



Longitudinal Principal Component Analysis With an Application to Marketing Data

Christopher Kinson, Xiwei Tang, Zhen Zuo & Annie Qu

To cite this article: Christopher Kinson, Xiwei Tang, Zhen Zuo & Annie Qu (2020) Longitudinal Principal Component Analysis With an Application to Marketing Data, Journal of Computational and Graphical Statistics, 29:2, 335-350, DOI: [10.1080/10618600.2019.1677244](https://doi.org/10.1080/10618600.2019.1677244)

To link to this article: <https://doi.org/10.1080/10618600.2019.1677244>



View supplementary material [↗](#)



Accepted author version posted online: 08 Oct 2019.
Published online: 05 Nov 2019.



Submit your article to this journal [↗](#)



Article views: 266



View related articles [↗](#)



View Crossmark data [↗](#)



Longitudinal Principal Component Analysis With an Application to Marketing Data

Christopher Kinson^a, Xiwei Tang^b, Zhen Zuo^c, and Annie Qu^a

^aDepartment of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL; ^bDepartment of Statistics, University of Virginia, Charlottesville, VA; ^cDepartment of Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, TX

ABSTRACT

We propose a longitudinal principal component analysis method for multivariate longitudinal data using a random-effects eigen-decomposition, where the eigen-decomposition uses longitudinal information through nonparametric splines and the multivariate random effects incorporate significant store-wise heterogeneity. Our method can effectively analyze large marketing data containing sales information for products from hundreds of stores over an 11-year time period. The proposed method leads to more accurate estimation and interpretation compared to existing approaches. We illustrate our method through simulation studies and an application to marketing data from IRI. Supplementary materials for this article are available online.

ARTICLE HISTORY

Received May 2018
Revised August 2019

KEYWORDS

Marketing data; Multivariate longitudinal data; Nonparametric spline; Principal component analysis; Random-effects model; Time-varying covariance matrix

1. Introduction

Technological innovation or stagnation can determine the fates of companies and their employees. In the first quarter of 2012, Kodak, a leading technological company well-known for producing photography supplies such as cameras and film, filed for bankruptcy and ceased production of digital cameras and photography accessories. Clearly, it is necessary and important to study consumer shopping behavior and marketing trends over time, to help companies avoid manufacturing products which are no longer attractive to consumers.

This article is motivated by the IRI marketing dataset, an immense collection of consumer panel data of grocery and drug store sales, pricing, and promotion strategies (Bronnenberg, Kruger, and Mela 2008; Kruger and Pagni 2008), which was created for marketing researchers to explore marketing trends and their impact on economics. The collected sales data contain weekly sales information on a variety of product categories in over 40 regional markets in the United States. In this article, we focus on grocery store sales for 20 representative products in 552 stores spanning the 11-year time period from 2001 to 2011, where the product categories include mostly consumer packaged goods, such as beer, cigarettes, household cleaners, soup, etc.

The presented multivariate longitudinal sales data contain a wide range of products with rich information regarding consumers' general shopping behavior over time. A marginal model separately applied on each individual product is obviously inadequate as it has to adjust for many other factors. Therefore, we are especially motivated to explore consumption behavior changes through investigating time-varying patterns of correlations between different products, so we can better understand marketing trends. One critical issue in analyzing

multivariate longitudinal data is to account for significant store-wise heterogeneity, which can be due to many factors such as geographic location and store size. For example, smaller stores tend to have a decreasing trend in sales of beer and peanut butter, while larger stores have an overall increase for these two product categories (see Figure 1). Store-wise variation can bring additional difficulty in modeling underlying correlations among products, as different sources of variation could be active in different stores.

In business and marketing, principal component analysis (PCA) is often employed to investigate the covariance structure of large-dimensional multivariate data and to reduce data dimensionality (e.g., Jain, Bass, and Chen 1990; Fader and Lattin 1993; Rossi and Allenby 1993; Bradlow 2002). However, PCA for multivariate longitudinal data is much less-studied. Performing PCA separately at each time-point is clearly inadequate, as it does not use any longitudinal information. In addition, it is crucial to incorporate within-subject correlation across different time-points, as well as among different variables from the same subject.

For time-dependent data, PCA models have been developed in the time series setting (e.g., Kim and Wu 1999; Jolliffe 2002). However, these approaches mainly target identification of similar time-points or extraction of common trends over time, which are not applicable for our problem. Moreover, the population covariance regression approaches (Hoff and Niu 2012; Fox and Dunson 2017) model the covariance for a scalar outcome observed at multiple time points, while the proposed model considers the covariance structure of multivariate outcomes changing over time.

In addition, the functional PCA (FPCA) has been proposed for univariate longitudinal data (Ramsay and Silverman 2005; Yao, Müller, and Wang 2005; Hall, Müller, and Wang 2006) and

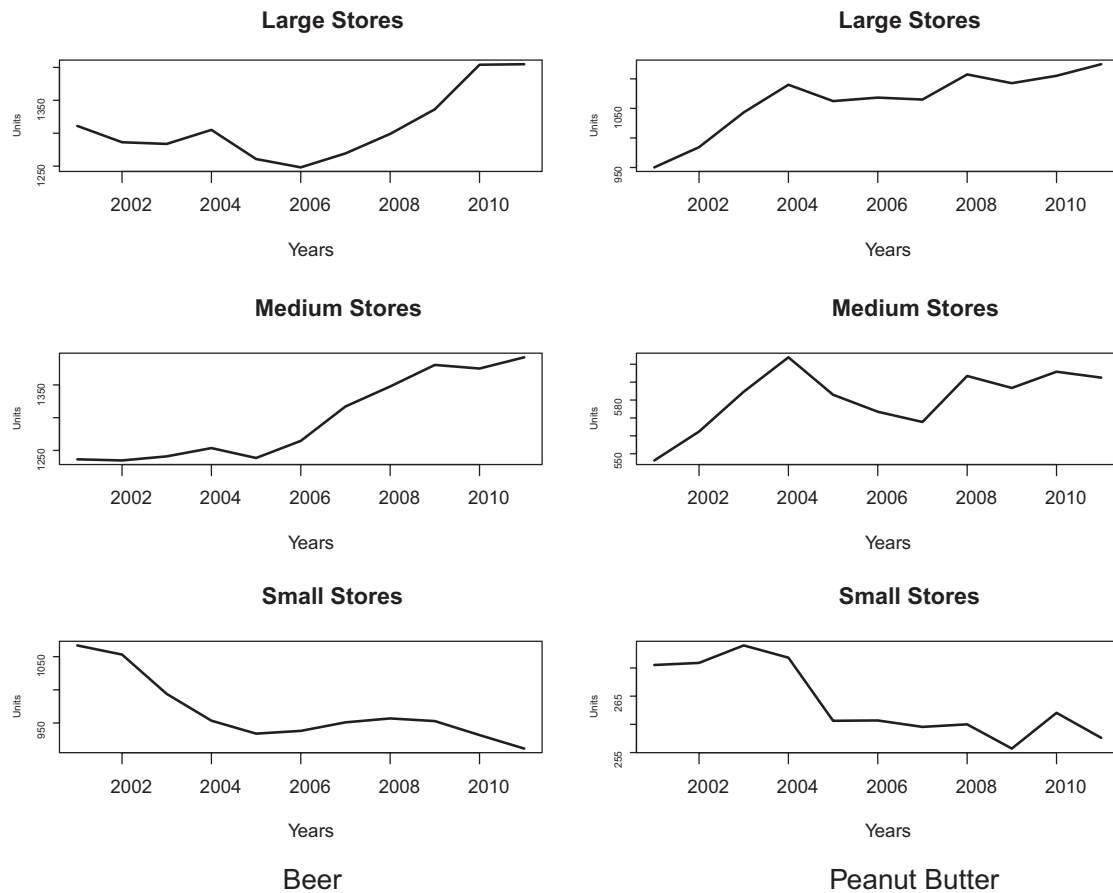


Figure 1. Average product sales for three store sizes for beer and peanut butter products.

further extended to multivariate data (MFPCA, Di et al. 2009; Jiang and Wang 2010; Berrendero, Justel, and Svarc 2011; Chiou, Chen, and Yang 2014; Happ and Greven 2018). The MFPCA models treat the whole multivariate longitudinal trajectory as one observation from each subject, where eigenfunctions of multivariate trajectories would have the same structure. Similarly, the longitudinal functional principal component analysis (LFPCA, Greven et al. 2010; Zipunnikov et al. 2014; Islam, Staicu, and Heugten 2018) investigates functional data which are observed repeatedly over time, where the functional eigenbases are assumed to be time-invariant. In contrast, the proposed approach models the time-varying covariance structure assuming that both the eigenvectors and eigenvalues change smoothly over time.

Recently, Miller and Bowman (2012) and Prvan and Bowman (2003) proposed a two-step smooth principal component analysis (SPCA) method to model changes in covariance over time. They first estimate the smooth mean function of each response variable, and then estimate the local covariance matrix of multivariate response variables with residuals from the first step at each time point. However, their approach does not account for any important longitudinal correlation among observations from the same subject. In addition, the smoothness of the pairwise covariance may not guarantee the smoothness of the principle components over time.

In this article, we propose a longitudinal PCA which decomposes correlation information arising from multivariate

observations over time while incorporating heterogeneity among stores. Specifically, store-specific multivariate random effects are implemented to capture heterogeneous variation among stores, while correlations among product sales are modeled through time-varying eigen-decomposition. In modeling the eigen-decomposition, we assume that both eigenvalues and eigenvectors are time-varying functions, which are estimated based on nonparametric splines.

The proposed method has three advantages over existing approaches. First, incorporating random effects accounts for the variation among different stores, thus improving the estimation of correlation structures of products. Second, modeling a small number of dominant eigen-components, rather than the entire covariance matrix over time, significantly reduces the dimensionality of unknown parameters. Third, time-varying eigenvectors are modeled and estimated via nonparametric splines, which are able to recover information from missing time-points by utilizing neighboring data-points. To implement the proposed method, we develop an iterative algorithm which updates time-varying eigenvalues and time-varying eigenvectors alternately. In addition, we adapt the Expectation-Substitution algorithm (Elashoff and Ryan 2004) to the proposed model to estimate the multivariate random-effects.

The article is outlined as follows. Section 2 presents the proposed model. Section 3 develops the algorithm and implementation. Section 4 illustrates the performance of the proposed method based on simulation studies. Section 5 demonstrates an

application to the IRI marketing dataset. The article concludes with a discussion in Section 6.

2. Methodology

2.1. Notation and Framework

We start with introducing some notation regarding multivariate longitudinal data. Let y_{ijt} be a response measured at a time point t for the j th variable from the i th subject, where $i = 1, 2, \dots, N$, $j = 1, 2, \dots, J$, and $t = 1, 2, \dots, T$. In our application, t represents the number of years, i represents the i th store, and j represents the j th product. Hence, $\mathbf{y}_{it} = (y_{i1t}, \dots, y_{iJt})^T$ denotes the $J \times 1$ random vector for the multivariate sales outcome at time t .

In general, for an unknown population quantity θ (e.g., a population covariance matrix $\mathbf{V}$), we use $\hat{\theta}$ to denote the analogous empirical statistic (e.g., a sample covariance matrix $\hat{\mathbf{V}}$), and use $\hat{\theta}$ to denote an estimate of θ based on a parametric model. Moreover, we use the superscript “bar” to denote a low-rank structure (e.g., a low-rank approximated covariance matrix $\bar{\mathbf{V}}$). In addition, we write a quantity θ that is assumed to change smoothly over time t as $\theta(t)$, but denote the discretized values at each time point with a subscript (e.g., an empirical statistic $\hat{\theta}_t$).

For the longitudinal outcome $\mathbf{y}_{it}$, we consider the eigen-decomposition for the population covariance matrix $\mathbf{V}_t$ at each time point,

$$\mathbf{V}_t = \sum_{k=1}^J \alpha_{kt} \mathbf{w}_{kt} \mathbf{w}_{kt}^T, \quad 1 \leq t \leq T, \quad (1)$$

where α_{kt} is the eigenvalue corresponding to the k th eigenvector $\mathbf{w}_{kt} = (w_{k1t}, \dots, w_{kJt})^T$ at time t .

Without using any neighboring time point information, a naive approach is to perform PCA on the sample covariance matrix at each time point separately, which is referred to as discretized principal component analysis (DPCA). Although the DPCA is easy to implement, it is not capable of incorporating changes in eigenvalues and eigenvectors gradually, as it does not utilize any neighboring time point information from the longitudinal measurements.

2.2. Time-Varying Eigen-Decomposition

To utilize longitudinal information, we assume that the time-varying eigenvalues and eigenvectors are continuous functions of t , which can be approximated by polynomial splines. Non-parametric splines have been extensively studied for longitudinal data by Anderson and Jones (1995), Huang, Wu, and Zhou (2004), Durbán et al. (2005), Liang and Xiao (2006), and Xue and Liang (2009). Additionally, semiparametric and nonparametric approaches have been proposed for covariance and correlation matrix estimation for longitudinal data (e.g., Diggle and Verbyla 1998; Wu and Pourahmadi 2003; Fan, Huang, and Li 2007; Sun, Zhang, and Tong 2007; Fan and Wu 2008; Maadooliat et al. 2013). In this article, we incorporate the splines to model the time-varying eigenvalues and eigenvectors simultaneously.

Let $\alpha_k(t)$ denote the time-varying eigenvalue function, and $\mathbf{w}_k(t) = (w_{k1}(t), \dots, w_{kJ}(t))^T$ denote the multivariate time-

varying eigenvector function, for $1 \leq k \leq J$. Hence, we represent the time-varying population covariance as

$$\mathbf{V}(t) = \sum_{k=1}^J \alpha_k(t) \mathbf{w}_k(t) \mathbf{w}_k^T(t),$$

and assume that the k th eigenvalue and the corresponding eigenvector can be modeled as

$$\begin{aligned} \alpha_k(t, \boldsymbol{\beta}'_k) &= \sum_{l=1}^{C_N} \beta'_{kl} b_{kl}(t), \\ \mathbf{w}_{kj}(t, \mathbf{v}_{kj}) &= \sum_{m=1}^{D_N} v_{kjm} g_{kjm}(t), \quad j = 1, \dots, J, \end{aligned} \quad (2)$$

where $\{b_{kl}(t)\}_{l=1}^{C_N}$ and $\{g_{kjm}(t)\}_{m=1}^{D_N}$ are sets of spline bases, and $\boldsymbol{\beta}'_k = (\beta'_{k1}, \dots, \beta'_{kC_N})^T$ and $\mathbf{v}_{kj} = (v_{kj1}, \dots, v_{kjD_N})^T$ are corresponding spline coefficients, respectively. Consequently, we denote $\mathbf{w}_k(t, \mathbf{v}_k)$ as the multivariate time-varying eigenvector function, where $\mathbf{v}_k = (\mathbf{v}_{k1}^T, \dots, \mathbf{v}_{kJ}^T)^T$.

In our model, we rescale time t to be in the interval $[0, 1]$. Therefore, the spline bases are constructed based on a partition ξ of the interval $[0, 1]$ with P_N interior knots

$$\xi = \{0 = \xi_0 < \xi_1 < \dots < \xi_{P_N} < \xi_{P_N+1} = 1\}.$$

Consequently, we have $C_N = P_N + M_1$ and $D_N = P_N + M_2$, where M_1 and M_2 are the orders of the polynomial splines for the eigenvalues and eigenvectors, respectively. Specifically, the employed spline bases are polynomial functions of degrees $M_1 - 1$ and $M_2 - 1$ on intervals $[\xi_i, \xi_{i+1}]$, for $i = 0, \dots, P_N - 1$, and $[\xi_{P_N}, \xi_{P_N+1}]$; and they are $M_1 - 2$ and $M_2 - 2$ continuously differentiable. Note that different sets of knots are allowed for eigenvalues and eigenvectors; however, for simplicity, we choose the same set of knots here. More details regarding the selection of knots and spline bases are discussed in Section 3.2.

One attractive feature of polynomial splines is that they approximate a smooth function sufficiently well without requiring a large number of knots. In general, the B-spline can be an alternative choice; however, in our setting, the truncated polynomial spline provides more straightforward interpretation on the time-varying eigen-functions. Indeed, the polynomial spline approach is equivalent to the B-spline approach for prediction, while the latter imposes constraints on construction.

2.3. Estimation of Eigenvalues and Eigenvectors

In this subsection, we provide time-varying eigenvalue and eigenvector estimations under the framework of generalized estimating equations (GEE, Liang and Zeger 1986). Consider the quasi-likelihood estimating function for clustered data $\mathbf{y}_{it}$,

$$\mathbf{g}_{it} = \dot{\boldsymbol{\mu}}_{it}^T \mathbf{V}_t^{-1} (\mathbf{y}_{it} - \boldsymbol{\mu}_{it}),$$

where $\boldsymbol{\mu}_{it} = \mathbf{E}[\mathbf{y}_{it}]$ and $\dot{\boldsymbol{\mu}}_{it}$ is the derivative of $\boldsymbol{\mu}_{it}$. In this study, without additional covariates, we have $\boldsymbol{\mu}_{it} = \boldsymbol{\mu}_t$ and $\dot{\boldsymbol{\mu}}_{it}$ is an identity matrix. For the estimation of the correlation structure in the GEE model, Qu, Lindsay, and Li (2000) proposed the quadratic inference function (QIF) assuming that the inverse of

the correlation matrix can be approximated by a linear combination of matrix bases; Zheng, Xue, and Qu (2018) extend the idea of the QIF to estimate a dynamic correlation structure by incorporating a time-varying coefficient model on coefficients of bases matrices. However, in their approaches, the basis matrices are prespecified, usually generated by eigenvectors, but they do not change over time. To investigate the dynamic correlation pattern of the multivariate longitudinal outcome, we assume a spline-based low-rank approximation $\bar{V}(t)$ to the covariance matrix with time-varying eigenvalues and eigenvectors as specified in (2). Following the spirit of the QIF, we estimate the model based on the generalized inverse of this approximation, which is also a linear combination of the first K ($1 \leq K \leq J$) time-varying eigen-components:

$$\bar{V}^{(-)}(t|\beta, \mathbf{v}) = \sum_{k=1}^K \alpha_k^{-1}(t, \beta_k) \mathbf{w}_k(t, \mathbf{v}_k) \mathbf{w}_k(t, \mathbf{v}_k)^T, \quad (3)$$

where $\alpha_k^{-1}(t, \beta_k)$ denotes a spline-based function for the inverse of time-varying eigenvalue, $\alpha_k(t, \beta_k')$, which share the same spline bases but with a different set of coefficients as modeled in (2), and $\beta = (\beta_1^T, \dots, \beta_K^T)^T$ and $\mathbf{v} = (\mathbf{v}_1^T, \dots, \mathbf{v}_K^T)^T$.

To estimate the time-varying covariance structure, we take the difference between the assumed low-rank model, $\bar{\mathbf{g}}_{it} = \bar{V}^{(-)}(t|\beta, \mathbf{v})(\mathbf{y}_{it} - \mu_t)$, and the sample version, $\tilde{\mathbf{g}}_{it} = \tilde{V}_t^{-1}(\mathbf{y}_{it} - \mu_t)$,

$$\mathbf{h}_{it}(\beta, \mathbf{v}; \mu_t) = \bar{\mathbf{g}}_{it} - \tilde{\mathbf{g}}_{it} = \left(\bar{V}^{(-)}(t|\beta, \mathbf{v}) - \tilde{V}_t^{-1} \right) (\mathbf{y}_{it} - \mu_t), \quad (4)$$

where $\tilde{V}_t$ is the sample covariance matrix. Consequently, analogous to the QIF approach, we estimate the model by minimizing the following loss function:

$$\begin{aligned} & \sum_{t=1}^T \sum_{i=1}^N \mathbf{h}_{it}^T(\beta, \mathbf{v}; \mu_t) \mathbf{h}_{it}(\beta, \mathbf{v}; \mu_t) \\ &= \sum_{t=1}^T \sum_{i=1}^N (\mathbf{y}_{it} - \mu_t)^T \left(\bar{V}^{(-)}(t|\beta, \mathbf{v}) - \tilde{V}_t^{-1} \right)^2 (\mathbf{y}_{it} - \mu_t). \end{aligned} \quad (5)$$

In addition, to ensure the identifiability and interpretability of the decomposed bases, the estimated eigen-vectors are required to be orthonormal. Therefore, we impose an L_2 -penalty on the inner-product between the bases vectors, and thus yield the following objective function:

$$\begin{aligned} L_\phi(\beta, \mathbf{v}; \mu) &= \sum_{t=1}^T \left(\sum_{i=1}^N \frac{\mathbf{h}_{it}^T(\beta, \mathbf{v}; \mu_t) \mathbf{h}_{it}(\beta, \mathbf{v}; \mu_t)}{N} \right. \\ &\quad \left. + \phi \sum_{k \neq k'} \|\mathbf{w}_k(t, \mathbf{v}_k)^T \mathbf{w}_{k'}(t, \mathbf{v}_{k'})\|_2^2 \right), \end{aligned} \quad (6)$$

where ϕ is a tuning parameter. It is worth noting that the L_2 -penalty cannot guarantee the exact orthogonality of the estimated bases vectors and thus the Gram-Schmidt process is further employed in the implementation procedure. Indeed, the L_2 -penalty can improve computational stability in practice. We provide more discussion about the L_2 -penalty in Section 3.2.

In our approach, the GEE-type objective function is adopted as it improves the estimation of the mean μ_t 's and thus further improves the estimation of the covariance structure. To illustrate an intuition of the proposed GEE-type loss function in (5), we provide an approximation of this loss function with respect to eigenvalue estimation. Given the mean value μ_t , the objective function in (5) can be rewritten as follows

$$\begin{aligned} & \sum_{t=1}^T \sum_{i=1}^N \frac{\mathbf{h}_{it}^T(\beta, \mathbf{v}; \mu_t) \mathbf{h}_{it}(\beta, \mathbf{v}; \mu_t)}{N} \\ &= \frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \text{Tr} \left\{ (\mathbf{y}_{it} - \mu_t)^T \left(\bar{V}^{(-)}(t|\beta, \mathbf{v}) - \tilde{V}_t^{-1} \right)^2 (\mathbf{y}_{it} - \mu_t) \right\} \\ &= \frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \text{Tr} \left\{ (\mathbf{y}_{it} - \mu_t)(\mathbf{y}_{it} - \mu_t)^T \left(\bar{V}^{(-)}(t|\beta, \mathbf{v}) - \tilde{V}_t^{-1} \right)^2 \right\} \\ &= \sum_{t=1}^T \text{Tr} \left\{ \frac{1}{N} \sum_{i=1}^N (\mathbf{y}_{it} - \mu_t)(\mathbf{y}_{it} - \mu_t)^T \left(\bar{V}^{(-)}(t|\beta, \mathbf{v}) - \tilde{V}_t^{-1} \right)^2 \right\} \\ &= \sum_{t=1}^T \text{Tr} \left\{ \tilde{V}_t \left(\bar{V}^{(-)}(t|\beta, \mathbf{v}) - \tilde{V}_t^{-1} \right)^2 \right\}, \end{aligned}$$

where $\text{Tr}\{\cdot\}$ indicates the trace, $\tilde{V}_t^{-1} = \sum_{k=1}^J \tilde{\alpha}_{kt}^{-1} \tilde{\mathbf{w}}_{kt} \tilde{\mathbf{w}}_{kt}^T$, where $\tilde{\alpha}_{kt}$ is the sample eigenvalue and $\tilde{\mathbf{w}}_{kt}$ is the sample eigenvector. Without loss of generality, we show the case when $K = J$ in the modeled $\bar{V}^{(-)}$ here. Focusing on eigenvalue estimation, by substituting sample eigenvectors $\tilde{\mathbf{w}}_{kt}$ into $\bar{V}_t^{(-)}$, the objective function in (5) can be approximated as

$$\begin{aligned} & \sum_{t=1}^T \text{Tr} \left\{ \left(\sum_{k'=1}^J \tilde{\alpha}_{kt'} \tilde{\mathbf{w}}_{k't} \tilde{\mathbf{w}}_{k't}^T \right) \left(\sum_{k=1}^J (\alpha_k^{-1}(t, \beta_k) - \tilde{\alpha}_{kt}^{-1}) \tilde{\mathbf{w}}_{kt} \tilde{\mathbf{w}}_{kt}^T \right)^2 \right\} \\ &= \sum_{t=1}^T \text{Tr} \left\{ \left(\sum_{k'=1}^J \tilde{\alpha}_{kt'} \tilde{\mathbf{w}}_{k't} \tilde{\mathbf{w}}_{k't}^T \right) \left(\sum_{k=1}^J (\alpha_k^{-1}(t, \beta_k) - \tilde{\alpha}_{kt}^{-1})^2 \tilde{\mathbf{w}}_{kt} \tilde{\mathbf{w}}_{kt}^T \right) \right\} \\ &= \sum_{t=1}^T \text{Tr} \left\{ \sum_{k=1}^J \sum_{k'=1}^J \tilde{\alpha}_{kt'} \left(\alpha_k^{-1}(t, \beta_k) - \tilde{\alpha}_{kt}^{-1} \right)^2 \tilde{\mathbf{w}}_{k't} \tilde{\mathbf{w}}_{k't}^T \tilde{\mathbf{w}}_{kt} \tilde{\mathbf{w}}_{kt}^T \right\} \\ &= \sum_{t=1}^T \text{Tr} \left\{ \sum_{k=1}^J \tilde{\alpha}_{kt} \left(\alpha_k^{-1}(t, \beta_k) - \tilde{\alpha}_{kt}^{-1} \right)^2 \tilde{\mathbf{w}}_{kt} \tilde{\mathbf{w}}_{kt}^T \right\} \\ &= \sum_{t=1}^T \sum_{k=1}^J \tilde{\alpha}_{kt} \left(\alpha_k^{-1}(t, \beta_k) - \tilde{\alpha}_{kt}^{-1} \right)^2, \end{aligned} \quad (7)$$

by noting that $\tilde{\mathbf{w}}_{kt}$'s are orthonormal bases. We can interpret (7) as a weighted square difference, which assigns more weights to the larger eigenvalues in estimation.

Furthermore, we compare the objective function in (5) to those based on the likelihood and the Frobenius norm in a similar fashion. The likelihood-based objective function is

$$\sum_{t=1}^T \left(\log \det(\bar{V}(t)) + \text{Tr} \left\{ \bar{V}^{(-)}(t) \tilde{V}_t \right\} \right),$$

where $\log \det(\cdot)$ indicates the logarithm of the determinant of a matrix. Similarly, the objective function based on the Frobenius norm of the difference between the true covariance and estimated covariance matrix is

$$\sum_{t=1}^T \left(\text{Tr} \left\{ (\bar{V}(t) - \tilde{V}_t)(\bar{V}(t) - \tilde{V}_t)^T \right\} \right).$$

Following a similar technique in obtaining (7) with respect to the eigenvalue estimation, we can simplify the likelihood-based objective function to be

$$\sum_{t=1}^T \sum_{k=1}^K \left(\frac{\tilde{\alpha}_{kt}}{\alpha_k(t, \beta'_k)} + \log(\alpha_k(t, \beta'_k)) \right),$$

and the objective function corresponding to the Frobenius norm turns to be

$$\sum_{t=1}^T \sum_{k=1}^K \left(\alpha_k(t, \beta'_k) - \tilde{\alpha}_{kt} \right)^2.$$

By comparing these simplified forms of the objective functions with respect to eigenvalues, it is straightforward that the proposed GEE-type objective function is able to improve the estimation of the more important eigen-components as it assigns more weights to the larger eigenvalues. The merit of the proposed GEE-type objective function is demonstrated via the eigenvalue estimation above, as there is no explicit “simple” formulation when eigenvalue functions and eigenvector functions are simultaneously estimated. In addition, we conduct a simple simulation study in the supplementary materials (Section 1.3) to show the estimation performance of all three different objective functions. The proposed norm in (5) produces smaller mean square errors of the estimations of the time-varying eigenvalue functions and the bases coefficients compared to the ones produced by the likelihood or Frobenius norm approaches.

2.4. Incorporation of Random Effects

Random-effects modeling (Laird and Ware 1982; Gardiner, Luo, and Roman 2009) is quite useful to accommodate store-specific effects. For the IRI marketing data application, the volumes of store sales show strong variations among different types of stores and products, as mentioned in Section 1. Therefore, we propose an eigen-decomposition equation with random effects (EERE) model to incorporate the store-wise heterogeneity into the time-varying PCA framework.

With store-specific random effects incorporated, the multivariate outcome $\mathbf{y}_{it}$ modeled in Section 2.2 cannot be observed directly. We assume that the observed multivariate response $\mathbf{y}_{it}^{\text{obs}}$ from the i th store at time t consists of two parts, that is, $\mathbf{y}_{it}^{\text{obs}} = \mathbf{y}_{it} + \mathbf{y}_i$, where $\mathbf{y}_i$ represents a J -dimensional store-specific random effect following a multivariate normal distribution $N(\mathbf{0}, \mathbf{D})$, and it is independent of the true variable of interest $\mathbf{y}_{it}$. Therefore, the corresponding variance of $\mathbf{y}_{it}^{\text{obs}}$ can be decomposed as

$$\text{var}(\mathbf{V}_t^{\text{obs}}) = \text{var}(\mathbf{y}_{it}) + \text{var}(\mathbf{y}_i), \quad (8)$$

and the targeted underlying time-varying covariance structure of $\mathbf{y}_{it}$ can be modeled by (2) and further estimated via (6).

It is clear that the store-wise heterogeneity induced by $\mathbf{y}_i$ brings additional variation to the covariance structure of interest. Incorporating random effects in the proposed model provides more accurate estimation for underlying associations among products. In contrast, modeling directly on the observed response $\mathbf{y}_{it}^{\text{obs}}$ is not able to differentiate different sources of correlations among products, and thus could lead to a biased

estimator of the covariance matrix $\text{var}(\mathbf{y}_{it})$ as it ignores the variation of the random effects $\mathbf{y}_i$.

Next, we introduce the estimation of multivariate random effects. Elashoff and Ryan (2004) developed an ES algorithm for the missing data in the GEE model which focuses on univariate longitudinal data; and Proudfoot et al. (2018) employ the ES algorithm to estimate a multivariate longitudinal data model with a mean regression form. In this article, we adapt the ES approach to the proposed objective functions to estimate the multivariate random effects, in a similar fashion as in Proudfoot et al. (2018).

Specifically, we estimate the random effects $\mathbf{y}_i$'s, the covariance matrix $\mathbf{D} = \text{cov}(\mathbf{y}_i)$, and the mean $\boldsymbol{\mu}_t$ of the interested underlying variable $\mathbf{y}_{it}$ iteratively. At the l th iteration, we update the random effects by

$$\hat{\mathbf{y}}^{(l)} = \hat{\mathbf{Z}}^{(l-1)} \left((\mathbf{1}_T \mathbf{1}_T^T) \otimes \hat{\mathbf{D}}^{(l-1)} + \hat{\mathbf{V}}^{(l-1)} \right)^{-1} (\mathbf{1}_T^T \otimes \hat{\mathbf{D}}^{(l-1)})^T, \quad (9)$$

where $\hat{\mathbf{Z}}^{(l-1)} = \left(\mathbf{Y}_1^{\text{obs}} - (\mathbf{1}_N \otimes \hat{\boldsymbol{\mu}}_1^{(l-1)T}), \mathbf{Y}_2^{\text{obs}} - (\mathbf{1}_N \otimes \hat{\boldsymbol{\mu}}_2^{(l-1)T}), \dots, \mathbf{Y}_T^{\text{obs}} - (\mathbf{1}_N \otimes \hat{\boldsymbol{\mu}}_T^{(l-1)T}) \right)$ is an $N \times JT$ matrix, $\mathbf{Y}_t^{\text{obs}} = (\mathbf{y}_{1t}^{\text{obs}}, \mathbf{y}_{2t}^{\text{obs}}, \dots, \mathbf{y}_{Nt}^{\text{obs}})^T$, $\hat{\mathbf{V}}^{(l-1)} = \text{diag}(\hat{\mathbf{V}}_1^{(l-1)}, \hat{\mathbf{V}}_2^{(l-1)}, \dots, \hat{\mathbf{V}}_T^{(l-1)})$ is a $JT \times JT$ block diagonal matrix and $\hat{\mathbf{V}}_t^{(l-1)}$ denotes the estimated sample covariance matrix of underlying variable $\mathbf{y}_t$ at $(l-1)$ th iteration by

$$\hat{\mathbf{V}}_t^{(l-1)} = \frac{1}{N} \sum_{i=1}^N (\mathbf{y}_{it}^{\text{obs}} - \hat{\mathbf{y}}_i^{(l-1)} - \hat{\boldsymbol{\mu}}_t^{(l-1)}) (\mathbf{y}_{it}^{\text{obs}} - \hat{\mathbf{y}}_i^{(l-1)} - \hat{\boldsymbol{\mu}}_t^{(l-1)})^T,$$

where $\hat{\boldsymbol{\mu}}_t^{(l-1)} = \frac{1}{N} \sum_{i=1}^N (\mathbf{y}_{it}^{\text{obs}} - \hat{\mathbf{y}}_i^{(l-1)})$. Moreover, the random effects' covariance matrix $\mathbf{D}$ is updated via

$$\hat{\mathbf{D}}^{(l)} = \hat{\mathbf{D}}^{(l-1)} - (\mathbf{1}_T^T \otimes \hat{\mathbf{D}}^{(l-1)}) \left((\mathbf{1}_T \mathbf{1}_T^T) \otimes \hat{\mathbf{D}}^{(l-1)} + \hat{\mathbf{V}}^{(l-1)} \right)^{-1} (\mathbf{1}_T^T \otimes \hat{\mathbf{D}}^{(l-1)})^T.$$

The updates to the random effects $\hat{\mathbf{y}}^{(l)}$ in Equation (9) and the random effects' covariance matrix $\hat{\mathbf{D}}^{(l)}$ are calculated as conditional moments under the multivariate normal distribution. See Equation (6) in Proudfoot et al. (2018) for more details.

Once the random effects are estimated through the iterative steps above until convergence at the L th iteration, we redefine the objective function in (6) by replacing $\mathbf{h}_{it}$ with $\mathbf{h}_{it}^*$ as

$$\mathbf{h}_{it}^*(\boldsymbol{\beta}, \mathbf{v}; \boldsymbol{\mu}_t) = \left(\bar{\mathbf{V}}^{(-)}(t|\boldsymbol{\beta}, \mathbf{v}) - \hat{\mathbf{V}}_t^{(L-1)} \right) (\mathbf{y}_{it}^{\text{obs}} - \boldsymbol{\mu}_t - \hat{\mathbf{y}}_i^{(L)}).$$

Then we minimize the following objective function

$$L_{\phi}^*(\boldsymbol{\beta}, \mathbf{v}; \boldsymbol{\mu}) = \sum_{t=1}^T \left(\sum_{i=1}^N \frac{\mathbf{h}_{it}^{*T}(\boldsymbol{\beta}, \mathbf{v}; \boldsymbol{\mu}_t) \mathbf{h}_{it}^*(\boldsymbol{\beta}, \mathbf{v}; \boldsymbol{\mu}_t)}{N} + \phi \sum_{k \neq k'} \|\mathbf{w}_k(t, \boldsymbol{\beta}_k)^T \mathbf{w}_{k'}(t, \boldsymbol{\beta}_{k'})\|_2^2 \right), \quad (10)$$

via an iterative algorithm as discussed in Section 3.

3. Algorithm and Implementation

In this section, we propose an efficient algorithm to solve the optimization problem for the proposed EERE model, that is, estimating the multivariate random effects via the ES algorithm, and updating the eigenvalues and eigenvectors iteratively. In addition, we provide detailed discussion about selecting important tuning parameters.

3.1. Algorithm for EERE

In this section, we provide the algorithm of estimating the EERE model which iterates through the Newton–Raphson and the Expectation-Substitution algorithms. To simplify the notation, we use $\alpha_k(t)^{(m)}$ and $\mathbf{w}_k(t)^{(m)}$ as shorthand for $\alpha_k(t, \hat{\boldsymbol{\beta}}_k^{(m)})$ and $\mathbf{w}_k(t, \hat{\mathbf{v}}_k^{(m)})$, respectively.

Algorithm: Estimated eigenanalysis with random effects (EERE)

- Step 1:** Initialize $\boldsymbol{\gamma}_i^{(0)}$, $\mathbf{D}^{(0)}$, $\boldsymbol{\mu}_t^{(0)}$, and $\mathbf{V}^{(0)}$;
- Step 2:** Update the random effects $\boldsymbol{\gamma}_i^{(l)}$ using the ES algorithm in (9);
- Step 3:** Update the parameters $\mathbf{D}^{(l)}$, $\boldsymbol{\mu}^{(l)}$, and $\mathbf{V}^{(l)}$ at the l th step;
- Step 4:** Repeat Steps 2 and 3 until $\|\boldsymbol{\gamma}_i^{(l)} - \boldsymbol{\gamma}_i^{(l-1)}\| < \epsilon_\gamma$, where ϵ_γ is a tolerance level.
- Step 5:** After removing the estimated random effects from the response, initialize the eigenvectors as the sample eigenvectors: $\mathbf{w}_k(t)^{(0)} = \tilde{\mathbf{w}}_{kt}$;
- Step 6:** Given $\mathbf{w}_k(t)^{(m-1)}$'s, update the eigenvalues $\alpha_k(t)^{(m)}$'s by minimizing (10), and update the spline coefficients $\hat{\boldsymbol{\beta}}^{(m)}$;
- Step 7:** Given $\alpha_k(t)^{(m)}$'s, update $\mathbf{w}_k(t)^{(m)}$'s using the Newton–Raphson algorithm, and update the spline coefficients $\hat{\mathbf{v}}^{(m)}$;
- Step 8:** Repeat Steps 6 and 7, stop if $d_\beta^{-1} \|\hat{\boldsymbol{\beta}}^{(m)} - \hat{\boldsymbol{\beta}}^{(m-1)}\| + d_v^{-1} \|\hat{\mathbf{v}}^{(m)} - \hat{\mathbf{v}}^{(m-1)}\| < \epsilon_e$, where d_β and d_v are the dimensions of $\boldsymbol{\beta}$ and $\mathbf{v}$, respectively, and ϵ_e is a chosen tolerance level.
-

The Newton–Raphson algorithm for the eigenvectors estimation in Step 6 is multivariate. Specifically, at the (m) th iteration, we update

$$\mathbf{w}(t)^{(m)} = \mathbf{w}(t)^{(m-1)} - \mathbf{J}_f^{-1} \left(\mathbf{w}(t)^{(m-1)} \right) \mathbf{f} \left(\mathbf{w}(t)^{(m-1)} \right),$$

where $\mathbf{w}(t)^{(m-1)} = (\mathbf{w}_1(t, \mathbf{v}_1)^{(m-1)T}, \dots, \mathbf{w}_K(t, \mathbf{v}_K)^{(m-1)T})^T$, K is the number of principal components, $\mathbf{f}(\cdot)$ is a vector of the first derivatives of the objective function in (10), and $\mathbf{J}_f(\cdot)$ is the Jacobian matrix of the second derivatives of (10). In detail, $\mathbf{f}$ and $\mathbf{J}_f$ have the following forms:

$$\begin{aligned} \mathbf{f}(\mathbf{w}(t)) &= \frac{2}{N} \sum_{i=1}^N \dot{\mathbf{h}}_{it}^T \mathbf{h}_{it} \\ &+ 2\phi \left(\sum_{l \neq 1} (\mathbf{w}_1(t)^T \mathbf{w}_l(t)) \mathbf{w}_l(t), \dots, \sum_{l \neq K} (\mathbf{w}_K(t)^T \mathbf{w}_l(t)) \mathbf{w}_l(t) \right)^T \end{aligned}$$

and

$$\begin{aligned} \mathbf{J}_f(\mathbf{w}(t)) &\approx \frac{2}{N} \sum_{i=1}^N \dot{\mathbf{h}}_{it}^T \dot{\mathbf{h}}_{it} + 2\phi \\ &\left(\begin{array}{ccc} \sum_{l \neq 1} \mathbf{w}_l(t) \mathbf{w}_l(t)^T & \dots & \mathbf{w}_1(t)^T \mathbf{w}_K(t) \mathbf{I} + \mathbf{w}_1(t) \mathbf{w}_K(t)^T \\ \vdots & \ddots & \vdots \\ \dots & \dots & \sum_{l \neq K} \mathbf{w}_l(t) \mathbf{w}_l(t)^T \end{array} \right), \end{aligned}$$

where $\dot{\mathbf{h}}_{it} = \frac{\partial \mathbf{h}_{it}}{\partial \mathbf{w}(t)}$, and $\mathbf{I}$ is a $J \times J$ identity matrix. Note that

$$\frac{1}{N} \sum_{i=1}^N \dot{\mathbf{h}}_{it}^T \dot{\mathbf{h}}_{it} \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

since $\mathbf{E}[\mathbf{h}_{it}] = \mathbf{0}$ and $\ddot{\mathbf{h}}_{it} = \frac{\partial \dot{\mathbf{h}}_{it}}{\partial \mathbf{w}(t)}$.

Although the regularization is employed in the objective function, the estimated eigenvectors are still not guaranteed to be exactly orthogonal due to the property of L_2 penalty. Hence, in Step 6, the Gram–Schmidt process is further applied on eigenvectors to ensure their pairwise orthonormality. Namely, the Gram–Schmidt process converts the estimated eigenvectors $\mathbf{w}_k(t)^{(m)}$'s into orthonormal vectors at each time point.

In conclusion, the proposed algorithm updates the eigenvalues and the eigenvectors iteratively, leading to a non-increasing sequence of the objective function, where the objective function at each iteration is differentiable and can be solved by the Newton–Raphson algorithm. Although the proposed algorithm cannot guarantee the global optimum, we implement multiple initial values to achieve an ideal solution.

3.2. Tuning Parameters Selection

In this subsection, we provide implementation strategies for selecting tuning parameters; for example, the number of components in PCA, the knots in nonparametric estimation for the eigenvalues and the eigenvectors, the order of the polynomial splines, and the tuning parameter ϕ for the L_2 penalty.

Choosing the number of components remains an open problem in PCA. One popular approach is to create a scree plot of the eigenvalues, in which the number of components is chosen at the elbow or the scree of the line drawn. This strategy is appealing since it provides a visual representation of the majority of variation to be retained from the data. For the IRI marketing data, which is collected longitudinally, analyzing the scree plots for each time point is inefficient, because different time points may result in different numbers of chosen components. Instead,

we propose calculating an overall variation by averaging over the T time points.

One such numeric criterion is to choose the percentage of overall variation of the data which determines the number of components to retain; that is, to increase the number of components until the desired cumulative percentage of variation from the data is attained. The cumulative percentage of variance contributed from the first K components is

$$\frac{\sum_{k=1}^K \sum_{t=1}^T \hat{\alpha}_{kt}}{\sum_{j=1}^J \sum_{t=1}^T \hat{\alpha}_{jt}}, \quad (11)$$

where $\hat{\alpha}_{jt}$ is the discretized sample eigenvalues (nonsmoothed) adjusted by the random effects.

Next, we illustrate how to choose the order of the polynomial splines and the knots. Although we can employ an additional smoothing penalty to select the number of bases (Ruppert 2002), it requires the selection of additional tuning parameters which can significantly increase the computation cost. In this article, we use the cubic spline for our numerical studies, which is widely adopted in many applications to achieve the desired smoothness (Hastie, Tibshirani, and Friedman 2009).

As suggested by Likhachev (2017), the AIC or BIC could be applied for the knots selection. In our framework, the information criterion is constructed based on the quasi-likelihood (Wedderburn 1974) analogous to the GEE model. For computational convenience, we recommend adopting the strategy suggested by Xue and Qu (2012), that is, using equally spaced knots with the number of interior knots $\lceil N^{1/(2M+3)} \rceil$, where $\lceil \cdot \rceil$ denotes the ceiling function, N is the sample size and $M + 1$ is the degree of the polynomial spline function. A similar knots selection strategy was also applied in Huang, Wu, and Zhou (2004) and Xue, Qu, and Zhou (2010). The proposed method allows the eigenvalues and eigenvectors to have different sets of knots. However, the need for using different sets of knots highly depends on the observed data. For example, it is useful to use different sets of knots if the eigenvalues change slightly, while the eigenvectors change more dramatically over the same time range. Since this type of pattern is not present in our numerical studies, we choose the same number of knots in estimating eigenvalues and eigenvectors in this article.

Another critical issue is to choose the tuning parameter ϕ associated with the orthogonality penalty. We choose ϕ based on a grid search for a given range of positive values to minimize the unpenalized loss function:

$$\frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \hat{h}_{it}^{*T} \left(\hat{\beta}(\phi), \hat{\nu}(\phi); \hat{\mu}(\phi) \right) \hat{h}_{it}^* \left(\hat{\beta}(\phi), \hat{\nu}(\phi); \hat{\mu}(\phi) \right),$$

where

$$\hat{h}_{it}^* \left(\hat{\beta}(\phi), \hat{\nu}(\phi); \hat{\mu}(\phi) \right) = \left(\widehat{\mathbf{V}}^{(-)}(t | \hat{\beta}(\phi), \hat{\nu}(\phi)) - \hat{\mathbf{V}}_t^{(L-1)} \right) (\mathbf{y}_{it}^{\text{obs}} - \hat{\mu}_t(\phi) - \hat{\gamma}_i^{(L)}),$$

and $\hat{\beta}(\phi)$, $\hat{\nu}(\phi)$, and $\hat{\mu}(\phi)$ are the estimations from the proposed objective function (10) given tuning parameter ϕ .

In fact, the L_2 -penalty is imposed to improve computational stability in practice, but does not influence the model estimation very much. In the supplementary materials, we provide additional simulation results to illustrate the estimation effects of the L_2 -penalty with different shrinkage levels as well as the estimation results without the L_2 -penalty. The simulation study indicates that the proposed method is robust against the shrinkage magnitudes of the L_2 -penalty within a reasonable range. In addition, with an appropriately selected ϕ for the L_2 -penalty, the average computation time for the proposed model is 19 min, while the model without the L_2 -penalty ($\phi = 0$) requires 23 min, which also shows that the L_2 -penalty can help to speed up computation.

In our implementation, the number of eigen-components, the smoothing parameters (the order of splines and the number of knots) and the L_2 -penalty coefficient are tuned sequentially, as the selection of each set of those parameters does not depend upon the selection of the others.

4. Simulation

In this section, we evaluate the proposed methodology with simulation studies, and compare it with other competitive methods. Specifically, we compare the proposed EERE method with the DPCA method, the DPCA incorporating marginal random effects (DPCARE), the smooth PCA (SPCA) method (Miller and Bowman 2012), and the SPCA incorporating marginal random effects (SPCARE).

For the EERE model, the tuning parameters are selected as described in Section 3.2 and the cubic splines are used for modeling both the eigenvalues and the eigenvectors. The DPCA performs the standard PCA on the sample covariance matrix at each time point separately and extracts the corresponding eigen-components. The DPCARE allows the random effects embedded into the DPCA method, for which we first estimate the marginal random effects with respect to each product through the statistical computing platform R (R Core Team 2019) and the R package “lme4” (Bates et al. 2015), and then perform DPCA after removing the estimated random effects. The SPCA is implemented by the R package “sm” (Bowman and Azzalini 2018), and the SPCARE adopts a similar strategy as the DPCARE, which applies the SPCA after removing the marginal random effects.

The simulation setting is designed to mimic the real data application of IRI products’ sales. We set the number of products as $J = 20$, the number of stores as $N = 200$ and the number of time points as $T = 11$. The eigenvectors, $\mathbf{w}_{kt}^0$ ’s ($k = 1, \dots, J$), are obtained from the real data. In particular, on real data, after removing the random effects estimated from the EERE model, we extract the eigenvector $\mathbf{w}_{kt}^0$ ’s ($k = 1, \dots, J$) from the sample covariance matrix at each time point.

Furthermore, we generate the first two eigenvalues as follows:

$$\alpha_{1t}^0 = \left(a_{10} + a_{11} \frac{t}{T} + a_{12} \left(\frac{t}{T} \right)^2 \right)^{-1} \quad \text{and} \quad \alpha_{2t}^0 = \left(a_{20} + a_{21} \frac{t}{T} + a_{22} \left(\frac{t}{T} \right)^2 \right)^{-1},$$

where $(a_{10}, a_{11}, a_{12})^T = (0.19, 0.20, 0.20)^T$, $(a_{20}, a_{21}, a_{22})^T = (0.30, 0.50, -0.50)^T$. The remaining $J - 2$ eigenvalues are gen-

erated from $N(1, 0.2^2)$ at each time point. Consequently, we generate response vector $\mathbf{y}_{it}^*$ ($i = 1, \dots, N$) from a multivariate normal distribution with mean zero and covariance matrix $\mathbf{V}_t^0 = \sum_{k=1}^J \alpha_{kt}^0 \mathbf{w}_{kt}^0 (\mathbf{w}_{kt}^0)^T$.

Table 1. The mean absolute deviation of error (MADE) for eigenvalue estimation and the mean cosine deviation error (MCDE) for eigenvector estimation, comparing the DPCA, the DPCARE, the SPCA, the SPCARE and the proposed EERE, based on 100 replications.

Trule	Comp	DPCA	DPCARE	SPCA	SPCARE	EERE
Eigenvalue (MADE)	First	5.36	1.94	5.22	1.96	0.37
	Second	3.97	2.71	3.76	2.31	0.34
Eigenvector (MCDE)	First	0.6165	0.3468	0.6191	0.4058	0.0318
	Second	0.6697	0.4316	0.35700	0.2647	0.0110

NOTE: The bold values emphasize our methodology's performance.

In addition, we generate the multivariate random effect $\mathbf{y}_i$ from a multivariate normal distribution with mean $\mathbf{0}$ and an exchangeable covariance matrix $V^2(\rho \mathbf{1}_{10}^T \mathbf{1}_{10} + (1 - \rho) \mathbf{I}_{10})$, where $V^2 = 3$ and $\rho = 0.25$. Consequently, the J -dimensional response variable is generated as $\mathbf{y}_{it} = \mathbf{y}_{it}^* + \mathbf{y}_i$, for $i = 1, \dots, N$ and $t = 1, \dots, T$.

To evaluate the estimation of the time-varying eigenvalues, we examine the plots of the average estimated eigenvalues versus time. Also, we calculate the mean absolute deviation of error (MADE) for the eigenvalues:

$$\text{MADE}_k = \frac{1}{T} \sum_{t=1}^T |\hat{\alpha}_{kt} - \alpha_{kt}^0| / \text{range}(\alpha_k^0),$$

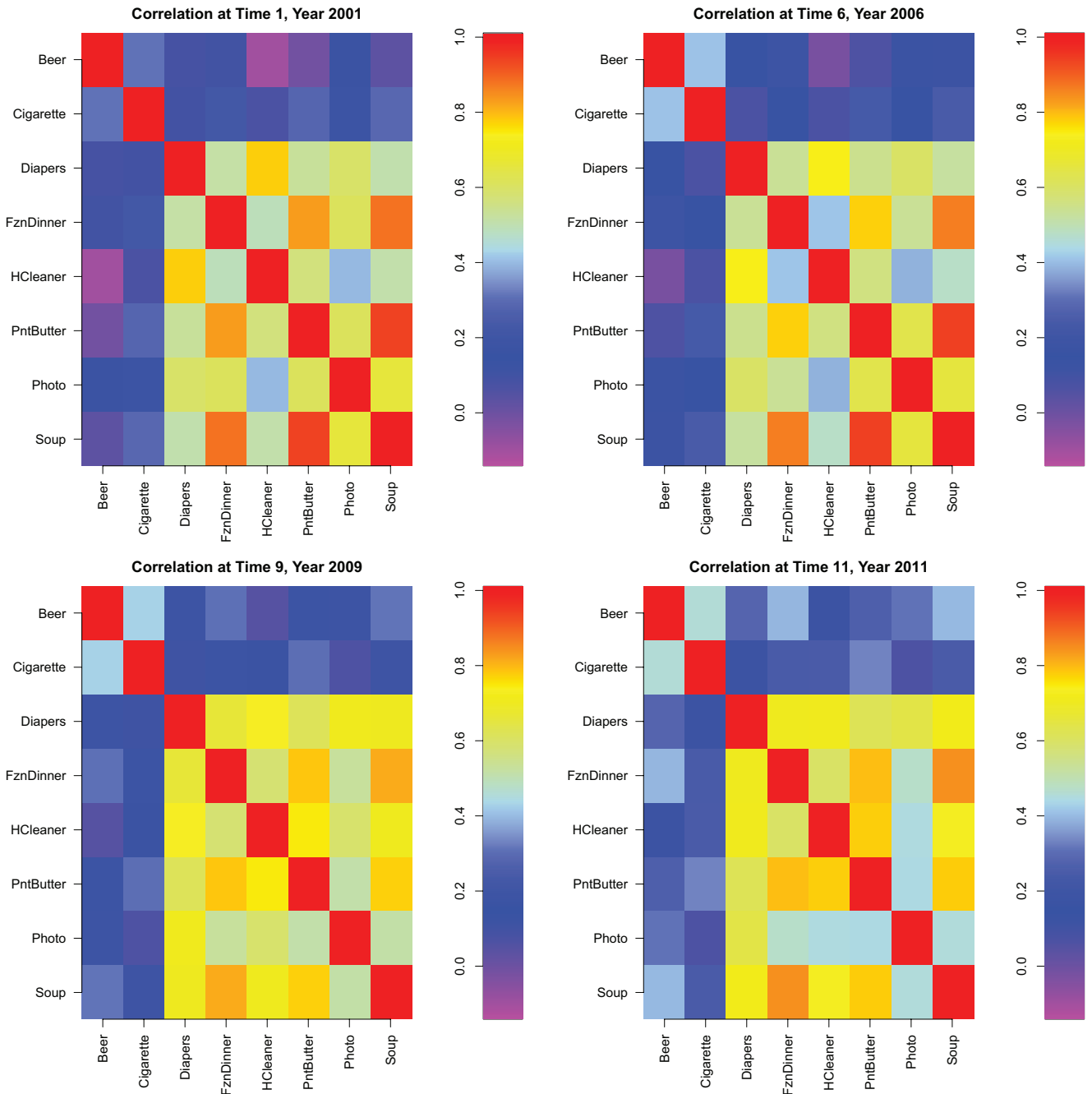


Figure 2. Correlations for eight selected products in the IRI dataset at the years 2001, 2006, 2009, and 2011.

where α_{kt}^0 is the k th true eigenvalue at time t and $k = 1, 2$. The range(α_k^0) = $\max(\alpha_{kt}^0) - \min(\alpha_{kt}^0)$, which covers the minimum and maximum among all time points. In general, it is not trivial to quantify the estimation errors of the time-varying eigenvectors. We adopt the mean cosine distance error (MCDE) of the estimated time-varying eigenvectors based on their angles with the true eigenvectors. In particular, the MCDE value is calculated as

$$\text{MCDE}_k = \frac{1}{T} \sum_t \left(1 - \cos(\hat{\mathbf{w}}_{kt}, \mathbf{w}_{kt}^0) \right) = \frac{1}{T} \sum_t \left(1 - \hat{\mathbf{w}}_{kt}^T \mathbf{w}_{kt}^0 \right),$$

for $k = 1, 2$, where $\hat{\mathbf{w}}_{kt}$ is the estimated k th eigenvector at time point t and $\mathbf{w}_{kt}^0$ is the corresponding true value. It is clear that the MCDE value is in a range of $[0, 2]$, and a smaller MCDE value suggests a more accurate eigenvector estimation. In addition, we evaluate time-varying eigenvector estimations through creating heatmaps corresponding to the average estimators of eigenvector loading over time. The heatmaps allow us to visualize group changes over time among different variables via color contrasts. Moreover, we also plot the time-varying curve of each loading of the estimated eigenvectors in the supplementary materials.

Table 1 summarizes the average MADE values and MCDE values for the estimations from the different models based on 100 replications. The results show that the EERE method yields the smallest MADE values and MCDE values compared to other competing methods. Figure 3 displays the estimated first and second time-varying eigenvalues obtained from various methods. In addition, it is evident that the smooth eigenvalue estimators from the EERE method are closest to the true eigenvalue curves. The DPCA and DPCARE methods produce non-smooth estimators as expected. Moreover, the SPCARE and DPCARE methods have better estimation results compared to the SPCA and the DPCA methods indicating the importance of incorporating the random effects associated with different stores. However, the proposed method still outperforms the

DPCA and the DPCARE significantly, suggesting that the proposed EERE has advantages from accounting for heterogeneity and neighboring information simultaneously.

The heatmaps in Figures 4 and 5 illustrate the estimated time-varying eigenvector estimators, where the EERE estimators are closest to the true eigenvectors compared to the other five methods. For the first eigenvector, Figure 4 shows that the estimators of the DPCA and SPCA have strong negative loadings for most of the products over all time points, implying that they are not able to capture the group changes over time for Products 1–3 and 12 at later time points. Furthermore, the DPCARE and SPCARE tend to group the first product with the last three products, which is very different far from the truth. For the second eigenvector, Figure 5 indicates that the estimator from the EERE method has the best approximation of the truth, while the SPCA and SPCARE methods lack power in capturing the grouping patterns at later time points, and the DPCA and DPCARE methods are not able to capture the smooth changes of the loading coefficients over time.

5. Analysis of IRI Marketing Data

In this article, we focus on grocery store sales spanning an 11-year time period, and apply the proposed method to the IRI marketing dataset, which consists of sales units of 552 grocery stores and 20 products collected over the years 2001–2011. Among the stores, 30% are located in the South, 28% in the West, 24% the Northeast, and 18% in the Midwest, and fewer than 1% of the stores do not belong to any chain. The product categories include beer, razor blades, carbonated beverages, cigarettes, cold cereals, deodorants, diapers, frozen dinners, frozen pizzas, hot dogs, household cleaners, laundry detergent, milk, mustards and ketchups, peanut butters, photography supplies, salty snacks, shampoos, soup, and toothbrushes. The 20 products represent a broad spectrum of consumer packaged

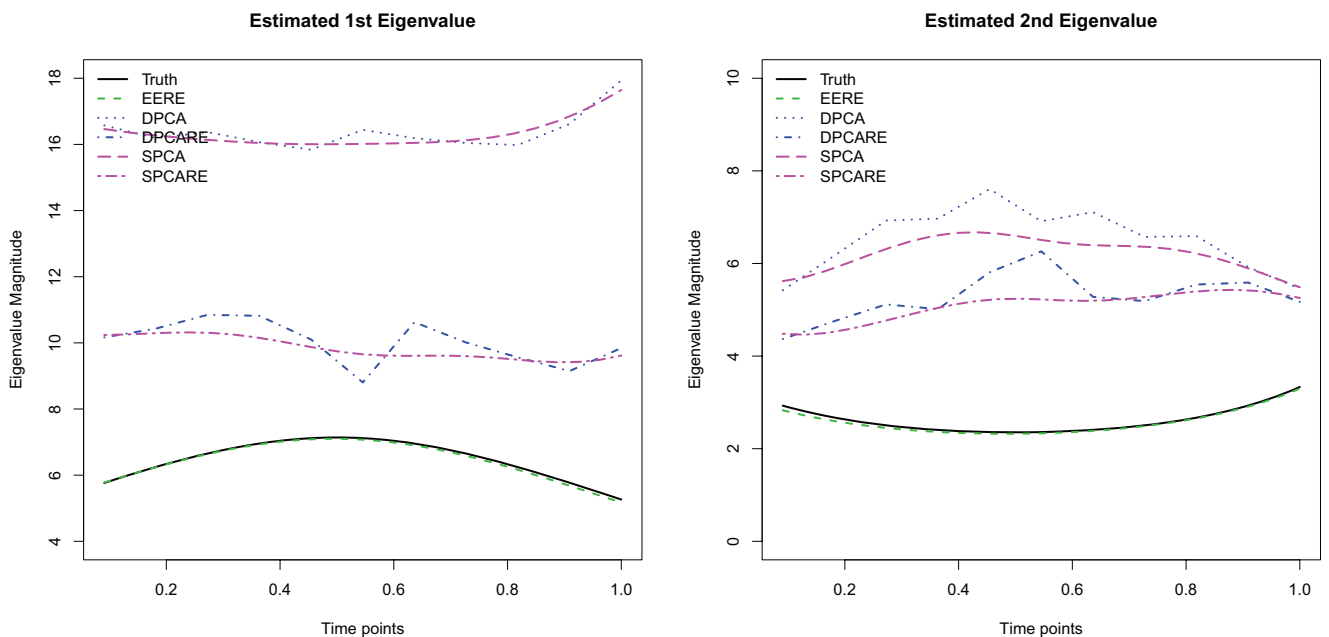


Figure 3. The estimated eigenvalue curves of the DPCA, the DPCARE, the SPCA, the SPCARE, and the EERE in the simulation study, based on 100 replications.

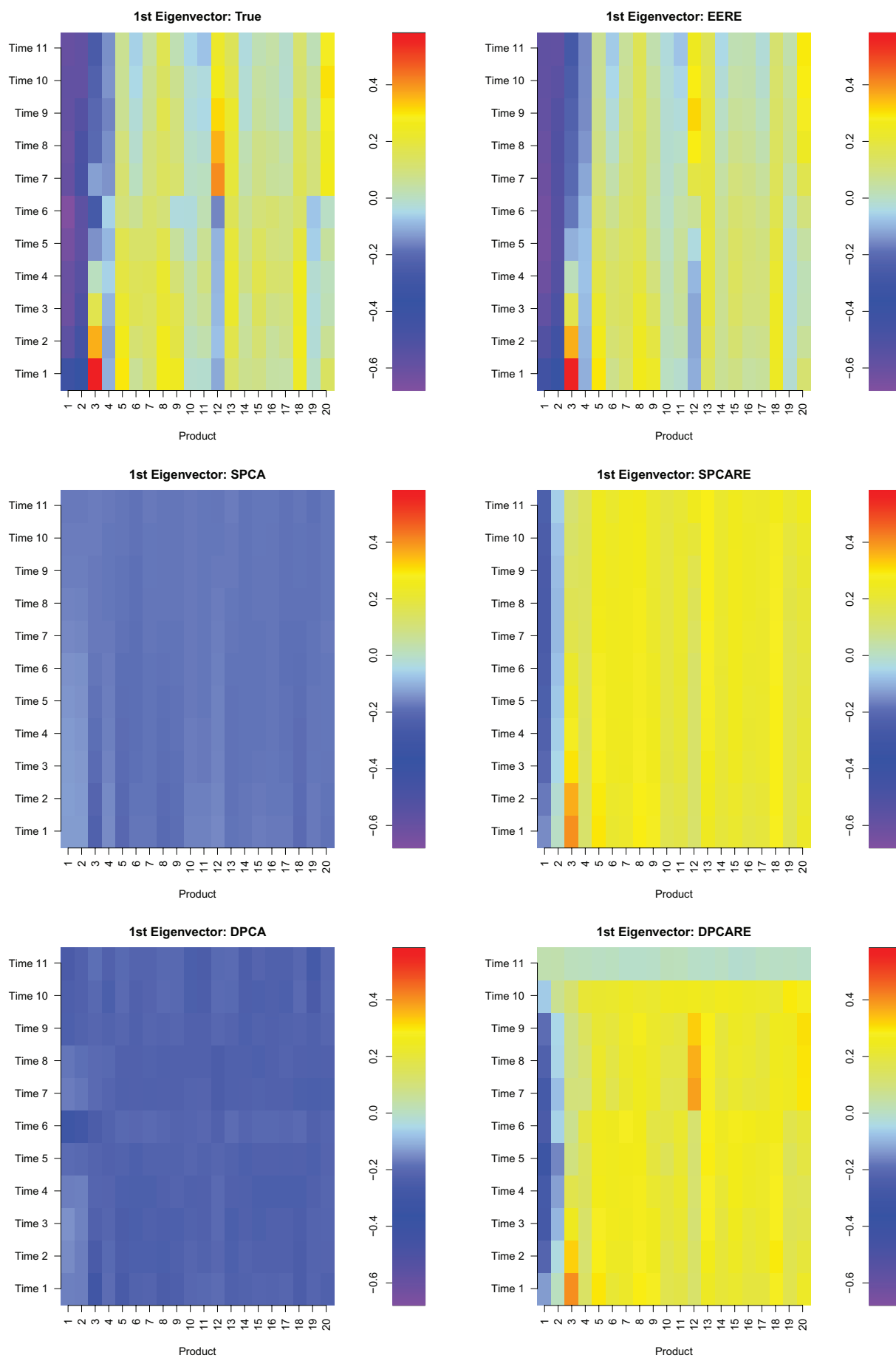


Figure 4. The heatmaps for the estimated first eigenvector for different models in the simulation study, based on 100 replications.

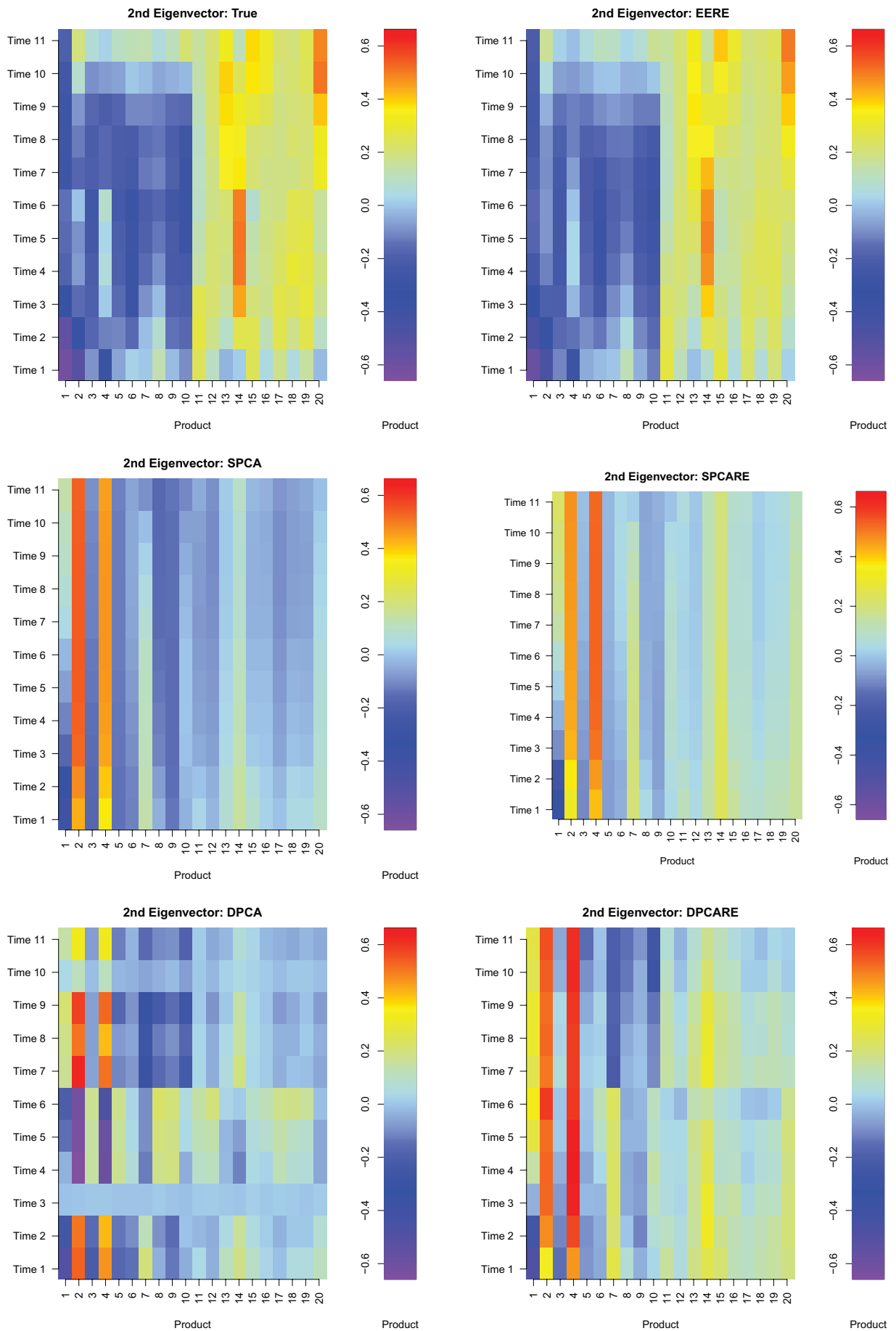


Figure 5. The heatmaps for the estimated second eigenvector for different models in the simulation study, based on 100 replications.

goods with varying amounts of sales over time. Among the products, milk has the largest volume of sales across time, and photography supplies have the smallest volume of sales over time. To overcome difficulties in different scales of sales volume, we rescale the data to achieve unit variance for the products over the 11 years.

This multivariate longitudinal sales data presents some interesting features, but is also challenging to analyze due to the sheer size, variability, and time-varying nature of the dataset. The variation among stores can be due to several extrinsic factors, such as geographic location and store size, or due to intrinsic factors, such as popularity and reputation of the stores. Figure 1 illustrates the average number of units sold at large, medium, and small stores for two products, beer and peanut butter. We notice that smaller stores tend to have a decreasing trend in sales of beer and peanut butter, while larger stores have an overall increase.

Since the number of products is large, it is essential to investigate correlations among product sales and capture the associations among them over time, and to reduce dimensionality. This enhances marketing researchers' ability to better understand consumer shopping behavior and marketing trends while considering a portion of the data that captures a reasonable amount of variation. Figure 2 illustrates the heatmaps for the correlation matrices among several selected products. We note that the magnitude among the pairwise correlations for photography supplies and other products changes over time, and the change is more obvious in later years. This phenomenon reflects the idea that consumers are likely to change their purchasing habits as technology progresses over time. We select the first two principal components, as they contribute about 80% of the total variation over the 11 years following the calculation in (11).

We implement the proposed EERE method and compare it to the DPCA, DPCARE, SPCA, SPCARE approaches. Figure 6 plots the estimated eigenvalues for different methods. Figures 7 and 8 show the first and second time-varying eigenvector

heatmaps, respectively. The heatmaps of DPCA, DPCARE, SPCA, and SPCARE have an overall averaging behavior for the first eigenvector and a grouping behavior for the second eigenvector. The first eigenvector describes an average sales volume among the products. The second eigenvector describes ingestible products versus noningestible products. The negative loadings with blue-purple hues of the second eigenvector for most years correspond to the following products: beer, carbonated beverages, cigarettes, cold cereal, frozen dinners, frozen pizzas, hot dogs, milk, mustards and ketchups, peanut butter, salty snacks, and soup. These products are those which consumers take into their bodies via swallowing or inhaling, hence "ingestible." The remaining noningestible products include razor blades, deodorants, diapers, household cleaners, laundry detergent, photography supplies, shampoo, and toothbrushes. The noningestible products are represented by positive loadings with yellow-orange hues.

Interestingly, the EERE heatmap in Figure 7 indicates that the beer and cigarette products, which are in yellow-orange hues, should be grouped together in contrast to the remaining products. The first eigenvector can be interpreted as the distinction between products for general consumers versus products for age-restricted consumers. Cigarettes and beer are age-restricted products for which consumers must be at least ages 18 and 21 years old, respectively, to consume and purchase in most regions and markets in the United States.

Figure 8 reveals that the heatmap of the proposed EERE method tends to group products differently from the other methods. In particular, the EERE makes a distinction between food and nonalcoholic beverages versus the remaining products. The foods and nonalcoholic beverages group includes carbonated beverages, cold cereal, frozen dinners, frozen pizzas, hot dogs, milk, peanut butter, salty snacks, and soup. The remaining product group includes beer, cigarettes, razor blades, deodorants, diapers, household cleaners, laundry detergent, photography supplies, shampoo, and toothbrushes. In this sense, the

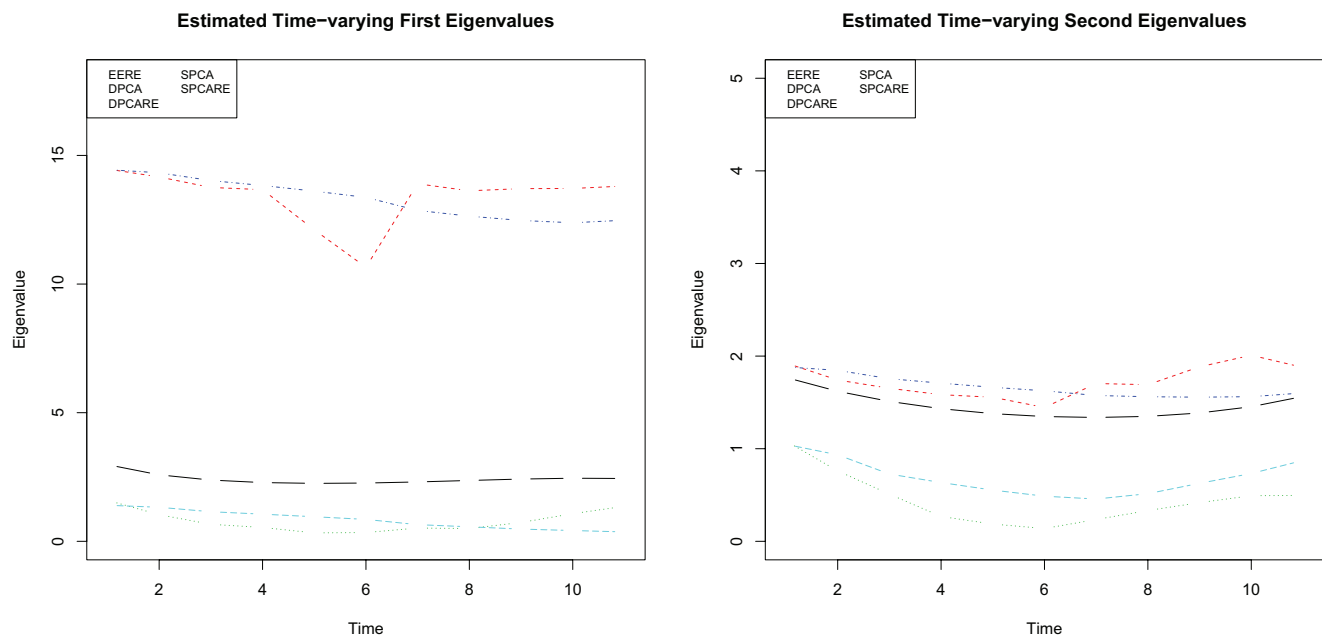


Figure 6. The estimated eigenvalue curves of the DPCA, the DPCARE, the SPCA, the SPCARE, and the EERE in the real data analysis.

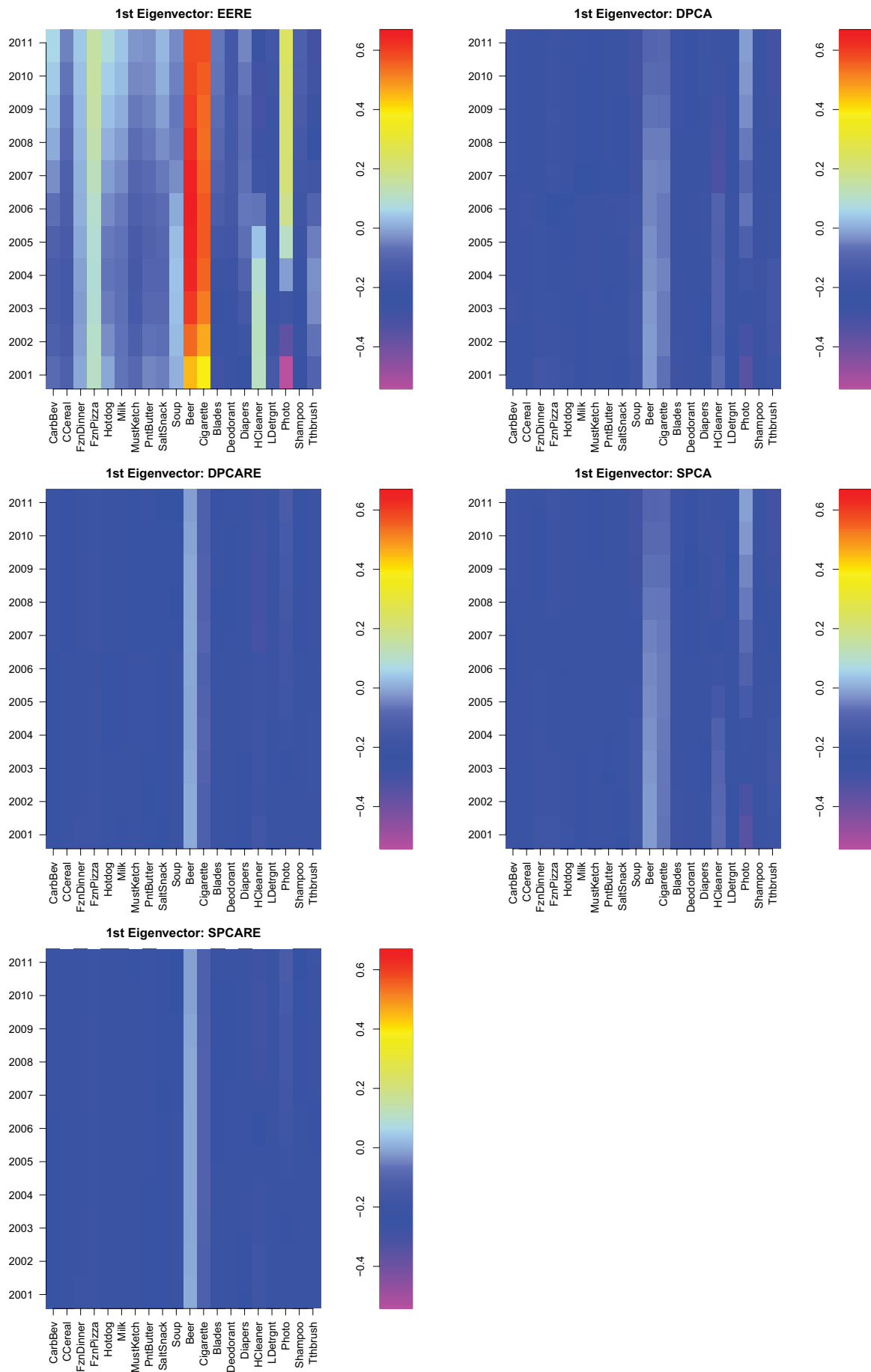


Figure 7. The heatmaps for the estimated first eigenvector for different models in IRI marketing data analysis containing 20 products.

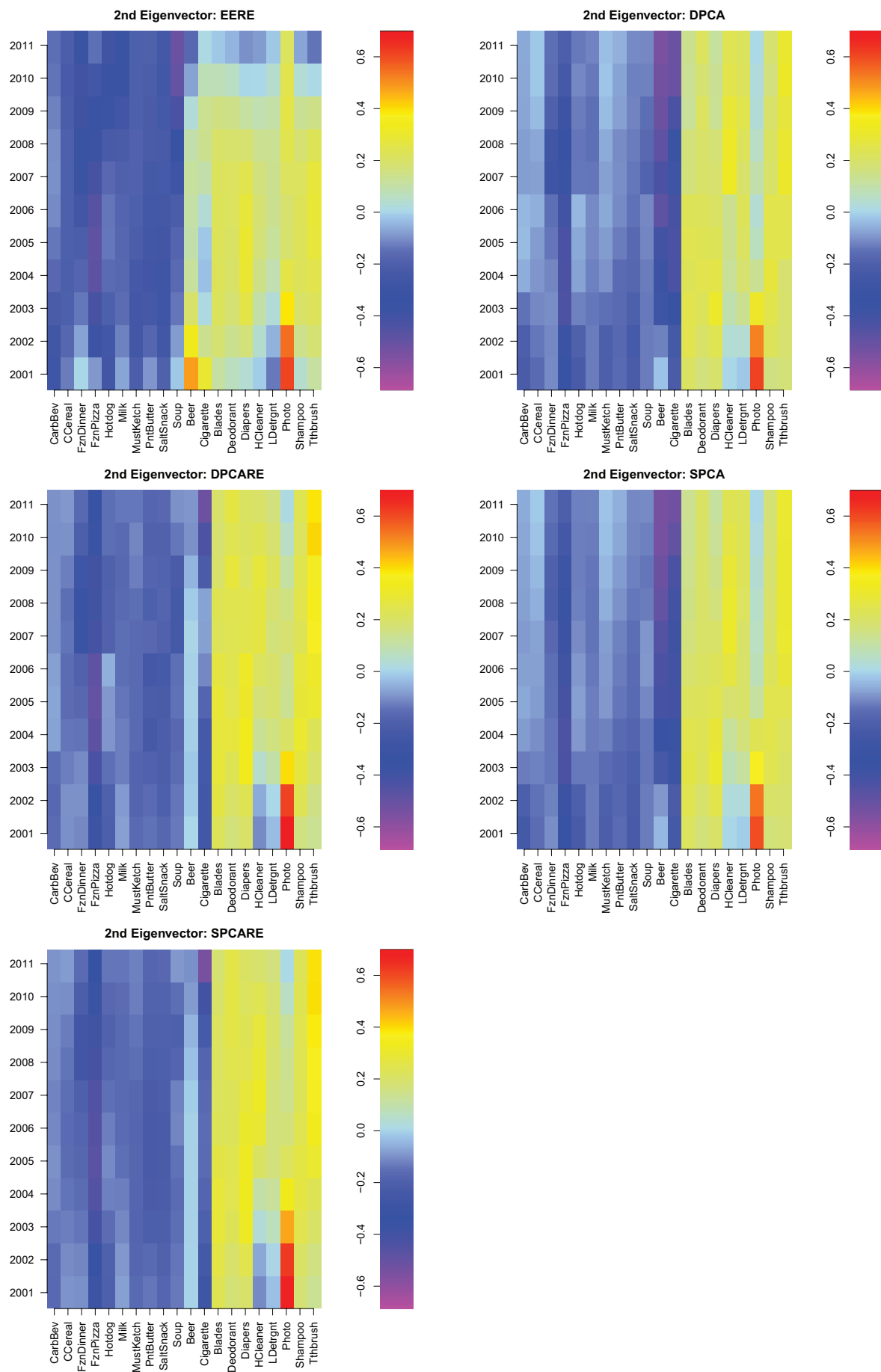


Figure 8. The heatmaps for the estimated second eigenvector for different models in IRI marketing data analysis containing 20 products.

proposed model allows for an alternative interpretation of the eigenvectors, which is more specific than the “ingestibles” definition of the DPCA, DPCARE, SPCA, and SPCARE methods.

In the heatmaps of Figure 7, the EERE displays photography supplies shifting from a moderately negative value to a positive value as the years progress from 2001 to 2011. This strong time-changing behavior is almost unnoticeable for the DPCA heatmap in Figure 7. Notice that the sale of photography supplies in grocery stores experienced a major decline nationwide after 2005, which is when camera phones and digital cameras gained popularity.

In summary, the EERE method, incorporating multivariate random effects from the stores, provides a more informative grouping strategy for products compared to existing methods. The analysis of the IRI data shows that the beer and the cigarettes can be grouped together, as both of them have a rather different pattern of sales from other consumer packaged goods. Therefore, the grocery stores might consider advertising beer and cigarettes in a similar fashion or arrange them in the same section of the store. In addition, it is recommended for grocery stores to cut down on photography supplies based on the above analysis.

6. Discussion

In this article, we propose an EERE model to incorporate random effects in modeling time-varying eigenvalues and eigenvectors under a longitudinal PCA framework. The proposed longitudinal PCA models time-varying eigen-components through nonparametric splines, and takes store variability into account via multivariate random effects. This leads to improved interpretation of the eigenvectors, which could be extremely useful in clustering grocery products that consumers purchase.

Our simulation studies indicate that the proposed method has a lower mean absolute deviation of errors for the time-varying eigenvalue estimation averaged over the simulations, and that the proposed time-varying eigenvector estimators match the true eigenvectors more closely compared to the other existing methods. In addition, the analysis of the IRI marketing data provides an illustration of how statistics and data analytics can play a role in business decision-making. We effectively utilize large marketing data over time to capture changes in consumer shopping behavior and extract intrinsic information about the associations among products.

Although this article does not focus on developing asymptotic theories, we make the following heuristic statements about the proposed methodology in general. Asymptotically, if the number of time points goes to infinity, we are able to obtain consistent estimators for both random effects and the parameters associated with the spline functions (Zhu and Qu 2018), while increasing time points will not benefit the DPCA model since it does not utilize any neighboring longitudinal information. In addition, the within-store correlation beyond the random effects is also allowed, which can be incorporated simultaneously with the random effects by adopting the method proposed by Wang, Tsai, and Qu (2012). However, the computation would be much more complicated.

This article tackles a real data problem involving the complex correlated nature of observations over time. We acknowledge that the proposed method is mostly suitable when the covariance structure and corresponding eigen-components are smooth functions of time, although it is worth exploring to incorporate abrupt jumps or changes for further extension. In addition, further research might be needed in incorporating an additional penalty term to allow the model to select the number of bases functions based on the data.

Supplementary Materials

The supplementary materials include two files: (1) “SUPPLEMENTARY MATERIAL.pdf,” is the appendix of the paper “Longitudinal Principal Component Analysis with an Application to Marketing Data” by Kinson, Tang, Zhuo and Qu, which includes additional simulation studies and some computational details; and (2) “EERE_rmarkdown_doc.html,” which includes the R codes for the LPCA model in the simulations.

Acknowledgments

The authors would like to thank Information Resources, Incorporated (IRI) for making the data available. All estimates and analysis in this article, based on data provided by IRI, are by the authors and not by IRI. Additionally, we thank Michael Kruger for his generosity on discussing details and updates of the dataset.

Funding

This research was supported by National Science Foundation Grants (DGE-1144245, DMS-1308227, DMS-1415308, DMS-1613190, and DMS-1821198).

References

- Anderson, S. J., and Jones, R. H. (1995), “Smoothing Splines for Longitudinal Data,” *Statistics in Medicine*, 14, 1235–1248. [337]
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015), “Fitting Linear Mixed-Effects Models Using lme4,” *Journal of Statistical Software*, 67, 1–48. [341]
- Berrendero, J. R., Justel, A., and Svarc, M. (2011), “Principal Components for Multivariate Functional Data,” *Computational Statistics & Data Analysis*, 55, 2619–2634. [336]
- Bowman, A. W., and Azzalini, A. (2018), “R Package sm: Nonparametric Smoothing Methods (Version 2.2-5.6),” University of Glasgow, UK and Università di Padova, Italia. [341]
- Bradlow, E. T. (2002), “Exploring Repeated Measures Data Sets for Key Features Using Principal Components Analysis,” *International Journal of Research in Marketing*, 19, 167–179. [335]
- Bronnenberg, B. J., Kruger, M. W., and Mela, C. F. (2008), “The IRI Marketing Data Set,” *Marketing Science*, 27, 745–748. [335]
- Chiou, J.-M., Chen, Y.-T., and Yang, Y.-F. (2014), “Multivariate Functional Principal Component Analysis: A Normalization Approach,” *Statistica Sinica*, 24, 1571–1596. [336]
- Di, C. Z., Crainiceanu, C. M., Caffo, B. S., and Punjabi, N. M. (2009), “Multilevel Functional Principal Component Analysis,” *The Annals of Applied Statistics*, 3, 458–488. [336]
- Diggle, P. T., and Verbyla, A. P. (1998), “Nonparametric Estimation of Covariance Structure in Longitudinal Data,” *Biometrics*, 54, 401–415. [337]
- Durbán, M., Harezlak, J., Wand, M. P., and Carroll, R. J. (2005), “Simple Fitting of Subject-Specific Curves for Longitudinal Data,” *Statistics in Medicine*, 24, 1153–1167. [337]

- Elashoff, M., and Ryan, L. (2004), "An EM Algorithm for Estimating Equations," *Journal of Computational and Graphical Statistics*, 13, 48–65. [336,339]
- Fader, P. S., and Lattin, J. M. (1993), "Accounting for Heterogeneity and Nonstationarity in a Cross-Sectional Model of Consumer Purchase Behavior," *Marketing Science*, 12, 304–317. [335]
- Fan, J., Huang, T., and Li, R. (2007), "Analysis of Longitudinal Data With Semiparametric Estimation of Covariance Function," *Journal of the American Statistical Association*, 102, 632–641. [337]
- Fan, J., and Wu, Y. (2008), "Semiparametric Estimation of Covariance Matrices for Longitudinal Data," *Journal of the American Statistical Association*, 103, 1520–1533. [337]
- Fox, E., and Dunson, D. (2017), "Bayesian Nonparametric Covariance Regression," arXiv no. 1101.2017. [335]
- Gardiner, J. C., Luo, Z., and Roman, L. A. (2009), "Fixed Effects, Random Effects, and GEE: What Are the Differences?," *Statistics in Medicine*, 28, 221–239. [339]
- Greven, S., Crainiceanu, C. M., Caffo, B. S., and Reich, D. (2010), "Longitudinal Functional Principal Component Analysis," *Electronic Journal of Statistics*, 4, 1022–1054. [336]
- Hall, P., Müller, H. G., and Wang, J. L. (2006), "Properties of Principal Component Methods for Functional and Longitudinal Data Analysis," *The Annals of Statistics*, 34, 1493–1517. [335]
- Happ, C., and Greven, S. (2018), "Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains," *Journal of the American Statistical Association*, 113, 649–659. [336]
- Hastie, T., Tibshirani, R., and Friedman, J. (2009), *The Elements of Statistical Learning*, New York: Springer. [341]
- Hoff, P., and Niu, X. (2012), "A Covariance Regression Model," *Statistica Sinica*, 22, 729–753. [335]
- Huang, J. Z., Wu, C. O., and Zhou, L. (2004), "Polynomial Spline Estimation and Inference for Varying Coefficient Models With Longitudinal Data," *Statistica Sinica*, 14, 763–788. [337,341]
- Islam, M., Staicu, A.-M., and Heugten, E. V. (2018), "Longitudinal Dynamic Functional Regression," arXiv no. 1611.01831. [336]
- Jain, D., Bass, F. M., and Chen, Y. M. (1990), "Estimation of Latent Class Models With Heterogeneous Choice Probabilities: An Application to Market Structuring," *Journal of Marketing Research*, 27, 94–101. [335]
- Jiang, C. R., and Wang, J. L. (2010), "Covariate Adjusted Functional Principal Component Analysis," *The Annals of Statistics*, 38, 1194–1226. [336]
- Jolliffe, I. T. (2002), *Principal Component Analysis* (2nd ed.), New York: Springer-Verlag. [335]
- Kim, K.-Y., and Wu, Q. (1999), "A Comparison Study of EOF techniques: Analysis of Nonstationary Data With Periodic Statistics," *Journal of Climate*, 12, 185–199. [335]
- Kruger, M. W., and Pagni, D. (2008), *IRI Academic Data Set Description* (Version 2.2), Chicago: Information Resources Incorporated. [335]
- Laird, N. M., and Ware, J. H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963–974. [339]
- Liang, H., and Xiao, Y. (2006), "Penalized Splines for Longitudinal Data With an Application in AIDS Studies," *Journal of Modern Applied Statistical Methods*, 5, 130–139. [337]
- Liang, K. Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22. [337]
- Likhachev, D. (2017), "Selecting the Right Number of Knots for B-Spline Parameterization of the Dielectric Functions in Spectroscopic Ellipsometry Data Analysis," *Thin Solid Films*, 636, 519–526. [341]
- Maadooliat, M., Pourahmadi, M., and Huang, J. Z. (2013), "Robust Estimation of the Correlation Matrix of Longitudinal Data," *Statistics and Computing*, 23, 17–28. [337]
- Miller, C., and Bowman, A. (2012), "Smooth Principal Components for Investigating Changes in Covariances Over Time," *Journal of the Royal Statistical Society, Series C*, 61, 693–714. [336,341]
- Proudfoot, J., Faig, W., Natarajan, L., and Xu, R. (2018), "A Joint Marginal-Conditional Model for Multivariate Longitudinal Data," *Statistics in Medicine*, 37, 813–828. [339]
- Prvan, T., and Bowman, A. W. (2003), "Nonparametric Time Dependent Principal Components Analysis," *ANZIAM Journal*, 44, 627–643. [336]
- Qu, A., Lindsay, B., and Li, B. (2000), "Improving Generalised Estimating Equations Using Quadratic Inference Functions," *Biometrika*, 87, 823–836. [337]
- R Core Team (2019), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [341]
- Ramsay, J. O., and Silverman, B. W. (2005), *Functional Data Analysis* (2nd ed.), New York: Springer. [335]
- Rossi, P. E., and Allenby, G. M. (1993), "A Bayesian Approach to Estimating Household Parameters," *Journal of Marketing Research*, 30, 171–182. [335]
- Ruppert, D. (2002), "Selecting the Number of Knots for Penalized Splines," *Journal of Computational and Graphical Statistics*, 11, 735–757. [341]
- Sun, Y., Zhang, W., and Tong, H. (2007), "Estimation of the Covariance Matrix of Random Effects in Longitudinal Studies," *The Annals of Statistics*, 35, 2795–2814. [337]
- Wang, P., Tsai, G., and Qu, A. (2012), "Conditional Inference Functions for Mixed-Effects Models With Unspecified Random-Effects Distribution," *Journal of the American Statistical Association*, 107, 725–736. [349]
- Wedderburn, R. W. M. (1974), "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss–Newton Method," *Biometrika*, 61, 439–447. [341]
- Wu, W. B., and Pourahmadi, M. (2003), "Nonparametric Estimation of Large Covariance Matrices of Longitudinal Data," *Biometrika*, 90, 831–844. [337]
- Xue, L., and Liang, H. (2009), "Polynomial Spline Estimation for a Generalized Additive Coefficient Model," *Scandinavian Journal of Statistics, Theory and Applications*, 37, 26–46. [337]
- Xue, L., and Qu, A. (2012), "Variable Selection in High-Dimensional Varying-Coefficient Models With Global Optimality," *Journal of Machine Learning Research*, 13, 1973–1998. [341]
- Xue, L., Qu, A., and Zhou, J. (2010), "Consistent Model Selection for Marginal Generalized Additive Model for Correlated Data," *Journal of the American Statistical Association*, 105, 1518–1530. [341]
- Yao, F., Müller, H. G., and Wang, J. L. (2005), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100, 577–590. [335]
- Zheng, X., Xue, L., and Qu, A. (2018), "Time-Varying Correlation Structure Estimation and Local-Feature Detection for Spatio-Temporal Data," *Journal of Multivariate Analysis*, 168, 221–239. [338]
- Zhu, X., and Qu, A. (2018), "Cluster Analysis of Longitudinal Profiles With Subgroups," *Electronic Journal of Statistics*, 12, 171–193. [349]
- Zipunnikov, V., Greven, S., Shou, H., Caffo, B., Reich, D., and Crainiceanu, C. (2014), "Longitudinal High-Dimensional Principal Components Analysis With Application to Diffusion Tensor Imaging of Multiple Sclerosis," *The Annals of Applied Statistics*, 8, 2175–2202. [336]