

Forecasting intermittent demand with dynamic Poisson mixtures: a score-driven approach

Rodrigo Sarlo

Department of Electrical Engineering, PUC-Rio
Rua Marquês de São Vicente, 225, Rio de Janeiro, RJ, 22451-900
rodrigo-sarlo@aluno.puc-rio.br

Cristiano Fernandes

Department of Electrical Engineering, PUC-Rio
cris@ele.puc-rio.br

Denis Borenstein

Management School, Federal University of Rio Grande do Sul
Rua Washington Luiz 855, Porto Alegre, RS, 90010-460
denis.borenstein@ufrgs.br

ABSTRACT

Forecasting intermittent demand has become increasingly important in retail, leading to the development and adoption of new forecasting methods like the score-driven (SD) approach. In this study, we propose a dynamic SD Poisson mixture model for intermittent time series forecasting, formulated within a general dynamic mixture model framework. The proposed model accommodates overdispersed data while preserving parameter recursions analogous to those of the standard SD Poisson model. We evaluate the model using real-world Walmart sales data, benchmarking its performance against existing SD-based forecasting models from the literature. Our results indicate that while the mixture model outperforms the basic SD Poisson model, it does not surpass the predictive accuracy of the SD negative binomial model.

KEYWORDS. Forecasting. Intermittent data. Score-driven model.

EST&MP - Statistics and Probabilistic Models, AD&GP - OR in Administration and Production Management, OA - Other Applications in OR

1. Introduction

Intermittent time series are characterized by sequences of positive values interspersed with zeros, and are commonly observed in the demand for spare parts and retail products at disaggregated levels. The seminal work of Croston [1972] has initiated the literature on forecasting and inventory control for such series, which proposed a method specifically designed to address the unique challenges of intermittent demand. Croston's method addresses the forecasting of intermittent time series by decomposing the problem into two separate components: the prediction of demand sizes (conditional on a positive observation) and the estimation of the number of periods between successive non-zero demands. Each component is forecasted independently using simple exponential smoothing (SES) [Brown, 1959], with the same smoothing parameter applied to both processes. Due to the high value of spare parts and the significant consequences that stock-out events can have for consumers, early applications of intermittent demand forecasting naturally concentrated on inventory control for spare parts. This emphasis is reflected in the frequent use of the Royal Air Force (RAF) dataset as a benchmark in early research, a tradition that continues to shape more recent studies [Eaves and Kingsman, 2004; Teunter and Duncan, 2009; Petropoulos and Kourentzes, 2015].

The expansion of intermittent demand forecasting into the retail sector has been driven by the increasing need to manage demand at highly granular levels, such as the stock-keeping unit (SKU)-store-day level. This shift reflects the growing recognition of the critical role that slow-moving items play in inventory management. As pointed out by Johnston and Boylan [1996] and Syntetos et al. [2016], a substantial portion of companies' total stock investments is tied to these items. More recently, Makridakis et al. [2022] demonstrated the scale of this phenomenon, showing that approximately 63% of Walmart's time series in the M5 competition could be classified as slow-moving, a proportion that increases to 90% at the SKU-store-day level.

The growing importance of intermittent demand forecasting methods is driven by their prevalence in typical supply chain settings and the financial impact that improvements in forecasting accuracy can yield. The relevance of this area in the broader forecasting landscape is evidenced by an entire subsection dedicated to it in the review by Petropoulos et al. [2022]. The authors divide the main contributions in this literature into *parametric* and *non-parametric* methods. In what follows, we focus on parametric methods, as the contribution of this paper builds upon and extends this family of approaches.

Most contributions to parametric methods for intermittent demand forecasting are, in some way, adaptations of simple exponential smoothing (SES), enhancing the original approach proposed by Croston. Syntetos and Boylan [2005] introduced a correction to Croston's method that achieves approximate unbiasedness under the assumed data-generating process (DGP) that motivated its development. Teunter et al. [2011] suggested forecasting the probability of demand rather than the interval between demands, which enables updates at every period; unlike earlier methods, forecasts are only updated following the occurrence of demand. Building on this, Snyder et al. [2012] proposed recurrently applying SES equations to update the mean of discrete probability distributions. The importance of generating full distributional forecasts, rather than solely point forecasts, is emphasized by Syntetos et al. [2016] and Kolassa [2016], particularly given that the primary application of these forecasts is in inventory control. Except for the approach by Snyder et al. [2012], the methods above require additional adaptations to produce quantile forecasts necessary for calculating safety stocks. The most commonly used technique for this purpose is based on a normal approximation [Teunter and Sani, 2009; Kourentzes, 2013].

A more recent contribution to probabilistic forecasting methods was made by Sarlo et al. [2023], who proposed the application of generalized autoregressive score (GAS) or score-driven

(SD) models to intermittent demand forecasting. Initially developed independently by Creal et al. [2013] and Harvey [2013], SD models provide a flexible framework for constructing dynamic models for non-Gaussian distributions, whether discrete or continuous. Using real sales data from a large retailer in Brazil, Sarlo et al. [2023] demonstrated the strong performance of their proposed models, particularly those based on the negative binomial distribution, as well as the hurdle Poisson and hurdle negative binomial models. Notably, the negative binomial model performed comparably to, and in some cases better than, alternative distributions, highlighting its suitability for intermittent demand settings. Unfortunately, the SD Poisson model did not perform well compared to other SD and traditional intermittent demand forecasting models.

Given the relevance of the Poisson distribution for modeling count data, as well as its flexibility and simplicity [Uglickich and Nagy, 2025], this study aims to enhance the score-driven Poisson model by investigating its mixture-based formulations. Specifically, we propose a dynamic Poisson mixture model within the framework introduced by Catania [2019], which develops a general setup for mixture models, referred to as dynamic mixture models (DAMMs). Our Poisson mixture score-driven (SD) model can handle overdispersed data - one of the key strengths of the negative binomial distribution - while retaining the parameter recursions similar to those of the standard Poisson model. The proposed model is evaluated using real-world data from slow-moving items at Walmart, which were part of the M5 Competition [Makridakis et al., 2022]. The idea of using Poisson mixtures to address overdispersion is well established in the literature, with earlier studies such as Wedel et al. [1993] and Viallefont et al. [2002] demonstrating the effectiveness of mixture models for predicting overdispersed count data, an advantage we also expect to be helpful in the context of intermittent demand forecasting.

This article is organized as follows. The next section presents the SD framework and its extension for mixture distributions proposed by Catania [2019]. The third section discusses the empirical exercise conducted to evaluate the proposed model; The fourth section compares the results with other SD-based models. Finally, Section 5 concludes the paper and suggests directions for future research.

2. Score-driven models

The score-driven or GAS framework [Creal et al., 2013; Harvey, 2013] provides a unified methodology to derive time-varying parameter models for non-Gaussian distributions. The key element of this framework is the (scaled) score of the assumed probability distribution, which drives the variation in the parameters of the models. This section presents SD models, focusing on intermittent demand forecasting, while more general formulations are indicated.

The first building block of a score-driven model is the assumed probability distribution for the time series of interest. Let y_t , where $t = 1, \dots, T$, be the time series of interest, which is assumed to follow a probability density/mass function, $y_t \sim p(y_t | f_{t|t-1}, \mathcal{F}_{t-1}; \theta)$. $f_{t|t-1}$ is a vector that collects the parameters of the conditional distribution of y_t assumed to be time-varying, $\mathcal{F}_{t-1}$ represents the information available at time t , and θ is a vector containing fixed parameters.

The recursive equations of the time-varying parameters in this work follow a local level model, one of the possible specifications in the unobserved components form, as in Caivano et al. [2016]:

$$h_f(f_{t|t-1}) = m_{t|t-1} \quad (1)$$

$$m_{t|t-1} = m_{t-1|t-2} + \kappa \tilde{s}_{f,t-1}, \quad \kappa > 0 \quad (2)$$

Equation (1) states that each element of the time-varying vector $f_{t|t-1}$, after suitable parametrization via h_f , is driven by a specific unobserved component, $m_{t|t-1}$, which in the present

context represents the level of the time series. More specifically, Equation (2) defines the law of motion assumed for the level component $m_{t|t-1}$, namely, a random walk. The element $\tilde{s}_{f,t}$ is the reparameterized scaled score that drives the dynamics of $f_{t|t-1}$, acting as an innovation in the model equations - further details on this will be provided in the subsequent paragraphs. In broader contexts, such as in Caivano et al. [2016], this same element is responsible for driving the dynamics of all unobserved components of the model, resembling the single-source error state space model discussed in Hyndman et al. [2008].

The fixed parameter κ and the initial value of the recursion in (2), m_0 , are collected in the vector θ and are estimated by maximum likelihood, which can be readily obtained since $p(y_t|f_{t|t-1}, \mathcal{F}_{t-1}; \theta)$ has a closed form. Alternatively, one can heuristically define initial values for the recursion and condition the estimation and filtered components on these values.

The model presented in Equations (1)-(2) is a specialized version of Sarlo et al. [2023], where the authors added more unobserved components and also explanatory variables. Another possible formulation uses ARIMA-like recursions for the time-varying parameters, as in Creal et al. [2013].

The final component of the specification of the SD framework is the reparameterized scaled score. We begin by defining the scaled score in Equation (3), which is the suitable driver of the components' dynamics when $h_f(\cdot)$ is the identity link. $\nabla_t = \frac{\partial(y_t|f_{t|t-1}, \mathcal{F}_{t-1}; \theta)}{\partial f_{t|t-1}}$ is the score of the assumed distribution for y_t , and $\mathcal{I}_{t|t-1} = E_{t-1}[\nabla_t \nabla_t']$ is Fisher information matrix. When $d = \frac{1}{2}$, $\mathcal{I}_{t|t-1}^{1/2} \stackrel{\text{def}}{=} \mathcal{J}_{t|t-1}$, the latter being obtained from the Cholesky decomposition $\mathcal{I}_{t|t-1} = \mathcal{J}_{t|t-1}' \mathcal{J}_{t|t-1}$.

$$s_t = \mathcal{I}_{t|t-1}^{-d} \cdot \nabla_t, \quad d = 0, 0.5, 1. \quad (3)$$

After defining the scaled score, the reparametrization follows from a direct application of the chain rule. The link function $h_f(\cdot)$ is employed to ensure that the elements of $f_{t|t-1}$ remain in the suitable region for the distribution of y_t . Natural use cases of link functions, other than the identity, are when the parameter we are interested in making time-varying is a probability, which must lie in the $[0, 1]$ interval, and the mean for count data distributions, which is typically greater than zero. In these cases, we define an auxiliary vector h , which contains the elements of the mapping $h_f(f_{t|t-1}) = \tilde{f}_{t|t-1}$, where $\tilde{f}_{t|t-1} \in \mathbb{R}$ and $f_{t|t-1}$ is a restricted parameter. Then, if we define a matrix $\dot{h}_t = \frac{\partial h(f_{t|t-1})}{\partial f_{t|t-1}}|_{f_{t|t-1}}$, it follows that $\tilde{\nabla}_t = \partial \ln p(y_t|\tilde{f}_{t|t-1}, \mathcal{F}_{t-1}; \theta) / \partial \tilde{f}_{t|t-1} = \dot{h}_t^{-1} \nabla_t$, and $\tilde{\mathcal{I}}_{t|t-1} = E_{t-1}[\tilde{\nabla}_t \tilde{\nabla}_t']$. Finally, we obtain the reparameterized scaled score as:

$$\tilde{s}_t = \begin{cases} \dot{h}^{-1} \nabla_t & \text{if } d = 0; \\ \mathcal{J}_{t|t-1} \nabla_t & \text{if } d = \frac{1}{2}; \\ \dot{h} \mathcal{I}_{t|t-1}^{-1} \nabla_t & \text{if } d = 1. \end{cases} \quad (4)$$

In the empirical exercise, we set $d = 0$, following the discussions in Sarlo [2021] and Sarlo et al. [2023]. Another important choice to derive the specific expressions of $\tilde{s}_t$ in Equation (4) is the assumed distribution for y_t . Natural candidates for intermittent data are the Poisson and negative binomial distributions, along with their hurdle and zero-inflated counterparts, which we use as the benchmarks to the proposed Poisson mixture model defined in the ongoing paragraphs. In the hurdle and zero-inflated specifications, the binomial trial that governs the occurrence of a zero demand also follows a random walk, allowing for time-varying probabilities. We refer the reader to Sarlo [2021] for the details of the derivation and final expressions of $\tilde{s}_t$.

2.1. Dynamic Poisson Mixture

Following the score-driven (SD) framework discussion, we present the extension to mixture models proposed by Catania [2019]. Mixture models arise in various contexts in statistics, including model-based clustering and density estimation [Bishop, 2006; Hastie et al., 2001]. As before, we focus our presentation on the model specification for intermittent demand forecasting. As in other applications of mixture distributions, we assume that the variable of interest is drawn from a convex combination of underlying probability distributions:

$$p(y_t | f_{1,t|t-1}, \dots, f_{J,t|t-1}, \mathcal{F}_{t-1}; \theta) = \sum_{j=1}^J \omega_j p_j(y_t | f_{j,t|t-1}, \mathcal{F}_{t-1}; \theta_j) \quad (5)$$

In Equation (5), we let each $p_j(y_t | f_{j,t|t-1}, \mathcal{F}_{t-1}; \theta_j)$ component follow a score-driven model. The mixture weights $\omega_j \in (0, 1)$ satisfy $\sum_{j=1}^J \omega_j = 1$, and define the full parameters set as $\theta = (\theta_j, \omega_j, j = 1 \dots J)$. Unlike the original formulation, we do not allow the mixture weights ω_j to vary over time; instead, they are estimated via maximum likelihood alongside the other elements of θ . Although not required, it is common practice to assume that all component distributions p_j in the mixture belong to the same family, differing only in their parameter values.

After defining the mixture distribution for our variable of interest y_t , it is easy to obtain the new score of the distribution of the mixture model:

$$\nabla_t^j = \omega_j \frac{p_j(y_t | f_{j,t|t-1}, \mathcal{F}_{t-1}; \theta_j)}{p(y_t | f_{1,t|t-1}, \dots, f_{J,t|t-1}, \mathcal{F}_{t-1}; \theta)} \nabla_{j,t} \quad (6)$$

where $\nabla_{j,t}$ is the standard score of the j th component. For the dynamic Poisson mixture, $\nabla_{j,t}$ has the same formula as that of the standard Poisson SD score: $\nabla_{j,t} = \frac{\partial \ln p_j(y_t | f_{j,t|t-1}, \mathcal{F}_{t-1}; \theta_j)}{\partial f_{j,t|t-1}} = \frac{y_t - f_{j,t|t-1}}{f_{j,t|t-1}}$, where $f_{j,t|t-1}$ is the mean of the j th Poisson component. The reader is referred to Catania [2019] for details on the Fisher information matrix and the reparameterization of the scaled score in mixture models.

The new time-varying term $\xi_{t,j} = \omega_j \frac{p_j(y_t | f_{j,t|t-1}, \mathcal{F}_{t-1}; \theta_j)}{p(y_t | f_{1,t|t-1}, \dots, f_{J,t|t-1}, \mathcal{F}_{t-1}; \theta)}$ in Equation (6) is responsible for weighting the parameter updates for each component of the mixture. In cases where one component captures regular demand patterns while others account for atypical or irregular events, this term helps to stabilize the updates by preventing large adjustments to the score of the regular-demand component when unusual observations occur. As a result, Equation (6) can be interpreted as a robust filtering mechanism, similar to the approach discussed in Blasques et al. [2024].

Previous studies successfully applied Poisson mixtures to overdispersed count data, such as Wedel et al. [1993] and Viallefont et al. [2002]. Building on this, we expect that the ability to handle overdispersion, combined with the robustness introduced by $\xi_{j,t}$ to the component's updates, can lead to improvements over the standard score-driven Poisson model.

In our empirical exercise, we follow the same specification as the benchmark models by setting $d = 0$ in Equation (4). When applying the score-driven mixture Poisson model to real data, we estimate versions with two to five mixture components. The final model used in the forecasting exercise is selected based on the Bayesian Information Criterion (BIC).

3. Empirical study

This section presents the experiment to assess whether the proposed score-driven (SD) Poisson mixture model improves upon the standard score-driven Poisson model. We first describe the dataset used, then discuss the evaluation of forecasting accuracy, and finally present the methodology for model comparison.

3.1. Data

To evaluate the performance of the dynamic Poisson mixture model, we conduct an empirical study using the dataset from the M5 competition [Makridakis et al., 2022]. This dataset, provided by Walmart, contains daily SKU-store sales and possible aggregations, encompassing 42,840 time series. The importance of generating forecasts at multiple aggregation levels for modern retail operations is highlighted by Fildes et al. [2022]. However, our study focuses on the most disaggregated level - SKU-store-day - where slow-moving series are highly prevalent.

At the chosen aggregation level, a total of 30,490 series are available. We applied filters to discard ill-behaved time series to select the final set of series for our model comparison. Following the classification proposed by Syntetos et al. [2005], we retained only those series with an average demand interval (ADI) greater than or equal to 1.32 days, thereby focusing exclusively on intermittent and lumpy demand patterns. As a final filtering step, we excluded any series in which the longest sequence of zero sales exceeded 28 consecutive days. Given the absence of explanatory variables capturing inventory levels, this last criterion is a proxy for identifying potential supply issues.

After applying all filtering criteria, the final dataset consists of 4,430 time series. We removed each series's initial sequence of zero sales to avoid estimating models on periods preceding the product's introduction at a given store. The final 28 days of each series were reserved for out-of-sample evaluation, consistent with the setup of the M5 competition, while the remaining observations were used for parameter estimation. This procedure results in in-sample periods ranging from 96 to 1,941 days, depending on the timing of the first recorded sale.

3.2. Forecasting accuracy

All models derived within the score-driven framework naturally lead to the generation of density forecasts, which are essential for inventory control since safety stock levels depend on the upper quantiles of predictive distributions. In this context, the variable of interest is typically the lead time (or cumulative) demand over a certain period. We assess forecast performance for lead times $L = 1$ and $L = 5$. Following the approach of Czado et al. [2009] and Kolassa [2016], we evaluate the quality of these forecasts using proper scoring rules. The definitions presented below are adapted from Kolassa [2016].

Scoring rules are a class of loss functions that map a density forecast $\hat{p}$ and the realization of a random variable y to a value. A scoring rule is considered proper if its expected value is minimized when the forecasted density matches the true distribution of y . It is strictly proper if this minimum is uniquely achieved when $\hat{p}$ equals the true distribution of y . We follow Sarlo et al. [2023] and use the Brier score when evaluating the predictive densities, as follows:

$$Brier(\hat{p}, y) = -2\hat{p}_y + \|\hat{p}\|^2 \quad (7)$$

where $\|\hat{p}\| = \sqrt{\sum_{k=0}^{\infty} \hat{p}_k^2}$. The term $\hat{p}_y$ represents the probability assigned to the value y by the predictive density $\hat{p}$.

3.3. Model comparison

We now describe the statistical procedure for comparing the proposed models across multiple time series. When evaluating two models on a single series, various comparison methods are available, with the Diebold-Mariano test for differences in forecast accuracy being among the most widely used. However, applying such tests repeatedly across many series leads to the multiple comparisons problem, increasing the likelihood of spurious significance. To address this issue, we adopt the methodology proposed by Demsar [2006], which is briefly outlined in the following paragraphs.

The procedure begins by calculating the forecasting accuracy metric for each dataset-model pair over the forecasting horizon of interest. For each dataset and horizon, models are ranked based on relative performance. The Friedman test, a nonparametric alternative to repeated-measures ANOVA, is applied to assess whether there are significant differences in average rankings across all datasets for a given forecasting horizon. If the null hypothesis of equal rankings is rejected, the Nemenyi post-hoc test is then used to identify which pairs of models differ significantly in performance. Ultimately, this procedure results in groups of models between which no statistically significant differences in performance can be detected.

4. Results

We begin the discussion of results by presenting the outcome of the multiple comparison procedure proposed by Demsar [2006]. For both lead time demand horizons evaluated, $L = 1$ and $L = 5$, the Friedman test rejected the null hypothesis of equal model rankings at the 5% significance level. This indicates that, for each lead time, at least one forecasting model performs statistically differently. The models under comparison are the score-driven Poisson (POISSON), negative binomial (NB), zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB) models, hurdle Poisson (HP), and hurdle negative binomial (HNB), along with the proposed Poisson mixture (MIXTURE) specification.

It is common to visualize the results of the Nemenyi post-hoc test in a plot, as shown in Figure 1. The horizontal axis displays the average ranking of each score-driven model across the 4,430 time series. A bold horizontal line connects models for which the null hypothesis of equal forecasting accuracy cannot be rejected at the 5% level, based on the Nemenyi test. For example, in Panel (a), although the ZINB model has a slightly better average ranking than the NB model, the difference is not statistically significant. Both the Friedman and Nemenyi tests, as well as the plot in Figure 1, were generated using the `scmamp` package [Calvo and Santafé, 2016].

The results from both panels of Figure 1 indicate that the dynamic Poisson mixture model does improve upon the standard Poisson model. However, somewhat unexpectedly, it still fails to outperform the simpler negative binomial (NB) model. For $L = 1$, the ZINB and NB models form a top-performing group with no statistically significant difference between them. For $L = 5$, all models except the standard Poisson perform similarly well.

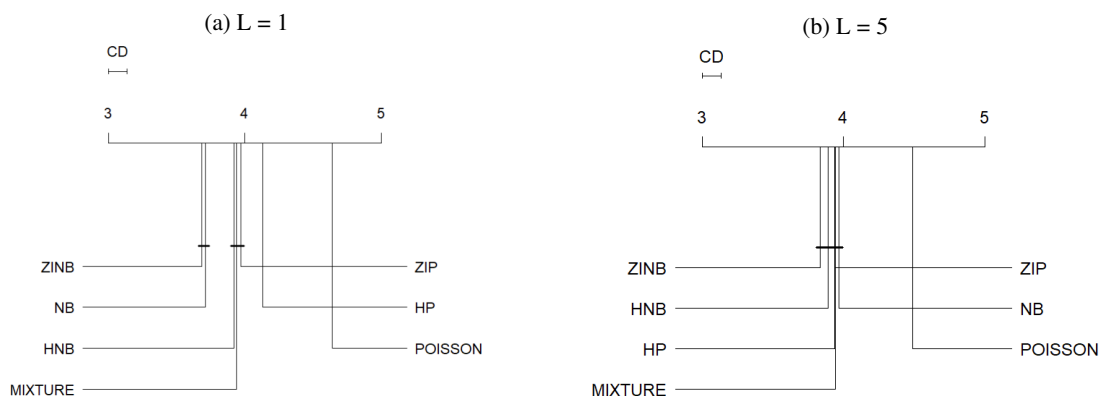


Figure 1: Comparison of model ranks

Two possible hypotheses may explain the superior performance of the negative binomial (NB) model compared to the dynamic Poisson mixture. First, the NB model benefits from its parsimony. In the current implementation, each component added to the Poisson mixture introduces

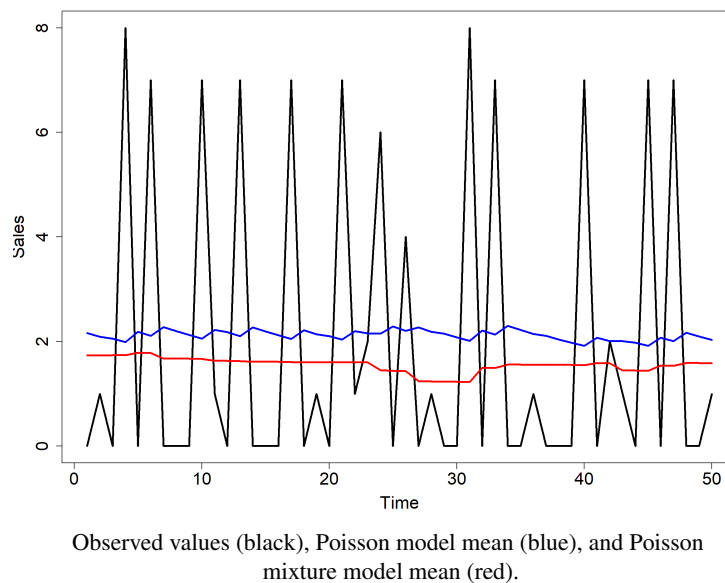
three additional parameters: (i) a mixture weight w_j ; (ii) a parameter associated to the reparameterized scaled score, κ in Equation (2); and (iii) a level initialization, m_0 . The parsimony of the NB specification may be particularly suited to modeling noisy time series, as is often the case in disaggregated sales data. In 4,410 of the 4,430 time series tested (more than 99% of the time), a model with $J = 2$ components was selected via BIC, supporting this hypothesis. A second possibility is that the NB model might be better equipped to handle the overdispersion observed in intermittent demand data.

We now focus on understanding when and why the dynamic Poisson mixture model outperforms the standard Poisson model. Figure 2 illustrates the last 50 in-sample days for one of the time series showing the largest out-of-sample performance gap, measured by the Brier score, between the two models. The figure also highlights two key hypotheses that may explain this difference in performance.

The first explanation relates to the role of the new time-varying term $\xi_{t,j}$ introduced by the mixture model - see Equation (6). As previously discussed, this term weights the parameter updates for each component, effectively adding robustness to the filtering mechanism. In Figure 2, this behavior is evident when contrasting the more erratic filtered mean from the Poisson model (blue) with the smoother, more stable estimates produced by the mixture model (red).

A second explanation for the improvements observed in the mixture model is the possibility that different mixture components capture distinct features of the true distribution of y_t . The standard Poisson model imposes strong constraints by tying together the mean, variance, and the probability of observing a zero through a single parameter, λ_t . This rigid structure can limit its ability to model intermittent and highly variable demand patterns accurately. In contrast, by combining multiple components, the mixture model offers greater flexibility to adjust for overdispersion and the frequency of zero observations, potentially leading to a better overall fit and more accurate forecasts. This decoupling between different characteristics of the distribution of y_t from λ_t likely explains the differences in the filtered mean levels observed in Figure (2).

Figure 2: Comparison of filtered means



To better understand the conditions under which the proposed model outperforms the standard Poisson SD model, we investigate the relationship between time series characteristics and the performance gains achieved by the mixture specification. Specifically, we focus on the two characteristics typically used to classify demand time patterns: the squared coefficient of variation (CV^2) and the average demand interval (ADI) [Syntetos et al., 2005]. Table 1 reports the estimated coefficients from a logistic regression model, where the dependent variable is an indicator that is equal to 1 if the Poisson mixture achieves a lower average Brier score than the standard Poisson model for $L = 1$. The regressors were log-transformed to improve the model fit, as measured by BIC.

Table 1: Logit Coefficient Estimates

	Intercept	$\log(CV^2)$	$\log(ADI)$
Estimate	1.162	0.951	0.386

Note: All regressors were statistically significant at the 1% level.

The results reported in Table 1 support our hypothesis that the empirical improvement achieved by using the mixture Poisson model is related to its ability to model different features of the true distribution of y_t . The likelihood of selecting the Poisson mixture model as the best one is increasing in both CV^2 and ADI. Higher values of CV^2 indicate greater dispersion relative to the mean, while higher ADI values are associated with more intermittent demand patterns. These are precisely the conditions under which the rigid structure of the standard Poisson model - where a single parameter simultaneously governs the mean, variance, and probability of zero demand - becomes most limiting. In contrast, the flexibility of the mixture model in capturing different distributional features allows it to better adapt to such data characteristics.

5. Conclusion

This paper proposed a score-driven Poisson mixture model, based on the developments of Catania [2019]. The score-driven framework from Creal et al. [2013] and Harvey [2013] was previously applied in the context of intermittent demand forecasting with success in Sarlo et al. [2023]. Motivated by the empirical findings of the last paper - that the score-driven negative binomial model substantially improves upon the standard Poisson model without the added complexity of hurdle or zero-inflated distributions - we explored whether a SD Poisson mixture model could offer similar or even superior benefits.

Although the negative binomial SD model remained the best-performing alternative overall, consistent with previous findings in the literature, the proposed dynamic Poisson mixture model showed meaningful improvements over the standard Poisson SD specification. The results demonstrate that introducing a flexible mixture structure within the SD framework can significantly enhance forecast accuracy, as measured by the Brier score. Discussing our empirical findings, we offered hypotheses for why the mixture model achieves substantial gains over the standard Poisson SD model. Further, we explained the mixture model limitations compared to the negative binomial distribution, highlighting the roles of model flexibility, parsimony, and the nature of overdispersion in the data.

While this study has been limited to the class of score-driven models, the results obtained suggest several promising avenues for future research. One direction is to incorporate covariates into the dynamic Poisson mixture model, potentially improving forecast accuracy by capturing additional sources of variation. This, however, poses a significant challenge, as the number of parameters grows with each added covariate and mixture component. Another possibility is to extend the

evaluation to an inventory simulation setting, which would provide a more direct assessment of the model's practical value for inventory control by linking forecast improvements to cost reductions and service level gains. Finally, the proposed dynamic Poisson mixture model should be compared with the extensive literature on intermittent demand forecasting. This includes recent extensions of the single-source error state space model to intermittent data, as proposed by Svetunkov and Boylan [2023], as well as machine learning approaches, such as those discussed in Lolli et al. [2017], to better guide practitioners in selecting models suited to their observed demand patterns.

References

- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg. ISBN 0387310738.
- Blasques, F., van Brummelen, J., Gorgi, P., and Koopman, S. J. (2024). Maximum likelihood estimation for non-stationary location models with mixture of normal distributions. *Journal of Econometrics*, 238(1):105575. ISSN 0304-4076.
- Brown, R. (1959). *Statistical Forecasting for Inventory Control*. McGraw-Hill.
- Caivano, M., Harvey, A., and Luati, A. (2016). Robust time series models with trend and seasonal components. *Journal of the Spanish Economic Association*, 7(4):99–120.
- Calvo, B. and Santafé, G. (2016). SCMAMP: Statistical comparison of multiple algorithms in multiple problems. *The R Journal*, 8:1–8.
- Catania, L. (2019). Dynamic adaptive mixture models with an application to volatility and risk*. *Journal of Financial Econometrics*, 19(4):531–564. ISSN 1479-8409.
- Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.
- Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly (1970-1977)*, 23(3):289–303.
- Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics*, 65 4:1254–61.
- Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30.
- Eaves, A. H. and Kingsman, B. G. (2004). Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society*, 55(4):431–437.
- Fildes, R., Ma, S., and Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*.
- Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Society Monographs. Cambridge University Press.
- Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc.

- Hyndman, R., Koehler, A., Ord, K., and Snyder, R. (2008). *Forecasting with Exponential Smoothing. The State Space Approach*. Springer Series in Statistics. Springer.
- Johnston, F. and Boylan, J. E. (1996). Forecasting for items with intermittent demand. *Journal of the Operational Research Society*, 47(1):113–121.
- Kolassa, S. (2016). Evaluating predictive count data distributions in retail sales forecasting. *International Journal of Forecasting*, 32(3):788–803.
- Kourentzes, N. (2013). Intermittent demand forecasts with neural networks. *International Journal of Production Economics*, 143(1):198–206.
- Lolli, F., Gamberini, R., Regattieri, A., Balugani, E., Gatos, T., and Gucci, S. (2017). Single-hidden layer neural networks for forecasting intermittent demand. *International Journal of Production Economics*, 183:116–128.
- Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2022). The M5 competition: Background, organization, and implementation. *International Journal of Forecasting*, 38(4):1325–1336.
- Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., De Baets, S., Dokumentov, A., Ellison, J., Fiszeder, P., Franses, P. H., Frazier, D. T., Gilliland, M., Góñi, M. S., Goodwin, P., Grossi, L., Grushka-Cockayne, Y., Guidolin, M., Gunter, U., Guo, X., Guseo, R., Harvey, N., Hendry, D. F., Hollyman, R., Januschowski, T., Jeon, J., Jose, V. R. R., Kang, Y., Koehler, A. B., Kolassa, S., Kourentzes, N., Leva, S., Li, F., Litsiou, K., Makridakis, S., Martin, G. M., Martinez, A. B., Meeran, S., Modis, T., Nikolopoulos, K., Ónkai, D., Paccagnini, A., Panagiotelis, A., Panapakidis, I., Pavão, J. M., Pedio, M., Pedregal, D. J., Pinson, P., Ramos, P., Rapach, D. E., Reade, J. J., Rostami-Tabar, B., Rubaszek, M., Sermpinis, G., Shang, H. L., Spiliotis, E., Syntetos, A. A., Talagala, P. D., Talagala, T. S., Tashman, L., Thomakos, D., Thorarindottir, T., Todini, E., Trapero Arenas, J. R., Wang, X., Winkler, R. L., Yusupova, A., and Ziel, F. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, 38(3): 705–871.
- Petropoulos, F. and Kourentzes, N. (2015). Forecast combinations for intermittent demand. *Journal of the Operational Research Society*, 66(6):914–924.
- Sarlo, R. (2021). Intermittent demand forecasting in retail: applications of the GAS framework. Master's thesis, Department of Electrical Engineering, PUC-RJ, Rio de Janeiro, Brazil.
- Sarlo, R., Fernandes, C., and Borenstein, D. (2023). Lumpy and intermittent retail demand forecasts with score-driven models. *European Journal of Operational Research*, 307(3):1146–1160.
- Snyder, R. D., Ord, J. K., and Beaumont, A. (2012). Forecasting the intermittent demand for slow-moving inventories: A modelling approach. *International Journal of Forecasting*, 28(2): 485–496.
- Svetunkov, I. and Boylan, J. E. (2023). iets: State space model for intermittent demand forecasting. *International Journal of Production Economics*, 265:109013. ISSN 0925-5273. URL <https://www.sciencedirect.com/science/article/pii/S0925527323002451>.

- Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., and Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, 252(1):1–26. ISSN 0377-2217.
- Syntetos, A. A. and Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting*, 21(2):303–314.
- Syntetos, A. A., Boylan, J. E., and Croston, J. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society*, 56(5):495–503.
- Teunter, R. and Sani, B. (2009). Calculating order-up-to levels for products with intermittent demand. *International Journal of Production Economics*, 118(1):82–86.
- Teunter, R., Syntetos, A., and Babai, M. Z. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research*, 214(3):606–615.
- Teunter, R. H. and Duncan, L. (2009). Forecasting intermittent demand: a comparative study. *Journal of the Operational Research Society*, 60(3):321–329.
- Uglickich, E. and Nagy, I. (2025). Poisson-based framework for predicting count data: Application to traffic counts in prague areas. *Journal of Computational Science*, p. 102534.
- Viallefont, V., Richardson, S., and and, P. J. G. (2002). Bayesian analysis of poisson mixtures. *Journal of Nonparametric Statistics*, 14(1-2):181–202.
- Wedel, M., Desarbo, W. S., Bult, J. R., and Ramaswamy, V. (1993). A latent class poisson regression model for heterogeneous count data. *Journal of Applied Econometrics*, 8(4):397–411.