main data used in this paper are simulated, with the details of the simulations provided within the paper itself.

**Conflict of Interest:** The author does not have any conflict of interest, hence declared no conflict of interest regarding this research work.

**Consent for Publication:** Not applicable.

**REFERENCES**

1. Guo, T., & Hauptmann, E. (2024). Fine-tuning large language models for stock return prediction using newsflow. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (Industry Track)*, pages 1028–1045.
2. Yan, R., Jin, J., & Han, K. (2024). Reinforcement learning for deep portfolio optimization. Electronic Research Archive, Vol. 32, No. 9, pages 5176–5200.
3. Rahmani, R., Parola, M., & Cimino, M. (2024). A machine learning workflow to address credit default prediction. arXiv. https://arxiv.org/abs/2403.03785
4. Kwan, A., & Philip, R. (2025). Reinforcement learning in a dynamic limit order market. *Working paper*, University of New South Wales and University of Sydney.
5. Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. *Artificial Intelligence Review*, Vol. 57, pages 216–260.
6. Broadbent, B. (2016), "The distributional implications of low structural interest rates and some remarks about monetary policy trade-offs", Speech given at the Society of Business Economists Annual Conference, 18 November. Available at: https://www.bankofengland.co.uk/-/media/boe/files/speech/2016/the-distributional-implications-of-low-structural-interest-rates.pdf
7. Tüzün, G. (2024), The importance of external shocks and global monetary conditions for a small-open economy: The case of Türkiye, *Central Bank Review*, Vol. 24, No. 3.

8. Tibshirani, R (1996), “Regression Shrinkage and Selection via the Lasso”, *Journal of the Royal Statistical Society, Series B (Methodological)* Vol. 58, No. 1, pages 267–88. http://www.jstor.org/stable/2346178.
9. Zou, H. and Hastie, T (2005), “Regularization and Variable Selection Via the Elastic Net”, *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, Vol. 67, No. 2 (April), pages 301–20. https://doi.org/10.1111/j.1467-9868.2005.00503.x
10. Hendry, D. F. (2024), “A Brief History of General-to-specific Modelling”, *Oxford Bulletin of Economics and Statistics*, Vol. 86, No. 1, pages 1-20. https://doi.org/10.1111/obes.12578
11. Christiano, L. J. (2012), “Christopher A. Sims and Vector Autoregressions”, Scandinavian Journal of Economics, Vol. 114, No. 4, pages 1082-1104.
12. Uhlig, H. (2005). “What are the effects of monetary policy on output? Results from an agnostic identification procedure”, *Journal of Monetary Economics*, Vol. 52, No. 2, pages 381–419.
13. Ellis, C., Mumtaz, H. and Zabczyk, P. (2014), “What lies beneath? A time-varying FAVAR model for the UK transmission mechanism”, *The Economic Journal*, Vol. 124, May, pages 668-699.
14. Ellis, C. and Price, S. (2004), “UK Business Investment and the User Cost of Capital”, *The Manchester School*, Vol. 72, pages 72-93. https://doi.org/10.1111/j.1467-9957.2004.00420.x
15. Ellis, C. (2006),”Elasticities, markups and technical progress: evidence from a state-space approach”, *Bank of England Working Paper* No. 300, July.
16. Ellis, C. (2017), “Scenario-Based Stress Tests: Are They Painful Enough?”, *Contemporary Economics*, Vol. 11, No. 2, pages 219-234.