Beyond Error Measures to the Utility And Cost of the Forecasts

Elizabeth Yardley and Fotios Petropoulos

Preview. The authors criticize the exclusive use of forecast-error metrics to evaluate forecast models, arguing for evaluations of *forecast value*, a construct that also incorporates considerations of the utility to the forecast user (effect on decision making), the cost of computation, and opportunity cost of understandability. They highlight evidence on forecast utility and cost from recent studies and explore the implications for forecast-method selection**.**

Key Points

- Traditional forecast-error metrics (based on the difference between forecast and actual values) fail to consider the cost of creating forecasts or the utility of the forecasts for decision making. As such, the common metrics are insufficient to evaluate the ultimate value of a forecasting method or the cost/benefit of an investment in a forecasting support system.
- We can extend traditional error metrics to account for *forecast utility* in decision making by incorporating operational information with the forecasts. Three examples are the constructs of *forecast value added, stochastic value added,* and *prediction intervals.*
- Forecasts of financial variables are those most often evaluated in economic terms, such as earnings from a portfolio of investments. What is apparent is that economic benefits are not necessarily correlated with forecast accuracy. Similarly, the accuracy of demand-forecasting methods is imperfectly correlated with inventory performance indicators.
- Few studies identify the costs associated with generating forecasts. Especially for cloud computing, such costs can be substantial. Additionally challenges in the understandability of a forecasting method provide impose an opportunity cost in the sense of failure to put the method into practice.

FORECAST VALUE AND FORECAST EVALUATION

There are dozens of metrics available to evaluate forecasting performance, such as MAD, MASE, and MAPE. But these common metrics consider just one aspect of performance – the error of the forecast compared to the actual that occurs. What they fail to consider is the cost of creating forecasts or the forecasts' usefulness in making better decisions. As such, the common metrics are insufficient to evaluate the ultimate *value* of a forecasting method.

Forecasting competitions illustrate this limitation. The M1-M5 forecasting competitions ranked forecasting methods based solely on forecast-error metrics. These rankings did not appraise the value of the forecasts in specific decision-making contexts, nor did they account for method computation costs.

In this paper we attempt to broaden the scope of forecast evaluation as a basis for forecast method selection. Early writings by Allan Murphy (1993) provide a start. Murphy defined forecast value in terms of the benefit and expense that forecasts generate for those who use them to make decisions. He emphasised that forecasts lack intrinsic worth and that it is only their role in decision-making that makes them valuable.

Murphy distinguishes three criteria for evaluation of "forecast goodness": consistency, quality and value. These are summarized in **Table 1**. Notice that he considers *accuracy* (how close are the forecasts to the actuals) to be only a part of forecast *quality* and quality to be only a part of forecast *value.*

Table 1. Types of Forecast Goodness (Murphy, 1993)

	Type 1: consistency	**Type 2: quality**	**Type 3: value**
Definition (p.283)	"Correspondence between forecasts and judgements"	"Correspondence between forecasts and observations"	"Incremental benefits of forecasts for users"
Elements	• By reflecting the uncertainty of forecasts accurately (by expressing the uncertainty and the correct degree of uncertainty). • By making sure that the spatial and temporal specificity matches the forecaster's best judgement.	• Bias • Association • **Accuracy** • Skill • Reliability • Resolution • Sharpness • Discrimination 1 • Discrimination 2 • Uncertainty	• The decision options available to the entity making the decision • The costs and benefits associated with the decision • The quality of the alternative decision-making information available • **Forecast quality**

The overarching purpose of forecasting is to deliver value and the evaluative question therefore is not "how accurate?" but rather, "how valuable?"

Determining forecast value is therefore a necessary precursor to forecast evaluation. We must address questions such as:

- How will the forecasts be used?
- How will forecasts impact the decision at hand?
- What are the expenses of forecast creation?

- What are the costs associated with forecast error?".

EXTENSIONS OF ERROR METRICS

We can extend traditional error metrics to account for *forecast utility* in decision making by incorporating operational information with the forecasts. Three examples are the constructs of *forecast value added, stochastic value added,* and *prediction intervals.*

Forecast Value Added

Mike Gilliland (2002, 2013)) proposed that each stage of a forecasting process be evaluated by analysing its *forecast value added,* FVA. Doing so can identify where resources are being wasted in the sense that money, time, and effort are not justified in terms of forecast improvement. In one case study he cited, a supplier applied its own cost of inaccuracy metric to the FVA analysis and concluded that certain refinements were not worth the resources committed.

Steve Morlidge (2014) used a relative absolute error RAE – an alternative form of the FVA - in combination with product volumes to identify best effort/reward ratios, such as high-volume products which exhibit large forecast errors. In this way, he suggested, forecast improvement efforts can be directed to where they have the greatest impact on the overall amount of error. (Note however that this criterion assumes that holding costs and service level penalties are equal across the product portfolio so that product volume is a proxy for profit contribution. This is which is not necessarily the case: with varying inventory costs across products, there may be greater commercial opportunity in improving forecasts of lower-volume products as well.)

Stochastic Value Added

Stefan de Kok (2017) criticised the use of traditional error metrics in inventory settings for failing to consider the broader probability distributions of possible demands, which provide a fuller accounting of the uncertainty being faced.

> *The way we traditionally measure forecast error prevents us from knowing how to drive improvement in forecast performance and hence reductions in obsolescence, fewer stock outs, and increased customer*

> *service levels. Rather we need forecast error metrics that measure more than the errors that occur on the average but address the full range of uncertainty.*

He proposes calculation of the *total percentile error*, a metric that views error not simply as a difference between an actual and forecast value but between and actual and forecast distributions. While Gilliland's FVA measures improvement (or deterioration) in an average error from application of a forecast method, the TPE offers an indicator of stochastic value added (SVA), which measures the gain or loss from a comparison of actual and forecasted distributions.

Prediction Intervals

The M5 competition also included a comparison of methods in terms of how well they estimated *uncertainty* in the forecasts (i.e. the width of prediction intervals). In practice, the width of prediction intervals serves as a proxy for inventory holding costs and provides valuable input for setting a target service level.

The M5 competition however did not link forecasts to specific decisions and so did not include an evaluation in terms of profit. Fittingly, in reference to forecast evaluation, Leitch and Tanner (1991) had earlier contended, "It is naturally better to examine profits directly than to examine a proxy that is at best indirectly related to profits" (p.580-581).

UTILITY EVALUATION

The methodologies presented within the remainder of this article seek to create a framework for assessing forecast utility for decision making and forecast value, inclusive of costs of forecast generation. We begin by summarising studies across various disciplines that have attempted to do so**.**

Unsurprisingly, forecasts of financial variables are those most often evaluated in economic terms, such as earnings from a portfolio of investments. What is apparent is that economic benefits are not necessarily correlated with forecast accuracy.

Financial markets

Cenesizoglu and Timmermann (2012) compared the statistical and economic performance of stock-return prediction models and found "dissonance" between the two. They assessed correlations between forecast error (RMSE) and economic performance, the latter measured in various ways such as the Certainty Equivalent Return, Sharpe Ratio and Risk Adjusted Return of a portfolio of stocks. In a majority of their cases, the forecasting models had better economic performance and concurrently worse forecast error than a benchmark. The authors concluded that statistical performance is a very weak predictor of economic performance.

Caldeira and colleagues (2016) evaluated combination methods in the context of interest rate forecasts used for portfolio decision-making. The study also used RMSE and the Sharpe Ratio for statistical and economic performance comparison and also concluded that correlations between RMSE and Sharpe Ratios were low.

To demonstrate a model's ability to forecast commodity prices, Dolatabadi and colleagues (2018) applied similar methods and assessed economic and statistical performance by implementing trading strategies based on the forecasts and calculated portfolio profits as well as Sharpe Ratios. While the study did not explicitly compare economic and statistical performance measures, as in the earlier two studies, the work employed both types of evaluation to support the superiority of the forecasting model under consideration and provides an additional example of how economic returns might be evaluated.

Leitch and Tanner (1991) found no systematic relationship between traditional error metrics, such as average absolute error (AAE) or root mean squared error (RMSE), and profits generated from trading decisions based on interest rate forecasts. Correlation analysis even showed that, for three of four trading rules, higher errors were associated with larger profits (p.586, Table 2). The authors suggest that this is noteworthy for forecast evaluation as well as for model estimation, given that models are often built by minimising errors.

Wing and Yi (1996) employed Leitch and Tanner's trading rules to evaluate the profitability of forecasts for the 3-month yen certificate of deposit. They analysed the statistical accuracy and profitability of using different types of forecasts: professional forecasts, the forward interest rate, an ARIMA model, and a naïve no-change model. They found that, on average, smaller forecasts errors were related to larger profits, except for the case of professional forecasts. Their findings suggest that the quality of *professional forecasts* should be judged against measures of profitability rather than traditional accuracy metrics.

Döpke and colleagues (2018) found disagreement between economic and statistical performance measures in an evaluation of business cycle forecasts, echoing Leitch and Tanner's findings as well as those of Caldeira and colleagues. They analysed risk-adjusted returns to evaluate economic performance and found that the performance ranking differed depending on whether statistical or economic measures of performance were used.

Using another similar method (by translating forecasts to financial market trading decisions), Kath and Ziel (2018) evaluate an electricity price forecasting model by comparing economic performance to that of benchmark portfolio based on the Sharpe ratio.

These financial studies point how forecasting models might be assessed monetarily, by employing set trading strategies on the financial markets. In general, they reveal disagreement between traditional error metrics and economic measures of performance and thus call into question the use of forecast error metrics to select forecasting methods in place of bottom line measures of profitability.

Supply Chain

Studies of demand forecasting for supply chains have attempted to account for key performance indicators such as inventory cost and service level, in addition to forecast error. John Boylan (2006) referred to such KPIs as *accuracy-implication metrics* and emphasized the distinction between forecast-accuracy metrics, which measure the errors resulting from a forecast method, and accuracy-implication metrics, which measure the achievement of the organization's stockholding and service-level goals. Both measurements he felt are important.

Peter Catt (2007) presented a Cost of Forecast Error (CFE) metric designed to improve decision-making by bridging the gap between error metrics and their cost implications. The CFE incorporates the costs of holding inventory and of poor service, enabling comparison of forecasting methods in economic terms and illustrating the cost impact of adjusting service levels.

Fotios Petropoulos and colleagues (2019) produced an evaluation of several forecasting methods based on their inventory performance, in order to supplement the forecast-error metrics reported in the M3 competition. They employed an inventory metric that accounts for holding costs, order variances, and service levels. While not directly translatable into economic performance, the result is a holistic measure of inventory performance. The study found some differences in the methods' rankings between the traditional and holistic measures; notably, that combination forecasting methods ranked better in terms of inventory performance relative to forecasting performance and that an approach based on multiple temporal aggregation levels performed (statistically significantly) better than all methods compared when considering the inventory metric.

Nada Sanders and Gregory Graman (2009) simulated a warehouse environment to investigate the cost of forecast errors in a unique service setting within the supply chain. Like Catt's CFE metric, they tie forecast errors to economic value by considering holding costs associated with errors. Their quantification also included the cost of labour and an alternative scenario incorporating a customer-imposed penalty for unprocessed inventory.

Betting

Fabian Wunderlich and Daniel Memmert (2020) found a "counterintuitive" relationship between forecast accuracy and profit in examining bets based on forecasts in betting markets. The authors recommended that both accuracy and profitability should be considered when fitting models and evaluating performance, but caution against equating the two, saying that "forecasting accurately and forecasting profitably are different tasks and should be treated as such" (p.719). They promote the findings as applicable across fields where betting odds exist, such as political elections, referendums, award-ceremonies and cultural events.

Meteorology

T.N. Palmer (2002) found limited interaction between those creating and those consuming meteorological forecasts, and believed this detachment to explain why models are often assessed separately using error metrics or economic performance. He proposed measuring the economic value of forecasts by considering a forecast user's cost/loss ratio, which reflects the cost of taking precautionary action and the potential losses associated with not taking action.

In what could be considered a cost/loss scenario, Florian Pappenberger and colleagues (2015) quantified the benefit of flood forecasting systems by calculating the cost savings in Europe associated with avoiding damages. This approach considered both forecast benefit and cost, which together, constitute the most fundamental appraisal of any investment.

SELF-FULFILLING PROPHECIES

A frequent by-product of forecasting is that the forecasts themselves influence the variable being forecast, changing decisions that affect the outcomes predicted. We often think of this phenomenon as a self-fulfilling prophecy.

A self-fulfilling prophecy can be triggered when forecasts are made widely available, a prime example of which is that the forecast of a recession alters behaviours in a way that weakens the economy. Similarly, in its reaction to a lower-than-expected sales forecast for a new month, a marketing team may implement a last-minute campaign to boost demand, ensuring that the original forecast is not realized. And forecasts made during a pandemic may be motivated mainly by changing the course of action. COVID-19 forecasts have been used to inform policies, such as lockdowns, that seek to limit infections and their serious human and financial consequences.

These forecasts are valuable precisely because they enable action to be taken to avoid the outcomes being forecast. John Ioannidis and colleagues (2020) examined the consequences of poor forecasts of the pandemic in terms of delays in treatment of other health conditions, patient avoidance of hospitals for necessary treatments and infection spikes within nursing homes.

Decisions based on faulty forecasts have measurable monetary costs; however, poor pandemic forecasts incur the far more significant cost of the loss of human life. So, the economic value of a human life should be part of a utility evaluation of pandemic forecasts. For example, in a study of the costs and benefits of UK lockdowns, Julian Jessop (2020) offers the concept of "the value of a prevented fatality" (p.139). However, the consequences of such actions may also have a deep emotional impact on those affected and, although difficult to quantify, these costs should not be overlooked when appraising the utility of pandemic forecasts.

Table 2 presents a summary of these studies on forecast utility.

Table 2. Indicative Studies on Forecast Utility

Reference	Domain	Monetary Evaluation	Study Purpose and Findings
Leitch and Tanner (1991)	Finance: interest rates	$	Compares statistical error measures and profits. Found only weak relationships between traditional error and profits from trading decisions.

Wing and Li (1996)	Finance: interest rates	$	Evaluates profit from forecasts for the 3-month yen certificates of deposit by employing the trading rules developed by Leitch and Tanner. Found that profits and errors from professional forecasts were not correlated.
Murphy (1993)	Meteorology		Defines the differences between forecast consistency, quality and value and explores their relationship under different conditions
Palmer (2002)	Meteorology	$	Proposes a connection between the potential economic value of forecasts and traditional error metrics leading to the author to recommend a generalised skill score to represent overall forecast value.
Catt (2007)	Inventory management	$	Defines Cost of Forecast Error metric and demonstrates how to perform a CFE calculation.
Sanders and Graman (2009)	Demand forecasting: service industry (warehouse)	$	Analyses impact of forecast errors on organisational costs such as inventory holding costs, labor costs, and penalties for unprocessed inventory,
Cenesizoglu and Timmermann (2012)	Stock returns	$	Compares statistical and economic performance measures for forecasting models. Found that models with larger forecast errors could still add economic value through trading decisions.
Pappenberger and colleagues. (2015)	Weather: flood forecasting	$	Measures the benefit of flood forecasts in monetary terms by comparing the cost of forecasts to the cost of damages avoided by acting on forecasts. Found that every 4 euros invested in the European Flood Awareness System saves 400 euros.
Caldeira and colleagues. (2016)	Interest rates	$	Considers both statistical and economic measures of model performance when forecasting Brazilian interest rate futures contracts using combination methods. Found a negative relationship between forecasting errors and economic performance.
Döpke and colleagues (2018)	Financial markets	$	Compares actively-managed vs passively managed portfolios. Found that forecast error metrics such as MAE were better for actively-managed portfolios economic returns were poorer.
Kath and Ziel (2018)	Energy	$	Proposes an approach for measuring the economic gains of an accurate model and demonstrates the model's effectiveness using trading decisions.
Petropoulos and colleagues. (2019)	Inventory management (M3 subset)		Extends evaluation of forecasting methods to include inventory performance and defined a holistic metric to combine inventory performance measures. Found differences when ranking methods by the inventory performance measure compared to the forecast error metric, MASE.

Wunderlich and Memmert (2020)	Betting	$	Distinguishes between accuracy and profitability from betting strategies when evaluating sports forecasts. Found that positive returns could be achieved without improved accuracy and therefore returns and accuracy should not be equated.
Ioannidis and colleagues (2020)	Healthcare	~	Summarises the adverse consequences of actions based on poor forecasts in the COVID-19 pandemic, including delayed treatment for other conditions, and hospital avoidance. Provides several reasons why forecasts might have failed.

COST EVALUATIONS

We could find few studies that identify the costs associated with generating forecasts. Especially for cloud computing, such costs can be substantial (Petropoulos and colleagues, 2021). Cost information enables you to make balanced assessments about the cost/benefit of a forecasting method and more considerably about whether a forecasting system is worth the investment.

Understandability can also be considered an element of cost. A forecast user who does not understand how the forecasts are made is unlikely to take ownership of them. So complexity in a model may not only increase computation cost but, by diminishing the likelihood that the forecasts will be used, increase the opportunity cost of forecast generation. Greater complexity can also increase the risk of miscommunication of forecast results.

Computational Cost

Some of the forecasting competitions included brief descriptions of the computation times associated with the applications of a forecasting method and offered general but non-quantitative judgments that cost increases with model complexity.

The M4 competition in 2020 Makridakis and colleagues (2020a) reported computational running time for most methods and indicated that the measure served as a proxy for model complexity. However, it did not count running time as a factor in the evaluation of the forecasting methods. Nor did the M5 competition, which referred only in general terms to the greater cost of ‘sophisticated’ and ‘complex’ methods. Makridakis and colleagues (2020b) recognized that the computational expense of ML methods could be large but believed that “the computational cost should not be prohibitively expensive to produce hundreds of thousands or even millions of forecasts on a weekly basis (Seaman, 2018, p.29)”.

Konstantinos Nikolopoulos and Fotios Petropoulos (2018) investigated the impact on forecasting performance of reducing computing time in parameter selection -- by limiting the search for optimal values -- and found that choosing suboptimal parameters did not necessarily impact forecast accuracy, which they called a "win-win"

In a similar study on the cost of cloud-computing, Petropoulos and colleagues (2021), used a subset of models as opposed to the whole family of statistical models such as exponential smoothing and found no loss of forecast accuracy or increase in forecast uncertainty. They estimated that using their reduced pool of exponential smoothing models would save Walmart $1.74 million per year in computational time (assuming that cloud-computing would be used) based on estimates of the combinations of fast-moving goods demanded in different postcodes, the CPU hours required to produce these forecasts on a monthly basis, and the cost of a CPU-hour,

These cost-of-forecasting studies have drawn attention as well to the environmental impact (unnecessary use) of computing power, as concern with social responsibility mandates elimination of waste and reduction of carbon footprint.

Understandability

"Learning requires some form of transparency, which forecasters can best achieve when they understand what they are doing." (Goldstein and Gigerenzer, 2009,p.770). Lack of understandability -- transparency, replicability, and simplicity -- can thus be considered a hindrance to and an additional cost of forecast improvement.

Simplicity Vs. Complexity

The Goldstein-Gigerenzer work, like numerous other studies, compared the forecasting accuracy of simple vs complex methods, with many of these finding advantages to simplicity across a variety of fields. Not only may simplicity improve the opportunity for understandability but may do so without necessarily sacrificing accuracy.

While the concept of forecast *utility* addresses the impact forecasts have on decision making, *understandability* is a characteristic of the forecasting method itself and hence a component of the cost of producing forecasts. Kesten Green and Scott Armstrong (2015) argued that in fact it is lack of understandability that creates model *complexity* rather than simply the number of variables and parameters in the model.

This element of cost is more difficult to measure but is surely related to model complexity. Further empirical study will be required, however, to monetize the costs associated with increasing complexity. Such research would also consider the costs in adjusting forecasts, especially if this is done manually. **Table 3** summarises various studies that address the costs of forecasting.

Table 3. Cost of Forecast Generation

Reference	**Domain**		**Study purpose and findings**
Makridakis et al. (1982) [M]	Forecasting competition – multiple data types: business & economic	~	Analyses simple and complex forecasting methods using traditional error metrics found that sophisticated methods do not always outperform simple ones and that accuracy can be improved by being selective or by combining forecasts. Computational cost is briefly mentioned but was not part of the evaluation.
Makridakis and Hibon (2000) [M3]	Forecasting competition – multiple data types: business & economic	~	Analyses 1001 time series of various types and frequency. Findings support the results of the earlier M competitions that simple methods can be as effective as complex. The cost of computation is briefly mentioned with a note that the forecaster should judge the trade-offs between accuracy and additional cost.
Green and Armstrong (2015)	Literature review		Analyses 32 studies comparing the accuracy of simple vs. complex forecasting methods. Of the 97 total comparisons, 84% showed that simple methods had maintained or improved accuracy over complex methods. The authors created a questionnaire to measure method simplicity.
Nikolopoulos and Petropoulos (2018)	Operations management (M3 subset)	$	Uses simplified selection of model parameters to demonstrate the accuracy vs. computation cost 'trade-off' on a subset of the M3 competition data Quantifies the annual computational cost savings in monetary terms; for example, given the cost of a CPU hour ($0.05) and an estimated number of SKUs per store (100,000) a specified number of stores (1,000), a retailer could save $950,000 from the proposed computation time savings.
Makridakis and colleagues (2020a) [M4]	Forecasting competition – multiple data types		Compares the computational time of running forecasting methods with the forecasting error (sMAPE). Found that as run time increases, accuracy improves (but so does computation cost)

Petropoulos and colleagues (2021)	Multiple data types: business & economic (M, M3 & M4)	$	Identifies subsets of families of forecasting models. Found that reducing "family size" does not negatively impact forecast accuracy or uncertainty estimation. Quantifies annual computational cost savings in monetary terms by using the cost of a CPU hour, the estimated number of forecasts produced annually and the time savings from sub-optimal parameter selection.

WHERE SHOULD WE BE HEADING?

While we found numerous examples of forecast appraisal based on *forecast utility*, there were barely a handful of examples that assessed the cost of the forecasts, and only one considered both utility and costs associated.. This is the 2015 article by Florian Pappenberger. Given that many studies have demonstrated the benefits of simple forecasting models, the two aspects of forecast value – utility and cost -- should be synthesized and possible tradeoffs between the two identified.

If the appraisals of utility and cost are made in monetary terms, a forecasting method can be assigned an economic value, enabling comparative evaluations of different methods in terms organizations can readily apply to improve efficiency. The benefits could extend to society at large, by improving environmental outcomes.

Perhaps most importantly, the studies cited here draw attention to the limitations of traditional forecast evaluation: there is abundant of evidence that improvements in forecast error measures don't necessarily correspond with economic benefits so a more practical means of forecast evaluation is required, one that focuses on forecast value with its components of utility and cost. In addition, consideration might be given to using economic rather than statistical criteria for fitting models to the historical data (Wunderlich and Memmert (2020)

In summary, we have highlighted the benefits of basing comparative evaluations of forecasting methods on *forecast value* rather than the more limited forecast-error measurement. These benefits include improvement in model selection and use, cost savings, and reductions in wasted resources, whether in finance, supply chain, meteorology, or other markets. It's clear that we need to go *beyond error measures* for method selection and evaluation .

References

1. Boylan, J., 2006. Accuracy and accuracy-implication metrics for intermittent demand. The International Journal of Applied Forecasting, 4, pp. 39-42.
2. Caldeira, J.F., Moura, G. v. and Santos, A.A.P., 2016. Predicting the yield curve using forecast combinations. Computational Statistics and Data Analysis [Online], 100, pp.79–98. Available from: https://doi.org/10.1016/j.csda.2014.05.008.
3. Catt, P.M., 2007. Assessing the cost of forecast error: a practical example. The International Journal of Applied Forecasting, 7 (Summer), pp.5–10.
4. Cenesizoglu, T. and Timmermann, A., 2012. Do return prediction models add economic value? Journal of Banking and Finance [Online], 36(11), pp.2974–2987. Available from: https://doi.org/10.1016/j.jbankfin.2012.06.008.
5. De Kok, S., 2017. The Quest for a Better Forecast Error Metric: Measuring More than the Average Error. Foresight: The International Journal of Applied Forecasting, Summer(46), pp.36–45.
6. Döpke, J., Müller, K. and Tegtmeier, L., 2018. The economic value of business cycle forecasts for potential investors – Evidence from Germany. Research in International Business and Finance [Online], 46(June), pp.445–461. Available from: https://doi.org/10.1016/j.ribaf.2018.06.001.
7. Dolatabadi, S., Narayan, P.K., Nielsen, M.Ø. and Xu, K., 2018. Economic significance of commodity return forecasts from the fractionally cointegrated VAR model. Journal of Futures Markets [Online], 38(2), pp.219–242. Available from: https://doi.org/10.1002/fut.21866.
8. Gilliland, M., 2002. Is Forecasting a Waste of Time? Supply Chain Management Review, 6(4), pp. 16-23. Available from: https://www.thefreelibrary.com/Is+forecasting+a+waste+of+time%3F+(Forecasting).-a089076989.
9. Mike Gilliland (2013): FVA: A Reality Check on Forecasting Practices, *Foresight,* Issue 29, 14-18.
10. Goldstein, D.G. and Gigerenzer, G., 2009. Fast and frugal forecasting. International Journal of Forecasting [Online], 25(4), pp.760–772. Available from: https://doi.org/10.1016/j.ijforecast.2009.05.010.
11. Green, K., C., and Armstrong, J., S., 2015. Simple versus complex forecasting: the evidence. Journal of Business research [Online], 68(8), pp.1678-1685. Available from: https://doi.org/10.1016/j.jbusres.2015.03.026
12. Ioannidis, J.P.A., Cripps, S. and Tanner, M.A., 2020. Forecasting for COVID-19 has failed. International Journal of Forecasting [Online]. Available from: https://doi.org/10.1016/j.ijforecast.2020.08.004.
13. Jessop, J., 2020. The UK lockdown and the economic value of human life. *Economic Affairs,* 40(2), pp. 138-147. Available from: https://doi.org/10.1111/ecaf.12417.
14. Kath, C. and Ziel, F., 2018. The value of forecasts: Quantifying the economic gains of accurate quarter-hourly electricity price forecasts. Energy Economics [Online], 76, pp.411–423. Available from: https://doi.org/10.1016/j.eneco.2018.10.005.
15. Koning, A.J., Franses, P.H., Hibon, M. and Stekler, H.O., 2005. The M3 competition: Statistical tests of the results. International Journal of Forecasting [Online], 21(3), pp.397–409. Available from: https://doi.org/10.1016/j.ijforecast.2004.10.003.

16. Leitch, G. and Tanner, J., 1991. Economic Forecast Evaluation: Profits versus the Conventional Error Measures. The American Economic Review [Online], 81(3), pp.580–590. Available from: https://doi.org/10.2307/2006520.
17. Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Parzen, E., Newton, J. and Winkler, R., 1982. The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition. Journal of Forecasting [Online], 1, pp.111–153. Available from: https://doi.org/10.2307/2345077.
18. Makridakis, S. and Hibon, M., 2000. The M3-Competition : results , conclusions and implications. 16, pp.451–476.
19. Makridakis, S., Spiliotis, E. and Assimakopoulos, V., 2020a. The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting [Online], 36(1), pp.54–74. Available from: https://doi.org/10.1016/j.ijforecast.2019.04.014.
20. Makridakis, S., Spiliotis, E. and Assimakopoulos, V., 2020b. The M5 Accuracy Competition: Results, Findings and Conclusions. [Online]. Available from: https://www.researchgate.net/publication/344487258.
21. Makridakis, S., Spiliotis, E., Assimakopoulos, V., Chen, Z., Gaba, A., Tsetlin, I. and Winkler, R.L., 2020c. The M5 Uncertainty competition: Results, findings and conclusions. [Online]. Available from: https://www.researchgate.net/publication/344487258.
22. Morlidge, S., 2014. Using Relative Error Metrics to Improve Forecast Quality in the Supply Chain. Foresight: The International Journal of Applied Forecasting, Summer(34), pp.39–47.
23. Murphy, A., 1993. What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. American Meteorological Society, pp.281–293.
24. Nikolopoulos, K. and Petropoulos, F., 2018. Forecasting for big data: Does suboptimality matter? Computers and Operations Research [Online], 98, pp.322–329. Available from: https://doi.org/10.1016/j.cor.2017.05.007.
25. Palmer, T., 2002. The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. Quarterly Journal of the Royal Meteorological Society, 128(581), pp.747–774.
26. Pappenberger, F., Cloke, H.L., Parker, D.J., Wetterhall, F., Richardson, D.S. and Thielen, J., 2015. The monetary benefit of early flood warnings in Europe. Environmental Science and Policy [Online], 51, pp.278–291. Available from: https://doi.org/10.1016/j.envsci.2015.04.016.
27. Petropoulos, F., 2020. The M4 competition, and a look to the future. *Foresight*, 57, pp.11–12.
28. Petropoulos, F., Grushka-Cockayne, Y., Siemsen, E and Spiliotis, E., 2021. Fast and Frugal Time Series Forecasting. SSRN Electronic Journal [Online], pp.1–26. Available from: https://doi.org/10.2139/ssrn.3792565.
29. Petropoulos, F., Wang, X. and Disney, S.M., 2019. The inventory performance of forecasting methods: Evidence from the M3 competition data. International Journal of Forecasting [Online], 35(1), pp.251–265. Available from: https://doi.org/10.1016/j.ijforecast.2018.01.004.
30. Sanders, N.R. and Graman, G.A., 2009. Quantifying costs of forecast errors: A case study of the warehouse environment. Omega [Online], 37(1), pp.116–125. Available from: https://doi.org/10.1016/j.omega.2006.10.004.

31. Wing, C.K and Yi, K.T.L, The Use of Profits as Opposed to Conventional Forecast Evaluation Criteria to Determine the Quality of Economic Forecasts.
32. Wunderlich, F. and Memmert, D., 2020. Are betting returns a useful measure of accuracy in (sports) forecasting? International Journal of Forecasting [Online], 36(2), pp.713–722. Available from: https://doi.org/10.1016/j.ijforecast.2019.08.009.