Discussion of “Multivariate Dynamic Modeling for Bayesian Forecasting of Business Revenue” by A.K.Yanchenko et al

Dawid Bernaciak, Jim Griffin, Ioanna Manolopoulou

November 2022

1 Introduction

We would like to congratulate the authors on this work, as well as their broader contributions to the field of business and retail analytics. Retail analytics is an application area with very rich and complex data which has lead to a number of recent works in the marketing literature (Dube and Rossi, 2019; Rooderkerk et al., 2022), but has not received sufficient attention from the statistics community despite posing numerous interesting methodological challenges.

Some of the primary challenges in retail data are the high dimensionality, the complex dependencies across time (Wang et al., 2019), space (Wang et al., 2021) and products (Ruiz et al., 2020), as well as the different levels of groupings both at the product level (brands, categories, sub-categories), the customer level (returning customers, households) and store level (both store type as well as geographical location). Leveraging information and dependencies across different levels of resolution while maintaining computational efficiency is methodologically difficult. The authors' approach of de-coupling and re-coupling in the Bayesian predictive synthesis framework in this and related works (Berry et al., 2020; Berry and West, 2020; West, 2020) balances the two objectives seamlessly and effectively.

The work opens up a variety of new questions and potential extensions.

2 The effects of aggregation

The de-coupling/re-coupling approach is used to estimate model parameters (such as regression coefficients) on data aggregated over all LSGs using a Dynamic Linear Model (DLM). These estimated parameters are included as multipliers of the regressors in LSG-specific DLMs. This approach works particularly well since the sales are fairly homogeneous across LSGs. However, aggregating time series with very different properties can have substantially different time series properties to individual time series (in an extreme the aggregate of AR(1) processes with different autoregressive parameters can have long memory). It would be interesting to know how the level of heterogeneity across the LSGs would affect the performance of the de-coupling/re-coupling approach.

Another drawback of the de-coupling/re-coupling approach is that there is aggregation across all products within each category using averages of generic covariates such as net price and promotion summaries at each LSG. However, product assortment (i.e., the set of available products at a particular store) may vary significantly from one store to another, reflecting marketing decisions based on local characteristics. Although full assortment details might be both computationally prohibitive as well as too high-resolution to be useful, simple summaries can easily be constructed; for example, proportion of products labelled “organic”, “luxury”, or “economy” would be expected to change in response to local demographics. Introducing assortment into the DLM can shed light onto how it interacts with LSG sales – potentially also linking sales and revenue to local characteristics.

3 Including demand modelling

The de-coupling/re-coupling approach fits a multivariate time series model of K LSGs, using $K+1$ univariate DLMs, which clearly has substantial computational benefits. However, the sales of each category of products are modelled separately. Prediction of revenue is closely linked to prediction of sales, which is often known as product choice modelling or demand estimation in the marketing literature.

There is a natural distinction between models which use aggregate store-level data on sales/revenue of each product and basket data of individual shopping trips. We will first describe models for aggregate store-level data before turning our attention to models for basket data. Models for store-level data have been successfully applied to large numbers of products and are usually regression-based, which fits naturally within the DLM framework favoured by the authors. Models for basket data allow a more complete and detailed understanding of customer behaviour but with the price of more complicated modelling and inference.

Regression models for store-level data fall in two main categories: logit linear models (where market share of each product is modelled using a multivariate logistic regression) or log-linear models (where the logarithm of sales are a linear function of price of the product, the prices of other products and other factors, such as promotion). These models have the benefit that regression coefficients of other product prices can be interpreted as elasticities. This allow cross-product effect to be estimated using the model rather than recovered post-hoc through residual analysis as described in the paper.

In a similar way to the paper, Bayesian hierarchical models provide natural ways to leverage information across stores, products or product categories. Hierarchical models to encourage shrinkage of elasticities across stores in log-linear models have been considered by Blattberg and George (1991), Montgomery (1997) and Boatwright et al. (1999). More recently, Smith et al. (2021) has shown how allowing economic theory to guide shrinkage can lead to better predictive performance than ad-hoc choices.

A recent strand of the literature on store-level data has considered how these models can be scaled to larger numbers of categories or products. Bajari et al. (2015) use a range of machine learning methods methods in logit and log linear models. Gabel and Timoshenko (2022) use deep learning for a wide assortments spanning multiple categories. Ruiz et al. (2020) use a word embedding approach whereas Donnelly et al. (2021) use matrix factorization. More in keeping with the DLM approach of the paper is Smith and Griffin (2022) who show how a high-dimensional log-linear model can be estimated using horseshoe-based shrinkage priors at a product rather than category level using a product classification tree. One benefit of shrinkage models in this context is that products whose price does not vary or very rarely varies in the sample can be included rather than removed from the analysis as is commonly done (see *e.g.* Donnelly et al., 2021).

The approach in the paper predicts revenue effectively without needing to reach the finest resolution of data, i.e., basket data. Although studying pairwise correlations at the category level is valuable, explicit modelling of purchasing choices can directly capture interaction between different product options. This is especially crucial in online grocery shopping as the retailer has the opportunity to suggest complementary items and rank potential substitutes. However, modelling baskets in a dynamic model directly, for example through substitution models (see Hoang and Breugelmans (2022) for a review), or topic models at category (Hruschka, 2014; Jacobs et al., 2016) or product level (Vega Carrasco et al., 2022), is currently not feasible; but a decouple-recouple approach of building a dynamic model on the coarse resolution and feeding it into a refined model on the high resolution might enable meaningful inclusion of temporal dynamics into the process.

Along with revenue forecasting, commercial interest also lies in understanding how different characteristics of products (such as brand; quantity; type; ingredients etc) impact sales. These have become particularly important as online shopping, where shoppers can filter products according to particular attributes, has gained significant traction in recent years. Including summaries of product characteristics should be possible within the proposed framework. Of course, studying the impact of assortment as well as product characteristics is sensitive to endogeneity, so requires careful exploration of possible effects at play (Ruiz et al., 2020).

Product characteristics play a larger role in retail sectors such as electronic goods (Sindhu et al., 2021), where products can be categorised based on common attributes (such as screen size; colour; resolution; brand; age of model). These sectors also see a more frequent introduction and removal of products into the market, which feeds into product assortment within each LSG. The hierarchical formulation naturally lends itself to introducing a new product at time t through borrowing of information across LSG and through the covariate coefficients of the new product. Studying the introduction of new products into the market also allows counterfactual exploration, which is commercially valuable when considering the launch of new products and the need to model potential cannibalization.

4 Full uncertainty modelling

Early in the text the authors list three important features that the model shall have. Two of those are: the possibility of the user intervention in informed ways and the full characterisation of the forecast uncertainties. In this section we will discuss how the approach outlined by the authors addresses these two points.

Throughout the analysis the authors choose to use the mean absolute percentage error (MAPE) which measures the percentage change is accuracy between the actual and the forecast values to assess the performance of the models. MAPE is a popular, intuitive and interpretable metric, however, it comes with some well-documented drawbacks such as asymmetry of negative and positive errors and distributional interpretation of the optimal values (see Hyndman and Koehler (2006) and Makridakis (1993) for more detailed studies and suggestion for alternative metrics). From the Bayesian point of view, MAPE is a point forecast loss metric, which means that it does not account for the forecast uncertainty output by the model. Extending assessment of model performance beyond MAPE would offer further insights into the business impact in terms of the revenue generated.

In terms of estimation of model uncertainty the paper suggests a "plug-in" method for the forecast propagation of the effect of in the aggregate model. The authors note that this MAP-like method understates the uncertainty in the revenue forecasts as it ignores the uncertainty about the aggregate discount effects. The authors note that this does not have a significant practical impact; however, it's not clear to what extent this would still hold under a different evaluation metric, using appropriate scores and coverage statistics.

Another interesting angle at which the uncertainty evaluation might be performed is the measurement of effect on the decision making. In light of the recent works on the forecasting using focused scoring (Loaiza-Maya et al., 2021) which reflects the business goals this evaluation could potentially pave a way of explicitly incorporating the probabilistic model forecasts in the business analysis.

We recall that the authors examine two years worth of weekly revenue data - which is the target variable - supplemented by the information about the prevalent discounts and promotions - which are used as the model features. The analysis of the revenue data clearly shows strong seasonality effects with revenue peaking around major holidays such as Thanksgiving and Christmas. Indeed, the model performance is always weakest around these major events as the model is not capable of learning the seasonality effect from the data. It is not surprising, just two years of data might not provide enough of seasonal information for a model to calibrate. Authors recognised this and proposed an expert intervention (West and Harrison (1989), West and Harrison (2006)) where each holiday information is encoded ahead of time and the outcomes are treated as missing observations. Such an intervention lead to a large reduction in MAPE which might indicate that the poor model performance around the major holidays in the paper is big enough to considerably skew the overall loss metric and mask the real performance differences. Therefore, one might argue that the results in the paper would have benefited from this adjustment for holidays from the outset.

Furthermore, the poor model performance around the major holidays causes spikes in the model volatility and which then take time to dissipate. This appears to be an unwanted artifact of the model which might have been remedied by the intervention approach outlined above. Further analysis of the model performance including the evaluation of the full forecast distribution would be of much interest.

References

Bajari, P., D. Nekipelov, S. P. Ryan, and M. Yang (2015). Demand estimation with machine learning and model combination. NBER Working Paper No. 20955.

Berry, L. R., P. Helman, and M. West (2020). Probabilistic forecasting of heterogeneous consumer transaction–sales time series. *International Journal of Forecasting 36*(2), 552–569.

Berry, L. R. and M. West (2020). Bayesian forecasting of many count-valued time series. *Journal of Business & Economic Statistics 38*(4), 872–887.

Blattberg, R. C. and E. I. George (1991). Shrinkage estimation of price and promotional elasticities: Seemingly unrelated equations. *Journal of the American Statistical Association 86*, 304–315.

Boatwright, P., R. McCulloch, and P. Rossi (1999). Account=level modeling for trade promotion: an application of a constrained parameter hierarchical model. *Journal of the American Statistical Association 94*, 1063–1073.

Donnelly, R., F. R. Ruiz, D. Blei, and S. Athey (2021). Counterfactual inference for consumer choice across many product categories. *Quantitative Marketing and Economics 19*, 369–407.

Dube, J.-P. and P. E. Rossi (2019). *Handbook of the Economics of Marketing*. North Holland.

Gabel, S. and A. Timoshenko (2022). Product choice with large assortments: a scalable deep-learning model. *Management Science 68*, 1808–1827.

Hoang, D. and E. Breugelmans (2022+). "sorry, the product you ordered is out of stock": Effects of substitution policy in online grocery retailing. *Journal of Retailing to appear*.

Hruschka, H. (2014). Linking multi-category purchases to latent activities of shoppers: analysing market baskets by topic models. *Journal of Research and Management 36*, 267–273.

Hyndman, R. J. and A. B. Koehler (2006). Another look at measures of forecast accuracy. *International journal of forecasting 22*(4), 679–688.

Jacobs, B. J., B. Donkers, and D. Fok (2016). Model-based purchase predictions for large assortments. *Marketing Science 35*, 389–404.

Loaiza-Maya, R., G. M. Martin, and D. T. Frazier (2021). Focused Bayesian prediction. *Journal of Applied Econometrics 36*, 517–543.

Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International journal of forecasting 9*(4), 527–529.

Montgomery, A. L. (1997). Creating micro-marketing pricing strategies using supermarket scanner data. *Marketing Science 16*, 315–337.

Rooderkerk, R. P., N. DeHoratius, and A. Musalem (2022). The past, present, and future of retail analytics: Insights from a survey of academic research and interviews with practitioners. *Production and Operations Management 31*(10), 3727–3748.

Ruiz, F. J., S. Athey, and D. M. Blei (2020). Shopper: A probabilistic model of consumer choice with substitutes and complements. *The Annals of Applied Statistics 14*(1), 1–27.

Sindhu, C., D. Mukherjee, and Sonakshi (2021). A joint sentiment-topic model for product review analysis of electronic goods. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 574–578.

Smith, A. N. and J. E. Griffin (2022+). Shrinkage priors for high-dimensional demand estimation. *Quantitative Marketing and Economics to appear*.

Smith, A. N., P. E. Rossi, and G. M. Allenby (2021). Inference for product competition and separable demand. *Marketing Science 38*, 690–710.

Vega Carrasco, M., I. Manolopoulou, J. O'Sullivan, R. Prior, and M. Musolesi (2022). Posterior summaries of grocery retail topic models: Evaluation, interpretability and credibility. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 71*(3), 562–588.

Wang, C., M. Zhang, W. Ma, Y. Liu, and S. Ma (2019). Modeling item-specific temporal dynamics of repeat consumption for recommender systems. In *The World Wide Web Conference*, pp. 1977–1987.

Wang, X., J. Pancras, and D. K. Dey (2021). Investigating emergent nested geographic structure in consumer purchases: a bayesian dynamic multi-scale spatiotemporal modeling approach. *Journal of Applied Statistics 48*(3), 410–433.

West, M. (2020). Bayesian forecasting of multivariate time series: scalability, structure uncertainty and decisions. *Annals of the Institute of Statistical Mathematics 72*(1), 1–31.

West, M. and J. Harrison (1989). Subjective intervention in formal models. *Journal of Forecasting 8*(1), 33–53.

West, M. and J. Harrison (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.