FORESIGHT

The International Journal of Applied Forecasting

ARTICLES *by* FORECASTERS *for* FORECASTERS: Q2:2024

SPECIAL FEATURE REVISITED: FVA CRITIQUE AND COMMENTARIES

Join the *Foresight* readership by becoming a member of the International Institute of Forecasters
forecasters.org/foresight/

SPECIAL FEATURE REVISITED: FVA CRITIQUE AND COMMENTARIES

A Critical Evaluation of the Assumptions of Forecast Value Added

CONOR E. DOHERTY

This article is based on a critique that originally appeared in the "Learn" section of ***www.lokad.com***

PREVIEW *Forecast Value Added (FVA) is a simple tool for evaluating the performance of each step (and contributor) in a forecasting process. Its goal is to eliminate waste by removing processes and activities (of any kind) that fail to increase forecast accuracy or reduce bias. FVA is predicated on the notion that better forecasting performance is worth pursuing, and that identifying the activities that increase it and eliminating those that do not is desirable. In this article, Conor Doherty argues that despite positive intentions, FVA demonstrates limited once-off utility and, if deployed on an ongoing basis, presents a multitude of drawbacks including faulty mathematical assumptions, misconceptions about the intrinsic value of increased forecasting accuracy, and the absence of a robust financial perspective.*

OVERVIEW OF FORECAST VALUE ADDED

Forecast value added aims to eliminate waste and increase demand forecasting accuracy by encouraging – and evaluating – inputs from multiple departments (including non-demand-planning teams, such as Sales, Marketing, Finance, Operations, etc.). By evaluating the value of each human touchpoint in the forecast process, FVA provides companies actionable data on the overrides that make the forecast worse, thus giving them an opportunity to identify and eliminate efforts and resources that do not contribute to better forecasting accuracy.

Michael Gilliland (2010), whose book *The Business Forecasting Deal* brought mainstream attention to the practice, argues that

> *FVA helps ensure that any resources invested in the forecasting process – from computer hardware and software to the time and energy of analysts and management – are making the forecast better... If these resources are not helping to forecast, they can be safely redirected to more worthwhile activities. (p. 91)*

Organizations often employ a multistage forecasting process where a statistical forecast is generated using their forecasting software. This computer-generated forecast is then subjected to manual changes (overrides) by each of the various departments engaged in their process. This adjusted forecast is then compared with a naïve, benchmark forecast (acting as a placebo), the computer-generated forecast, and the real, observed data.

If these departmental changes made the statistical forecast more accurate (compared to the untouched statistical forecast), they contributed positive value. If they made it less accurate, they contributed negative value. Similarly, if

Key Points

- FVA has no mechanism for confirming the forecast was more accurate because of the insight behind an override (e.g., demand was higher/lower for the reason believed). Thus, "positive" and "negative value" are determinations based on correlation, not causation. As such, there is no positive/negative value, only lucky/unlucky tweaks.
- Machine learning (ML) models have been shown to eclipse non-ML demand forecasting models in retail scenarios. FVA measuring the rightness or wrongness of manual override is thus questionable – particularly when FVA cannot distinguish correlation from causation.
- FVA prioritizes increasing forecasting accuracy without consideration of reducing financial error. From an overall risk management perspective, the former is arguably less significant than the latter (and vulnerable to classes of cognitive bias and manipulation that a purely financial perspective is not).
- FVA is predicated on a deterministic, time series (point forecast) perspective – a method that ignores the implications of future uncertainty (i.e., the full range of possible future outcomes).
- FVA can effectively demonstrate how flawed human – particularly collaborative – override is, hence it does have utility if employed as a once-off reality check for overly confident demand forecasters and planners.

the statistical forecast was more accurate than the placebo, it added positive value (and opposite if it was less accurate).

FVA is, thus, "…[a measure of] the change in a forecasting performance metric that can be attributed to a particular step or participant in the forecasting process" (Gilliland, 2010, p. 82).

Supporters of forecast value added argue that it is an essential tool in modern supply chain management. By identifying which parts of the forecasting process are beneficial and which are not, organizations can optimize their forecast accuracy. The overarching rationale is improved forecasting leads to better inventory management, smoother production planning, and more efficient resource allocation.

This, consequently, should reduce costs, minimize stockouts, and reduce overstocks, all while increasing customer satisfaction and generating a more inclusive forecasting and corporate ethos. The process has proved remarkably popular, with FVA having been applied at several notable companies in exceptionally competitive industries, including Intel, Yokohama Tire Canada, and Nestlé (Gilliland, 2015).

PERFORMING A FORECAST VALUE ADDED ANALYSIS

Performing a forecast value added analysis involves several intuitive steps, typically a close version of the following:

- **Define** the process by identifying the individual steps or components, i.e., the list of departments that will be consulted, and the order of consultation.
- **Generate** a benchmark forecast, which typically takes the form of a naïve forecast. A statistical forecast is also generated, as per the normal forecasting process within the company, using the same dataset utilized in the generation of the benchmark. This statistical forecast serves as the foundation for all subsequent adjustments.
- **Collect** insights (adjustments to the forecast) from the designated contributors. These might be based on market trend insights, promotional plans, operational constraints, etc.
- **Calculate** the FVA for each contributor by comparing the accuracy of the statistical forecast before and after that contributor's input. In turn, the accuracy of the statistical forecast is contrasted with that of the simple benchmark forecast. Contributions that enhance forecast accuracy receive positive FVA, while those that diminish accuracy receive negative FVA.

- **Optimize** by improving or eliminating contributions with negative FVA, while preserving or enhancing those with positive FVA.

These steps form an ongoing process that is iteratively improved in pursuit of greater forecast accuracy.

Figure 2 is based on processes described by Gilliland (2010) and Chybalski (2017). In contrast with **Figure 1**, there are multiple stages of human override (EO1, EO2, and CF). Complex forecasting processes (like Figure 2) can feature even more stages of human intervention, including a final phase of executive management override – all of which can be evaluated using an FVA analysis (Gilliland, 2010, pp. 93 and 97).

In **Table 1**, the evaluating metric is MAPE (mean absolute percentage error). This example indicates the Statistical Forecast improved forecasting accuracy (added value by reducing forecast error by 5 percentage points) compared to the Naïve Forecast. Furthermore, human override was unhelpful, with a significant increase in the forecasting error introduced at the Expert Override 2 stage.

Consider an apple seller. Paul (Demand Forecasting/Planning) informs management that the company sold eight apples in each of the last three months. A naïve forecast says the company will sell eight again next month, but Paul has advanced statistical software that predicts 10 apples will be sold (statistical forecast). John (Marketing) chimes in and says he intends on releasing a snazzy new slogan this month and sales are likely to be higher. George (Sales) intends to bundle apples together and lower prices slightly, stimulating sales even further and increasing demand. Richard (Operations) is initially stumped but then revises the forecasted demand to reflect an upcoming downtime in crucial apple-sorting machinery that he believes will adversely impact the company's ability to meet demand. The statistical forecast has, so far, been manually tweaked three times. The departments

Figure 1. A Simple Forecasting Process, Utilizing Minimal Human Override (Reserved for EO stage)

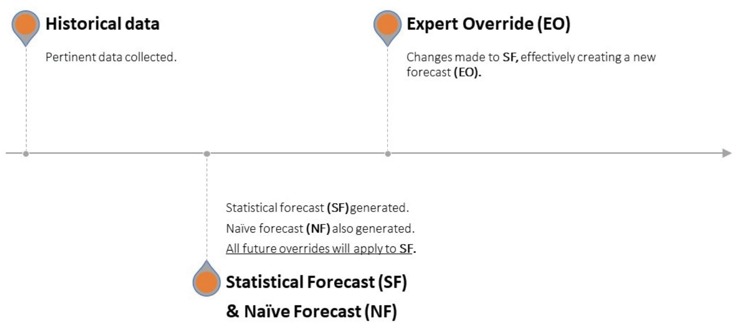

Figure 2. A Complex, Collaborative Forecasting Process (Each relevant step will be subsequently evaluated using the FVA framework)

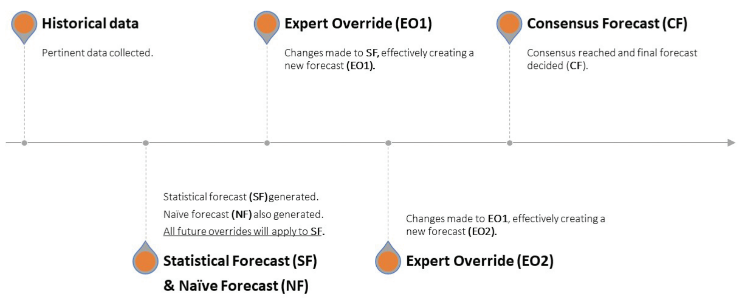

Table 1. Model Stairstep Report (Example Adapted from Schubert and Rickard, 2011) Demonstrating the Value (Positive or Negative) Added at Each Step of the FVA Process

Forecasting Stage	MAPE	FVA vs. NF	FVA vs. SF	FVA vs. EO1	FVA vs. EO2
Naïve Forecast (NF)	50%	-	-	-	-
Statistical Forecast (SF)	45%	+5%	-	-	-
Expert Override 1 (EO1)	52%	-2%	-7%	-	-
Expert Override 2 (EO2)	58%	-8%	-13%	-6%	-
Consensus Forecast (CF)	55%	-5%	-10%	-3%	+3%

next congregate to verbally reach a consensus forecast.

One month later, the company performs a backtest to confirm how great the delta was at each step of this forecasting relay – i.e., how "off" each department's contribution was. This is not difficult as they now possess the actual sales data for the previous month, and Paul can isolate, beat by beat, how much error was introduced by John, George, and Richard, respectively, as well as the consensus forecast stage.

As per Gilliland (2010, p. 98), "measuring FVA over one time period is not enough to draw any valid conclusions," hence one "…must make sure that time frame has been sufficient to produce meaningful results." This allows organizations to "be reasonably certain the observed results are not due to chance." For large organizations forecasting thousands of SKUs, an FVA analysis will require considerable data collection.

Contrary to forecasting processes or analysis tools that demand an advanced knowledge of mathematics and statistical reasoning, FVA "is a common-sense approach that is easy to understand."

THE MATHEMATICAL PERSPECTIVE ON FORECAST VALUE ADDED

Under the hood, forecast value added is a remarkably straightforward and deliberately uncomplicated process. Contrary to forecasting processes or analysis tools that demand an advanced knowledge of mathematics and statistical reasoning, FVA "is a common-sense approach that is easy to understand. It expresses the results of doing something versus having done nothing" (Gilliland, 2015, p. 1).

Expressing the results of having done something versus nothing, however, still requires mathematical intervention, and this typically takes the form of a simple time series – the backbone of traditional forecasting methods. The primary goal of time series analysis is conveniently and intuitively representing future demand as a single, actionable value (the "point forecast"). In the context of FVA, the baseline time series serves as a placebo or control, against which all the analyst overrides (detailed in the previous section) are compared. A baseline time series can be generated through various methods, commonly including various forms of naïve forecasting. These are commonly evaluated using metrics such as MAPE, MAD, and MFE.

Choosing a Benchmark Forecast

The choice of baseline forecast will vary depending on the goals or constraints of the company in question; however, it is advised that whatever model one chooses should be something that could legitimately be used for forecasting purposes (Gilliland, 2010).

Typically, a naïve ("no change") forecast is recommended to/adopted by FVA consumers. Naïve forecasts are easy to calculate and understand, as they are predicated on the assumption that previous data will be repeated in the future (as per the naïve forecast used in the apple-selling analogy). That said, there are alternative flavors of naïve forecast, such as moving-average models, and there is no strict guideline on what benchmark should be selected – other than the injunction to select one model and apply it consistently and transparently across the duration of one's FVA analysis (Gilliland, 2010).

Evaluating Forecast Value Added Results

* **MFE (Mean Forecast Error)** can be used to assess whether a forecast tends to overestimate or underestimate actual results (also known as bias). This could be a useful metric in a situation where it is more costly to overforecast than to underforecast, or vice versa.
* **MAD (Mean Absolute Deviation) and MAPE (Mean Absolute Percentage Error)** provide measures of forecast accuracy that consider both over- and underforecasting demand. They might be used as gauges of accuracy when it is

important to minimize the overall size of forecast errors, regardless of whether they result in over- or underforecasting.

Though MAPE is commonly featured in FVA-related sources, there is no strict formalism, and consensus varies between academics and vendors as to which forecast-metric configuration to utilize in an FVA analysis (see Gilliland, 2010 & 2015; Chybalski, 2017; Vandeput, 2021; Goodwin, 2018).

LIMITATIONS TO FVA

Forecast value added, despite its inclusive approach, noble goals, and low barrier for entry, is arguably subject to an extensive array of limitations and false premises. These deficiencies span a wide range of fields, including mathematics, modern forecasting theory, and economics.

Lack of Scientific Formalism

Gilliland (2010, p. 100) argues "FVA analysis lets you take an objective, scientific, and data-driven approach to process analysis." Unfortunately, this is simply not the case when FVA is applied to the measurement of anything that involves making manual – and even collaborative – overrides based on *human insight*. This criticism is also rooted in an overall lack of scientific formalism or standardization when it comes to how the insights/overrides are collected and, importantly, how they are interpreted.

In a section titled "Collecting the Data," Gilliland (2010, pp. 94-95) presents a very intuitive process for experts to submit their overrides (in a simple forecasting process, as per Figure 1). The elements required are remarkably straightforward, amounting to people simply entering either a higher, lower, or unchanged value next to the values provided by the naïve and statistical forecasts. There is no suggestion that contributors should provide any justification for their overrides and, fundamentally, it would be incredibly misguided to do so.

The reason is simple: FVA has no mechanism for confirming the forecast was more accurate *because of the insight behind an override* (e.g., demand was higher/lower for the reason held by the person who made the override). Apple demand, to return to an earlier example, rises and falls for any number of reasons. John's snazzy apple slogan could have influenced demand to rise (thus appearing to validate his upward override to the statistical forecast), or demand could have spiked for any one (or more) of a million (or more) factors. In reality, there is simply no way to make this determination. Thus, FVA's designations of *positive* and *negative value* are based on correlation, not causation. As such, in the context of forecasting processes like Figure 1 and Figure 2, there is no positive/negative value with each touchpoint, only lucky/unlucky tweaks.

This stands in stark contrast to Gilliland's advice in a section titled "Reporting the Results" (2010, p. 96): "In some cases, it may be possible to improve performance with education or technical training for the participants." However, one would not improve the performance of the participant(s), only sink more resources into an activity that is engineered to solicit lucky or unlucky tweaks – tweaks that cannot be demonstrated (at least not by FVA) to be anything more than measures of correlation and (good/bad) luck.

Importantly, even if more sophisticated software is available for the extraction of FVA overrides, it will still be subject to the same criticism listed above – as well as several others to come.

Forecasting Should Be Automated, Not Collaborative

FVA is, at best, ambivalent when it comes to the merits of collaborative forecasting. As mentioned above, FVA, by design, only tells you if an activity appears to have added value (and not whether or not the insight behind the override was correct or just coincidence). This design tacitly endorses collaborative forecasting by providing data to support the practice of soliciting ongoing human override *if it appears to add value*.

For example, if multiple human overrides – including a consensus stage as per Figure 2 – to the statistical forecast were demonstrated to add (or subtract) value, Gilliland advises caution: one should solicit overrides *for long enough* until one is convinced they truly add (or subtract) value (2010, p. 98):

> *Measuring FVA over one time period is not enough to draw any valid conclusions. Period to period, FVA may go up or down, and over short time frames FVA may be particularly high or low due to chance.... No one should receive adulation (or be fired!) for numbers that are* not valid indicators of their performance [emphasis mine], *so it is essential to be reasonably certain the observed results are not due to chance.*

However, as mentioned in the previous section, FVA has no mechanism for determining if the results of an FVA analysis – in this case a collaborative one – are due to the insight behind the override or complete chance. It should be noted that no advice is provided in "Interpreting the Results" (Gilliland, 2010, pp. 98-101), or the entire chapter dedicated to FVA (Gilliland 2010, pp. 81-110) on how FVA consumers are supposed to discern if the results are valid (not owing to chance). As such, collecting and measuring collaborative overrides (based on unprovable and untestable insights) from disparate departments (including nonspecialists) is arguably quite wasteful if they cannot be tested in any meaningful, direct way (as per FVA's goal of objectivity and scientific inquiry).

Despite this, and even if FVA does not explicitly advocate it, its very architecture endorses collaborative forecasting as long as it *appears* to add value. This is evident in the terminology itself: marketing/sales/operations/finance/consensus overrides can "add positive value" despite the fact FVA has no mechanism to demonstrate that the underlying insight(s) actually contributed the value as opposed to mere coincidence.

Furthermore, recent literature indicates that the most effective forecasting process does not feature additional touchpoints at all. An extensive review (Makridakis and colleagues, 2022, pp. 1361-1362) of the fifth Makridakis forecasting competition demonstrated that

> *All 50 top-performing methods were based on ML (machine learning). Therefore, M5 is the first M-competition in which all of the top-performing methods were both ML methods and better than all of the other statistical benchmarks and their combinations.*

The M5 competition was based on forecasting sales using historical data for Walmart, the largest retail company in the world by revenue. Furthermore, "the winning model was developed by a student with little forecasting knowledge and little experience in building sales forecasting models" (Makridakis and colleagues, 2022, p. 1359), thus casting doubt on how vital the market insights of disparate departments truly are in a forecasting context.

This is not to claim that more complex forecasting models are inherently desirable. Rather, there is a strong case to make that models predicated upon multiple touchpoints are less than desirable compared to ones that are not. By extension, measuring (e.g., with FVA) the efficacy of such models – particularly if they feature multiple and even collaborative overrides – is becoming increasingly difficult to justify given they are not built on solid foundations to begin with.

Granted, FVA could still theoretically be applied to an ML-based forecasting model to test how efficiently it (the ML model) performs relative to a simple naïve forecast. This, however, would in no way validate or justify the idea of dedicating company resources to the *ongoing* extraction of multiple overrides from disparate collaborators (be they specialists, such as data scientists, or nonspecialists, such as salespeople). Instead, it would support the idea of dedicating more resources to the fine-tuning of the algorithm underpinning the ML model (and then leaving it to run automatically in the background). This is certainly true if resources are limited (which they invariably are) and one must decide between the two options.

Ignores Future Uncertainty

FVA evaluates the accuracy and value added by a series of overrides to a time series forecast (see Gilliland, 2010, as well as any typical FVA vendor). Thus, FVA implicitly (if not explicitly) accepts that knowledge of the future (in this case, demand) *can be* represented in the form of a time series. This is flawed for two reasons.

First of all, the future, be that in general or forecasting terms, is irreducibly uncertain. As such, expressing it as a single value is an inherently misguided approach (even if augmented with a safety stock formula). Presented with the irreducible uncertainty of the future, the most sensible approach is determining a range of likely future values, evaluated with respect to each one's potential financial return. This trumps, from a *risk management perspective*, attempting to identify a single value as per a traditional time series – something that entirely ignores the problem of future uncertainty.

Secondly, the insights (however useful they may seem) of collaborators are typically of the kind not easily (if at all) translated to a time series forecast. Consider a situation in which a company knows ahead of time that a rival is about to enter the market. Alternatively, imagine a world in which competitive knowledge indicates that one's fiercest competitor is planning to release an impressive new line of summer clothes. The proposition that these kinds of insights can be collaboratively folded by non-specialists into a single value expressed in a time series is fanciful.

As previously argued, in reality, any similarities to actual future values (e.g., positive value added) will be difficult (if not impossible) to differentiate from chance. In this sense, human overrides, be they rounding demand up or down, are equal expressions of the same faulty input. A person who contributes negative value is thus no more right or wrong – from a logical perspective – than a person who contributes positive value. Furthermore, even if this were not the case, FVA has no mechanism for confirming it, regardless of how long the data are collected. FVA merely confirms the presence of correlation (or chance), not that the insight behind the override was in fact correct.

At its core, FVA aims to evaluate the accuracy of multidimensional properties (human insights) being crudely thrust onto a two-dimensional surface (a time series). It may look right from a certain angle, but that does not mean it is right. This gives FVA a rather misleading appearance of statistical rigor.

Even if a company uses a simple forecasting process with minimal human touchpoints (as per Figure 1), if the underlying forecast being analyzed by FVA is a time series, the analysis itself is an exercise in futility.

Ironically Wasteful

As a once-off demonstration of overconfidence and biased decision making, FVA has utility. Nobel prizes have been awarded on the depth, breadth, and endurance of cognitive biases in human decision making (Kahneman, 2011; Karelse, 2022), yet it is entirely conceivable some teams fail to accept just how faulty human override typically is until they are emphatically shown.

However, as an ongoing management tool, FVA is inherently flawed and arguably contradictory. If one's statistical forecasts are outperformed by a naïve forecast and collaborative tinkering, one should really consider the following question: Why are the statistical models failing?

One reason is that the statistical forecast overfits the data, thus it will fail to generalize properly when presented with new data. However, this situation is an effective litmus test for spotting professional incompetence (i.e., failure to perform sufficient model backtesting, cross-validation, etc.), rather than an explicit argument in support of naïve forecasts and FVA. (For more on the problem of generalization in forecasting, see Vermorel [2023].) Another possible explanation is a systemic shock/shift in the market that is not reflected in the historical data, hence any overperformance of a naïve forecast is pure coincidence. For example, if stores suddenly limit sales of a critical item during an unforeseen pandemic, this might produce artificially constant "demand." In such a scenario, a naïve forecast might work exceptionally well.

FVA, unfortunately, has no answer for the problem of statistical forecasts underperforming because it fundamentally is not designed to. It does not yield insights into *why* statistical models might underperform, simply *that* they underperform. FVA is thus not so much a diagnostic tool as a magnifying glass.

At its core, FVA presumes that increased forecast accuracy is worth pursuing in isolation, and proceeds on this basis as if this were self-evidently true.

While a magnifying glass can be useful, it does not provide actionable insights into what the underlying problems with the statistical forecasting software actually are. Understanding why one's statistical forecasts underperform has far greater direct and indirect (as well as short- and long-term) value, and is something FVA does not bring into sharper focus.

Not only does FVA software not yield this important insight, it formalizes waste in other ways. Gilliland (2010) presents a theoretical situation in which a consensus forecast is outperformed in 11 out of 13 weeks (85% fail rate), averaging 13.8 percentage points of error. Rather than warranting immediate discontinuation, the advice is to

> *...bring these findings to your management and try to understand why the consensus process is having this effect. You can start to investigate the dynamics of the consensus meeting and the political agendas of the participants. Ultimately, management must decide whether the consensus process can be fixed to improve the accuracy of the forecast, or whether it should be eliminated. (p. 100)*

In this scenario, not only does the FVA software not diagnose the underlying problem of statistical forecast performance, but the layer of FVA instrumentation merely increases bureaucracy and resource allocation by dissecting activities that manifestly do not contribute value.

As such, installing a layer of FVA software ensures that one continues to get similar low-resolution images of an ongoing problem and directs valuable resources to understanding faulty inputs that could have been ignored right from the outset.

This, arguably, is not the most prudent allocation of company resources that have alternative uses.

Overestimating the Value of Accuracy

At its core, FVA presumes that increased forecast accuracy is worth pursuing in isolation, and proceeds on this basis as if this were self-evidently true. The notion that increased forecast accuracy is desirable is an understandably appealing one, but from a business perspective it presumes that greater accuracy translates into greater profitability. This is patently not the case.

This is not to claim that an accurate forecast is not worth having. Rather, an accurate forecast should be tightly tethered to a *purely financial perspective* measured in dollars (or other currency) of expected returns (margin) and expected losses (costs). A forecast might be more accurate, but the associated cost means the company makes less profit overall while increasing the overhead and complexity associated with more accurate forecasts (this is commonplace for large companies deploying in-house data science teams – to discover several years down the road there is no significant financial gain achieved). The forecast, though appreciably more accurate (positive value added), has not reduced dollars of error. This violates the core tenet of business: make more money, or at least do not waste it. This is easier to say than to do, but in this process it is better to be approximately right (by assessing expected ROI and iteratively improving the assessment) rather than exactly wrong (optimizing nonfinancial metrics, namely percentages of forecast accuracy).

In terms of FVA, it is entirely conceivable that the positive value added by one department is a net loss to a company, whereas the negative value added by another is imperceptible. While Gilliland does acknowledge that some activities might increase accuracy without adding financial worth, this angle is not followed to its logical endpoint: a purely financial perspective. Gilliland (2010) uses the example of an analyst increasing forecast accuracy by a single percentage point:

> *The mere fact that a process activity has positive FVA does not necessarily mean it should be kept in the process. We need to compare the overall benefits of the improvement to the cost of that activity. Is the extra accuracy increasing revenue, reducing costs, or making customers happier? In this example, the analyst override did reduce error by one percentage point. But having to hire an analyst to review every forecast can be expensive, and if the improvement is only one percentage point, is it really worth it?* (p. 83)

In other words, a one percent increase might not be worth pursuing, but a greater increase in forecast accuracy could be. This presumes that financial value is tied to greater forecast accuracy, which is not necessarily true.

Thus, there is an ineluctable financial dimension to forecasting that is at best understated in FVA and, at worst, barely noticed. This purely financial perspective really ought to be the foundation upon which a tool aimed at reducing waste is built.

Vulnerable to Manipulation and Bias

FVA also presents an obvious opportunity for gaming and forecast manipulation, especially if forecast accuracy is used as a measure of departmental performance. This is the spirit of Goodhart's law, which states that once an indicator becomes the chief measure of success (accidentally or deliberately), that indicator ceases to be useful. This phenomenon can often open the door to misinterpretation and/or manipulation.

Suppose the sales team is tasked with making short-term adjustments to the demand forecast based on their interactions with customers. The sales department might view this as an opportunity to signal their value and start making changes to the forecast even when not necessary, in an attempt to demonstrate a positive FVA. They might overstate demand, making them appear to be generating value, or recalculate demand downward, making them appear to be correcting an overly sanguine projection from a previous department. Either way, the sales department may appear more valuable to the company. As a result, the marketing department might then feel pressured to appear to be generating value, too, and the team starts making similarly arbitrary tweaks to the forecast – and so on and so forth.

In this scenario, the FVA measure, originally intended to improve forecast accuracy, becomes merely a (costly) political mechanism for departments to signal value rather than adding any, a criticism even FVA advocates acknowledge (Vandeput, 2021). These examples demonstrate the potential dangers of Goodhart's law when it comes to FVA.

Moreover, even if the underlying motivations of forecasters were trustworthy, the reliability of their inputs is very questionable. On the topic of forecaster override and motivations during FVA analysis, Fildes and colleagues (2023) conclude,

> *When discussing the potential value added to the system forecast through judgmental adjustment, the focus has been on adjusting based on factors not included in the data or model, such as special events, rumours or political motivations. However, our analysis suggests that it is not this unmodelled component or information on external factors that is given the most weight .... What does influence the adjustment is information typically displayed in forecasting support systems, such as the current system forecast, the previous forecast error and the previous adjustment – factors that are inappropriate for supporting the adjustment decision. Indeed, this information had a stronger effect on decisions to adjust than unobserved information (i.e., factors that were driving*

demand which were not recorded in our data). (p. 30)

Fildes and colleagues further note that "forecasters' [override] behavior proved to be predictable, with the previous adjustment stimulating a further adjustment, mostly in the same direction." This suggests forecaster overrides are – to a large degree – both predictable and unmotivated by additional relevant insights, raising further doubts regarding the merits of extracting and measuring them in an FVA analysis.

It should be noted that the conclusions drawn by the researchers are based on their interpretation(s) of the results, given "no data sets have documented justification for the adjustments made" (Fildes and colleagues, 2023, p. 9) – something that would be incredibly difficult, if not impossible, to analyze (in terms of causation) even if documented.

Supporters of FVA might argue these psychological criticisms are the entire point of FVA, namely the identification of *valuable versus junk inputs*. However, given the biases associated with human override in forecasting are *so predictable and well-understood across several domains of study*, the resources spent dissecting these bias-laden inputs would likely be better allocated to a process that avoids (as much as possible) these inputs in the first place.

Local Solution to a Systemic Problem

Implicitly, the attempt to optimize demand forecasting in isolation presupposes that the problem of demand forecasting is separate from other supply chain problems *and* these problems are not interrelated. In reality, demand forecasting is complex due to the interaction of a wide array of systemic supply chain causes, including the influence of varying supplier lead times, unexpected supply chain disruptions, stock allocation choices, pricing strategies, cannibalizations, substitutions, etc. Importantly, none of these issues is remedied by pursuing better demand forecasting accuracy in isolation (Vermorel, 2020a). Furthermore, if left unresolved, these other sources of flux will (obviously) continue to undermine one's demand forecasting efforts – including the utility of one's FVA measurements.

Attempting to optimize demand forecasting in isolation – "local optimization" (Vermorel, 2020a) – is a misguided approach given the system-level problems – the true root causes – are not properly understood and addressed. Supply chain problems – of which demand forecasting is certainly one – are like people standing on a trampoline: moving one person produces disequilibrium for everyone else (to paraphrase Carol Gilligan's public remarks, originally made in the context of human action). For this reason, holistic, end-to-end optimization is better than attempting to cure symptoms (demand forecasting) in isolation. The baseline for this end-to-end optimization is a data-driven perspective that measures the financial impact of supply chain decisions in totality, rather than the accuracy of isolated ones, such as demand forecasting (Vermorel, 2020b).

From this perspective, FVA has no system-wide utility, as it is not tethered to a strict financial perspective and pursues – and overestimates the value of – increased forecast accuracy in isolation.

CONCLUSIONS AND RECOMMENDATIONS

Forecast Value Added, despite its noble aims, appears to be a suboptimal approach to a systemic problem. It lacks rigid scientific formalism; it extracts overrides indistinguishable from chance; it is predicated upon a dated time series perspective to future uncertainty; it facilitates wasteful resource allocation (if deployed long-term); it lacks a robust financial perspective in its DNA; it is vulnerable to costly manipulation and predictably biased inputs; and it overestimates the value of forecast accuracy.

Rather than employing FVA, a more sophisticated strategy would be to look beyond the entire concept of forecasting accuracy and opt instead for a risk management policy that reduces dollars (or euros) of error by design. In conjunction with a probabilistic forecasting approach, this mindset moves away from arbitrary KPIs – such as increasing forecasting accuracy in isolation – and factors the totality of one's economic drivers, constraints, and potential supply chain shocks into one's inventory decision making. These kinds of risk (and waste) vectors cannot be effectively quantified (and eliminated) in the forecast value added framework.

Furthermore, by separating demand forecasting from overall supply chain optimization, FVA (perhaps unintentionally) augments the accidental complexity of the demand forecasting process (see Vermorel, 2021 for commentary on accidental versus intentional complexity). Accidental complexity is artificial (man-made) and results from the gradual accretion of unnecessary noise in a process. Adding evaluation stages to the forecasting process, as FVA does, is a prime example of accidental complexity and can make the problem at hand significantly more complex.

Demand forecasting is an intentionally complex problem, which is to say it is an inherently puzzling and resource-intensive task. This complexity is an immutable trait of the problem and represents a much more troubling class of challenge than accidentally complex issues. For this reason, it is best to avoid attempts at solutions that oversimplify and fundamentally misconstrue the problem at hand. To echo the medical rhetoric of FVA literature (where naïve forecasts are referred to as "placebos"), this is the difference between curing an underlying illness and constantly treating symptoms as they arise.

In short, FVA exists in the space between cutting-edge supply chain theory and the public's awareness of it. Greater education in the underlying causes of demand uncertainty – and its roots in the evolving supply chain discipline – is recommended.

REFERENCES

Chybalski, F. (2017). Forecast Value Added (FVA) Analysis as a Means to Improve the Efficiency of a Forecasting Process, *Operations Research and Decisions*, 27(1), 5-19.

Fildes, R., Goodwin, P. & De Baets, S. (2023). Forecast Value Added in Demand Planning (September 1, 2023). **papers.ssrn.com/sol3/papers.cfm?abstract_id=4558708**

Gilliland, M. (2010). *The Business Forecasting Deal*, Wiley.

Gilliland, M. (2015). Forecast Value Added Analysis: Step-by-Step, SAS. **sas.com/content/dam/SAS/en_us/doc/whitepaper1/forecast-value-added-analysis-106186.pdf**

Goodwin, P. (2018). *Profit from Your Forecasting Software*, Wiley.

Kahneman, D. (2011). *Thinking, Fast and Slow*, Macmillan.

Karelse, J. (2022). *Histories of the Future*, Forbes Books.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2022). M5 Accuracy Competition: Results, Findings, and Conclusions, *International Journal of Forecasting*, 38(4), 1346-1364.

Schubert, S. & Rickard, R. (2011). Using Forecast Value Added Analysis for Data-driven Forecasting Improvement, IBF Best Practices Conference.

Vandeput, N. (2021). Forecast Value Added, Medium. **nicolas-vandeput.medium.com/forecast-value-added-ebc163d7ccd**

Vermorel, J. (2020a). Quantitative Principles for Supply Chain. Lokad. **lokad.com/tv/2021/1/20/quantitative-principles-for-supply-chains/**

Vermorel, J. (2020b). The Quantitative Supply Chain in a Nutshell. Lokad. **lokad.com/tv/2020/12/02/the-quantitative-supply-chain-in-a-nutshell/**

Vermorel, J. (2021). 21st Century Trends in Supply Chain. Lokad. **lokad.com/tv/2021/1/6/21st-century-trends-in-supply-chain/**

Vermorel, J. (2023). Generalization (Forecasting), Lokad. **lokad.com/generalization/**

Conor E. Doherty is Head of Communication at Lokad, an engineering company specializing in quantitative supply chain optimization. His duties include composing technical documentation and critiquing supply chain theory and forecasting practices. He also conducts in-depth interviews with industry and academic experts across multiple fields, including forecasting and supply chain planning. Additionally, Conor is a lecturer at SciencesPo Paris's Centre of Rhetoric and Writing, where he delivers master's-level courses in rhetoric and composition.

conor.doherty@lokad.com