Biometrika (2025), **112**, 3, asaf028 https://doi.org/10.1093/biomet/asaf028
Advance Access publication 12 April 2025

Dynamic factor analysis of high-dimensional recurrent events

By F. CHEN

Department of Statistics, Columbia University,
1255 Amsterdam Avenue, New York, New York 10027, U.S.A.
fc2630@columbia.edu

Y. CHEN

Department of Statistics, London School of Economics and Political Science,
Houghton Street, London WC2A 2AE, U.K.
y.chen186@lse.ac.uk

Z. YING

Department of Statistics, Columbia University,
1255 Amsterdam Avenue, New York, New York 10027, U.S.A.
zying@stat.columbia.edu

and K. ZHOU

Department of Statistics, Columbia University,
1255 Amsterdam Avenue, New York, New York, 10027, U.S.A.
kz2326@columbia.edu

Summary

Recurrent event time data arise in many studies, including in biomedicine, public health, marketing and social media analysis. High-dimensional recurrent event data involving many event types and observations have become prevalent with advances in information technology. This article proposes a semiparametric dynamic factor model for the dimension reduction of high-dimensional recurrent event data. The proposed model imposes a low-dimensional structure on the mean intensity functions of the event types while allowing for dependencies. A nearly rate-optimal smoothing-based estimator is proposed. An information criterion that consistently selects the number of factors is also developed. Simulation studies demonstrate the effectiveness of these inference tools. The proposed method is applied to grocery shopping data, for which an interpretable factor structure is obtained.

Some key words: Counting process; Factor analysis; Information criterion; Kernel smoothing; Marginal modelling.

1. Introduction

As information technology advances, high-dimensional recurrent event data are becoming increasingly common. For example, such data are commonly seen in market basket

© 2025 Biometrika Trust
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025

analysis, which often tracks customers' purchasing behaviour over time to develop personalized recommendation strategies. In this case, each customer can be viewed as an observation unit, and their shopping history can be viewed as a multivariate counting process, wherein the elements of the process correspond to a large number of merchandise items and the event times correspond to the times at which the items are purchased. Another example is text data from social media platforms (e.g., Liang et al., 2018; Bogdanowicz & Guan, 2022). In such data, a user's dynamics correspond to a multivariate counting process, where event times record the occurrence of words or phrases in posts. The user dynamics are often analysed for user profiling, opinion mining, or understanding and predicting the information cascade on a social media platform. High-dimensional recurrent event data also arise in human-computer interactions such as simulated problem-solving tasks in educational assessment (Chen, 2020), where event times are the time stamps of different types of actions. Data of a similar structure also occur in medicine and public health, finance, and insurance (e.g., Sun, 2006; Cook & Lawless, 2007; Yang, 2022).

We propose a dynamic factor model for analysing high-dimensional recurrent event time data. The model introduces low-dimensional time-varying factors in a continuous time domain to capture the dynamic trends underlying a multivariate counting process while keeping the constant event-type-specific parameters, known as the loadings, to strengthen interpretability. The model is specified based on only the mean rate functions (Lin et al., 2000), allowing for a flexible conditional dependence structure among the processes. This is crucial in applications such as consumer shopping behaviour analysis, where recurrent events could be highly dependent owing to population heterogeneity. Model identification is studied, based on which rotation methods for exploratory factor analysis (Browne, 2001; Rohe & Zeng, 2023) can be applied to the model to obtain an interpretable factor structure. Simultaneous estimation of factors and loadings is proposed based on a kernel-smoothed pseudolikelihood function. We further propose an information criterion for determining the number of factors. Desirable asymptotic properties are established as the number of event types and the sample size grow to infinity. In particular, we show that the proposed information criterion consistently selects the number of factors, and the estimation is consistent and nearly rate-optimal. The proposed method is applied to a large grocery shopping dataset. The analysis finds interpretable customer factors that provide insight into grocery shopping behaviours.

The proposed method is related to frailty models for recurrent event data (e.g., Abu-Libdeh et al., 1990; Chen et al., 2005). These models introduce correlated event-type-specific random effects (frailties) into the intensity functions to capture the dependence among events. With many event types, the traditional frailty model has to introduce many random effects and specify their joint distribution, making the model specification and parameter estimation challenging. The proposed model is also related to dynamic factor models for irregularly spaced longitudinal data (Lu et al., 2015; Tang et al., 2017; Chen & Zhang, 2020), where the dynamic factors are treated as stochastic processes and Bayesian or empirical Bayesian inferences are performed. The proposed method may also be viewed as an extension of high-dimensional factor analysis methods (Bai & Li, 2012; Wang et al., 2019; Chen et al., 2020, 2021; He et al., 2023; Liu et al., 2023). In these methods, the latent factors are treated as unknown parameters rather than random variables during parameter estimation, which avoids distributional assumptions on the latent factors and makes the estimation computationally more affordable. Based on this estimation framework, information criteria are developed for determining the number of factors (Bai & Ng,

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025

2002; Chen & Li, 2022). The present work is similar in spirit but involves a more challenging task of estimating low-dimensional functions of dynamic factors.

For a matrix $\mathbf{X} = (x_{ij})_{N\times J}$, let $||\mathbf{X}||_{\mathrm{F}} = (\sum_{i,j} x_{ij}^2)^{1/2}$ and $||\mathbf{X}||_{2\to\infty} = \sup_{||\alpha||_2=1} ||X\alpha||_\infty$ denote its Frobenius norm and two-to-infinity norm, respectively. For two real numbers a and b, we write $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$. For two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n \ll b_n$ or, equivalently, $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$, write $a_n = O(b_n)$ (or $a_n \lesssim b_n$) if there is a positive constant M independent of n such that $|a_n| \leqslant M|b_n|$ for all n, and write $a_n \asymp b_n$ if there are two positive constants M_1 and M_2 independent of n such that $M_1|b_n| \leqslant |a_n| \leqslant M_2|b_n|$. We use the standard $O_{\mathrm{p}}(\cdot)$ notation for stochastic boundedness in probability. We let $L^2_{N\times J}[0,1] = \{(f_{ij}(t))_{N\times J} : 0 \leqslant t \leqslant 1, ||f_{ij}||_{L^2[0,1]} < \infty \text{ for all } i \text{ and } j\}$ be the space of $(N \times J)$-dimensional square-integrable matrix-valued functions on $[0, 1]$.

2. Proposed method

2.1. *Model*

Consider multivariate recurrent event data from N independent observation units on a standardized time interval $[0, 1]$. The data from observation unit i can be described by $\mathbf{Y}_i(t) = (Y_{i1}(t), \ldots, Y_{iJ}(t))^{\mathrm{T}}$, where J is the number of event types and each component $Y_{ij}(t)$ is a right-continuous counting process. We introduce a factor model to reduce the dimensionality of the data and to further identify and interpret the factors underlying the observed processes. A marginal modelling approach (Lin et al., 2000) is adopted to accommodate a more flexible conditional dependence structure among the processes. This approach specifies the mean rate function for each event type j as

$$E\{\mathrm{d}\,Y_{ij}(t)\} = f\{X_{ij}(t)\}\,\mathrm{d}t, \tag{1}$$

where $f : \mathbb{R} \to [0, \infty)$ is a prespecified link function and $X_{ij}(t)$ is an unknown function with a low-dimensional structure. Specifically, $X_{ij}(t)$ is parameterized as

$$X_{ij}(t) = \sum_{k=1}^{r} a_{jk}\theta_{ik}(t), \tag{2}$$

where the $\theta_{ik}(\cdot)$ are functions that may be interpreted as unobserved dynamic factors, the a_{jk} are referred to as the loading parameters, and r is the number of factors. We let $\mathbf{\Theta}(t) = (\theta_{ik}(t))_{N\times r}$, $\mathbf{A} = (a_{jk})_{J\times r}$ and $\mathbf{X}(t) = (X_{ij}(t))_{N\times J}$. Rewriting (2) in matrix form, we have $\mathbf{X}(t) = \mathbf{\Theta}(t)\mathbf{A}^{\mathrm{T}}$, where both $\mathbf{\Theta}(\cdot)$ and $\mathbf{A}$ are to be estimated.

Remark 1 (*Link function*). The link function f is needed to ensure that the mean rate function is nonnegative. For simplicity, we let f be known and set $f(x) = \exp(x)$ in the numerical analysis. Extensions to the setting with unknown f can be obtained by estimating the link function nonparametrically, for example via nonnegative basis function approximations.

Remark 2 (*Intensity formulation*). As an alternative to the mean rate specification (1), one can model the intensity functions as $E\{\mathrm{d}\,Y_{ij}(t)|\mathcal{F}_t\} = f\{X_{ij}(t)\}\,\mathrm{d}t$ for a suitable right-continuous filtration $\{\mathcal{F}_t\}_{0\leqslant t\leqslant 1}$ that leads to a martingale structure (Andersen et al., 1993). As pointed out by Lin et al. (2000), the mean rate specification (1) is more versatile than the intensity specification in that it allows arbitrary dependence structures among recurrent

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025

events. For example, in the analysis of customer purchasing behaviour, multiple merchandise items may be purchased simultaneously and thus have the same event time. When analysing user dynamics on a social media platform, multiple words or phrases often appear in the same post and thus have the same event time. The intensity specification implies independent increments, i.e., $\mathrm{d}Y_{ij}(t)$ $(j = 1, \ldots, J)$ are conditionally independent given $\mathcal{F}_t$. As a result, $\mathrm{d}Y_{ij}(t) = 1$ can only occur to one of the J event types for a specific time t, which does not align with the real-world situations mentioned previously. The mean rate specification does not have this restriction.

Remark 3 (*Connection with factor models*). The proposed model is closely related to the Poisson factor model for count data. Consider a special case of (1) where the $Y_{ij}(t)$ $(j = 1, \ldots, J)$ are independent Poisson processes with static factors θ_{ik}, i.e., $f\{X_{ij}(t)\} = f\big(\sum_{k=1}^{r} a_{jk}\theta_{ik}\big)$. In this case, the counts $\{Y_{ij}(1) : i = 1, \ldots, N; j = 1, \ldots, J\}$ constitute a sufficient statistic for the unknown parameters, with the $Y_{ij}(1)$ following a Poisson distribution with rate $f\big(\sum_{k=1}^{r} a_{jk}\theta_{ik}\big)$. This model for count data is known as the Poisson factor model (Wedel & Kamakura, 2001; Chen et al., 2020), where the a_{jk} are known as the loading parameters and the θ_{ik} are interpreted as the unobserved factors. In this sense, the proposed model (1) and (2) can be viewed as an extension of the Poisson factor model. The Poisson factor model can be estimated by a constrained joint maximum likelihood estimator (Chen et al., 2020), which is consistent and minimax rate-optimal under suitable regularity conditions. Our model is also closely related to matrix factor models (Wang et al., 2019; He et al., 2023) in that the data at each time can be viewed as a matrix. A major difference is that our model is posed on a continuous time domain, with the observed data being very sparse at each time. In contrast, the matrix factor models assume a discrete time domain.

Remark 4 (*Indeterminacy of* $\boldsymbol{\Theta}(\cdot)$ *and* $\mathbf{A}$ *and a rotated solution*). In our model, $\boldsymbol{\Theta}(\cdot)$ and $\mathbf{A}$ are not determined, in the sense that for any $r \times r$ invertible matrix $\mathbf{Q}$, the model remains unchanged if we replace the factors by $\boldsymbol{\Theta}(t)(\mathbf{Q}^{\mathrm{T}})^{-1}$ and the loadings by $\mathbf{AQ}$. Similar indeterminacies occur in other factor models (see, e.g., Bai & Li, 2012). To interpret the factor structure, one must fix the transformation $\mathbf{Q}$, which may be done by using an analytic rotation method (Browne, 2001; Rohe & Zeng, 2023). However, the current setting is slightly different from standard exploratory factor analysis settings, as the factors here are functions of time t. To apply existing analytic rotation methods, we could first aggregate the factors by calculating $\bar{\boldsymbol{\Theta}} = \int_0^1 \boldsymbol{\Theta}(t)\,\mathrm{d}t$ and then apply an analytic rotation method to $\bar{\boldsymbol{\Theta}}\mathbf{A}^{\mathrm{T}}$. In the real data analysis in § 5, a varimax rotation method (Kaiser, 1958; Rohe & Zeng, 2023) is applied to fix the transformation.

Remark 5 (*Time-varying loadings*). The flexibility of the model can be further enhanced by letting the loading parameters be time-varying, i.e., $X_{ij}(t) = \sum_{k=1}^{r} a_{jk}(t)\theta_{ik}(t)$. However, this model is far less determined than the current model as $\mathbf{X}(t) = \boldsymbol{\Theta}(t)\{\mathbf{A}(t)\}^{\mathrm{T}} = \boldsymbol{\Theta}(t)\{\mathbf{Q}(t)^{\mathrm{T}}\}^{-1}\{\mathbf{A}(t)\mathbf{Q}(t)\}^{\mathrm{T}}$ for any $r \times r$ invertible matrix-valued function $\mathbf{Q}(t)$. Determining this transformation function $\mathbf{Q}(t)$ is more challenging than determining the time-independent transformation discussed in Remark 4. Consequently, it is hard to identify and interpret the factor structure. In our grocery shopping application, each event type corresponds to a merchandise item, and each observation corresponds to a customer. In this context, the loading parameters can be viewed as a summary of item characteristics, and the

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025

factors can be interpreted as a summary of customer preferences. Because item characteristics tend to be stable while customer preferences often vary over time, treating the loading parameters as static and the factors as dynamic is sensible. Therefore, this work focuses on the static loading and dynamic factor setting.

2.2. *Estimation*

We introduce a kernel-based approach to estimating the unknown parameters. Kernel smoothing borrows information from nearby time-points because the observed events are very sparse at each single time-point in the continuous time domain. Let $K(x)$ be a kernel function with sufficient smoothness, satisfying $K(x) \geqslant 0$, $K(-x) = K(x)$ and $\int_{-\infty}^{\infty} K(x)\,\mathrm{d}x = 1$. For a smoothing bandwidth $h > 0$, we further define $K_h(x) = (1/h)K(x/h)$. We consider the following kernel-smoothed pseudolikelihood function:

$$\mathcal{L}_h(\Theta, \mathbf{A}) = \sum_{i=1}^{N}\sum_{j=1}^{J}\int_h^{1-h}\left[\frac{\int_0^1 K_h(t-s)\,\mathrm{d}Y_{ij}(s)}{\int_0^1 K_h(t-s)\,\mathrm{d}s}\log f\{X_{ij}(t)\} - f\{X_{ij}(t)\}\right]\mathrm{d}t, \tag{3}$$

where $X_{ij}(t)$ is a function of $\boldsymbol{\Theta}$ and $\mathbf{A}$ as defined in (2). We consider the parameter space $\mathcal{G} = \{(\Theta, \mathbf{A}) : \sup_{t\in[0,1]} ||\Theta(t)||_{2\to\infty} \leqslant M^{1/2},\ ||\mathbf{A}||_{2\to\infty} \leqslant M^{1/2}\}$, where $M > 0$ is a prespecified constant. We define $(\hat{\Theta}, \hat{\mathbf{A}})$ as a constrained maximizer of (3),

$$(\hat{\Theta}, \hat{\mathbf{A}}) \in \arg\max \mathcal{L}_h(\Theta, \mathbf{A}) \text{ such that } (\Theta, \mathbf{A}) \in \mathcal{G}. \tag{4}$$

Since the parameter space $\mathcal{G}$ is compact and $\mathcal{L}_h(\Theta, \mathbf{A})$ is continuous with respect to the norm $||(\Theta, \mathbf{A})|| := \sup_{t\in[0,1]} ||\Theta(t)||_{2\to\infty} \vee ||\mathbf{A}||_{2\to\infty}$, the existence of at least one solution is guaranteed. Therefore, $(\hat{\Theta}, \hat{\mathbf{A}})$ is well-defined.

Remark 6. The pseudolikelihood (3) ignores the possible dependence between event types that is allowed under the mean rate specification (1). If (1) is replaced by an intensity specification, i.e., $E\{\mathrm{d}Y_{ij}(t)|\mathcal{F}_t\} = f\{X_{ij}(t)\}\,\mathrm{d}t$, and we let h go to 0, then (3) becomes the loglikelihood function for recurrent event time data (Cook & Lawless, 2007).

Remark 7. In practice, we can only obtain an approximate solution to (4), as the optimization involves infinite-dimensional functions. When the resolution of the approximation is carefully chosen, this approximate solution can achieve the same error rate as that of (4). More specifically, the approximate solution is obtained by a two-step procedure. In the first step, we discretize the interval $[h, 1-h]$ by equally spaced grid points $t_1, \ldots, t_q$ and solve

$$\begin{aligned}(\tilde{\boldsymbol{\Theta}}(t_1), \ldots, \tilde{\Theta}(t_q), \tilde{\mathbf{A}}) \in \arg\max\ & \mathcal{L}_h\{\boldsymbol{\Theta}(t_1), \ldots, \boldsymbol{\Theta}(t_q), \mathbf{A}\} \\ & \text{such that } ||\boldsymbol{\Theta}(t_l)||_{2\to\infty} \leqslant M,\ ||\mathbf{A}||_{2\to\infty} \leqslant M\ (l = 1, \ldots, q),\end{aligned}$$

where

$$\begin{aligned}&\mathcal{L}_h\{\boldsymbol{\Theta}(t_1), \ldots, \boldsymbol{\Theta}(t_q), \mathbf{A}\} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{l=1}^{q}\left[\frac{\int_0^1 K_h(t_l-s)\,\mathrm{d}Y_{ij}(s)}{\int_0^1 K_h(t_l-s)\,\mathrm{d}s}\log f\{X_{ij}(t_l)\} - f\{X_{ij}(t_l)\}\right]\end{aligned}$$

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025

is the pseudolikelihood defined on the grid points. In the second step, based on $\tilde{\boldsymbol{\Theta}}$ we find an approximation to $\hat{\boldsymbol{\Theta}}$ on $[h, 1-h]$ by interpolation, such as a linear interpolation. By choosing the number of grid points to be inversely proportional to the error rate of (4), the approximate solution is guaranteed to achieve the same error rate. An efficient projected gradient descent algorithm is developed to obtain the approximate solution. This algorithm handles the constraints based on the two-to-infinity norm with an easy-to-compute projection operator. The details of the algorithm and its convergence properties are given in the Supplementary Material.

2.3. *Determining the number of factors*

In practice, the number of factors r is unknown and needs to be chosen. We propose an information criterion for choosing r. To avoid ambiguity, we let $(\hat{\boldsymbol{\Theta}}^{(r)}, \hat{\mathbf{A}}^{(r)})$ denote the estimator (4) to emphasize its dependence on the number of factors. The proposed information criterion takes the form $\text{IC}(r) = -2\mathcal{L}_h(\hat{\boldsymbol{\Theta}}^{(r)}, \hat{\mathbf{A}}^{(r)}) + v(N, J, r)$, where $v(N, J, r)$ is a penalty term that increases with N, J and r. The conditions on $v(N, J, r)$ for consistent model selection will be determined in § 3.2. Given $v(N, J, r)$, we choose the number of factors by $\hat{r} = \arg\min_{r \in \mathcal{R}} \text{IC}(r)$, where $\mathcal{R} \subset \mathbb{N}$ is a candidate set for the number of factors. As shown in § 3, under suitable conditions on the penalty term and additional regularity conditions, $\hat{r}$ consistently selects the number of factors. In our implementation, the pseudolikelihood $\mathcal{L}_h(\hat{\boldsymbol{\Theta}}^{(r)}, \hat{\mathbf{A}}^{(r)})$ in $\text{IC}(r)$ is replaced by its discretized version as discussed in Remark 7.

3. THEORETICAL PROPERTIES

3.1. *Consistency and rate of convergence*

We present our main theoretical results about the estimator proposed in § 2.2. When deriving these results, the number of factors is assumed to be correctly specified. To avoid ambiguity of notation, we let $\boldsymbol{\Theta}^*(\cdot)$ and $\mathbf{A}^*$ denote the true parameters and further let $\mathbf{X}^*(t) = \boldsymbol{\Theta}^*(t)(\mathbf{A}^*)^{\mathrm{T}}$. To avoid the complications arising from the indeterminacy of $\boldsymbol{\Theta}(\cdot)$ and $\mathbf{A}$, we focus on evaluating the estimation accuracy of $\hat{\mathbf{X}}(t) := \hat{\boldsymbol{\Theta}}(t)\hat{\mathbf{A}}^{\mathrm{T}}$. Let $m \geqslant 1$ be a positive integer. We assume the following regularity conditions.

Condition 1. The link function f is m times continuously differentiable. Moreover, for $x \in [-M, M]$, $f(x)$ and $f'(x)$ are bounded away from 0.

Condition 2. The matrix function $\mathbf{X}^*(\cdot) \in \mathcal{G}$ is m times continuously differentiable on $[0, 1]$.

Condition 3. The kernel function K satisfies the following properties: (i) it is a Lipschitz function of order m with compact support on $[-1, 1]$; (ii) it attains its unique maximum at $x = 0$; (iii) it is twice continuously differentiable in a neighbourhood of 0 and $(\log K)''(0) < 0$.

Condition 4. (i) The multivariate point processes $\{Y_{1j}(\cdot) : j \in [J]\}, \ldots, \{Y_{Nj}(\cdot) : j \in [J]\}$ are independent. (ii) There exists $\lambda > 0$ such that for any i, j, k and $0 < s_1 < \cdots < s_k < 1$, $E\{\mathrm{d}Y_{ij}(s_1) \cdots \mathrm{d}Y_{ij}(s_k)\} \leqslant \lambda^k \, \mathrm{d}s_1 \cdots \mathrm{d}s_k$. (iii) For any i, there exists a partition $B_{i,1}, \ldots, B_{i,W_i}$ of $\{1, \ldots, J\}$ and a function $\phi(J) = o(J)$ satisfying $\max_{i=1,\ldots,N} \max_{k=1,\ldots,W_i} |B_{i,k}| \leqslant \phi(J)$ such that $\{Y_{ij}(\cdot) : j \in B_{i,1}\}, \ldots, \{Y_{ij}(\cdot) : j \in B_{i,W_i}\}$ are independent.

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025

Remark 8. In Condition 1, both $f(x)$ and $f'(x)$ are assumed to be nonzero in $[-M, M]$. This requirement is rather mild. In particular, it is automatically satisfied when $f(x)$ is strictly positive and monotone increasing (or decreasing), including when $f(x) = \exp(x)$.

Remark 9. Condition 2 requires the true model to lie in the same parameter space $\mathcal{G}$ as the one used to regularize our estimator (4). This requirement, together with Condition 1, implies that the mean rate function $f\{X_{ij}(t)\}$ is nonnegative and uniformly bounded away from zero for all i, j and t. In the context of our grocery shopping application, it means that the proposed method is most suitable for analysing customers who shop frequently and items that are frequently purchased. The size of this parameter space plays a key role in the theory about model estimation and selection. Condition 2 also imposes a smoothness requirement that is standard in nonparametric regression models (Györfi et al., 2002).

Remark 10. Parts (i) and (ii) of Condition 3 are standard assumptions for kernel functions. Condition 3(iii) assumes log-concavity at the maximum point.

Remark 11. Condition 4(i) assumes that the observed data $\mathbf{Y}_i(t) = (Y_{i1}(t), \ldots, Y_{iJ}(t))^{\mathrm{T}}$ are independent across observational units $i = 1, \ldots, N$. Condition 4(ii) assumes nondegeneracy of the counting process. Condition 4(iii) assumes a blockwise-independent structure, which substantially relaxes the independence assumption among different event types. We restrict the maximum block size rather than assuming the blockwise-independent structure to be the same across observations $i = 1, \ldots, N$. Condition 4(iii) can be further relaxed. Instead of requiring the processes in all the blocks to be independent, the results in Theorems 1 and 3 are still valid if only $\{Y_{ij}(\cdot) : j \in B_{i,1}\}, \ldots, \{Y_{ij}(\cdot) : j \in B_{i,W_i-1}\}$ are independent. This relaxed condition allows the processes in B_{i,W_i} to be dependent on all the rest of the processes. Condition 4(iii) can be seen as a condition for the low-rank structure $\mathbf{X}(t)$ to be identifiable, as it excludes the noise in the data having a low-rank structure through the independence or blockwise-independence assumption. The assumption of blockwise independence is similar in spirit to the weakly dependent error assumption adopted in approximate factor models (e.g., Chamberlain & Rothschild, 1983; Bai & Ng, 2023) and may be seen as a version of the weakly dependent error assumption for factor models of high-dimensional recurrent event data.

THEOREM 1 (UPPER BOUND). *Under Conditions 1–4, the following results hold.*

(i) *(Dependent case) Assume $J = O(N)$ and recall $\phi(J)$ from Condition 4(iii). For any $\delta > 0$, choose $h \asymp \{J/\phi(J)\}^{-1/(2m+1)+\delta/m}$. Then, as N and J go to infinity,*

$$\frac{1}{NJ}\int_h^{1-h} \left\|\hat{\mathbf{X}}(t) - \mathbf{X}^*(t)\right\|_{\mathrm{F}}^2 \mathrm{d}t = O_{\mathrm{p}}\left[\{J/\phi(J)\}^{-m/(2m+1)+\delta}\right].$$

(ii) *(Independent case) Assume that $\phi(J) = 1$ in Condition 4(iii) and $\log(N \vee J) \ll N \wedge J$. For any $\delta > 0$, choose $h \asymp [(N \wedge J)/\{\log^2(N \wedge J)\}]^{-1/(2m+1)}$. Then, as N and J go to infinity,*

$$\frac{1}{NJ}\int_h^{1-h} \left\|\hat{\mathbf{X}}(t) - \mathbf{X}^*(t)\right\|_{\mathrm{F}}^2 \mathrm{d}t = O_{\mathrm{p}}\left\{(N \wedge J)^{-2m/(2m+1)+\delta}\right\}.$$

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025

To show the near optimality of the proposed estimator, we derive the minimax lower bound under the independent Poisson process setting in the following theorem.

THEOREM 2 (LOWER BOUND). *Assume that the Y_{ij} are independent Poisson point processes and $\phi(J) = 1$ in Condition 4(iii). Further assume that $\sup_{|x| \leqslant M} f'(x)^2/f(x) < \infty$. Then there is an absolute constant $C > 0$ such that for any estimator $\hat{\mathbf{X}}(t) \in L^2_{N\times J}[0,1]$ there exists an $\mathbf{X}^*(t)$ satisfying Condition 2 such that*

$$\mathrm{pr}\left\{\frac{1}{NJ}\int_0^1 \left\|\hat{\mathbf{X}}(t) - \mathbf{X}^*(t)\right\|_{\mathrm{F}}^2 \mathrm{d}t \geqslant C(N \wedge J)^{-2m/(2m+1)}\right\} \geqslant \frac{1}{2}.$$

Hence, this information-theoretic lower bound matches the upper bound in Theorem 1 only up to an arbitrarily small exponent under the independence assumption, which implies the near minimax optimality of our estimator.

Remark 12. When the blockwise-independent structure is the same across observations (i.e., $W_i = W$ and $B_{i,w} = B_w$ for $i = 1, \ldots, N$ and $w = 1, \ldots, W$), we can sharpen the rate in Theorem 1(i) from $-m/(2m+1) + \delta$ to $-2m/(2m+1) + \delta$ and establish its near minimax optimality.

Remark 13. Because of the rotational indeterminacy mentioned in Remark 4, the estimated loading matrix $\hat{\mathbf{A}}$ is not guaranteed to converge to the true loading matrix $\mathbf{A}^*$. However, it can be shown that the maximum principal angle between the column spaces of $\hat{\mathbf{A}}$ and $\mathbf{A}^*$ converges to zero in probability under the same conditions as in Theorem 1 and an additional regularity condition on the singular values of $\mathbf{X}^*(t)$. See the Supplementary Material for more details.

3.2. *Model selection consistency*

As introduced in § 2.3, $v(N, J, r)$ is the penalty function in the information criterion. As $v(N, J, r)$ is required to be increasing in r, we write $u(N, J, r) = v(N, J, r) - v(N, J, r-1) > 0$. Further, for any $t \in [0, 1]$, let $\sigma_{1,t} \geqslant \sigma_{2,t} \geqslant \cdots \geqslant \sigma_{r^*,t}$ be the nonzero singular values of $\mathbf{X}^*(t)$. Theorem 3 provides sufficient conditions on $u(N, J, r)$ for consistent model selection.

THEOREM 3 (MODEL SELECTION CONSISTENCY). *Assume that the candidate set $\mathcal{R}$ has a finite number of elements and $r^* \in \mathcal{R}$. Under Conditions 1–4, the following results hold.*

(i) *(Dependent case) Assume $J = O(N)$ and that the function u satisfies $u(N, J, r) = o\big(\int_h^{1-h} \sigma^2_{r^*,t}\,\mathrm{d}t\big)$ and $NJ\{J/\phi(J)\}^{-m/(2m+1)+\delta} = o\{u(N, J, r)\}$ for any sufficiently small $\delta > 0$ and any $r \in \mathcal{R}$ as N and J go to infinity. Choose $h \asymp \{J/\phi(J)\}^{-1/(2m+1)+\delta/m}$. Then $\lim_{N,J\to\infty} \mathrm{pr}(\hat{r} = r^*) = 1$.*

(ii) *(Independent case) Assume that $\phi(J) = 1$ in Condition 4(iii), $\log(N \vee J) \ll N \wedge J$, and the function u satisfies $u(N, J, r) = o\big(\int_h^{1-h} \sigma^2_{r^*,t}\,\mathrm{d}t\big)$ and $NJ\,(N \wedge J)^{-2m/(2m+1)+\delta} = o\{u(N, J, r)\}$ for any sufficiently small $\delta > 0$ and any $r \in \mathcal{R}$ as N and J go to infinity. Choose $h \asymp [(N \wedge J)/\{\log^2(N \wedge J)\}]^{-1/(2m+1)}$. Then $\lim_{N,J\to\infty} \mathrm{pr}(\hat{r} = r^*) = 1$.*

Remark 14. The two conditions on $u(N, J, r)$ in both the dependent and the independent cases are needed to ensure the existence of a suitable penalty that guards against both

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025

overselection and underselection of the number of factors. For such a function u to exist, $\int_h^{1-h} \sigma^2_{r^*,t}\,\mathrm{d}t$ cannot be too small. The first condition $u(N,J,r) = o\big(\int_h^{1-h} \sigma^2_{r^*,t}\,\mathrm{d}t\big)$ requires that $u(N,J,r)$ be smaller than the integral of the gap between nonzero singular values and zero singular values of $\mathbf{X}^*(\cdot)$. It ensures that the probability of underselecting the number of factors will be small. The second condition requires that $u(N,J,r)$ grow faster than the upper bound of the estimation error, which guarantees that, with high probability, we do not overselect the number of factors.

Remark 15. The results in Theorems 1 and 3 can be extended if it is of interest to use a kernel function supported on the whole real line, such as the Gaussian kernel. In such cases, Condition 3 needs to be modified. The details are given in the Supplementary Material.

Remark 16. The results of Theorems 1 and 3 can also be extended to a missing data setting under an ignorable missingness assumption. Let ω_{ij} be a binary random variable indicating the missingness of $\{Y_{ij}(t) : t \in [0,1]\}$, where $\omega_{ij} = 1$ means that $\{Y_{ij}(t) : t \in [0,1]\}$ is observed and $\omega_{ij} = 0$ means that $\{Y_{ij}(t) : t \in [0,1]\}$ is missing. We can still establish results corresponding to Theorems 1 and 3 under suitable conditions based on a pseudolikelihood function that replaces the summations over all i and j in (3) by summations over i and j such that $\omega_{ij} = 1$.

4. Simulation study

We evaluate the proposed estimator and information criterion with a simulation study. In this study, we generate data from the proposed model, where the number of factors is set to $r^* = 3$ and the numbers of observation units and event types satisfy $N = 2J$. We consider three patterns for the dynamic component $\Theta^*(t)$, denoted by C1, C2 and C3, in which $\Theta^*(t)$ is constant, changes linearly and changes periodically, respectively. We further consider two different settings for generating $\mathbf{A}^*$, denoted by S1 and S2, resulting in two different signal-to-noise levels, where setting S1 has a stronger signal than setting S2. We vary the number of event types J by setting $J = 100, 200, 400$ and 800. Finally, we consider data generation under the dependent and independent settings in Theorem 1. The factors discussed above lead to a total of 24 simulation settings. For each setting, 50 independent replications are generated. The proposed method is compared with the Poisson factor model discussed in Remark 3, which ignores the dynamic nature of the process and is concerned only with the total event counts on the entire time interval. Following a similar proof to that for Theorem 1, the likelihood-based estimator under the Poisson factor model is consistent even under the dependent-event-type settings where $\Theta^*(t)$ is constant. As the Poisson factor model involves fewer parameters, it is expected to be statistically more efficient than the proposed estimator in the settings where $\Theta^*(t)$ is constant. In the other settings, the Poisson factor model has biases as it ignores the dynamic nature of the event data.

We now elaborate on the data generation and results in some settings with dependent event types. Further details about the simulations are given in the Supplementary Material. More simulations are performed in the Supplementary Material under additional settings, including those with independent event types, more event types than observation units, and modified specifications for $\Theta^*(t)$ and $\mathbf{A}^*$ that lead to even weaker signals. While the results vary across settings, their patterns are consistent with the results of the simulations reported here.

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025

Table 1. *Mean estimation error among 50 independent replications based on the proposed estimator and the estimator under the Poisson factor model in 24 simulation settings*

		S1			S2	
Kernel-based method	C1	C2	C3	C1	C2	C3
$J = 100$	0.1006	0.1174	0.1048	0.1371	0.1606	0.1407
$J = 200$	0.0536	0.0630	0.0562	0.0692	0.0806	0.0727
$J = 400$	0.0291	0.0350	0.0308	0.0378	0.0437	0.0398
$J = 800$	0.0159	0.0190	0.0170	0.0205	0.0240	0.0217
Poisson factor model	C1	C2	C3	C1	C2	C3
$J = 100$	0.0154	0.9743	0.7518	0.0192	0.6928	0.5530
$J = 200$	0.0073	0.9785	0.7611	0.0091	0.6830	0.5458
$J = 400$	0.0036	0.9773	0.7442	0.0046	0.6911	0.5442
$J = 800$	0.0018	1.0012	0.7542	0.0023	0.6955	0.5491

We set $\phi(J) = J^{1/3}$ and generate data $\{Y_{ij}(t) : t \in [0, 1]\}$ as follows. First, we divide the event types $j = 1, \ldots, J$ into $\lfloor J/\phi(J) \rfloor$ blocks of approximately equal sizes $B_1, \ldots, B_{\lfloor J/\phi(J) \rfloor}$, where $\lfloor J/\phi(J) \rfloor$ denotes the greatest integer less than or equal to $J/\phi(J)$. Second, for the kth block, we generate a Poisson process with intensity function $f_k(t) := \max_{j \in B_k} f\{X^*_{ij}(t)\} = f\{\max_{j \in B_k} X^*_{ij}(t)\}$ and denote the generated event times by $0 < t_{k,1} < \cdots < t_{k,p_k} < 1$. Finally, using a thinning algorithm (Chen, 2016), for each $i = 1, \ldots, N$ and each $j \in B_k$ we accept $t_{k,1}, \ldots, t_{k,p_k}$ with probabilities $f\{X^*_{ij}(t_{k,1})\}/f_k(t_{k,1}), \ldots, f\{X^*_{ij}(t_{k,p_k})\}/f_k(t_{k,p_k})$ independently and let the accepted time-points be the event times of $Y_{ij}(t)$. The resulting processes are guaranteed to follow the proposed model. We choose the Epanechnikov kernel function $K(x) = 0.75(1 - x^2)$ for $-1 \leqslant x \leqslant 1$, with kernel order $m = 2$. It is easy to verify that the chosen kernel function satisfies Condition 3. The link function is $f(x) = \exp(x)$. We set $h = 0.1\{J/\phi(J)\}^{-0.19}$ and $M = 36$. Our estimation is based on a discretized likelihood with 31 evenly distributed time-points $t_1, \ldots, t_{31}$ on $[h, 1 - h]$. A sensitivity analysis is performed in the Supplementary Material with respect to the number of grid points, which suggests that the choice of 31 is sufficient for our simulation settings. After the estimation, we obtain $\hat{\mathbf{X}}(t)$ for $t \in [h, 1 - h]$ by a linear interpolation. The estimation error is evaluated by $\int_h^{1-h} ||\mathbf{X}^*(t) - \hat{\mathbf{X}}(t)||_{\mathrm{F}}^2 \, \mathrm{d}t/(NJ)$. Under the Poisson factor model, we obtain $\hat{\mathbf{A}}$ and $\hat{\mathbf{\Theta}}$. We compute $\int_h^{1-h} ||\mathbf{X}^*(t) - \hat{\mathbf{X}}(t)||_{\mathrm{F}}^2 \, \mathrm{d}t/(NJ)$ as its estimation error, where $\hat{\mathbf{X}}(t) = \hat{\mathbf{\Theta}}\hat{\mathbf{A}}^{\mathrm{T}}$ is constant over time.

The results regarding the estimation errors are given in Table 1. They show that for each combination of S_i and C_j, with $i = 1, 2$ and $j = 1, 2, 3$, the estimation error of the proposed method decays as N and J grow. In settings where $\Theta^*(t)$ is constant (i.e., C1), the estimator given by the Poisson factor model has smaller errors than the proposed estimator. In the rest of the settings, the proposed estimator yields substantially smaller estimation errors than those under the Poisson factor model. In the Supplementary Material, the two models are also compared in terms of recovering the loading matrix $\mathbf{A}^*$. Because of the rotational indeterminacy mentioned in Remark 4, we measure the accuracy by the principal angles between the subspace spanned by the column vectors of $\mathbf{A}^*$ and that spanned by those of $\hat{\mathbf{A}}$. The results show that the proposed method provides substantially more accurate estimates of $\mathbf{A}^*$ in settings where the Poisson factor model is misspecified and yields similar but slightly less accurate estimates when the Poisson factor model is correctly specified.

Finally, we evaluate the accuracy in selecting the number of factors. We set the penalty term to $v(N, J, r) = 40rNJh^{1.99}$ and select r from the candidate set $\{1, 2, 3, 4, 5\}$. According to

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025

Table 2. *Products with large positive factor loadings for each of the three factors*

Factor 1	yogurt (10), salad (3), herbs (parsley, cilantro) (3), organic fruit/vegetable (3), blueberry (3), mushroom (2), tropical fruit (mango, pineapple) (2), beans (2), pepper (2), cheese (2)
Factor 2	soft drink (11), cold cereal (5), hot sauce (5), refrigerated drink (4), chicken wings (4), frozen meat (3), dinner sausage (3), candy (3), frozen pizza (2), cigarette (2), potato chips (2), canned pasta (2)
Factor 3	cheese (7), milk (5), white bread (4), fruit (banana, grape, strawberry) (4), egg (4), vegetable (cucumber, celery, cabbage, corn) (4), onion (4), salad (3), soft drink (2), hamburger bun (2), beef (2), tomato (2), potato (2)

our simulation results, the number of factors is always correctly selected in all the simulation settings, except for three settings where $J = 100$ and the signal-to-noise level follows S2. For these three cases (i.e., C1–C3), 13, 29 and 10 out of 50 replications mis-select the number of factors. Overall, the proposed information criterion shows effective performance.

5. Application to grocery shopping data

5.1. *Background, data processing and analysis*

We apply the proposed method to a grocery shopping dataset available at https://www.dunnhumby.com/source-files. It contains transaction records collected by a retailer over two years about its frequent shoppers. We discard the first 15% of the observation period since the number of transactions is significantly lower than in the rest of the period, likely due to late entries. The remaining period is then standardized to the interval $[0, 1]$. After pre-processing, we obtain a dataset with $N = 1978$ shoppers and $J = 2000$ products. The dataset contains information on each product regarding its type (e.g., cheese, chips); it also contains demographic information on 796 shoppers, including age, income and whether or not they have children. This information is not used in the proposed model but is used for validating and interpreting our results. Here, the matrix-valued function $\mathbf{\Theta}(\cdot)$ may be interpreted as the dynamic customer factors, and the matrix $\mathbf{A}$ may be interpreted as the attributes of the products. We apply the proposed information criterion with the candidate set $\{1, 2, 3, 4, 5\}$, which selects $r = 3$ factors. Following the discussion in Remark 4, we apply a varimax rotation for the selected three-factor model to obtain an interpretable factor structure.

5.2. *Interpretation of factors*

We interpret the factors based on the estimated loading matrix after rotation. Specifically, let $\tilde{\mathbf{A}} = (\tilde{a}_{ij})_{J\times r}$ be the loading matrix after rotation. We say that a product j dominantly loads on factor k if $\tilde{a}_{jk}^2/(\sum_{l=1}^{r} \tilde{a}_{jl}^2)$ is large, i.e., $\tilde{a}_{jk}$ is dominant in magnitude over the rest of the loadings. We investigate the top 60 products that dominantly load on each factor. Table 2 lists the types of these products. Many products with dominant loadings on the same factor tend to be of a small number of types. These types are presented only once in the table, followed by the corresponding number of products of this type in parentheses. Product types that appear only once for each factor are omitted for brevity.

Table 2 shows that the items with dominant loadings on the first factor are mostly fresh and healthy food products suitable for vegetarian diets. Items with dominant loadings on the second factor contain unhealthy (e.g., soft drinks, candy, potato chips), fast food (e.g., cold cereal, frozen pizza) or budget-friendly (e.g., frozen meat) products. Finally, items that load dominantly on the third factor are mostly basic food products of daily need, including bread,

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025

Table 3. *Coefficients when regressing the factors on demographic variables*

	Factor 1 ($R^2 = 0.13$)	Factor 2 ($R^2 = 0.17$)	Factor 3 ($R^2 = 0.02$)
age		–0.007 ($p = 0.00$)	0.006 ($p = 0.00$)
income1	0.006 ($p = 0.00$)	–0.005 ($p = 0.02$)	
income2	0.015 ($p = 0.00$)	–0.021 ($p = 0.00$)	
child	–0.007 ($p = 0.01$)	0.010 ($p = 0.00$)	
income2 × child	0.009 ($p = 0.01$)		

eggs, milk and beef. While these products include many fresh and healthy food products similar to those loading on the first factor, they tend to be more budget-friendly. Given these features, we may interpret the three factors as healthy food consumption, unhealthy food consumption and basic food consumption factors, respectively.

We further investigate the three factors by regressing them on the three demographic variables of age, income and child. Here, 'age' is an ordinal variable referring to the estimated age range of the shopper. For simplicity, we transform it into a binary variable which takes value 1 if the shopper is aged over 55 and 0 otherwise. The variable 'income' is an ordinal variable recording the shopper's household income level. We simplify it to a variable with three categories: under \$35 000 (income1 = 0, income2 = 0), \$35 000–\$75 000 (income1 = 1, income2 = 0) and above \$75 000 (income1 = 0, income2 = 1). Finally, the variable 'child' is a binary variable indicating whether the shopper's household has children. We run a linear regression model for each factor by regressing the factor scores on age, income1, income2, child, and the interactions between child and income1 and income2. The interaction terms are added because it is suspected that the child effect differs between high- and low-income households.

The results from these regression models are reported in Table 3, where the statistically significant coefficients and their p-values are presented and the R-squared values of the three models are given. As the coefficients for the interaction between the dummy variable income1 and child are insignificant in all three models, the corresponding row is not presented. All the terms are statistically significant for the first factor, except for age and the interaction between income1 and child. In particular, the coefficients associated with the summary variables for income are all positive, meaning that the consumption of healthy food increases with household income, controlling for the rest of the variables. In addition, the coefficient for child is negative, and the coefficient for the interaction between income2 and child is positive and larger in absolute value than the coefficient for child. This means that households with lower income (up to \$75 000) tend to buy less healthy food when they have children, while those with higher income (above \$75 000) tend to buy more healthy food when they have children.

All the coefficients are significant for the second factor, except those associated with the two interaction terms. The coefficient for age is negative, suggesting that the older group tends to consume less unhealthy food than the younger one, controlling for the rest of the variables. The coefficients for income are also negative, suggesting that households with higher income tend to consume less unhealthy food when controlling for the rest of the variables. On the other hand, the coefficient for child is positive, meaning that households with children tend to consume more unhealthy food. This may be because this food category contains most soft drinks and snacks such as candy and potato chips that children often favour.

Regarding the third factor, only the coefficient for age is statistically significant, and the R-squared value is quite low. The positive coefficient means that older people consume more basic food products. Taken together with the results for the second factor, this may indicate that older people tend to have a healthier lifestyle. Although they do not seem to consume

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025

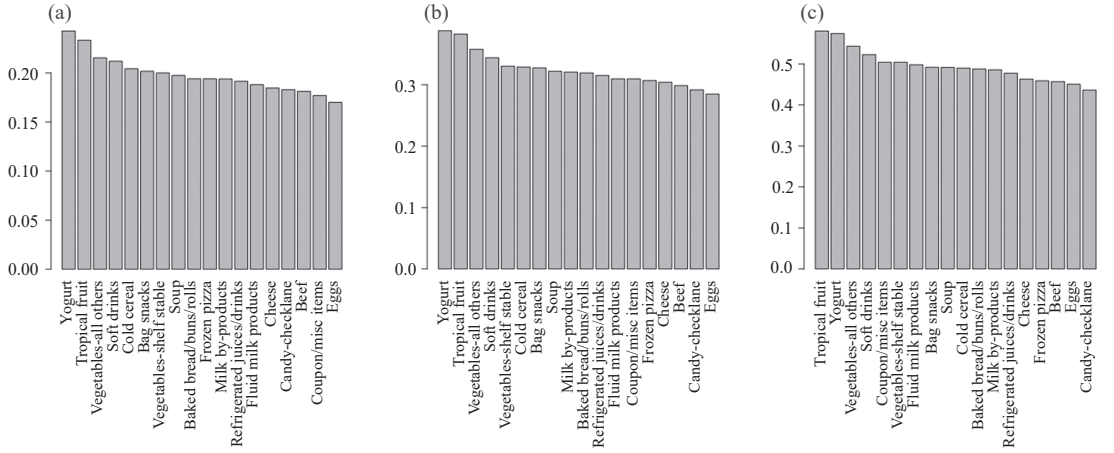

Fig. 1. Quartiles of the variability of the most frequently purchased product types: (a) first quartile; (b) second quartile, i.e., median; (c) third quartile.

more healthy food associated with the first factor, they may cook more frequently using basic food products and eat less unhealthy food than the younger group.

5.3. *Investigating purchase dynamics*

We further investigate the dynamic trend that the model captures. In particular, for each pair consisting of consumer i and product j, we measure the variability in the personal purchasing rate by the total variation of $X_{ij}(\cdot)$, i.e., $\int_0^1 |X'_{ij}(t)|\,\mathrm{d}t$, where $X'_{ij}(t)$ denotes the derivative of $X_{ij}(t)$. A larger total variation implies a higher variability. Under the estimated three-factor model, we estimate this variability based on the finite differences between $\hat{X}_{ij}(t)$ for time t at adjacent grid points.

The variability measure is computed for 1019 products of 18 product types that are most frequently purchased. For each product type, we look at the empirical distribution of the estimated total variations based on all the shoppers and all the products of this type and compute its quartiles, i.e., the 25%, 50% (i.e., median) and 75% percentiles. The results are displayed in Fig. 1, where the 18 product types are organized in descending order for each quartile. The ranking of product types is reasonably stable across the three quartiles and consistent with our understanding of their sales pattern. We remark that dimension reduction is important for the proposed method to produce these results. One cannot obtain sensible results by averaging the sales of the products over shoppers because of the high level of noise in the data.

Vegetables, tropical fruits, yogurt and soft drinks are product types with consistently high variability scores across all three quartiles. The price and quality of many vegetables and fruits depend on their growing seasons. In addition, tropical fruits are imported products whose price and supply depend on additional factors that fluctuate over time, such as transportation costs. Owing to the previously mentioned factors, these products show higher variability in their sales. On the other hand, the higher variability of yogurt and soft drinks may be due to seasonal shifts in consumer demand. The demand for these products tends to increase during warmer months and decrease during colder months when warming foods and drinks are preferred.

Dairy products, eggs, beef and candy displayed at the checkout lane are product types with consistently low variability. These are staples in many people's diets. Their supply and

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025

demand are typically stable throughout the year. The sales of candies displayed in the checkout lane are expected to be stable because of their constant high visibility, accessibility and affordability, which are barely affected by economic conditions or other seasonal factors.

6. Discussion

The theoretical results for the proposed estimator in the dependent event setting could be improved. There is a gap between the error rates in the dependent and independent settings in Theorem 1, and in particular the convergence rate is slower when $\phi(J)$ is of a constant order in the upper bound for the dependent setting than for the independent setting. This could be an artefact of our proof strategy, as certain random matrix results that are key to establishing the upper bound for the independent setting do not apply to the dependent setting. As discussed in Remark 12, the gap can be filled when the blockwise-independent structure does not vary across individuals. Under the more general individual-specific blockwise structure in Condition 4, this gap may still be filled with a more refined analysis. We leave this problem for future investigation.

The current method is not specifically designed for forecasting, though it still has some prediction power. For example, one could predict events associated with the existing observation units and event types at a future time-point (i.e., $t > 1$) based on $f\{\hat{\mathbf{X}}(1-h)\}$. Such a prediction would be sensible if the model still holds after time 1, and $\mathbf{X}^*(1-h)$ and $\mathbf{X}^*(t)$ are close to each other owing to the smoothness of the function. We can improve the prediction power of the proposed method by further assuming a stochastic model, such as a Gaussian random field model, for the latent process $\mathbf{X}(t)$ and estimating it based on our estimate $\hat{\mathbf{X}}(t)$. Such a model may allow us to better predict future events, even if they are associated with new observation units or event types not used in the model training, as long as the new observation units and event types are from the same populations as the existing ones.

A useful application of the proposed method is for detecting changes in each observation unit, which may be of interest in many applications. For example, in the grocery shopping application, a change in the dynamic factor of a household may imply a structural change in consumer behaviour, based on which an individualized marketing strategy may be developed. Although we currently require each $\theta_{ik}(t)$ to be sufficiently smooth, this requirement can be relaxed to allow each $\theta_{ik}(t)$ to be a piecewise-smooth function. Using the proposed method, changes can be detected based on the estimated functions, which is closely related to change-point detection in the nonparametric regression literature (e.g., Xia & Qiu, 2015). Methods and theories remain to be developed for optimally localizing the changes based on the estimated functions.

Acknowledgment

We are grateful to the editor, associate editor and two referees for their careful reviews and valuable comments.

Supplementary material

The Supplementary Material includes proofs of the theoretical results, the computational algorithm, details of the simulation settings and additional simulation results. The code for our simulation study and real data analysis is available at `https://github.com/Fangyi-Chen98/CountingProcessFactor`.

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025

References

Abu-Libdeh, H., Turnbull, B. W. & Clark, L. C. (1990). Analysis of multi-type recurrent events in longitudinal studies: Application to a skin cancer prevention trial. *Biometrics* **46**, 1017–34.

Andersen, P. K., Borgan, O., Gill, R. D. & Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.

Bai, J. & Li, K. (2012). Statistical analysis of factor models of high dimension. *Ann. Statist.* **40**, 436–65.

Bai, J. & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191–221.

Bai, J. & Ng, S. (2023). Approximate factor models with weaker loadings. *J. Economet*. **235**, 1893–916.

Bogdanowicz, A. & Guan, C. (2022). Dynamic topic modeling of Twitter data during the COVID-19 pandemic. *PLoS One* **17**, 1–22.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Mult. Behav. Res*. **36**, 111–50.

Chamberlain, G. & Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* **51**, 1281–304.

Chen, B. E., Cook, R. J., Lawless, J. F. & Zhan, M. (2005). Statistical methods for multivariate interval-censored recurrent events. *Statist. Med.* **24**, 671–91.

Chen, L., Dolado, J. J. & Gonzalo, J. (2021). Quantile factor models. *Econometrica* **89**, 875–910.

Chen, Y. (2016). *Thinning algorithms for simulating point processes*. Florida State University, Tallahassee, Florida.

Chen, Y. (2020). A continuous-time dynamic choice measurement model for problem-solving process data. *Psychometrika* **85**, 1052–75.

Chen, Y. & Li, X. (2022). Determining the number of factors in high-dimensional generalized latent factor models. *Biometrika* **109**, 769–82.

Chen, Y., Li, X. & Zhang, S. (2020). Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications. *J. Am. Statist. Assoc*. **115**, 1756–70.

Chen, Y. & Zhang, S. (2020). A latent Gaussian process model for analysing intensive longitudinal data. *Br. J. Math. Statist. Psychol*. **73**, 237–60.

Cook, R. J. & Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. New York: Springer.

Györfi, L., Kohler, M., Krzyzak, A. & Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. New York: Springer.

He, Y., Kong, X., Trapani, L. & Yu, L. (2023). One-way or two-way factor model for matrix sequences? *J. Economet*. **235**, 1981–2004.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **23**, 187–200.

Liang, S., Zhang, X., Ren, Z. & Kanoulas, E. (2018). Dynamic embeddings for user profiling in Twitter. In *Proc. 24th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining*. New York: Association for Computing Machinery, pp. 1764–73.

Lin, D. Y., Wei, L.-J., Yang, I. & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. R. Statist. Soc. B* **62**, 711–30.

Liu, W., Lin, H., Zheng, S. & Liu, J. (2023). Generalized factor model for ultra-high dimensional correlated variables with mixed types. *J. Am. Statist. Assoc*. **118**, 1385–401.

Lu, Z.-H., Chow, S.-M., Sherwood, A. & Zhu, H. (2015). Bayesian analysis of ambulatory blood pressure dynamics with application to irregularly spaced sparse data. *Ann. Appl. Statist*. **9**, 1601–20.

Rohe, K. & Zeng, M. (2023). Vintage factor analysis with varimax performs statistical inference. *J. R. Statist. Soc. B* **85**, 1037–60.

Sun, J. (2006). *The Statistical Analysis of Interval-Censored Failure Time Data*. New York: Springer.

Tang, N., Chow, S.-M., Ibrahim, J. G. & Zhu, H. (2017). Bayesian sensitivity analysis of a nonlinear dynamic factor analysis model with nonparametric prior and possible nonignorable missingness. *Psychometrika* **82**, 875–903.

Wang, D., Liu, X. & Chen, R. (2019). Factor models for matrix-valued high-dimensional time series. *J. Economet*. **208**, 231–48.

Wedel, M. & Kamakura, W. A. (2001). Factor analysis with (mixed) observed and latent variables in the exponential family. *Psychometrika* **66**, 515–30.

Xia, Z. & Qiu, P. (2015). Jump information criterion for statistical inference in estimating discontinuous curves. *Biometrika* **102**, 397–408.

Yang, L. (2022). Nonparametric copula estimation for mixed insurance claim data. *J. Bus. Econ. Statist*. **40**, 537–46.

[*Received on* 2 *April* 2024. *Editorial decision on* 31 *March* 2025]

Downloaded from https://academic.oup.com/biomet/article/112/3/asaf028/8111905 by guest on 22 July 2025