Adjusted principal component estimation for binary factor model

Wang, Peng and Wang, Xi

Hong Kong University of Science and Technology, Institute for Economic and Social Research, Jinan University

7 March 2025

Adjusted Principal Component Estimation for Binary Factor Model *

Peng Wang

Department of Economics, Hong Kong University of Science and Technology

Xi Wang

Institute for Economic and Social Research, Jinan University

March 7, 2025

Abstract

In economic decision-making, the binary factor model is widely employed to characterize decision processes and capture individuals' exposures to various factors. This paper reveals that when the binary response is factorized, additional factors emerge, including an augmented time-invariant item that can lead to overestimation of the individual effect. These findings explain why the principal component method often produces misleading estimates when applied to binary data. To address this issue, we develop an adjusted principal component (APC) method, which modifies the eigenvalue ratio test to determine factor numbers, estimates factors in the transformed model, and recovers estimates for the original binary model. It avoids parametric error distribution specifications and initial value selection, overcoming limitations of existing iterative methods. Extensive Monte Carlo experiments confirm APC's robustness. We then apply APC to analyze dividend initiation factors using S&P 500 data (1998-2016), demonstrating its practical effectiveness.

*We gratefully acknowledge financial support from the Hong Kong Research Grants Council under Project 16604120 and Guangdong Basic and Applied Basic Research Foundation 2022A1515011890.
Corresponding author: Peng Wang, E-mail address: pwang@ust.hk, Xi Wang, E-mail address: theawang@connect.ust.hk.

1 Introduction

Panel data analysis has been increasingly popular in recent decades due to its ability to leverage rich information across time periods for a large number of individuals, thereby mitigating potential endogeneity issues. Among the various panel data models, the linear factor model, $Y_{it} = \lambda_i' f_t + \varepsilon_{it}$, plays an important role in accounting for individuals' heterogenous responses to common factors. In earlier literature, researchers often relied on observed proxy variables, such as factor returns and aggregate market shocks, to represent these common factors. However, as Ludivigson and Ng (2009) highlighted, the proxy method can produce biased estimates when certain factors are missing or subject to measurement errors. In addition, a limited set of observed variables is typically insufficient to fully capture the information set of common factors, particularly for predictive purposes.

To address these challenges, a growing body of literature in high-dimensional analysis treats both factors and factor loadings as unobserved. Bai and Ng (2002) introduced information criteria, such as AIC and BIC, to determine the number of factors. Bai (2003) further developed the inferential theory for the principal components estimator in the context of linear factor models. Subsequent work by Onatski (2010) utilized differenced eigenvalues to construct an edge distribution estimator for determining the number of factors. To avoid the need for overly cautious choices of penalty functions or tuning parameters, Ahn and Horenstein (2013) proposed using the ratio of eigenvalues to estimate the number of factors. Additionally, Bai and Li (2016) introduced a quasi-likelihood estimation method tailored for dynamic factor models. These advancements have significantly enhanced the robustness and flexibility of factor model estimation in high-dimensional settings.

Discrete outcomes are commonplace in empirical research. For instance, firms make decisions on whether to initiate dividends, split stocks, or choose particular capital structures, as explored in studies such as Bates et al. (2009), Baker et al. (2009), and Grahm et al. (2015). These phenomena call for the use of binary response models to characterize economic relationships. Similarly, ordered response models are often required for issues such as bond credit ratings for companies, as examined by Badoer and Demiroglu (2018), and creditworthiness assessments by rating agencies, as discussed in Baker and Mansi (2002). Traditionally, scholars have employed methods such as probit (oprobit), logit (ologit), or OLS regressions to analyze binary or ordered data, using observed proxy variables to substitute for unobserved influencing factors. In recent years, the literature on factor analysis has emphasized that latent factor models are more effective for characterizing panel datasets, monitoring economic activities, and improving forecasting accuracy. As a result, binary and ordered factor models have become increasingly appealing to practitioners and warrant significant academic attention.

Despite significant progress in the research on linear factor models, the estimation of binary and ordered factor models remains an ongoing challenge. Kolenikov and Angeles (2009) and Perez et al. (2015) specified tetrachoric or polychoric correlations and introduced multi-step estimation strategies. Ng (2015) estimated polychoric correlations and extended principal component analysis but acknowledged that determining the number of factors to control for is irregular and challenging, which compromises the consistent estimation of factors. Boneva and Linton (2017) proposed a common correlated effects maximum likelihood estimator by specifying the error distribution and estimating unobserved factors as averages of regressors. However, their method relies on strong assumptions about the error distribution and factors, limiting its applicability. To relax restrictions on unobserved factors, Wang (2022) treated both factors and factor loadings as fixed effects and proposed a maximum likelihood estimation method. Nevertheless, potential misspecification and lack of identification make his estimator sensitive to the parametric specification of the error distribution and the initial values of factors, potentially leading to inconsistent estimates.

Vermunt and Magidson (2005) argued that ordinary linear regression for binary data is attractive to applied researchers. However, directly applying traditional principal component analysis to binary data can lead to issues due to the underlying nonlinearity and latent factor structure inherent in binary factor models. To overcome the limitations of existing methods, we employ a transformation to provide an interpretation for the irregular number of factors encountered by Ng (2015). Subsequently, we develop an adjusted method to robustly determine the number of factors and estimate the factors for the binary factor model. The main contributions of this paper are as follows.

The primary insight is to distinguish between the binary factor model and the transformed factor model. The transformed factor model is an expanded version of the binary factor model and accounts for a larger number of factors than the binary factor model. A naive application of principal component analysis to the original binary data, without considering the expanded factor space, will yield inconsistent estimates.

Practically, this paper provides a guideline for implementing the principal component method for binary data. Specifically, we discover the underlying reason for the frequent underestimation of the number of factors in binary data. Based on this theoretical insight, we extend the eigenvalue ratio test to robustly determine the number of factors for the transformed factor model and propose an adjusted principal component estimation procedure for both binary and ordered factor models.

Unlike existing maximum likelihood methods, our adjusted principal component method does not impose parametric restrictions on the error distribution or rely on iterative algorithms. Importantly, the adjusted principal component estimator is robust to the error distribution and does not require initial estimates of factors. From an applied perspective, these advantages are crucial for practitioners to avoid misleading estimates caused by misspecification and arbitrary choices of initial values. Additionally, our estimation algorithm is straightforward to implement and highly tractable.

The rest of this paper is organized as follows. Section 2 focuses on binary factor models, both with and without a time-invariant factor, and compares them with transformed factor models. Based on these theoretical findings, Section 3 proposes the adjusted principal component method, which involves multiple steps: robustly determining the number of factors, estimating the factors for the transformed factor model, and then recovering the factors in the original binary factor model. Section 4 extends the adjusted principal component method to the ordered factor model. Section 5 presents the asymptotic properties of the adjusted principal component estimator. Section 6 conducts Monte Carlo experiments to validate the method. Section 7 applies the adjusted principal component estimator to an empirical example of dividend initiation using S&P 500 data, demonstrating its performance. Section 8 concludes the paper.

2 Binary Factor Models

Throughout this paper, $P_\varepsilon(\cdot)$ and $p_\varepsilon(\cdot)$ denote the cumulative distribution function and the density function of ε respectively. Additionally, $diag(v_1, \cdots, v_r)$ represents a diagonal matrix with elements $v_1, \cdots, v_r$.

In many empirical studies, applied researchers are particularly interested in estimating and predicting outcomes using binary factor models, where the dependent variable is binary. For example, what factors influence a firm's decision to initiate dividends or engage in corporate acquisitions? These questions motivate our focus on binary factor models. In this context, we consider the following two scenarios.

Case I. Binary factor model without individual effects

Consider the following binary response model for N individuals over T periods,

$$Y_{it} = 1\{\lambda_i^{*\prime} f_t^* - \varepsilon_{it} > 0\},\ i = 1, \ldots, N,\ t = 1, \ldots, T, \tag{1}$$

where Y_{it} is the observed dummy outcome, $1\{\cdot\}$ is the indicator function, $f_t^* = (f_{1t}^*, \cdots, f_{r^*t}^*)'$ and $\lambda_i^* = (\lambda_{1t}^*, \cdots, \lambda_{r^*t}^*)'$ are $r^* \times 1$ vectors of latent time-varying factors and latent factor loadings, ε_{it} is the error term. Denote $Y_{it}^* = \lambda_i^{*\prime} f_t^* - \varepsilon_{it}$. In the present setup, there is no time-invariant factor. The number of common factors in Y_{it}^* is $r = r^*$. As in most studies of binary response models, ε_{it} is assumed to be independent of f_t^* and λ_i^*. It follows that

$$E(Y_{it}|\lambda_i^*, f_t^*) = P_\varepsilon(\lambda_i^{*\prime} f_t^*).$$

Define the transformation $t(\lambda_i^{*\prime} f_t^*)$ to be an s-th order Taylor expansion of $P_\varepsilon(\lambda_i^{*\prime} f_t^*)$ around μ_i, a certain value of $\lambda_i^{*\prime} f_t^*$, e.g., $\mu_i = \frac{1}{T}\sum_{t=1}^{T} \lambda_i^{*\prime} f_t^*$[1]. We have,

$$\begin{aligned} t(\lambda_i^{*\prime} f_t^*) = {} & [P_\varepsilon(\mu_i) - p_\varepsilon(\mu_i)\mu_i + \frac{1}{2}p_\varepsilon'(\mu_i)\mu_i^2 + \cdots + \frac{(-1)^s}{s!} p_\varepsilon^{(s-1)}(\mu_i)\mu_i^s] + [p_\varepsilon(\mu_i) - \mu_i \\ & \times p_\varepsilon'(\mu_i) + \cdots + \frac{(-1)^{s-1}}{(s-1)!} p_\varepsilon^{(s-1)}(\mu_i)\mu_i^{s-1}]\lambda_i^{*\prime} f_t^* + \cdots + \frac{1}{s!} p_\varepsilon^{(s-1)}(\mu_i)\left(\lambda_i^{*\prime} f_t^*\right)^s. \end{aligned} \tag{2}$$

[1] Here μ_i is a single value for easy analysis. In general, μ_i can take multiple values.

Then, we obtain the following representation

$$E(Y_{it}|\lambda_i^*, f_t^*) = t(\lambda_i^{*\prime} f_t^*) + \omega_{it}, \tag{3}$$

where $\omega_{it} = P_\varepsilon(\lambda_i^{*\prime} f_t^*) - t(\lambda_i^{*\prime} f_t^*)$ is the approximation error. Observe that $t(\lambda_i^{*\prime} f_t^*)$ has a product form, then we shall write $t(\lambda_i^{*\prime} f_t^*) = \Lambda_i' F_t$ with F_t and Λ_i being $R \times 1$ vectors of common factors and factor loadings. Let $F_t = (1, F_t^{*\prime})'$ and F_t^* be the $R^* \times 1$ of time-varying factors of $t(\lambda_i^{*\prime} f_t^*)$ with the corresponding factor loadings Λ_i^*. For example, if $s = 2$, $F_t^* = (f_{1t}^*, \cdots, f_{r^*t}^*, f_{1t}^{*2}, f_{1t}^* f_{2t}^*, \cdots, f_{r^*t}^{*2})'$ and $\Lambda_i^* = ((p_\varepsilon(\mu_i) - \mu_i p_\varepsilon'(\mu_i))\lambda_{1i}^*, \cdots, (p_\varepsilon(\mu_i) - \mu_i p_\varepsilon'(\mu_i))\lambda_{r^*i}^*, \frac{p_\varepsilon'(\mu_i)}{2}\lambda_{1i}^{*2}, p_\varepsilon'(\mu_i)\lambda_{1i}^*\lambda_{2i}^*, \cdots, \frac{p_\varepsilon'(\mu_i)}{2}\lambda_{r^*i}^{*2})'$ with $R^* = \frac{r^*(r^*+3)}{2}$. Based on (3), (1) can be written as

$$Y_{it} = \Lambda_i' F_t + \nu_{it}. \tag{4}$$

In the transformed factor model (4), the error term consists of the approximation error and the deviation of Y_{it} from $E(Y_{it}|\lambda_i^*, f_t^*)$, i.e., $\nu_{it} = \eta_{it} + \omega_{it}$ where $\eta_{it} = Y_{it} - E(Y_{it}|\lambda_i^*, f_t^*)$. The transformation demonstrates that applying the principal components method to data generated by (1) is equivalent to conducting factor analysis based on the regression specification (4).

Without loss of generality, the following discussion uses $s = 1$ as an example to illustrate the distinction between the transformed factor model and the original binary factor model:

$$Y_{it} = P_\varepsilon(\mu_i) - p_\varepsilon(\mu_i)\mu_i + p_\varepsilon(\mu_i)\lambda_i^{*\prime} f_t^* + \nu_{it}, \tag{5}$$

where $F_t = (1, f_t^{*\prime})'$ is an $R \times 1$ vector of factors and $\Lambda_i = (P_\varepsilon(\mu_i) - p_\varepsilon(\mu_i)\mu_i, p_\varepsilon(\mu_i)\lambda_i^{*\prime})'$ is an $R \times 1$ vector of factor loadings with $R = r^* + 1$. Notably, an additional time-invariant factor emerges in the transformed factor model, even though no such factor exists in Y_{it}^* in the original binary factor model. This observation has an important implication to applied researchers. If the common component $\lambda_i^{*\prime} f_t^*$ moves around μ_i such that the moments of $Y_{it} - t(\lambda_i^{*\prime} f_t^*)$ exist up to the eighth order, it is possible to apply principal component analysis to the binary data using an adjusted number of factors. Specifically, the number of factors specified should be $R > r^*$; otherwise, the principal component analysis would yield biased estimates. If one applies Bai and Ng (2002)'s or Ahn and Horenstein (2013)'s methods to binary response data, the estimator will indicate the number of factors in the transformed factor model, i.e., R, rather than the number of factors in the original binary factor model r^*.

Case II. Binary factor model with individual effect

From the above analysis, applying the principal components method to binary response data can yield consistent estimates of F_t if the number of factors $R > r^*$ is correctly specified. Consider another popular binary factor specification that incorporates both unobserved time-invariant individual effects and latent time-varying factors.

$$Y_{it} = 1\{\alpha_i^* + \lambda_i^{*\prime} f_t^* - \varepsilon_{it} > 0\}, \tag{6}$$

where f_t^* and λ_i^* are $r^* \times 1$ vectors of latent time-varying factors and factor loadings with $E(f_t^*) = 0$. Here α_i^* can be viewed as the factor loading for a time-invariant factor. Note that if $E(f_t^*) = c \neq 0$, one can normalize latent time-varying factors and the unobserved individual effect to $\tilde{f}_t^* = f_t^* - c$ and $\tilde{\alpha}_i^* = \alpha_i^* + \lambda_i^{*\prime} c$. Clearly, the resulting model $Y_{it} = 1\{\tilde{\alpha}_i^* + \lambda_i^{*\prime} \tilde{f}_t^* - \varepsilon_{it} > 0\}$ reverts to the setup of (6). Let $f_t = (1, f_t^{*\prime})'$ be the vector of common factors and $\lambda_i = (\alpha_i^*, \lambda_i^{*\prime})'$ be the vector of latent factor loadings in $Y_{it}^* = \alpha_i^* + \lambda_i^{*\prime} f_t^* - \varepsilon_{it}$. The number of common factors in Y_{it}^* is denoted by $r = r^* + 1$.

Following (3), we use an s-th order Taylor expansion around μ_i to approximate the cumulative distribution function of ε_{it} to obtain

$$E(Y_{it}|\alpha_i^*, \lambda_i^*, f_t^*) = P_\varepsilon(\alpha_i^* + \lambda_i^{*\prime} f_t^*) = t(\alpha_i^* + \lambda_i^{*\prime} f_t^*) + \omega_{it}.$$

The transformation results in a representation in product form $t(\alpha_i^* + \lambda_i^{*\prime} f_t^*) = \Lambda_i' F_t$, where $F_t = (1, F_t^{*\prime})'$, and F_t^* is the $R^* \times 1$ vector of time-varying factors with factor loadings Λ_i^*. If $s = 2$, then $F_t^* = (f_{1t}^*, \cdots, f_{r^*t}^*, f_{1t}^{*2}, f_{1t}^* f_{2t}^*, \cdots, f_{r^*t}^{*2})'$ and $\Lambda_i^* = ((p_\varepsilon(\mu_i) + (\alpha_i - \mu_i) p_\varepsilon'(\mu_i))\lambda_{1i}^*, \cdots, (p_\varepsilon(\mu_i) + (\alpha_i - \mu_i) p_\varepsilon'(\mu_i))\lambda_{r^*i}^*, \frac{p_\varepsilon'(\mu_i)}{2}\lambda_{1i}^{*2}, p_\varepsilon'(\mu_i)\lambda_{1i}^*\lambda_{2i}^*, \cdots, \frac{p_\varepsilon'(\mu_i)}{2}\lambda_{r^*i}^{*2})'$. Once the transformation is applied, (6) can be represented as

$$Y_{it} = \Lambda_i' F_t + \nu_{it}, \tag{7}$$

where the composite error $\nu_{it} = \eta_{it} + \omega_{it}$ with $\eta_{it} = Y_{it} - P_\varepsilon(\alpha_i^* + \lambda_i^{*\prime} f_t^*)$. Applying the principal components method to the data generated by (6) is equivalent to the factor analysis based on the regression specification (7).

Consider the case $s = 1$ for a comparison with (5). According to (7),

$$Y_{it} = P_\varepsilon(\mu_i) + p_\varepsilon(\mu_i)(\alpha_i - \mu_i) + p_\varepsilon(\mu_i)\lambda_i^{*\prime} f_t^* + \nu_{it}. \tag{8}$$

Under the transformation, $F_t = (1, f_t^{*\prime})'$ is an $R \times 1$ vector of common factors and $\Lambda_i = (P_\varepsilon(\mu_i) + p_\varepsilon(\mu_i)(\alpha_i - \mu_i), p_\varepsilon(\mu_i)\lambda_i^{*\prime})'$ is an $R \times 1$ vector of factor loadings with $R = r^* + 1$. Unlike model (1), the transformation does not generate an additional time-invariant factor as both model (6) and (8) already include it. Despite $F_t = f_t$ and $R = r^* + 1$ in the transformed factor model, determining the number of factors in (8) poses a challenge for implementing the principal components method. In particular, the generated term, associated with the time-invariant factor by the transformation, complicates the estimation of the number of factors using existing methods such as Bai and Ng (2002), Onatski (2010), Ahn and Horenstein (2013). For example, Ahn and Horenstein (2013)'s approach relies on the ratio of adjacent eigenvalues of the outer-product of the data (i.e., their ER criterion) or the ratio of residual variances from recursive factor-based predictive regressions (i.e., their GR criterion). However, in (8), the variation of the individual effect relative to that of the product associated with time-varying factors is likely to be much larger than in (6) due to the additional term $P_\varepsilon(\mu_i) - p_\varepsilon(\mu_i)\mu_i$ introduced by the transformation. As a result, the number of factors may be underestimated to be one, i.e., $\hat{R} = 1$.

In general, the transformed factor model not only expands the factor space that may include high order terms of time-varying factors and arisen time-invariant factor, but also may exaggerate the explanatory power of the unobserved individual fixed effects in the total variation of Y_{it}, such that the eigenvalue corresponding to the time-invariant factor dominates those of time-varying factors (f_t^*). Once the number of factors is misspecified, the principal components method yield misleading results when applied to binary response data.

Remark 1 *Ng (2015) estimates common factors from categorical data but does not explain overestimation or underestimation of the number of factors. Here we provide* *a theoretical explanation for the irregularity stemming from the expanded factor space. It serves as the basis of our estimation strategy.*

3 Adjusted Principal Components Estimation

When a time-invariant factor is dominant in determining the variation in the binary response data Y_{it}, standard criteria may lead to misleading estimates for the number of factors. For example, Monte Carlo simulations in Section 6 demonstrate that Ahn and Horenstein's (2013) criteria tend to underestimate the true number of factors in the transformed factor model. In addition, the transformed factor model may contain the high order terms of time-varying factors up to an unknown order. All these motivate us to develop a robust procedure to estimate common factors for the data of binary response consistently. Our adjusted principal components (APC) estimation algorithm is described below.

Step 1: Determine the number of common factors R in the transformed factor model.

For linear factor models, Bai and Ng (2002) estimated the number of factors by minimizing the PC or IC criterion function. Onatski (2010) estimated the number of factors using the differenced eigenvalues. In the finite sample, these estimates could be sensitive to the threshold value specified in the penalty function. In subsequent work, Ahn and Horenstein (2013) proposed the eigenvalue ratio and growth ratio criteria to determine the number of factors that are free of any penalty terms. However, in the case of a dominant time-invariant factor, the eigenvalue or growth ratio test yield misleading estimates for the number of factors in the binary response data.

Stack Y_{it} into an $N \times T$ matrix $\mathbf{Y}$ and let $u_{NT,k}$ be the k-th largest eigenvalue of $\mathbf{Y}'\mathbf{Y}/(NT)$. Define the criterion functions

$$BER(\tilde{r}) = \frac{u_{NT,\tilde{r}}}{u_{NT,\tilde{r}+1}} \text{ and } BGR(\tilde{r}) = \frac{\ln[V(\tilde{r}-1)/V(\tilde{r})]}{\ln[V(\tilde{r})/V(\tilde{r}+1)]},$$

where $V(\tilde{r}) = \sum_{m=\tilde{r}+1}^{T} u_{NT,m}$ is the sample mean of the squared residuals from regressing Y_{it} on the first $\tilde{r}$ principal components of $\mathbf{Y}'\mathbf{Y}/(NT)$. The terms BER and

BGR refer to "binary eigenvalue ratio" and "binary growth ratio" respectively. Then, we propose the adjusted eigenvalue ratio test for R, the number of common factors in Y_{it}. The estimators are defined as

$$\hat{R}_{BER} = arg \max_{2 \leq \tilde{r} \leq R_{\max}} BER(\tilde{r}) \text{ and } \hat{R}_{BGR} = arg \max_{2 \leq \tilde{r} \leq R_{\max}} BGR(\tilde{r}). \tag{9}$$

Note that (9) differs from Ahn and Horenstein (2013)'s ER or GR method in that (9) searches for maxima of $BER(\tilde{r})$ and $BGR(\tilde{r})$ over $\{2, \cdots, R_{\max}\}$ rather than the traditional $\{1, 2, \cdots, R_{\max}\}$, so the possible dominant influence from the augmented time-invariant factor is ruled out. This adjustment is crucial for determining the number of factors and thus providing factor estimates.

Step 2: Estimate the common factors F_t in the transformed factor model

Once a consistent estimate of R is available, we estimate the common factors in the transformed factor model using principal components method with $\hat{R}_{BER}$(or $\hat{R}_{BGR}$) factors. Hereafter the BER criterion is adopted for easy notations. (Λ_i, F_t) can be estimated by the solution to the minimizing the following criterion function

$$S_{NT}(\tilde{\Lambda}_i, \tilde{F}_t) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \tilde{\Lambda}_i' \tilde{F}_t \right)^2,$$

where $\tilde{\Lambda}_i$ and $\tilde{F}_t$ are $\hat{R}_{BER} \times 1$ vectors of parameters subject to $\tilde{\mathbf{F}}'\tilde{\mathbf{F}} \Big/ T = I_{\hat{R}_{BER}}$ with $\tilde{\mathbf{F}} = (\tilde{F}_1, ..., \tilde{F}_T)'$ and $\tilde{\mathbf{\Lambda}}'\tilde{\mathbf{\Lambda}}$ being diagonal with distinct entries with $\tilde{\mathbf{\Lambda}} = (\tilde{\Lambda}_1, ..., \tilde{\Lambda}_N)'$. Concentrating out $\tilde{\mathbf{\Lambda}}$, the estimator for common factors in the transformed factor model is

$$\hat{\mathbf{F}}_{BER} = \arg\max_{\tilde{\mathbf{F}} \in \mathcal{F}} \text{tr}(\tilde{\mathbf{F}}'(\mathbf{Y}'\mathbf{Y})\tilde{\mathbf{F}}), \tag{10}$$

where $\hat{\mathbf{F}}_{BER} = (\hat{F}_{BER,1}, \cdots, \hat{F}_{BER,T})'$ and $\mathcal{F} = \{\tilde{\mathbf{F}} : \tilde{\mathbf{F}}'\tilde{\mathbf{F}} \Big/ T = I_{\hat{R}_{BER}}\}$. Subsequently, the estimator for factor loadings in the transformed factor model is

$$\hat{\mathbf{\Lambda}}_{BER} = \hat{\mathbf{F}}_{BER}' \mathbf{Y} \Big/ T. \tag{11}$$

It is well known that $\mathbf{F} = (F_1, ..., F_T)'$ is identified up to a rotation. Namely, there exists an invertible $\hat{R}_{BER} \times \hat{R}_{BER}$ matrix $\mathbf{H}$ such that $\hat{\mathbf{F}}_{BER}$ is an estimator of $\mathbf{FH}$ and $\hat{\mathbf{\Lambda}}_{BER}$ is an estimator of $\mathbf{\Lambda H}'^{-1}$.

Step 3: Determine the order of expansion s in the transformation.

Based on Step 1, it is straightforward to estimate the number of time-varying factors in the transformed factor model by

$$\hat{R}^*_{BER} = \hat{R}_{BER} - 1 \text{ and } \hat{R}^*_{BGR} = \hat{R}_{BGR} - 1.$$

Although the order of expansion to the distribution function of ε_{it} is unknown to researchers, we have shown that the time-varying factors in (4) or (7) are polynomial functions of the factors in (1) or (6). This finding allows us to determine s ($\leq \hat{R}^*_{BER}$) by regressing each of time-varying factor on the $\tilde{s}$-th power of other factors sequentially for $\tilde{s} = \hat{R}^*_{BER}, \hat{R}^*_{BER} - 1, \cdots, 2$.

Specifically, For $\tilde{R} = 1, \cdots, \hat{R}_{BER}$,

$$\hat{F}_{BER,\tilde{R}t} = \hat{F}^{\tilde{s}\prime}_{BER,\tilde{R}^-t}\gamma_{\tilde{R},\tilde{s}} + v_t \tag{12}$$

where $\hat{F}_{BER,\tilde{R}t}$ is the $\tilde{R}$-th argument of $\hat{F}_{BER,t}$ and $\hat{F}^{\tilde{s}}_{BER,\tilde{R}^-t}$ denotes a $(\hat{R}_{BER} - 1) \times 1$ vector of $\left(\hat{F}_{BER,1t}\right)^{\tilde{s}}, \cdots, \left(\hat{F}_{BER,(\tilde{R}-1)t}\right)^{\tilde{s}}, \left(\hat{F}_{BER,(\tilde{R}+1)t}\right)^{\tilde{s}}, \cdots, \left(\hat{F}_{BER,\hat{R}_{BER}t}\right)^{\tilde{s}}$. Apparently we shall conduct a F-test or observe R^2 to examine whether $\hat{F}_{BER,\tilde{R}t}$ is a linear combination of $\tilde{s}$-th power of other factors. Let $\hat{\gamma}_{\tilde{R},\tilde{s}}$ be the least square estimate from the regression specification (12). We repeat the above procedure in sequence for $\tilde{s} = \hat{R}^*_{BER}, \hat{R}^*_{BER} - 1, \cdots, 2$. The estimate of s, $\hat{s}$, is the maximum of $\tilde{s}$ such that $\gamma_{\tilde{R},\tilde{s}}$ for some $\tilde{R}$ significantly differs from zero.

For example, if $\hat{R}_{BER} = 3$, then we regress $\hat{F}_{BER,1t}$ on $(\hat{F}^2_{BER,2t}, \hat{F}^2_{BER,3t})$, $\hat{F}_{BER,2t}$ on $(\hat{F}^2_{BER,1t}$,
$\hat{F}^2_{BER,3t})$ and $\hat{F}_{BER,3t}$ on $(\hat{F}^2_{BER,1t}, \hat{F}^2_{BER,2t})$ respectively[2]. If any slope coefficient estimate is statistically significant, $\hat{s} = 2$; otherwise, $\hat{s} = 1$.

[2] One of $\hat{F}_{BER,1t}$, $\hat{F}_{BER,2t}$ and $\hat{F}_{BER,3t}$ is the estimate of the time invariant factor.

Step 4: Estimate the common factor f_t in the binary factor model.

The estimates of factors and factor loadings in the transformed factor model will be useful for forecast. In the meantime, researchers may be interested in learning about the economic factors that influence the individual's decision. Therefore, our ultimate goal is estimation of f_t.

Our analysis indicates that F_t^* consists of f_t^* and its high order terms. Denote $\hat{F}_{BER,t} = (\hat{F}_{BER,1t}, \hat{F}^*_{BER,t})$ where $\hat{F}_{BER,1t}$ is the estimate of the time invariant factor and $\hat{F}^*_{BER,t}$ is the estimate of F_t^*. In practice, it is tractable to determine $\hat{F}_{BER,1t}$ since it has very small sample variance. Once $\hat{F}^*_{BER,t}$ is recognized, regress each estimated factor, $\hat{F}^*_{BER,\tilde{R}^*t}$, on high order terms of other factors up to $\hat{s}$, i.e.

$$\hat{F}^*_{BER,\tilde{R}^*t} = \hat{F}^{*(\hat{s})\prime}_{\hat{s},\tilde{R}^*-t}\delta_{\tilde{R}^*} + v_t$$

where $\hat{F}_{BER,\tilde{R}^*t}$ is the $\tilde{R}^*$th argument of $\hat{F}^*_{BER,t}$, and $\hat{F}^{*(\hat{s})}_{\tilde{R}^*-t}$ denotes a vector that includes higher order terms of other factors up to the $\hat{s}$-th order. For example, if $\hat{s}=2$ and $\hat{R}^*_{BER}=2$, then we regress $\hat{F}^*_{BER,1t}$ on $\left(\hat{F}^*_{BER,2t}\right)^2$ and $\hat{F}^*_{BER,2t}$ on $\left(\hat{F}^*_{BER,1t}\right)^2$. $\hat{F}^*_{BER,\tilde{R}^*t}$ is classified as the generated factor by the transformation if the least square estimate $\hat{\delta}_{\tilde{R}^*}$ is significantly different from zero.

According to the above rule, we determine r and obtain the matrix S_f which selects the $\hat{r}\times 1$ subvector from a $\hat{R}_{BER}\times 1$ vector. Subsequently, $\hat{f}_t$, the estimate of f_t, has the following representation,

$$\hat{f}_t = S_f \hat{F}_{BER,t}.$$

For example, if the last $\hat{R}_{BER}-\hat{r}$ elements of $\hat{F}_{BER,t}$ is a linear combination of some order terms of the rest elements, then $S_f = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \end{pmatrix}_{\hat{r}\times\hat{R}_{BER}}$.

Remark 2 *In contrast with Wang's (2022) maximum likelihood estimation of non-linear factor models, our estimation is free of the error's distribution, and avoid the inconsistency caused by arbitrary choices of initial values.*

Remark 3 *Although the target is the common factor in the binary factor model, our estimation strategy starts from the transformed factor model rather than the original model, which is different from the factor analysis in the linear model.*

Remark 4 *Ng (2015) suggested removing individual effects using the overtime de-mean method. However, it does not make sense to apply the demean approach to higher order terms.*

4 Extension: Ordered Factor Models

This section extends the proposed APC method to ordered response data. We consider a classical ordered factor model with K responses,

$$Y_{it} = \begin{cases} 1, & \text{if } Y_{it}^* < a_1 \\ \vdots & \vdots \\ k, & a_{k-1} \leq Y_{it}^* < a_k \\ \vdots & \vdots \\ K \quad , & Y_{it}^* \geq a_{K-1} \end{cases}, \tag{13}$$

where the latent outcome is $Y_{it}^* = \lambda_i^{*\prime} f_t^* - \varepsilon_{it}$ with λ_i^* and f_t^* defined as before. $\{a_1 < \cdots < a_{K-1}\}$ is the sequence of threshold values for a given positive integer $K \geq 2$.

Provided that ε_{it} is independent of $\lambda_i^*$and f_t^*, we obtain the following condition,

$$\begin{aligned} E(Y_{it}|\lambda_i^*, f_t^*) &= 1 - P_\varepsilon(\lambda_i^{*\prime} f_t^* - a_1) + \cdots + k \cdot [P_\varepsilon(\lambda_i^{*\prime} f_t^* - a_{k-1}) - P_\varepsilon(\lambda_i^{*\prime} f_t^* - a_k)] + \cdots \\ &\quad + K \cdot P_\varepsilon(\lambda_i^{*\prime} f_t^* - a_{K-1}) \\ &= 1 + \sum_{k=1}^{K} P_\varepsilon(\lambda_i^{*\prime} f_t^* - a_k) = t(\lambda_i^{*\prime} f_t^*) + \omega_{it}. \end{aligned} \tag{14}$$

The transformation $t(\lambda_i^{*\prime} f_t^*)$ is an s-th order expansion of $P_\varepsilon(\lambda_i^{*\prime} f_t^* - a_k)$ around $\lambda_i^{*\prime} f_t^* = \mu_i$. Similar to the binary factor model, the transformation has a product form i.e., $t(\lambda_i^{*\prime} f_t^*) = \Lambda_i' F_t$ where Λ_i and F_t are $R \times 1$ vectors of factors and factor loadings. For example, when $K = 3$ and $s = 2$, $\Lambda_i = (1 + P_\varepsilon(\mu_i - a_1) + P_\varepsilon(\mu_i - a_2) - (p_\varepsilon(\mu_i - a_1) + p_\varepsilon(\mu_i - a_2))\mu_i + \frac{1}{2}(p'_\varepsilon(\mu_i - a_1) + p'_\varepsilon(\mu_i - a_2))\mu_i^2, (p_\varepsilon(\mu_i - a_1) + p_\varepsilon(\mu_i - a_2))\lambda_i^{*\prime}, \frac{1}{2}(p'_\varepsilon(\mu_i - a_1) + p'_\varepsilon(\mu_i - a_2))\lambda_{1i}^{*2}, (p'_\varepsilon(\mu_i - a_1) + p'_\varepsilon(\mu_i - a_2))\lambda_{1i}^*\lambda_{2i}^*, \cdots, \frac{1}{2}(p'_\varepsilon(\mu_i - a_1) + p'_\varepsilon(\mu_i - a_2))\lambda_{r^*i}^{*2})'$ and $F_t = (1, f_{1t}^*, \cdots, f_{r^*t}^*, f_{1t}^{*2}, f_{1t}^* f_{2t}^*, \cdots, f_{r^*t}^{*2})'$ with $R = 1 + \frac{r^*(r^*+3)}{2}$.

Based on (14), (13) is transformed to be

$$Y_{it} = \Lambda_i' F_t + \nu_{it}. \tag{15}$$

where $\nu_{it} = \eta_{it} + \omega_{it}$ with $\eta_{it} = Y_{it} - E(Y_{it} | \lambda_i^*, f_t^*)$. It seems to make sense to apply the traditional eigenvalue ratio test, but just like the case for binary data, the expansion give rise to high order terms of the true time-varying factors, and additional time-invariant items in (14) may exaggerate the explanatory power of the time-invariant factor to the extent that it dominates other factors. Once the dominant factor arises, the ER or GR test tends to underestimate of the number of factors. This motives us to use the adjusted principle components estimation.

The estimation algorithm for the categorical data factor model follows the same procedure as in the binary factor model. First, we determine the number of factors in (15) according to

$$\hat{R}_{OER} = arg \max_{2 \le \tilde{r} \le R_{\max}} \frac{u_{NT,\tilde{r}}}{u_{NT,\tilde{r}+1}} \text{ and } \hat{R}_{OGR} = arg \max_{2 \le \tilde{r} \le R_{\max}} \frac{\ln[V(\tilde{r}-1)/V(\tilde{r})]}{\ln[V(\tilde{r})/V(\tilde{r}+1)]},$$

where $u_{NT,\tilde{r}}$ and $V(\tilde{r})$ are defined as before. The domain of $\tilde{r}$ is $\{2, \cdots, R_{\max}\}$ in order to avoid the dominant influence of the time invariant factor. Next, we apply the principal components method by controlling for $\hat{R}_{OER}$ factors. The estimator for the factors in the transformed factor model $\mathbf{F} = (F_1, \cdots, F_T)$ is

$$\hat{\mathbf{F}}_{OER} = \arg\max_{\tilde{\mathbf{F}}} \operatorname{tr}(\tilde{\mathbf{F}}'(\mathbf{Y}'\mathbf{Y})\tilde{\mathbf{F}}),$$

where $\tilde{\mathbf{F}}$ is a $T \times \hat{R}_{OER}$ matrix subject to $\tilde{\mathbf{F}}'\tilde{\mathbf{F}} \Big/ T = I_{\hat{R}_{OER}}$. The estimator for the factor loadings $\Lambda = (\Lambda_1, \cdots, \Lambda_N)$ is

$$\hat{\mathbf{\Lambda}}_{OER} = \hat{\mathbf{F}}'_{OER}\mathbf{Y}' \Big/ T.$$

Third, determine the order of expansion (s) by sequentially regressing each factor on certain higher order terms of other factors. For $\tilde{R} = 1, \cdots, \hat{R}_{OER}$, run the regression

$$\hat{F}_{OER,\tilde{R}t} = \hat{F}^{\tilde{s}\prime}_{OER,\tilde{R}-t}\gamma_{\tilde{R},\tilde{s}} + v_t,$$

where $\hat{F}_{OER,\tilde{R}t}$ denotes the $\tilde{R}$th argument of $\hat{F}_{OER,t}$ and $\hat{F}^{\tilde{s}\prime}_{OER,\tilde{R}-t}$ is a $(\hat{R}_{OER} - 1) \times 1$ vector of $\left(\hat{F}_{OER,1t}\right)^{\tilde{s}}, \cdots, \left(\hat{F}_{OER,(\tilde{R}-1)t}\right)^{\tilde{s}}, \left(\hat{F}_{OER,(\tilde{R}+1)t}\right)^{\tilde{s}}, \cdots, \left(\hat{F}_{OER,\hat{R}_{BER}t}\right)^{\tilde{s}}$. One can obtain a sequence of coefficient estimates $\hat{\gamma}_{\tilde{R},\tilde{s}}$ for $\tilde{s} = \hat{R}^*_{OER}, \hat{R}^*_{OER} - 1, \cdots, 2$. Then the order of expansion is determined by the maximum of $\tilde{s}$ such that $\gamma_{\tilde{R},\tilde{s}}$ significantly differs from zero for some $\tilde{R}$. Decompose $\hat{F}_{OER,t} = (\hat{F}_{OER,1t}, \hat{F}^*_{OER,t})$ where $\hat{F}^*_{OER,t}$ is the estimate of F^*_t. Subsequently, regress each estimated factor, $\hat{F}^*_{OER,\tilde{R}^*t}$, on high order terms of other factors up to $\hat{s}$, i.e.

$$\hat{F}^*_{OER,\tilde{R}^*t} = \hat{F}^{*(\hat{s})\prime}_{\hat{s},\tilde{R}^*-t}\delta_{\tilde{R}^*} + v_t$$

where $\hat{F}^*_{OER,\tilde{R}^*t}$ is the $\tilde{R}^*$th argument of $\hat{F}^*_{OER,t}$ and $\hat{F}^{*(\hat{s})}_{\tilde{R}^*-t}$ is a vector composed of second order up to $\hat{s}$ order terms of other factors. If the least square estimate $\hat{\delta}_{\tilde{R}^*}$ is significantly different from zero, $\hat{F}^*_{OER,\tilde{R}^*t}$ does not belong to the class of factors in the original model. The selection rule determines the selection matrix S_f which selects the $\hat{r} \times 1$ subvector from a $\hat{R}_{OER} \times 1$ vector. It follows that the APC estimator of f is

$$\hat{f}_t = S_f\hat{F}_{OER,t}.$$

5 Asymptotic Properties

This section will develop the asymptotic theory for our proposed APC estimator. Following Ahn and Horenstein (2013)'s Theorem 1, we establish the consistency of $\hat{R}_{BER}$. As Bai (2003) has pointed out, the asymptotic distribution of the factor estimator is not affected by the estimated number of factors if it is consistent. Likewise, the consistent estimate of s does not influence the asymptotic distribution of $\hat{f}_t$, which is essentially the subvector of $\hat{F}_{BER,t}$. Hence, the asymptotic distribution of $\hat{\mathbf{F}}_{BER}$ is derived under the premise that R and s are treated as known.

Assumption 1. ε_{it} is independent of (λ_i^*, f_t^*), and the common density function of ε_{it}, $p_\varepsilon(\mu_i)$, is s-order continuously differentiable.

Assumption 2. $E\left\|F_t\right\|^4 \leq M$ and $E\left\|\Lambda_i\right\|^4 \leq M$ for a positive number M and all (i,t). There exists positive definite matrix Σ_Λ and Σ_F such that $\frac{1}{T}\sum_{t=1}^T F_t F_t' \xrightarrow{p} \Sigma_F$ and $\frac{1}{N}\sum_{i=1}^N \Lambda_i \Lambda_i' \xrightarrow{p} \Sigma_\Lambda$.

Assumption 3. (i)

$$E(\nu_{it}) = 0 \text{ and } E(|\nu_{it}|^8) \leq M;$$

(ii) Let $\gamma_{st} = E[N^{-1}\sum_{i=1}^N \nu_{is}\nu_{it}]$. Assume that $|\gamma_{tt}| \leq M$ and $\sum_{s=1}^T |\gamma_{st}| \leq M$ for all t uniformly in T;

(iii) Let $\sigma_{ij,ts} = E(\nu_{it}\nu_{js})$.

$$\sum_j |\sigma_{ij,tt}| \leq M$$

for all (i,t) and

$$\frac{1}{NT}\sum_i \sum_j \sum_s \sum_t |\sigma_{ij,ts}| \leq M$$

uniformly in N and T.

(iv) For all (s,t),

$$E\left|N^{1/2}\sum_{i=1}^N [\nu_{is}\nu_{it} - E(\nu_{is}\nu_{it})]\right|^4 \leq M.$$

Assumption 4.

$$E\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Lambda_i\nu_{it}\right\|^2\right)\leq M \text{ and } E\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t\nu_{it}\right\|^2\right)\leq M.$$

Assumption 5. The eigenvalues of the matrix $\Sigma_\Lambda \cdot \Sigma_F$ are distinct with a finite R.

Assumption 1 is a standard smoothness condition made for the expansion in the transformation. In Assumption 2, moment restrictions are imposed on F_t and Λ_i to ensure the nontrivial contribution of each factor in the transformed factor model. This assumption implies a higher order moment requirement on the factor in the binary factor model if $s \geq 2$. Assumption 3 (i) sets $E(\nu_{it}) = 0$, otherwise one can redefine the error term to be $\tilde{\nu}_{it} = \nu_{it} - E(\nu_{it})$ and $E(\nu_{it})$ is absorbed into the time-invariant factor. Assumption 3 (ii)-(iv) are high-level conditions to restrict the dependence of ν_{it} over time and across individuals. Despite the same format as Bai (2003), Assumption 3 has a distinct implication since the composite error ν_{it} consists of two components. One is η_{it} that represents the distance between Y_{it} and the propensity score similar to the error term in the linear factor model. The other component ω_{it} characterizes the discrepancy between the propensity score and the transformation. In light of the expansion, ω_{it} is the Lagrange remainder of the form $\frac{1}{(s+1)!}p_\varepsilon^{(s)}(\bar{\mu}_{it})\left(\lambda_i^{*\prime}f_t^* - \mu_i\right)^{s+1}$ where $\bar{\mu}_{it}$ lies between $\lambda_i^{*\prime}f_t^*$ and μ_i. Hence Assumption 3 (ii)-(iv) not only require the weak dependence of η_{it}, but also restrict the dynamic properties of f_t^*. This restriction is not strong due to the existence of $p_\varepsilon^{(s)}(\bar{\mu}_{it})$. Assumption 4 allows for weak correlation of F_t and ν_{it}. In the linear factor model, the error term may be independent of factors and factor loadings, however, ν_{it} is not because the remainder ω_{it} contains $\lambda_i^{*\prime}f_t^*$ up to $(s+1)$-th order. The s-th order derivative of $p_\varepsilon(\cdot)$ which is nested in ω_{it} plays an important role as the weight to automatically "eliminate" the influence of outlier factor observations to guarantee the weak correlation. The idea that utilizes the weight to tackle outlier observations is also applied by Dong et al. (2021) to the context of sieve time series model. Ours differs from their strategy in that the former gives rise to the weight automatically in the process of transformation, whereas the latter constructs the weight manually. In the APC estimation, the determination of the common factors in the first step ensures that Assumption 4 holds naturally. Assumption 5 helps to identify F_t up to a rotation $\mathbf{H}$. It is similar to the familiar positive definiteness condition and rules out any collinearity among F_t. This explains why the same factors in the transformation (2) are merged.

For the asymptotic theory, stack Y_{it} into a $T \times N$ matrix $\mathbf{Y}$, and let $\mathbf{U}_{NT} = diag(u_{NT,1}, \cdots, u_{NT,R})$ where $u_{NT,1} > \cdots > u_{NT,R}$ are the R largest eigenvalues of $\frac{1}{NT}\mathbf{Y}\mathbf{Y}'$. The next theorem shows the consistency of $\hat{\mathbf{F}}_{BER}$.

Theorem 1 *Suppose Assumption 1-5 hold, then (i)* $plim_{T,N\to\infty}\frac{\hat{\mathbf{F}}'_{BER}\mathbf{F}}{T} = \mathbf{Q}$. *In particular,* $\mathbf{Q} = \mathbf{U}^{1/2}\Upsilon'\Sigma_\Lambda^{-1/2}$ *is invertible, where* $\mathbf{U} = diag(u_1, \cdots, u_R)$ *is a diagonal matrix with* $u_1 > \cdots > u_R > 0$ *as the eigenvalues of* $\Sigma_\Lambda^{1/2}\Sigma_F\Sigma_\Lambda^{1/2}$*, and* Υ *is the corresponding eigenvector matrix such that* $\Upsilon'\Upsilon = I_R$. *(ii)* $\frac{1}{T}\left\|\hat{\mathbf{F}}_{BER} - \mathbf{F}\mathbf{H}\right\|^2 = O_p\left(\frac{1}{\min\{N,T\}}\right)$, *where* $\mathbf{H} = \left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)\left(\frac{\mathbf{F}'\hat{\mathbf{F}}_{BER}}{T}\right)\mathbf{U}_{NT}^{-1}$.

Theorem 1 (i) shows that $\frac{\hat{\mathbf{F}}'_{BER}\mathbf{F}}{T}$ has a unique probability limit. It implies that the spaces spanned by $\hat{\mathbf{F}}_{BER}$ and $\mathbf{F}$ are asymptotically the same. In light of infinite dimension, Theorem 1 (ii) provides the average norm consistency of $\hat{\mathbf{F}}_{BER}$. There are two sources of nonlinearity in the binary factor model. One comes from the limited dependent variable, the other is the factor form in the latent outcome. Maximum likelihood estimation (MLE) has been widely applied in traditional studies of models with limited dependent variable. Nevertheless, it may fail to consistently estimate the binary factor model due to the identification issue. As a result, the MLE estimates are sensitive to the choice of initial values in practice. In addition, the maximum likelihood estimation relies on the correct specification of the distribution of the error term ε_{it}. To address such issues, especially the nonlinear problem, this paper employs the transformation (expansion) for the nonlinear model and applies the principal components analysis for the transformed model after adjusting the number of factors. It is notable that adjustment of the number of factors is crucial for our consistent estimation of the factors. In addition, the (estimated) factor in Theorem 1 is for the transformed factor model but not for the original binary factor model, although the latter is a subvector of the former.

Assumption 6. (i)

$$E\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s[\nu_{is}\nu_{it}-E(\nu_{is}\nu_{it})]\right\|^2\leq M$$

for any t;

(ii)

$$E\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}F_t\Lambda_i'\nu_{it}]\right\|^2\leq M;$$

(iii) For any t,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Lambda_i\nu_{it}\xrightarrow{d}N(0,\Omega_t),$$

where $\Omega_t = plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\Lambda_i\Lambda_j'\nu_{it}\nu_{jt}$. For any i,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t\nu_{it}\xrightarrow{d}N(0,\Omega_i),$$

where $\Omega_i = plim_{T\to\infty}\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}F_sF_t\nu_{is}\nu_{it}$.

Assumption 6 (i)-(ii) are sufficient but not necessary moment conditions for deriving the limiting distribution. Assumption 6 (iii) holds by the central limit theorem for mixing processes or martingale difference processes under various circumstances. In contrast with the traditional condition that $E\left(\Lambda_i\nu_{it}\right)=0$ and $E\left(F_t\nu_{it}\right)=0$, Assumption 6 (iii) only requires that asymptotic bias is negligible, that is, $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Lambda_i\nu_{it}=o_p(1)$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t\nu_{it}=o_p(1)$. The next theorem provides the limiting distribution for the APC estimator.

Theorem 2 *Suppose Assumptions 1-6 hold. (i) If* $\frac{\sqrt{N}}{T}\to 0$*, then for each* t*,*

$$\sqrt{T}(\hat{F}_{BER,t}-\mathbf{H}'F_t)=\mathbf{U}_{NT}^{-1}\left(\frac{\hat{\mathbf{F}}_{BER}'\mathbf{F}}{T}\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Lambda_i\nu_{it}+o_p(1)\xrightarrow{d}N(\mathbf{0},\mathbf{U}^{-1}\mathbf{H}\Omega_t\mathbf{H}'\mathbf{U}^{-1}).$$

(ii) If $\frac{\sqrt{T}}{N} \to 0$*, then for each* i*,*

$$\sqrt{N}(\hat{\Lambda}_{BER,i} - \mathbf{H}^{-1}\Lambda_i) = \mathbf{U}_{NT}^{-1}\left(\frac{\hat{\mathbf{F}}'_{BER}\mathbf{F}}{T}\right)\left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)\frac{1}{\sqrt{T}}\sum_{i=1}^{N} F_t\nu_{it} + o_p(1)$$
$$\xrightarrow{d} N(\mathbf{0}, (\mathbf{H}')^{-1}\Omega_i(\mathbf{H})^{-1}).$$

Theorem 2 provides the limiting distributions for $\hat{F}_{BER,t}$ and $\hat{\Lambda}_{BER,i}$, the estimated factors and factor loadings in the transformed factor model. The covariance matrix of $\hat{F}_{BER,t}$ can be estimated by its sample analogue:

$$\hat{\Pi}_t = \mathbf{U}_{NT}^{-1}\frac{1}{N}\sum_{i=1}^{N}\hat{\Lambda}_{BER,i}\hat{\Lambda}'_{BER,i}\hat{\nu}_{it}^2\mathbf{U}_{NT}^{-1},$$

where $\mathbf{U}_{NT}$ is the diagonal matrix consisting of R eigenvalues of $\mathbf{Y}'\mathbf{Y}/(NT)$ and $\hat{\nu}_{it} = Y_{it} - \hat{\Lambda}'_{BER,i}\hat{F}_{BER,t}$. To derive the limiting distribution for $\hat{f}_t$, we employ the fact that $\hat{f}_t$ is a subvector of $\hat{F}_{BER,t}$. With slight abuse of notations, replace $\hat{R}_{BER}$ and $\hat{r}$ in S_f with the true value R_{BER} and r. Write $\mathbf{H}_s = S_f\mathbf{H}S'_f$. We obtain

$$\sqrt{N}(\hat{f}_t - \mathbf{H}'_s f_t) = S_f\mathbf{U}_{NT}^{-1}\left(\frac{\hat{\mathbf{F}}'_{BER}\mathbf{F}}{T}\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Lambda_i\nu_{it} + o_p(1) \xrightarrow{d} N(0, S_f\mathbf{U}^{-1}\mathbf{H}\Omega_t\mathbf{H}'\mathbf{U}^{-1}S'_f).$$

In practice, applied researchers are usually interested in estimating the propensity score, $P_\varepsilon(\lambda_i^{*\prime}f_t^*)$, since it indicates the probability of individual i's action at time period t. In this case, we recommend controlling a larger number of factors ($\tilde{R} >> R$) for the principal components analysis, and then employ the resulting estimates $\tilde{\Lambda}'_{BER,i}\tilde{F}_{BER,t}$ for $\tilde{\Lambda}'_i\tilde{F}_t$ to consistently estimate $P_\varepsilon(\lambda_i^{*\prime}f_t^*)$. We summarize the the limiting distribution of the propensity score estimator in the following theorem.

Theorem 3 *Suppose Assumption 1-6 hold and* $\delta_{NT}^2 E\left[\max_{\lambda_i^{*\prime}f_t^*}\left(\tilde{\Lambda}'_i\tilde{F}_t - P_\varepsilon(\lambda_i^{*\prime}f_t^*)\right)^2\right]$

$= o(1)$ *with* $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$ *and* $\tilde{R}\delta_{NT}^{-1} \to 0$. *Then,*

$$\left(\frac{1}{N}W_{it}^{(1)} + \frac{1}{T}W_{it}^{(2)}\right)^{-1/2} \left(\tilde{\Lambda}'_{BER,i}\tilde{F}_{BER,t} - P_{\varepsilon}(\lambda_i^{*\prime} f_t^*)\right) \xrightarrow{d} N(0,1).$$

In particular, $W_{it}^{(1)} = \tilde{\Lambda}'_i \Sigma_{\tilde{\Lambda}}^{-1}\tilde{\Omega}_t\Sigma_{\tilde{\Lambda}}^{-1}\tilde{\Lambda}_i$ *and* $W_{it}^{(2)} = \tilde{F}'_t\Sigma_{\tilde{F}}^{-1}\tilde{\Omega}_i\Sigma_{\tilde{F}}^{-1}\tilde{F}_t$ *where* $\Sigma_{\tilde{\Lambda}} = plim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \Lambda_i\Lambda'_i$, $\tilde{\Omega}_t = plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \tilde{\Lambda}_i\tilde{\Lambda}'_j\tilde{\nu}_{it}\tilde{\nu}_{jt}$, $\Sigma_{\tilde{F}} = plim_{T\to\infty}\frac{1}{N}\sum_{t=1}^T \tilde{F}_t\tilde{F}'_t$ *and* $\tilde{\Omega}_i = plim_{T\to\infty}\frac{1}{T}\sum_{s=1}^T\sum_{t=1}^T \tilde{F}_s\tilde{F}_t\tilde{\nu}_{is}\tilde{\nu}_{it}$ *with* $\tilde{\nu}_{it}$ *defined analogous to* v_{it}.

6 Monte Carlo Simulations

This section carries out a set of experiments to provide numerical evidence for the findings in Section 2 and verifies the robust performance of the proposed APC estimator. We consider the following data generating processes for binary outcomes.

DGP I: $Y_{it} = 1\{\lambda^* f_t^* - \varepsilon_{it} > 0\}$ where ε_{it} follows the logistic distribution with mean zero and variance one;

DGP II: $Y_{it} = 1\{\lambda_i^* f_t^* - \varepsilon_{it} > 0\}$ where ε_{it} follows the standard normal distribution;

DGP III: $Y_{it} = 1\{\lambda_i^* f_t^* - \varepsilon_{it} > 0\}$ where $\varepsilon_{it} = \sqrt{0.8}\varepsilon_{1it} + \sqrt{0.2}\varepsilon_{2it}$. ε_{1it} follows a gamma distribution with mean zero and variance one, and ε_{2it} follows a standard normal distribution;

DGP IV: $Y_{it} = 1\{\alpha_i^* + \lambda_i^* f_t^* - \varepsilon_{it} > 0\}$ where ε_{it} follows the logistic distribution with mean zero and variance one;

DGP V: $Y_{it} = 1\{\alpha_i^* + \lambda_i^* f_t^* - \varepsilon_{it} > 0\}$ where ε_{it} follows the standard normal distribution;

DGP VI: $Y_{it} = 1\{\alpha_i^* + \lambda_i^* f_t^* - \varepsilon_{it} > 0\}$ where $\varepsilon_{it} = \sqrt{0.8}\varepsilon_{1it} + \sqrt{0.2}\varepsilon_{2it}$. ε_{1it} follows a gamma distribution with mean zero and variance one, and ε_{2it} follows a standard normal distribution.

DGP VII: $Y_{it} = 1\{\lambda_i^{*\prime} f_t^* - \varepsilon_{it} > 0\}$ where ε_{it} follows the logistic distribution with mean zero and variance one;

DGP VIII: $Y_{it} = 1\{\lambda_i^{*\prime} f_t^* - \varepsilon_{it} > 0\}$ where ε_{it} follows the standard normal distribution;

DGP IX: $Y_{it} = 1\{\lambda_i^{*\prime} f_t^* - \varepsilon_{it} > 0\}$ where $\varepsilon_{it} = \sqrt{0.8}\varepsilon_{1it} + \sqrt{0.2}\varepsilon_{2it}$. ε_{1it} follows a gamma distribution with mean zero and variance one, and ε_{2it} follows a standard normal distribution.

In DGP I-VI, λ_i^* and f_t^* are i.i.d $N(0,1)$ for all i and t. The DGP I-III contain no time invariant factor with $r = 1$, whereas DGP IV-VI have the individual effects $\alpha_i = 0.5w_i$ with $r = 2$ where $w_i \tilde{} N(0,1)$. In DGP VII-IX, $\lambda_i^* = (\lambda_{1i}^*, \lambda_{2i}^*)'$ and $f_t^* = (f_{1t}^*, f_{2t}^*)'$ where $\lambda_{1i}^*, \lambda_{2i}^*, f_{1t}^*, f_{2t}^*$ are i.i.d $N(0,1)$ for all i and t. The number of factors is $r = 2$. We consider various error specifications including the logistic distribution, the normal distribution, and mixture of normal and gamma distributions. The sample size is $(N,T) \in \{(50,50),(100,50),(50,100),(100,100)\}$. Each experiment is replicated 1000 times ($REP = 1000$).

Table 1 compares the proposed BER test with Ahn and Horenstein (2013)'s ER test. The results from comparing BGR and GR tests are similar and are not reported here. On one hand, the numerical results verify our theoretical conjecture that the number of factors controlled in the principal components analysis should be the number of factors R in the transformed factor model rather than the number of factors r in the original binary factor model. What the test determines is R instead of r. On the other hand, according to the first two columns in each panel of Table 1, ER method underestimates R to be one with a large probability for all DGPs, and the probability of misjudgment substantially increases when more factors are involved,

Table 1. Coverage Rate: ER vs. BER

	ER			Demeaned ER			BER		
$\hat{R}$	1	2	3	1	2	3	1	2	3
	$(N,T) = (50,50)$								
DGP I	0.328	0.672	-	1.000	0.000	-	0.000	1.000	-
DGP II	0.489	0.511	-	1.000	0.000	-	0.000	1.000	-
DGP III	0.429	0.571	-	1.000	0.000	-	0.000	1.000	-
DGP IV	0.577	0.423	-	1.000	0.000	-	0.000	1.000	-
DGP V	0.673	0.327	-	1.000	0.000	-	0.000	1.000	-
DGP VI	0.597	0.403	-	1.000	0.000	-	0.000	1.000	-
DGP VII	0.820	0.001	0.179	0.006	0.9940	0.000	0.000	0.010	0.990
DGP VIII	0.919	0.000	0.081	0.014	0.986	0.000	0.000	0.014	0.986
DGP IX	0.842	0.001	0.157	0.005	0.995	0.000	0.000	0.007	0.993

Table 1 Con't. Coverage Rate: ER vs. BER

	ER			Demeaned ER			BER		
$\hat{R}$	1	2	3	1	2	3	1	2	3
				$(N,T) = (100,50)$					
DGP I	0.088	0.912	-	1.000	0.000	-	0.000	1.000	-
DGP II	0.189	0.811	-	1.000	0.000	-	0.000	1.000	-
DGP III	0.137	0.863	-	1.000	0.000	-	0.000	1.000	-
DGP IV	0.273	0.727	-	1.000	0.000	-	0.000	1.000	-
DGP V	0.404	0.596	-	1.000	0.000	-	0.000	1.000	-
DGP VI	0.301	0.699	-	1.000	0.000	-	0.000	1.000	-
DGP VII	0.423	0.000	0.577	0.000	1.000	0.000	0.000	0.000	1.000
DGP VIII	0.609	0.000	0.3910	0.0010	0.999	0.000	0.000	0.0010	0.999
DGP IX	0.483	0.000	0.517	0.0010	0.999	0.000	0.000	0.0010	0.999
				$(N,T) = (50,100)$					
DGP I	0.088	0.912	-	1.000	0.000	-	0.000	1.000	-
DGP II	0.181	0.819	-	1.000	0.000	-	0.000	1.000	-
DGP III	0.135	0.865	-	1.000	0.000	-	0.000	1.000	-
DGP IV	0.283	0.717	-	1.000	0.000	-	0.000	1.000	-
DGP V	0.408	0.592	-	1.000	0.000	-	0.000	1.000	-
DGP VI	0.315	0.685	-	1.000	0.000	-	0.000	1.000	-
DGP VII	0.425	0.000	0.575	0.000	1.000	0.000	0.000	0.001	0.999
DGP VIII	0.610	0.000	0.390	0.001	0.999	0.000	0.000	0.001	0.999
DGP IX	0.474	0.000	0.526	0.000	1.000	0.000	0.000	0.000	1.000
				$(N,T) = (100,100)$					
DGP I	0.002	0.998	-	1.000	0.000	-	0.000	1.000	-
DGP II	0.010	0.990	-	1.000	0.000	-	0.000	1.000	-
DGP III	0.004	0.996	-	1.000	0.000	-	0.000	1.000	-
DGP IV	0.028	0.972	-	1.000	0.000	-	0.000	1.000	-
DGP V	0.070	0.930	-	1.000	0.000	-	0.000	1.000	-
DGP VI	0.031	0.969	-	1.000	0.000	-	0.000	1.000	-
DGP VII	0.016	0.000	0.984	0.000	1.000	0.000	0.000	0.000	1.000
DGP VIII	0.052	0.000	0.948	0.000	1.000	0.000	0.000	0.000	1.000
DGP IX	0.030	0.000	0.970	0.000	1.000	0.000	0.000	0.000	1.000

especially with small or moderate sample sizes. In contrast, the last two columns in each panel illustrate that the BER method consistently estimates R and the coverage rate increases as the sample size increases. Also, the middle two columns report the ER test on the overtime demeaned Y_{it} and the result verifies our conjecture that the additional/dominant term comes from the time-invariant factor.

Next, we examine the numerical performance of the APC estimator. In all experiments, the APC method works well without pre-specifying the error's distribution. For example, in the case where errors follow a mixture of normal and gamma distributions, the property of MLE is unknown due to the unknown density of ε_{it}, but the APC method performs well. Unlike Bai (2003) and Wang (2022), there are two time-varying factors in DGP VI-IX, and then it is not straightforward to compare correlation coefficients among DGPs. Instead, Table 2 regresses each factor on the estimate $\hat{F}_t$ i.e., $f^*_{r,t} = \hat{F}'_{BER,t}\beta_r + error$ and computes the corresponding R-squared that essentially plays the same role as the correlation coefficient. The large value of R-squared verifies our theoretical conjecture that $\hat{F}_t$ estimates an orthogonal transformation of $f^*_{r,t}$.

Table 2. R-square of the APC Estimator

(N,T)	$(50,50)$		$(100,50)$		$(50,100)$		$(100,100)$	
	$R^2_{f^*_1}$	$R^2_{f^*_2}$	$R^2_{f^*_1}$	$R^2_{f^*_2}$	$R^2_{f^*_1}$	$R^2_{f^*_2}$	$R^2_{f^*_1}$	$R^2_{f^*_2}$
DGP I	0.8812	-	0.9036	-	0.8844	-	0.9055	-
DGP II	0.8857	-	0.9110	-	0.8882	-	0.9125	-
DGP III	0.8847	-	0.9047	-	0.8855	-	0.9062	-
DGP IV	0.8887	-	0.9136	-	0.8922	-	0.9151	-
DGP V	0.8910	-	0.9179	-	0.8926	-	0.9192	-
DGP VI	0.8907	-	0.9125	-	0.8914	-	0.9137	-
DGP VII	0.8712	0.8716	0.9010	0.9023	0.8768	0.8753	0.9024	0.9024
DGP VIII	0.8729	0.8724	0.9053	0.9059	0.8792	0.8778	0.9081	0.9083
DGP IX	0.8736	0.8740	0.8996	0.9016	0.8781	0.8766	0.9021	0.9026

Table 3 presents the mean squared error of $\hat{f}^*_{r,t} = \hat{F}'_{BER,t}\hat{\beta}_r$ to reveal the consistency of the APC estimator. It is easy to see that the mean squared errors in all experiments are low and vanish as N increases. For example, in DGP VIII, the mean squared errors are 0.1185 and 0.1200 when $(N, T) = (50, 100)$ and decreases to around 0.0908 and 0.0906 if the sample size increases to $(N, T) = (100, 100)$. Figure 1 plots the true factor process (solid line) along with the 95% confidence intervals (dashed line) for $f_t^* = (f_{1t}^*, f_{2t}^*)$ for the first 20 time periods when $(N, T) = (100, 100)$. The true factors rarely fall outside the confidence intervals.

Table 3. Mean Squared Error of the APC Estimator

(N, T)	$(50, 50)$		$(100, 50)$		$(50, 100)$		$(100, 100)$	
	$MSE_{f_1^*}$	$MSE_{f_2^*}$	$MSE_{f_1^*}$	$MSE_{f_2^*}$	$MSE_{f_1^*}$	$MSE_{f_2^*}$	$MSE_{f_1^*}$	$MSE_{f_2^*}$
DGP I	0.1166	-	0.0955	-	0.1148	-	0.0941	-
DGP II	0.1120	-	0.0883	-	0.1110	-	0.0871	-
DGP III	0.1133	-	0.0947	-	0.1138	-	0.0935	-
DGP IV	0.1089	-	0.0856	-	0.1070	-	0.0845	-
DGP V	0.1067	-	0.0814	-	0.1064	-	0.0803	-
DGP VI	0.1071	-	0.0868	-	0.1077	-	0.0859	-
DGP VII	0.1237	0.1246	0.0968	0.0956	0.1211	0.1230	0.0965	0.0964
DGP VIII	0.1215	0.1234	0.0925	0.0917	0.1185	0.1200	0.0908	0.0906
DGP IX	0.1214	0.1225	0.0984	0.0962	0.1198	0.1216	0.0969	0.0963

Table 4. Mean Squared Error of the Estimated Propensity Score

(N, T)	$(50, 50)$	$(100, 50)$	$(50, 100)$	$(100, 100)$
DGP I	0.0147	0.0111	0.0111	0.0076
DGP II	0.0151	0.0114	0.0113	0.0077
DGP III	0.0147	0.0114	0.0113	0.0080
DGP IV	0.0150	0.0118	0.0117	0.0087
DGP V	0.0150	0.0117	0.0116	0.0084
DGP VI	0.0149	0.0120	0.0120	0.0090
DGP VII	0.0209	0.0163	0.0163	0.0120
DGP VIII	0.0211	0.0163	0.0163	0.0119
DGP IX	0.0213	0.0171	0.0171	0.0131

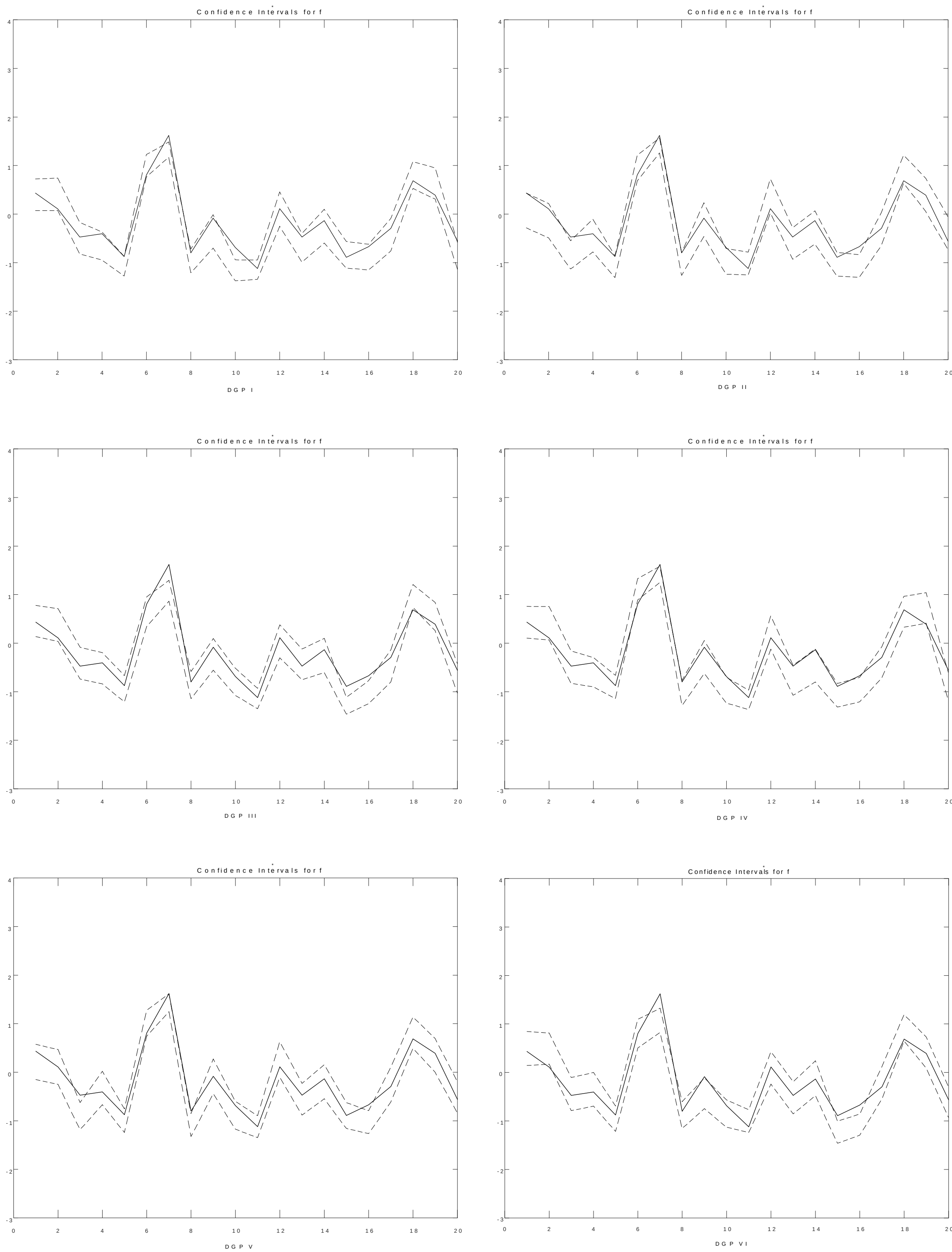

Figure 1. Confidence Intervals for Factors

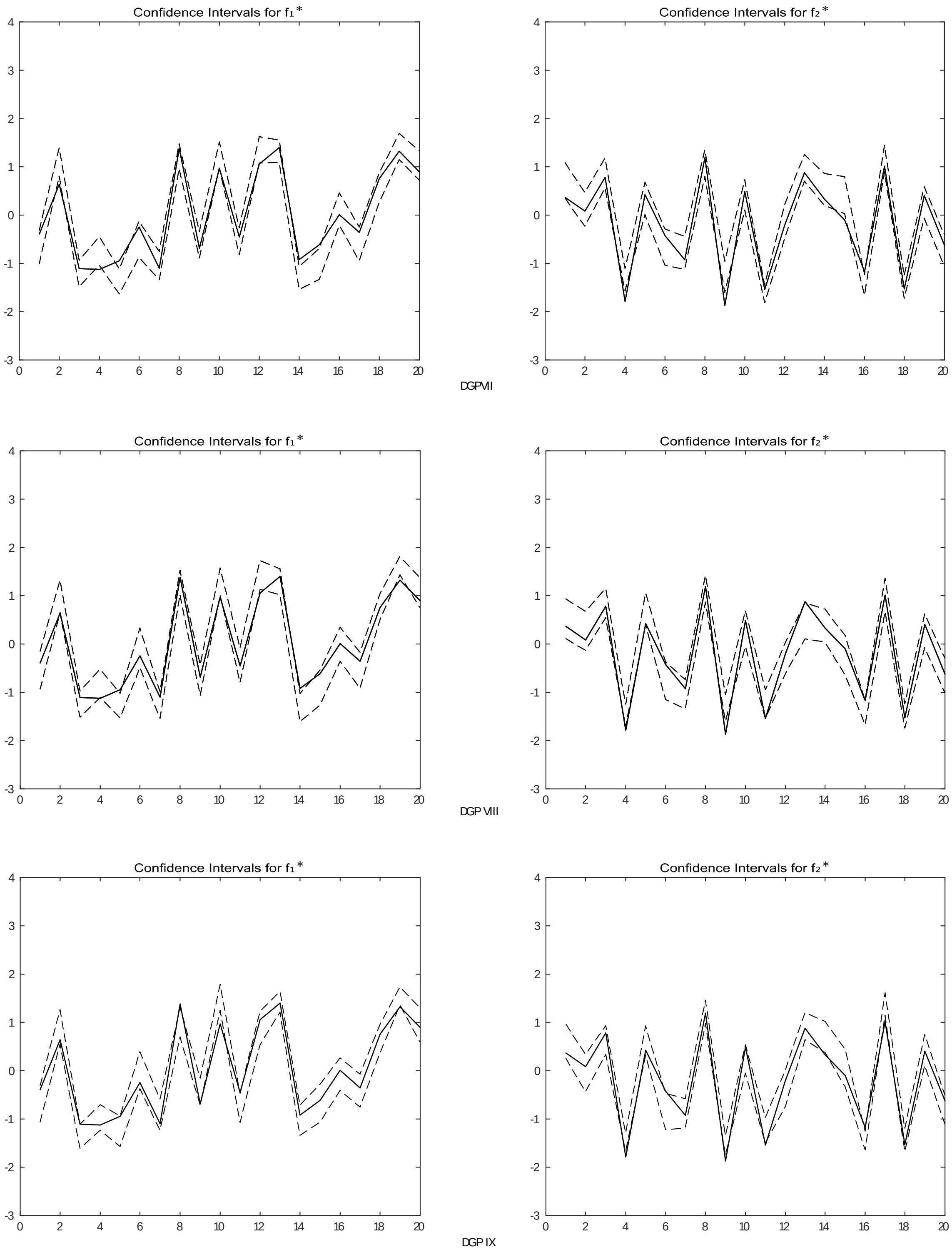

Figure 1 con't. Confidence Intervals for Factors

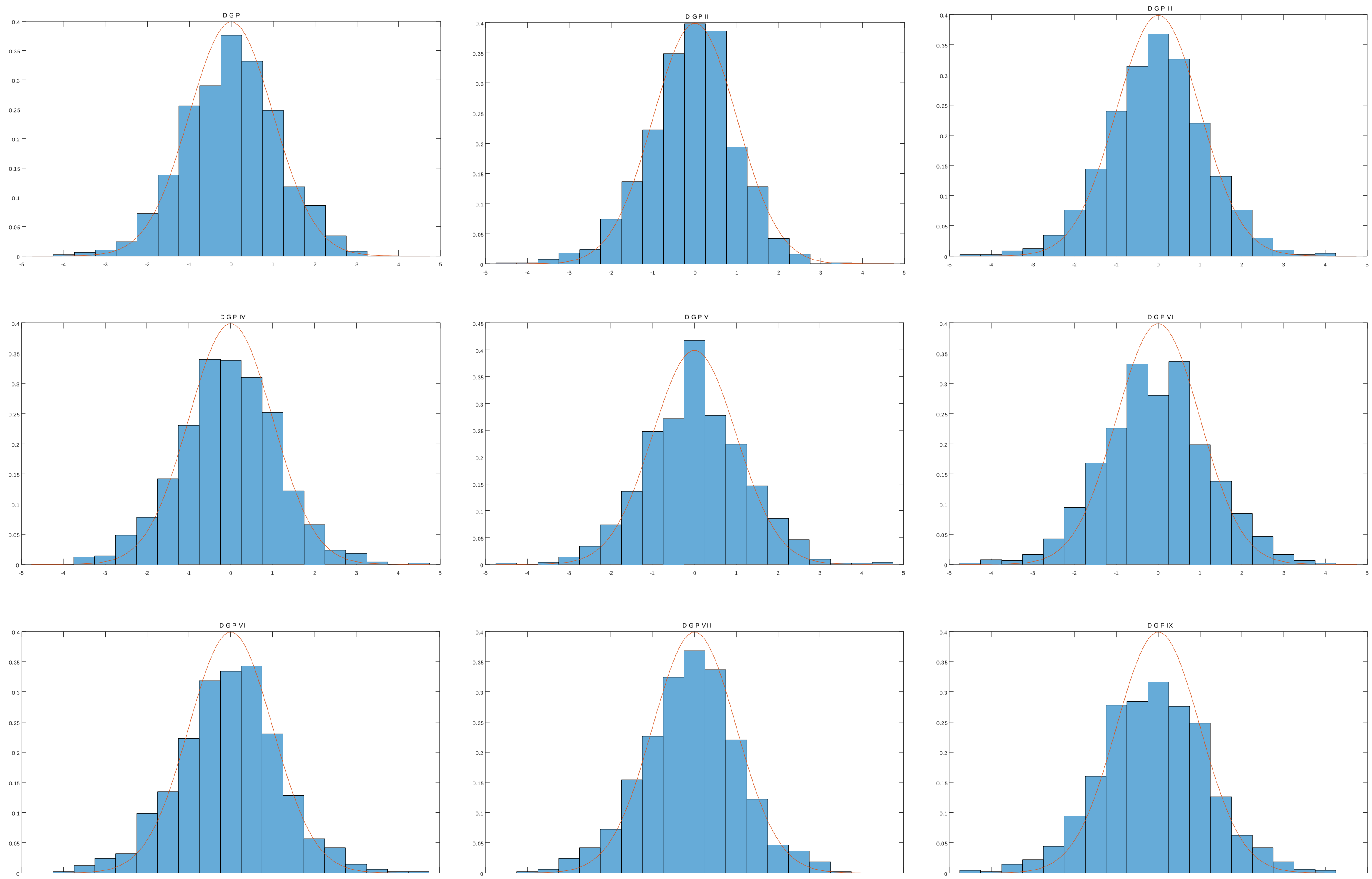

Figure 2. Histogram for Estimated Propensity Scores

For the binary model, one is often interested in an individual's propensity to take an action. Table 4 calculates the mean squared error of the estimated propensity score by $\frac{1}{1000}\sum_{rep=1}^{1000}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\Lambda}'_{BER,i}\hat{F}_{BER,t} - P_{\varepsilon}(\lambda_i^{*\prime}f_t^*))^2$. The small mean squared error numerically verifies the consistency of $\hat{\Lambda}'_{BER,i}\hat{F}_{BER,t}$. In addition, Figure 2 plots normalized histograms of the estimated propensity score and the true normal density (solid line) for $(i,t) = (N/2, T/2)$. Figure 2 graphically shows that the normal density is a good approximation of the normalized propensity score estimates.

7 Application: Determinants of Dividend Initiation

Whether to pay a dividend is a common decision that companies need to make in the routine operational management. Among investment and financing policies, dividend decision is an important research area of corporate finance. Researchers are often interested in identifying factors that influence firms' dividend decision.

A vast literature has considered effects of various influencing factors on a firm's dividend decision, such as tax, risk information asymmetry, debt, corporate governance (Miller & Scholes (1982), Gordon (1963), Black & Scholes (1974), Grullon, Michaely & Swaminathan (2002), Kania & Bacon (2005), Jo & Pan (2009)). A common practice is to conduct an OLS regression of dividend payout, with less attention being paid to the underlying reasons of the dividend initiation decision (i.e., positive or zero dividend). To address this issue, Fama and French (2001) and Grinstein and Michaely (2005) studied the importance of time varying factors in the propensity of paying dividends. The binary regression plays a crucial role even if the dividend payout is of interest and treated as the censored data, e.g. Forti et al. (2015), since the propensity score lays the foundation to restrict the subpopulation that chooses to pay dividends. Furthermore, it is well known that the factor structure in the latent outcome is more flexible than the traditional linear structure with observed factors. The companies are allowed to have different exposures to common factors. Such observations motivate us to concentrate on a binary factor analysis of the dividend decision. To deal with the misspecification caused by parametric specification of the error's distribution, we apply the proposed APC method to identify determinants of the dividend decision, and utilize the estimated factors to seek for possible influencing factors.

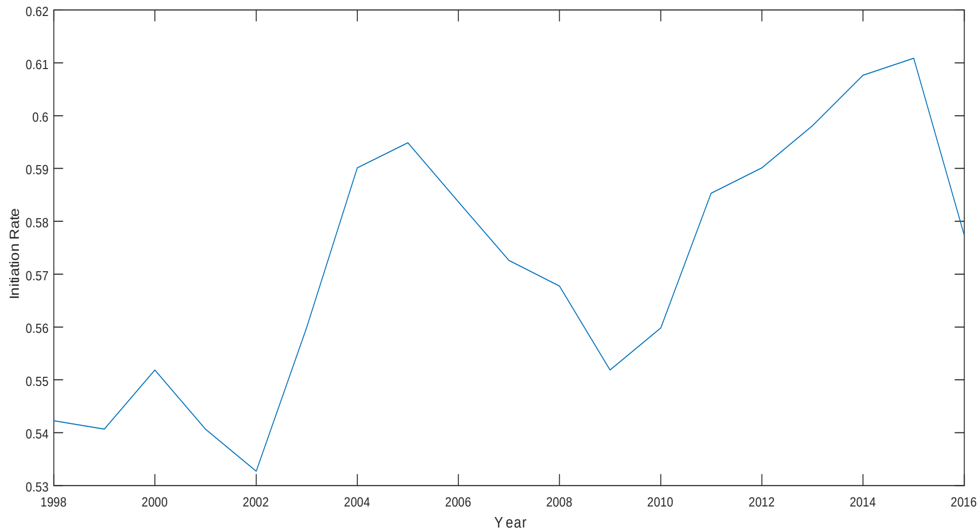

Figure 3. Rate of Dividend Initiation

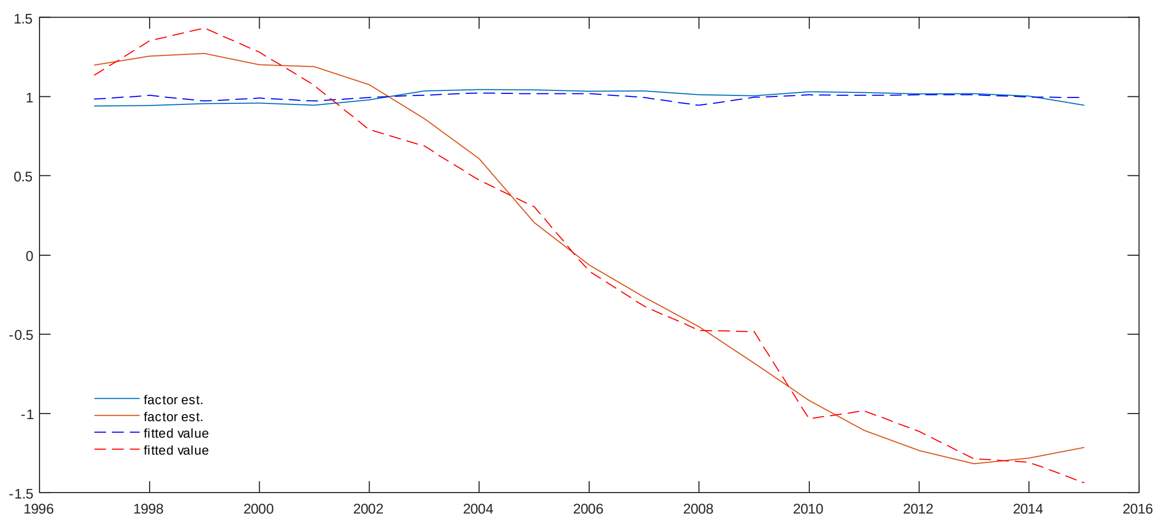

Figure 4. Estimated Factors and Fitted Values

The dataset we consider is the stock dividend initiation by S&P500 firms from 1998 to 2016 on an annual frequency. We drop an observation if the firm enters the

S&P500 database no more than 4 years. The total number of firms is 627, that is, $(N, T) = (627, 19)$. In the dataset, 86.28% of firms have initiated dividends during the past 19 years. The overall proportion of initiation is 57.15%. The initiation rate over year is plotted in Figure 3. Overall, the rate of initiation is low from 2000 to 2004, and hit the bottom in 2002. It is called "lost period" caused by dot-com bubble in March 2000. The rate fluctuates afterwards due to the rebound of equity market and the financial crisis of 2008.

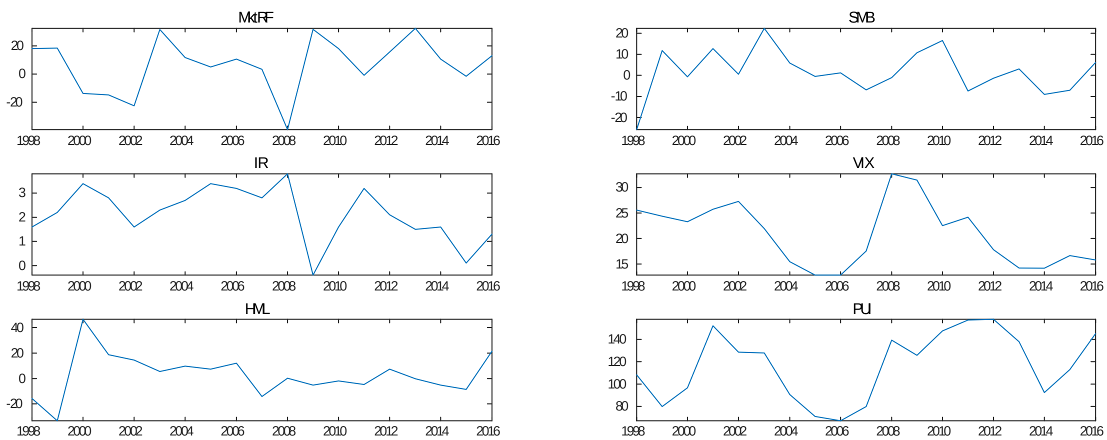

Figure 5. Possible Influencing Factors

The dependent variable Y_{it} takes one if firm i in year t issues the dividend, and zero otherwise. The number of common factors in the transformed factor model, as being determined by the proposed BER method, is found to be $\hat{R} = 2$. In contrast, the traditional ER test underestimates the number of factors to be one. Next, after controlling two factors, we apply the APC method to obtain factor estimates. Figure 4 plots the factors at the bottom panel in solid lines. The estimation result further verifies our theoretical conjecture. One estimated factor $(\hat{F}_1)$ characterizes the time-invariant factor, the other $(\hat{F}_2)$ estimates the time-varying factor. Next, we try to match our factor estimates to a few popular observed influencing factors in the financial literature. Fama-French factors are widely adopted in asset pricing and portfolio management studies to describe the stock market returns. We treat Market Risk-free Return Rate $(MktRF)$, Small Minus Big (SMB), and High Minus

Low (HML)[3] as possible influencing factors of the dividend initiation. Moreover, Basse and Reddemann (2013) pointed out that managers are inclined to take the inflationary environment into account of the dividend initiation. Hence, we include the inflation rate (IR) as an explanatory variable. We also consider the policy uncertainty index (PUI) and the volatility index (VIX) to control the uncertainty of economic environment. The six observed series are plotted in Figure 5. We regress the estimated time-varying factor on these observed factors along with a time trend variable $Year$, i.e.,

$$\hat{F}_2 = \beta_1 MktRF + \beta_2 SMB + \beta_3 HML + \beta_4 IR + \beta_5 VIX + \beta_6 PUI + \beta_7 Year + error,$$

where all explanatory variables are normalized to have zero mean in the OLS regression.

Table 5. OLS Regression of Time-varying Factor

	Coef.	Coef.	Coef.	Coef.
MktRF	-0.0016	-0.0022	-0.0038**	-0.0072***
	(0.0020)	(0.0023)	(0.0018)	(0.0021)
SMB	0.0087***	0.0087***	0.0115***	0.0127***
	(0.0030)	(0.0031)	(0.0031)	(0.0031)
HML	0.0053**	0.0055***	0.0040	0.0036
	(0.0024)	(0.0024)	(0.0027)	(0.0024)
IR		-0.0280		-0.0901***
		(0.0438)		(0.0285)
VIX			-0.0119**	-0.0217***
			(0.0060)	(0.0058)
PUI			-0.0013	-0.0008
			(0.0013)	(0.0012)
Year	-0.1754***	-0.1772***	-0.1765***	-0.1867***
	(0.0070)	(0.0063)	(0.0071)	(0.0063)
R-squared	0.9704	0.9710	0.9777	0.9823
Adj R-squared	0.9645	0.9627	0.9691	0.9735

[3] Fama-French factors source: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html

The regression result reported in Table 5 shows that $MktRF$, SMB, IR and VIX have strong explanatory power for the estimated factors $\hat{F}_2$ with an R-squared 98.23%. The bottom panel Figure 4 plots the fitted value in dashed lines. It shows that a linear combination of $MktRF$, SMB, IR and VIX can account for the time-varying factor well.

8 Conclusion

Decisions and ratings are prevalent economic behaviors across various economic fields, such as finance and international trade. To address these issues, this paper explores binary and ordered factor models, where factors and factor loadings are treated as fixed effects to account for unobserved heterogeneous effects. We introduce an adjusted principal components (APC) method that eliminates the need for parametric specifications of the error distribution or the selection of initial values for factors. Moreover, our estimation approach is computationally straightforward and accessible to applied researchers. Extensive Monte Carlo experiments validate our theoretical conjectures and demonstrate the robust performance of the APC estimator across various empirical settings. Using an example of dividend initiation among S&P 500 firms from 1998 to 2016, we illustrate the practical application of our APC procedure.

Several potential extensions for future research emerge from this work. Boneva and Linton (2017) and Chen et al. (2021) have explored maximum likelihood estimation for discrete response models and nonlinear models with interactive effects. Unlike factor models, these approaches incorporate observed regressors and shift the focus from prediction to estimating marginal and average effects of explanatory variables. However, the former relies on the common correlated effects method, which imposes restrictive assumptions on the relationship between observed regressors and unobserved factors, while the latter is sensitive to initial values due to identification challenges. Investigating how to consistently estimate parameters of interest in binary or general nonlinear models with interactive fixed effects remains an intriguing area for future research. We leave this discussion for subsequent work.

9 Appendix

Proof of Theorem 1. (i) From (10), we obtain the identity $\frac{1}{NT}\mathbf{Y}\mathbf{Y}'\hat{\mathbf{F}}_{BER} = \hat{\mathbf{F}}_{BER}\mathbf{U}_{NT}$. Let ν is a $T \times N$ matrix composed of v_{it}. In view of $\mathbf{Y} = \mathbf{F}\mathbf{\Lambda}' + \nu$, multiplying $T^{-1}\left(\mathbf{\Lambda}'\mathbf{\Lambda}/N\right)^{1/2}\mathbf{F}'$ on both sides of the identity yields

$$\left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)^{1/2}\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)\left(\frac{\mathbf{F}'\hat{\mathbf{F}}_{BER}}{T}\right)+a_{NT} = \left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)^{1/2}\left(\frac{\mathbf{F}'\hat{\mathbf{F}}_{BER}}{T}\right)\mathbf{U}_{NT}, \quad (16)$$

where

$$a_{NT} = \left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)^{1/2}\left[\frac{\mathbf{F}'\mathbf{F}}{T}\frac{\mathbf{\Lambda}'\nu'\hat{\mathbf{F}}_{BER}}{NT} + \frac{\mathbf{F}'\nu\mathbf{\Lambda}}{NT}\frac{\mathbf{F}'\hat{\mathbf{F}}_{BER}}{T} + \frac{\mathbf{F}'\nu}{NT}\frac{\nu'\hat{\mathbf{F}}_{BER}}{T}\right] \equiv \sum_{l=1}^{3} a_{NT,l}.$$

By the Cauchy-Schwarz inequality under Assumption 4 and the property that $\frac{1}{T}\left\|\hat{\mathbf{F}}_{BER} - \mathbf{F}\mathbf{H}\right\|^2$
$= o_p(1)$ which is proven in Theorem 1 of Bai and Ng (2002), we have

$$a_{NT,1} = \left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)^{1/2}\frac{\mathbf{F}'\mathbf{F}}{T}\frac{\mathbf{\Lambda}'\nu'\mathbf{F}\mathbf{H}}{NT} + \left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)^{1/2}\frac{\mathbf{F}'\mathbf{F}}{T}\frac{\mathbf{\Lambda}'\nu'(\hat{\mathbf{F}}_{BER} - \mathbf{F}\mathbf{H})}{NT} = o_p(1)$$

$$a_{NT,2} = \left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)^{1/2}\frac{\mathbf{F}'\nu\mathbf{\Lambda}}{NT}\frac{\mathbf{F}'\mathbf{F}\mathbf{H}}{T} + \left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)^{1/2}\frac{\mathbf{F}'\nu\mathbf{\Lambda}}{NT}\frac{\mathbf{F}'(\hat{\mathbf{F}}_{BER} - \mathbf{F}\mathbf{H})}{T} = o_p(1)$$

$$a_{NT,3} = \left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)^{1/2}\frac{\mathbf{F}'\nu}{\sqrt{N}T}\frac{\nu'\mathbf{F}\mathbf{H}}{\sqrt{N}T} + \left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)^{1/2}\frac{\mathbf{F}'\nu}{\sqrt{N}T}\frac{\nu'(\hat{\mathbf{F}}_{BER} - \mathbf{F}\mathbf{H})}{\sqrt{N}T} = o_p(1).$$

After a simple calculation, (16) becomes

$$\left[\left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)^{1/2}\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)^{1/2} + a_{NT}\left(\left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)^{1/2}\left(\frac{\mathbf{F}'\hat{\mathbf{F}}_{BER}}{T}\right)\right)^{-1}\right]\Upsilon_{NT} = \Upsilon_{NT}\mathbf{U}_{NT}, \quad (17)$$

where $\Upsilon_{NT} = \left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)^{1/2}\left(\frac{\mathbf{F}'\hat{\mathbf{F}}_{BER}}{T}\right)\mathbf{U}_{NT}^{*-1/2}$ with $\mathbf{U}_{NT}^{*}$ being a diagonal matrix composed of diagonal elements of $\left(\frac{\mathbf{F}'\hat{\mathbf{F}}_{BER}}{T}\right)'\left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)\left(\frac{\mathbf{F}'\hat{\mathbf{F}}_{BER}}{T}\right)$. The left hand side of (17) converges to $\Sigma_\Lambda^{1/2}\Sigma_F\Sigma_\Lambda^{1/2}$ in probability since each term is positive definite by Assumption 2. Under Assumption 5, Υ_{NT} is uniquely determined with the probability limit Υ which is a unique eigenvector matrix of $\Sigma_\Lambda^{1/2}\Sigma_F\Sigma_\Lambda^{1/2}$. By the definition of Υ_{NT},

$$\frac{\mathbf{F}'\hat{\mathbf{F}}_{BER}}{T} = \left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)^{-1/2}\Upsilon_{NT}\mathbf{U}_{NT}^{*1/2}.$$

It follows that

$$\frac{\mathbf{F}'\hat{\mathbf{F}}_{BER}}{T} \xrightarrow{p} \Sigma_\Lambda^{-1/2}\Upsilon\mathbf{U}.$$

(ii) According to Theorem 1 of Bai and Ng (2002),

$$\begin{aligned}\hat{F}_{BER,t} - \mathbf{H}'F_t &= \mathbf{U}_{NT}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\hat{F}_{BER,s}\gamma_{st} + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_{BER,s}\left(\frac{v_s'v_t}{N} - \gamma_{st}\right)\right.\\ &\quad\left. + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_{BER,s}F_s'\mathbf{\Lambda}'v_t\Big/ N + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_{BER,s}F_t'\mathbf{\Lambda}'v_s\Big/ N\right)\\ &\equiv \sum_{l=1}^{4} b_{t,l}. \end{aligned} \tag{18}$$

Notice that

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_{BER,t} - \mathbf{H}'F_t\right\|^2 \lesssim \frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{4}\|b_{t,l}\|^2.$$

In view of $\left\|\mathbf{U}_{NT}^{-1}\right\| = O_p(1)$, by the Cauchy-Schwarz inequality and Assumption 3

$$\frac{1}{T}\sum_{t=1}^{T}\|b_{t,1}\|^2 \lesssim T^{-2}\sum_{t=1}^{T}\left(T^{-1}\sum_{s=1}^{T}\left\|\hat{F}_{BER,s}\right\|^2\right)\left(T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_{st}^2\right) = O_p(T^{-1});$$

$$\frac{1}{T}\sum_{t=1}^{T}\|b_{t,2}\|^2 \lesssim \frac{1}{T}\sum_{t=1}^{T}\left[T^{-2}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\left(\frac{v_s'v_t}{N} - \gamma_{st}\right)\left(\frac{v_s'v_t}{N} - \gamma_{st}\right)\right)^2\right]^{1/2}$$

$$\times \left[T^{-2}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\hat{F}_{BER,s}\hat{F}_{BER,u}\right)^2\right]^{1/2} = O_p(N^{-1});$$

$$\frac{1}{T}\sum_{t=1}^{T}\|b_{t,3}\|^2 \lesssim \left\|\frac{1}{T}\sum_{s=1}^{T}\hat{F}'_{BER,s}F_s\right\|\left(\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{\Lambda}'v_t/N\|^2\right) = O_p(N^{-1});$$

$$\frac{1}{T}\sum_{t=1}^{T}\|b_{t,4}\|^2 \lesssim \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_{BER,s}\right\|^2\right)^{1/2}\left\|\frac{1}{T}\sum_{t=1}^{T}F_s\right\|\left(\frac{1}{T}\sum_{s=1}^{T}\|\mathbf{\Lambda}'v_s/N\|^2\right)^{1/2} = O_p(N^{-1}).$$

It follows that $\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_{BER,t}-\mathbf{H}'F_t\right\|^2 = O_p(\delta_{NT}^{-2})$. Consequently, the asserted claim holds.

Proof of Theorem 2. Based on (18) and Theorem 1 under Assumption 2-4 and Assumption 6, the following terms hold: First

$$b_{t,1} = \mathbf{U}_{NT}^{-1}\frac{1}{T}\sum_{s=1}^{T}\left[\left(\hat{F}_{BER,s}-\mathbf{H}'F_s\right)\gamma_{st}+\mathbf{H}'F_s\gamma_{st}\right] \lesssim T^{-1/2}O_p(\delta_{NT}^{-1}),$$

since $T^{-1}\sum_{s=1}^{T}F_s\gamma_{st} = O_p(T^{-1})$ by the Markov inequality and $E\left\|T^{-1}\sum_{s=1}^{T}F_s\gamma_{st}\right\| \lesssim O(T^{-1})$, and

$$\left\|\frac{1}{T}\sum_{s=1}^{T}\left(\hat{F}_{BER,s}-\mathbf{H}'F_s\right)\gamma_{st}\right\| \leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_{BER,s}-\mathbf{H}'F_s\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\gamma_{st}^2\right)^{1/2}$$
$$= T^{-1/2}O_p(\delta_{NT}^{-1}).$$

Next,

$$b_{t,2} = \mathbf{U}_{NT}^{-1}\frac{1}{T}\sum_{s=1}^{T}\left[\left(\hat{F}_{BER,s}-\mathbf{H}'F_s\right)+\mathbf{H}'F_s\right]\left(\frac{v_s'v_t}{N}-\gamma_{st}\right) \lesssim N^{-1/2}O_p(\delta_{NT}^{-1}),$$

because of $\left\|\frac{1}{T}\sum_{s=1}^{T}\left(\hat{F}_{BER,s}-\mathbf{H}'F_s\right)\left(\frac{v_s'v_t}{N}-\gamma_{st}\right)\right\| \leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_{BER,s}-\mathbf{H}'F_s\right\|^2\right)^{1/2}$

$\times \left(\frac{1}{T} \sum_{s=1}^{T} \left(\frac{v_s' v_t}{N} - \gamma_{st} \right)^2 \right)^{1/2} = N^{-1/2} O_p(\delta_{NT}^{-1})$ and $\frac{1}{T} \sum_{s=1}^{T} \mathbf{H}' F_s \left(\frac{v_s' v_t}{N} - \gamma_{st} \right) = O_p((NT)^{-1/2})$. In addition,

$$b_{t,3} = \mathbf{U}_{NT}^{-1} \frac{1}{NT} \sum_{s=1}^{T} \left[\left(\hat{F}_{BER,s} - \mathbf{H}' F_s \right) + \mathbf{H}' F_s \right] F_s' \mathbf{\Lambda}' v_t \lesssim O_p(N^{-1/2})$$

due to $\left\| \frac{1}{NT} \sum_{s=1}^{T} \left(\hat{F}_{BER,s} - \mathbf{H}' F_s \right) F_s' \mathbf{\Lambda}' v_t \right\| \leq \left(\frac{1}{T} \sum_{s=1}^{T} \left\| \hat{F}_{BER,s} - \mathbf{H}' F_s \right\|^2 \right)^{1/2}$ $\times \left(\frac{1}{N^2 T} \sum_{s=1}^{T} \| F_s' \mathbf{\Lambda}' v_t \|^2 \right)^{1/2} = N^{-1/2} O_p(\delta_{NT}^{-1})$. Finally,

$$b_{t,4} = \mathbf{U}_{NT}^{-1} \frac{1}{NT} \sum_{s=1}^{T} \left[\left(\hat{F}_{BER,s} - \mathbf{H}' F_s \right) + \mathbf{H}' F_s \right] F_t' \mathbf{\Lambda}' v_s = N^{-1/2} O_p(\delta_{NT}^{-1}),$$

since $\left\| \frac{1}{T} \sum_{s=1}^{T} \left(\hat{F}_{BER,s} - \mathbf{H}' F_s \right) v_s' \mathbf{\Lambda} F_t \right\| \leq \frac{1}{\sqrt{N}} \left(\frac{1}{T} \sum_{s=1}^{T} \left\| \hat{F}_{BER,s} - \mathbf{H}' F_s \right\|^2 \right)^{1/2}$ $\times \left(\frac{1}{T} \sum_{s=1}^{T} \left\| \frac{v_s' \mathbf{\Lambda}}{\sqrt{N}} \right\| \right)^{1/2} \| F_t \| = N^{-1/2} O_p(\delta_{NT}^{-1})$ and $\frac{1}{NT} \sum_{s=1}^{T} F_s v_s' \mathbf{\Lambda} F_t = O_p((NT)^{-1/2})$.

Collecting all above terms, we obtain

$$\hat{F}_{BER,t} - \mathbf{H}' F_t = \mathbf{U}_{NT}^{-1} \frac{1}{NT} \sum_{s=1}^{T} \hat{F}_{BER,s} F_s' \mathbf{\Lambda}' v_t + O_p(\delta_{NT}^{-2}).$$

It follows that

$$\sqrt{N} \left(\hat{F}_{BER,t} - \mathbf{H}' F_t \right) = \mathbf{U}_{NT}^{-1} \left(\frac{\hat{\mathbf{F}}_{BER}' \mathbf{F}}{T} \right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Lambda_i \nu_{it} + o_p(1) \tag{19}$$

given that $\frac{\sqrt{N}}{T} \to 0$. Subsequently, we obtain the asserted claim by the central limit theorem.

(ii) By (11),

$$\hat{\Lambda}_{BER,i} = \frac{1}{T}\hat{\mathbf{F}}'_{BER}\mathbf{F}\Lambda_i + \frac{1}{T}\hat{\mathbf{F}}'_{BER}\nu_i$$
$$= \mathbf{H}^{-1}\Lambda_i + \frac{1}{T}\sum_{t=1}^{T}\mathbf{H}'F_t\nu_{it} - \frac{1}{T}\sum_{t=1}^{T}(\hat{F}_{BER,t} - \mathbf{H}'F_t)\hat{F}'_{BER,t}\mathbf{H}^{-1}\Lambda_i \quad (20)$$
$$+ \frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_{BER,t} - \mathbf{H}'F_t\right)\nu_{it}.$$

On the one hand,

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{F}_{BER,t} - \mathbf{H}'F_t)F'_t$$
$$= \mathbf{U}_{NT}^{-1}\left(T^{-2}\sum_{s=1}^{T}\sum_{t=1}^{T}\hat{F}_{BER,s}F'_t\gamma_{st} + T^{-2}\sum_{s=1}^{T}\sum_{t=1}^{T}\hat{F}_{BER,s}F'_t\left(\frac{v'_s v_t}{N} - \gamma_{st}\right)\right.$$
$$\left. + T^{-2}\sum_{s=1}^{T}\sum_{t=1}^{T}\hat{F}_{BER,s}F'_s\mathbf{\Lambda}'v_tF'_t \Big/ N + T^{-2}\sum_{s=1}^{T}\sum_{t=1}^{T}\hat{F}_{BER,s}v'_s\mathbf{\Lambda}F_tF'_t \Big/ N\right)$$
$$\equiv \sum_{l=1}^{4}c_l$$

where $c_1 = T^{-1/2}O_p(\delta_{NT}^{-1})$ and $c_2 = c_4 = N^{-1/2}O_p(\delta_{NT}^{-1})$ similar to $b_{t,1}$, $b_{t,2}$ and $b_{t,4}$. Moreover,

$$c_3 = \mathbf{U}_{NT}^{-1}N^{-1}T^{-2}\sum_{s=1}^{T}\sum_{t=1}^{T}\left[\left(\hat{F}_{BER,s} - \mathbf{H}'F_s\right)F'_s\mathbf{\Lambda}'v_tF'_t + \mathbf{H}'F_sF'_s\mathbf{\Lambda}'v_tF'_t\right]$$
$$= (NT)^{-1/2}O_p(\delta_{NT}^{-1}) + O_p((NT)^{-1/2}).$$

Together with Theorem 1, the third term of (20) is

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{F}_{BER,t} - \mathbf{H}'F_t)(\hat{F}_{BER,t} - \mathbf{H}'F_t)'\mathbf{H}^{-1}\Lambda_i + \frac{1}{T}\sum_{t=1}^{T}(\hat{F}_{BER,t} - \mathbf{H}'F_t)F'_t\Lambda_i = O_p(\delta_{NT}^{-2}).$$

On the other hand, write

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_{BER,t}-\mathbf{H}'F_t\right)\nu_{it}\\
&=\mathbf{U}_{NT}^{-1}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{F}_{BER,s}\gamma_{st}v_{it}+\mathbf{U}_{NT}^{-1}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{F}_{BER,s}\left(\frac{v_s'v_t}{N}-\gamma_{st}\right)v_{it}\\
&+\mathbf{U}_{NT}^{-1}\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{F}_{BER,s}F_s'\mathbf{\Lambda}'v_tv_{it}+\mathbf{U}_{NT}^{-1}\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{F}_{BER,s}F_t'\mathbf{\Lambda}'v_sv_{it}\\
&=\sum_{l=1}^{4}d_l.
\end{aligned}
$$

Under Assumption 1-6, it is easy to show

$$
\begin{aligned}
d_1&=\mathbf{U}_{NT}^{-1}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{F}_{BER,s}-\mathbf{H}'F_s)\gamma_{st}v_{it}+\mathbf{U}_{NT}^{-1}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{H}'F_s\gamma_{st}v_{it}\\
&=T^{-1/2}O_p(\delta_{NT}^{-1});\\
d_2&=\mathbf{U}_{NT}^{-1}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{F}_{BER,s}-\mathbf{H}'F_s)\left(\frac{v_s'v_t}{N}-\gamma_{st}\right)v_{it}+\mathbf{U}_{NT}^{-1}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{H}'F_s\left(\frac{v_s'v_t}{N}-\gamma_{st}\right)v_{it}\\
&=N^{-1/2}O_p(\delta_{NT}^{-1});\\
d_3&=\mathbf{U}_{NT}^{-1}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{F}_{BER,s}-\mathbf{H}'F_s)F_s'\mathbf{\Lambda}'v_tv_{it}+\mathbf{U}_{NT}^{-1}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{H}'F_sF_s'\mathbf{\Lambda}'v_tv_{it}\\
&=N^{-1/2}O_p(\delta_{NT}^{-1});\\
d_4&=\mathbf{U}_{NT}^{-1}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{F}_{BER,s}-\mathbf{H}'F_s)F_t'\mathbf{\Lambda}'v_sv_{it}+\mathbf{U}_{NT}^{-1}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{H}'F_sF_t'\mathbf{\Lambda}'v_sv_{it}\\
&=N^{-1/2}O_p(\delta_{NT}^{-1}).
\end{aligned}
$$

Then, the fourth term of (20) is $O_p(\delta_{NT}^{-2})$.

Based on (20),

$$\sqrt{N}(\hat{\Lambda}_{BER,i} - \mathbf{H}^{-1}\Lambda_i) = \mathbf{U}_{NT}^{-1}\left(\frac{\hat{\mathbf{F}}'_{BER}\mathbf{F}}{T}\right)\left(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\right)\frac{1}{\sqrt{T}}\sum_{i=1}^{N} F_t\nu_{it} + O_p(\delta_{NT}^{-2}) \quad (21)$$

$$\xrightarrow{d} N(\mathbf{0}, (\mathbf{H}')^{-1}\Omega_i(\mathbf{H})^{-1})$$

given that $\frac{\sqrt{T}}{N} \to 0$.

Proof of Theorem 3. Write $\tilde{\mathbf{\Lambda}} = (\tilde{\Lambda}_1, \cdots, \tilde{\Lambda}_N)'$, $\tilde{\mathbf{F}} = (\tilde{F}_1, \cdots, \tilde{F}_T)'$ and $\tilde{\mathbf{F}}_{BER} = (\tilde{F}_{BER,1}, \cdots, \tilde{F}_T)'$. Let $\tilde{\mathbf{U}}_{NT} = diag(\tilde{u}_{NT,1}, \cdots, \tilde{u}_{NT,\tilde{R}})$ where $\tilde{u}_{NT,1} > \cdots > \tilde{u}_{NT,\tilde{R}}$ are the $\tilde{R}$ largest eigenvalues of $\frac{1}{NT}\mathbf{YY}'$, and $\tilde{\mathbf{H}} = \left(\frac{\tilde{\mathbf{\Lambda}}'\tilde{\mathbf{\Lambda}}}{N}\right)\left(\frac{\tilde{\mathbf{F}}'\tilde{\mathbf{F}}_{BER}}{T}\right)\tilde{\mathbf{U}}_{NT}^{-1}$.

Extending proofs of Theorem 1 to $\tilde{R} \to \infty$ yields $\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_{BER,t} - \tilde{\mathbf{H}}'\tilde{F}_t\right\|^2 = O_P(\tilde{R}\delta_{NT}^{-2})$, $\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\Lambda}_{BER,i} - \tilde{\mathbf{H}}^{-1}\tilde{\Lambda}_i\right\|^2 = O_P(\tilde{R}\delta_{NT}^{-2})$. Similar to (19) and (21), one can show that

$$\delta_{NT}\left(\tilde{F}_{BER,t} - \tilde{\mathbf{H}}'\tilde{F}_t\right) = \tilde{\mathbf{U}}_{NT}^{-1}\left(\frac{\tilde{\mathbf{F}}'_{BER}\tilde{\mathbf{F}}}{T}\right)\frac{\delta_{NT}}{N}\sum_{i=1}^{N}\tilde{\Lambda}_i\tilde{\nu}_{it} + O_p(\tilde{R}\delta_{NT}^{-1})$$

and

$$\delta_{NT}(\tilde{\Lambda}_{BER,i} - \tilde{\mathbf{H}}^{-1}\tilde{\Lambda}_i) = \tilde{\mathbf{U}}_{NT}^{-1}\left(\frac{\tilde{\mathbf{F}}'_{BER}\tilde{\mathbf{F}}}{T}\right)\left(\frac{\tilde{\mathbf{\Lambda}}'\tilde{\mathbf{\Lambda}}}{N}\right)\frac{\delta_{NT}}{T}\sum_{i=1}^{N}\tilde{F}_t\tilde{\nu}_{it} + O_p(\tilde{R}\delta_{NT}^{-1}).$$

After a simple calculation,

$$\begin{aligned}\tilde{\Lambda}'_{BER,i}\tilde{F}_{BER,t} - P_\varepsilon(\lambda_i^{*\prime}f_t^*) =& \tilde{\Lambda}'_{BER,i}\tilde{F}_{BER,t} - \tilde{\Lambda}'_i\tilde{F}_t + \tilde{\omega}_{it}\\ &= \tilde{\Lambda}'_i\tilde{\mathbf{H}}'^{-1}(\tilde{F}_{BER,t} - \tilde{\mathbf{H}}'F_t) + F'_t\tilde{\mathbf{H}}(\tilde{\Lambda}_{BER,i} - \tilde{\mathbf{H}}^{-1}\tilde{\Lambda}_i)\\ &+ O_p(\tilde{R}\delta_{NT}^{-2}) + \tilde{\omega}_{it}.\end{aligned}$$

where $\tilde{\omega}_{it} = \tilde{\Lambda}'_i\tilde{F}_t - P_\varepsilon(\lambda_i^{*\prime}f_t^*)$. Combing all above equations together with the condition in Theorem 3, we obtain

$$\delta_{NT}\left(\tilde{\Lambda}'_{BER,i}\tilde{F}_{BER,t}-P_{\varepsilon}(\lambda_i^{*\prime}f_t^*)\right)=\frac{\delta_{NT}}{N}\tilde{\Lambda}'_i\left(\frac{\tilde{\mathbf{\Lambda}}'\tilde{\mathbf{\Lambda}}}{N}\right)^{-1}\sum_{i=1}^{N}\tilde{\Lambda}_i\tilde{\nu}_{it}+\frac{\delta_{NT}}{T}\tilde{F}'_t\tilde{\mathbf{H}}\tilde{\mathbf{H}}'\sum_{i=1}^{N}\tilde{F}_t\tilde{\nu}_{it}+o_p(1).$$

By the central limit theorem, we have

$$\left(\frac{1}{N}W_{it}^{(1)}+\frac{1}{T}W_{it}^{(2)}\right)\left(\hat{\Lambda}'_{BER,i}\hat{F}_{BER,t}-P_{\varepsilon}(\lambda_i^{*\prime}f_t^*)\right)\xrightarrow{d}N(0,1).$$

Reference

[1] Ahn, S. C., & Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. Econometrica, 81(3), 1203-1227.

[2] Bai, J. (2003). Inferential theory for factor models of large dimensions. Econometrica, 71(1), 135-171.

[3] Bai, J., & Li, K. (2016). Maximum likelihood estimation and inference for approximate factor models of high dimension. Review of Economics and Statistics, 98(2), 298-309.

[4] Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. Econometrica, 70(1), 191-221.

[5] Badoer, D. C., & Demiroglu, C. (2018). The relevance of credit ratings in transparent bond markets. The Review of Financial Studies, 32(1), 42-74.

[6] Baker, H. K., & Mansi, S. A. (2002). Assessing credit rating agencies by bond issuers and institutional investors. Journal of Business Finance & Accounting, 29(9-10), 1367-1398.

[7] Baker, M., Greenwood, R., & Wurgler, J. (2009). Catering through nominal share prices. The Journal of Finance, 64(6), 2559-2590.

[8] Basse, T., & Reddemann, S. (2013). Inflation and the dividend policy of us firms. , 37(1), 34-46.

[9] Bates, T. W., Kahle, K. M., & Stulz, R. M. (2009). Why do US firms hold so much more cash than they used to?. The journal of finance, 64(5), 1985-2021.

[10] Black, F., & Scholes, M. (1974). The effects of dividend yield and dividend policy on common stock prices and returns. Journal of financial economics, 1(1),

1-22.

[11] Boneva, L., & Linton, O. (2017). A discrete-choice model for large heterogeneous panels with interactive fixed effects with an application to the determinants of corporate bond issuance. Journal of Applied Econometrics, 32(7), 1226-1243.

[13] Chen, M., Fernández-Val, I., & Weidner, M. (2021). Nonlinear factor models for network and panel data. Journal of Econometrics, 220(2), 296-324.

[14] Dong, C., Linton, O., & Peng, B. (2021). A weighted sieve estimator for nonparametric time series models with nonstationary variables. Journal of Econometrics, 222(2), 909-932.

[15] Fama, E. F., & French, K. R. (2001). Disappearing dividends: changing firm characteristics or lower propensity to pay?. Journal of Financial economics, 60(1), 3-43.

[16] Forti, C. A. B., Peixoto, F. M., & Alves, D. L. (2015). Determinant Factors of Dividend Payments in Brazil. Revista Contabilidade & Financas, 26(68), 167-180.

[17] Gordon, M. J. (1963). Optimal investment and financing policy. The Journal of finance, 18(2), 264-272.

[18] Graham, J. R., Leary, M. T., & Roberts, M. R. (2015). A century of capital structure: The leveraging of corporate America. Journal of Financial Economics, 118(3), 658-683.

[19] Grinstein, Y., & Michaely, R. (2005). Institutional holdings and payout policy. The Journal of Finance, 60(3), 1389-1426.

[20] Grullon, G., Michaely, R., & Swaminathan, B. (2002). Are dividend changes a sign of firm maturity?. The journal of Business, 75(3), 387-424.

[21] Jo, H., & Pan, C. (2009). Why are firms with entrenched managers more likely to pay dividends?. Review of Accounting and Finance, 8(1), 87-116.

[22] Kania, S. L., & Bacon, F. W. (2005). What Factors Motivate The Corporate Dividen Policy.

[23] Kolenikov, S., & Angeles, G. (2009). Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer?. Review of Income and Wealth, 55(1), 128-165.

[24] Miller, M. H., & Scholes, M. S. (1982). Dividends and taxes: Some empirical evidence. Journal of Political Economy, 90(6), 1118-1141.

[25] Ng, S. (2015). Constructing common factors from continuous and categorical data. Econometric Reviews, 34(6-10), 1141-1171.

[26] Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. The Review of Economics and Statistics, 92(4), 1004-1016.

[27] Perez, M. F., Shkilko, A., & Sokolov, K. (2015). Factor models for binary financial data. Journal of Banking & Finance, 61, S177-S188.

[28] Vermunt, J. K. , & Magidson, J. . (2005). Factor analysis with categorical indicators: a comparison between traditional and latent class approaches. Circuits Systems & Signal Processing, 137(1), 1-18.

[29] Wang, F. (2022). Maximum likelihood estimation and inference for high dimensional generalized factor models with application to factor-augmented regressions. Journal of Econometrics, 229(1), 180-200.