Received: 28 January 2023 | Accepted: 31 January 2023

DOI: 10.1002/asmb.2753

COMMENTARY

WILEY

Discussion of “Multivariate dynamic modeling for Bayesian forecasting of business revenue”

Dimitris Korobilis | **Santiago Montoya-Blandón**

Adam Smith Business School, University of Glasgow, Glasgow, UK

Correspondence
Dimitris Korobilis, Adam Smith Business School, University of Glasgow, G12 8QQ Glasgow, UK
Email: dimitris.korobilis@glasgow.ac.uk

1 | INTRODUCTION

The paper by Yanchenko et al.[1] presents a long-term business revenue forecast effort for a large retail company. We congratulate the authors for elaborating on a framework that is of theoretical interest while responding to specific needs of business revenue forecasting and the retail sector. Forecasting approaches in academic spheres usually focus on developing flexible and general models with desirable statistical properties. However, models for forecasting in practical settings further require considerations of computational feasibility, scalability, replicability and interpretability. Most of these concerns are addressed with simple and transparent Bayesian techniques that afford the firms a lot of control over the data and inputs used, as well as interpretation of the results. Using information on pricing strategies, promotions and discounts, and taking into the account the multi-scale nature of stores and categories of items that affect these pricing decisions, the authors present 12-week ahead forecasts on revenue at the local store group and item levels. Their analysis first shows how using multi-scale information on pricing and discount decisions can improve forecasts for some categories and stores, but that sharing information across scales does not uniformly lead to forecast improvements. In particular, smaller store groups tend to benefit more than larger groups. The analysis then focuses on how additional modeling of price and seasonal effects can improve performance and explore these gains in the context of their available data. Finally, the authors explore the correlation structure across categories of goods to try and understand how these can inform better pricing strategies and the forecasting specification itself.

The methods in the paper build upon the classic dynamic linear model (DLM) framework of West and Harrison[2] and their extensions in Berry and West.[3] As perfectly exhibited in the paper, the DLM framework is tremendously diverse and flexible, with many modeling and fine-tuning options that can be suited to a wide range of applications. On the other hand, the sheer amount of flexibility in these models means there can still be strategies that lead to potential forecasting gains. Consequently, in this comment we explore possible extensions to the specification presented in the paper that might add value to this and future business revenue forecasting applications. The remainder of the comment is structured as follows. Section 2 discusses extensions to the covariates included in the model, both in terms of additional pricing variables and use of mixed-frequency information in DLMs; Section 3 discusses other technical extensions, for example one that could allow for recursively incorporating expert information in the prior structure, or one to the coupling/decoupling idea viewed as a problem of variable selection and dimensionality reduction; Section 4 discusses the use of other application-relevant objective functions and their possible gains for forecasting; finally, Section 5 presents our concluding remarks and further advice for future research.

2 | ADDITIONAL PREDICTORS AND MIXED-FREQUENCY DATA

Despite the fact that the authors claim that their dataset has several “breadth of discount” measures, they only consider three as potential predictors of sales. It would be interesting to explore more predictors, and also allow different predictors to have different importance for forecasting each week's sales. Within the context of DLM that the authors present (see Yanchenko et al.,[1] equation (1)) our suggestion implies that some of the coefficients $\boldsymbol{\theta}_t$ may be zero in some time periods but not in others. This is the so-called *dynamic variable selection* problem in statistics, and recently Koop and Korobilis[4] have developed a fast variational Bayes algorithm that allows to select among hundreds of predictors each time period.

Related to the previous point, many times there are relevant predictors that are not measured at the same frequency as the target variable. A prime example is regional macroeconomic data, such as disposable income of families in the geographical area of the Local Store Groups considered in the study, which most probably will not be available at the weekly frequency used in the analysis. Conversely, financial data and other indicators might be available at the daily level. The DLM is an appropriate framework for forecasting with mixed-frequency data, as it allows to treat low-frequency observations (e.g., monthly) as high-frequency observations (e.g., weekly) that have missing values, without having to dramatically alter the forecasting specification or estimation algorithms; see Harvey[5] for an early exposition of this idea.

3 | EXPERT OPINIONS AND DECOUPLING/RECOUPLING

Even if one argues that additional predictor variables may not necessarily provide valuable information for long-term forecasting, it is hard to understate that expert opinion from decision-makers or professional forecasters is invaluable and cannot be fully replaced by any model-based forecast. This additional information from the decision-maker may come in the form of experience, private information, or knowledge that simply cannot be measured and incorporated into a statistical model. Here we would like to stress again the flexibility of the DLM, which, when estimated with variance discounting methods, allows to incorporate expert opinion in a numerically explicit way. In particular, Kulhavỳ and Zarrop[6] provide details of a generalized variance discounting framework (also known as “forgetting” in the signal processing jargon) for updating the regression coefficients $\boldsymbol{\theta}_t$ in equation (1) of Yanchenko et al.[1] This scheme allows to define both the mean and variance of the regression coefficients at time t as a mixture of past information (i.e., estimates obtained using data up to time $t-1$), as well as dynamically updated prior information that is available each period t. In contrast, standard DLM estimation only allows to choose an initial condition at time $t = 0$, and then parameters are updated solely by information in the data. While in general estimation problems data information provides an objective way of tracking parameters, when forecasting it might be beneficial to allow “experts” to inject private information and beliefs into parameter updates. The generalized variance discounting approach of Kulhavỳ and Zarrop[6] allows for the incorporation of expert opinion each week in the form of recursive priors on the regression coefficients; we refer the authors to this paper for technical details.

We find particularly interesting how the decoupling/recoupling strategy operates in the context of DLMs. The decoupling aspect is simple as it begins by modeling the univariate time series conditional on the appropriate states. Both the recoupling and multi-scale analyses operate through these states, whether they be latent or observed. As outlined in Zhou et al.[7] and West,[8] one needs to decide on a way to incorporate contemporaneous relationships between the forecasted time series in order to account for all relevant information and increase forecasting performance. Then, the recoupling aspect relates to accounting for these relationships in the posterior distributions of DLM coefficients and variances. A similar structure arises for sharing multi-scale information across different levels. In this case, however, the contemporaneous information that is included in the modeling for each time series is through a common aggregate factor that impacts those lower levels of aggregation. As noted in the paper, accounting for both of these sources of information using the recoupling approach has an advantage over standard hierarchical models, given the extra computational burden of smoothing after each posterior sampling iteration.

These models attempt to account for cross-correlation structures in two different ways. One is through the allowed components in the variance matrices for the DLM updates. The second is through the recoupling approach. As outlined in West,[8] this turns out to be similar as trying to learn a relevant network or graph relating the variables at each point in time but without directly incorporating information on their correlation structure. The work in Gruber and West[9,10] shows how to use tools from graphical analysis in order to better inform this decision or construct models with more general coupling strategies. Exploring the use of further graphical tools that are used in network analysis (and that can potentially be included within a Bayesian structure; see, e.g., Fan et al.[11]) would be an interesting additional extension for the paper and the decoupling/recoupling framework in general.

4 | ALTERNATIVE LOSS FUNCTIONS

One key takeaway of the paper is that 12-week forecasting of the particular sales data is hard. Even though this conclusion sounds pessimistic, this is the truth in many hard forecasting problems: even the most elaborate statistical models and algorithms may not be able to beat a naive forecast. We quote as examples the problems of forecasting exchange rates, oil prices, and the yield curve (interest rates) for which the random walk forecast is notoriously hard to beat. In such cases, it might be worth reconsidering what the modeler and decision-maker consider is a reasonable loss. The mean absolute percentage error (MAPE) metric used in the paper is a benchmark criterion for many forecasting problems, but it is not always representative of the ability of a statistical model to provide useful output for decision-making. The forecasting literature has numerous examples of other loss functions and targets useful for sales forecasting. For example, one could attempt to forecast only the sign (whether sales will go up or down), or follow a mixed approach of point-forecasting sales within a specified range of interest and forecasting the sign beyond that range. Depending on the exact target of the client, such loss functions might provide more realistic input to decision-making.

5 | CONCLUSION

In summary, we believe that the DLM is an excellent choice of statistical framework for forecasting count data, such as sales. The paper by Yanchenko et al.[1] strikes a great balance between the desire to have a model that is flexible enough and being able to communicate results easily to interested stakeholders. Due to the fact that forecasting economic and financial data comes with many challenges, this comment provides some thoughts/suggestions that cover various aspects of the statistical modeling process. These suggestions have more general applicability to other forecasting problems and can be summarized as follows. First, the statistician might want to consider expanding the primary inputs (data) to their statistical model by considering a larger number of exogenous predictors. Second, they might want to expand their statistical model (and/or estimation algorithms) drawing from the numerous advances in statistical machine learning. Third, they can use their prior distribution as a means of incorporating expert opinion, such that they rigorously account for information that is missing in the data or likelihood function. Finally, if all else fails and forecasts cannot be improved substantially, the statistician can adjust their expectations about the part of time series data that is truly forecastable (e.g., by using different loss functions when evaluating forecasts).

ORCID

Dimitris Korobilis https://orcid.org/0000-0001-9146-3008

REFERENCES

1. Yanchenko AK, Tierney G, Lawson J, Hellmayr C, Cron A, West M. Multivariate dynamic modeling for bayesian forecasting of business revenue. *Appl Stoch Model Bus Ind*. 2023 (forthcoming).
2. West M, Harrison J. *Bayesian Forecasting and Dynamic Models*. Springer Science & Business Media; 1996.
3. Berry LR, West M. Bayesian forecasting of many count-valued time series. *J Bus Econ Stat*. 2020;38(4):872-887.
4. Koop G, Korobilis D. Bayesian dynamic variable selection in high dimensions. *Int Econom Rev*. 2023 (forthcoming).
5. Harvey AC. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press; 1990.
6. Kulhavý R, Zarrop MB. On a general concept of forgetting. *Int J Control*. 1993;58(4):905-924.
7. Zhou X, Nakajima J, West M. Bayesian forecasting and portfolio decisions using dynamic dependent sparse factor models. *Int J Forecast*. 2014;30(4):963-980.
8. West M. Bayesian forecasting of multivariate time series: scalability, structure uncertainty and decisions. *Ann Inst Stat Math*. 2020;72(1):1-31.
9. Gruber L, West M. GPU-accelerated Bayesian learning and forecasting in simultaneous graphical dynamic linear models. *Bayesian Anal*. 2016;11(1):125-149.
10. Gruber LF, West M. Bayesian online variable selection and scalable multivariate volatility forecasting in simultaneous graphical dynamic linear models. *Econom Stat*. 2017;3:3-22.
11. Fan J, Feng Y, Wu Y. Network exploration via the adaptive LASSO and SCAD penalties. *Ann Appl Stat*. 2009;3(2):521-541.

How to cite this article: Korobilis D, Montoya-Blandón S. Discussion of “Multivariate dynamic modeling for Bayesian forecasting of business revenue”. *Appl Stochastic Models Bus Ind*. 2023;39(3):315-317. doi: 10.1002/asmb.2753