科技部補助專題研究計畫成果報告
期末報告

近似因子模型的有效估計-經由懲罰最小平方法

計 畫 類 別 ： 個別型計畫
計 畫 編 號 ： MOST 103-2410-H-004-213-
執 行 期 間 ： 103年12月01日至104年11月30日
執 行 單 位 ： 國立政治大學國際經營與貿易學系

計 畫 主 持 人 ： 顏佑銘

報 告 附 件 ： 出席國際會議研究心得報告及發表論文

處 理 方 式 ：
1.公開資訊：本計畫可公開查詢
2.「本研究」是否已有嚴重損及公共利益之發現：否
3.「本報告」是否建議提供政府單位施政參考：否

中 華 民 國 105 年 02 月 15 日

中文摘要：近似因子模型及由其衍生出來的各種計量方法，目前被廣泛地應用在各種預測及經濟
分析上。究其原因，乃是近似因子模型可以幫助研究者有效地從大量相關變量中提取
對研究有用的訊息。在近似因子模型中，我們通常假設預測因子之間要有一定的共同
性。在這個計畫裏，我們將著重於有效地估計一種近似因子模型，其中的預測因子除
了受到共同性因素的影響之外，另外也受到一些非共同性因素，如不尋常的巨大異常
值的影響。以下我們列出本計畫會從事的工作項目:(1)我們將發展一個可行的計量方
法來估計這種近似因子模型，而該計量法方法將基於以下的假設:預測因子間的非共
同性因素的出現頻率非常的低; (2)在此假設下，我們將提出了一種懲罰最小平方估
計法(penalized least squares) 來同時分解並估計預測因子的共同及非共同性因
素; (3)為了解決這個估計問題，我們將會開發一個有效率及具彈性的演算程序，而
這項工作將有賴於一些最近提出的優化方法; (4)之後我們會經由大量的蒙地卡羅模
擬，來比較我們所提出的方法和傳統的主成分分析法，在有限樣本下，何者比較能有
效地估計這種近似因子模型; (5)最後我們會將我們所提出的方法用於預測重要總體
經濟指標的年成長率及探討隱性變量如何影響橫斷面預期資產收益率。

中文關鍵詞：近似因子模型；主成分分析法；懲罰最小平方估計法；預測；隱性變量

英文摘要：Approximate factor models and their extensions are widely used in forecasting
and economic analysis due to their ability to extracting useful information from
a large number of relevant variables. In these models, candidate predictors are
typically subject to some common components. In this project, we will focus on
e ciently estimating an approximate factor model in which the candidate predic-
tors are additionally subject to idiosyncratic large uncommon components such as
jumps or outliers. We outline our plan for the project as follows: (1) We will de-
velop a viable econometric method to estimate such an approximate factor model.

The econometric method will be based on the assumption that occurrences of the
uncommon components are rare; (2) Under this assumption, we will propose a
penalized least squares estimation procedure to simultaneously disentangle and
estimate the common and uncommon components; (3) To solve the estimation
problem, we will develop an e cient and
exible algorithm, which will rely on
some recently developed optimization methods; (4) We will conduct an intensive
Monte-Carlo simulation study to compare nite-sample e ciency of the proposed
method and traditional PCA method; (5) We also will demonstrate performances
of the proposed method with empirical applications on forecasting yearly growths
of important macroeconomic indicators and investigating how latent factors a ect
cross sectional expected asset returns.

英文關鍵詞： Approximate Factor Model; PCA; Penalized Least Squares; Forecast; Latent Factors

Efficient Estimations of Approximate Factor Models via Penalized Leases Squares

February 15, 2016

Abstract

Approximate factor models and their extensions are widely used in risk evaluations due to their ability to extracting useful information from a large number of relevant variables. In these models, candidate predictors are typically subject to some common components. In this paper we evaluated risks by proposing a new method for robustly estimating the approximate factor models. We considered a class of approximate factor models in which the candidate predictors are additionally subject to idiosyncratic large uncommon components such as jumps or outliers. By assuming that occurrences of the uncommon components are rare, we developed an estimation procedure to simultaneously disentangle and estimate the common and uncommon components. Through intensive simulations, we compared the proposed method and traditional PCA method in terms of their finite-sample efficiency. We then used the proposed method to investigate whether risks from the latent factors are priced for expected returns of Fama and French 100 size and book-to-market ratio portfolios. We found that while the risk from the common factor is priced for the 100 portfolios, the risks from the idiosyncratic factors are not. However, we also found that model uncertainty risks of the idiosyncratic factors are priced, suggesting that with effective diversifications, only the predictable idiosyncratic risks can be reduced, but the unpredictable ones may still exist.

KEYWORDS: Approximate Factor Model, PCA, Norm Penalty, Common factor, Idiosyncratic Risk

1 Introduction

Approximate factor models and their extensions are widely used in economic analysis and forecasting due to their ability to extracting useful information from a large number of relevant variables (Stock and Watson, 2002; Bernanke et al., 2005; Ludvigson and Ng, 2009). In these models, data generating processes are often specified as a linear combination of relevant common factors and error terms. Estimating these models can pose some difficulties when the relevant common factors are unobservable. An important goal for estimating such models is therefore to identify the latent common factors and their factor loadings. Popular methods for carrying out this estimation task include the maximum likelihood method (MLE), Markov Chain Monte Carlo (MCMC) and Principal Component Analysis (PCA). Nevertheless researchers in econometrics often estimate these models with high dimensional data. Therefore the PCA method, which is less computational intensive than MLE and MCMC, is often more preferred in practice.

Although the PCA method has a computational advantage, it is widely known that the method may fail to yield accurate estimations of the latent factors and factor loadings when large idiosyncratic uncommon components are present in the data (Jolliffe, 2002). To overcome this difficulty, we propose a simple and efficient method for estimating latent factors and factor loadings. This estimation method aims to reduce estimation biases in the latent common factors and their factor loadings by simultaneously disentangling and estimating the common and uncommon components. To develop this method, we first formulate the estimation problem as a penalized least squares problem in which a norm penalty function is imposed on the uncommon components. We then solve the estimation problem by building an algorithm to iteratively solve a principal component analysis (PCA) problem and a one dimensional shrinkage estimation problem. The algorithm can flexibly incorporate with the methods for selecting the number of common components. We call the proposed estimation method the P-PCA method (*Penalized least squares plus PCA method*).

Recently many different approximate factor models and their corresponding estimation procedures were developed. Moench et al. (2013) proposed a multilevel factor model for large panel data with between-block variations and idiosyncratic noise, and developed an estimation procedure that can both separate block-level shocks and genuinely common factors and achieve dimension reduction. Ando and Bai (2013) proposed a multifactor model for data with a large number of observable factors and unobservable common and group-specific pervasive factors. Their estimation procedure for the model can simultaneously select relevant observable factors and determine the number of common and group-specific unobservable factors. Cheng et al. (2014) proposed a factor model in which both factor loadings and number of factors can have a behavior of structure break. They adopted a shrinkage estimator that can simultaneously and consistently estimate the number of common factors before and after the structure break. Their estimation procedure is carried out by solving a convex optimization with principal components of the data matrix as its inputs.

A main difference between the aforementioned research and our proposed estimation method is that we develop our method by considering a data generating process in which certain large idiosyncratic uncommon components are present. This means that in the data generating process observations are occasionally blurred by extremely large noises such as asset price jumps. They are not broken by a permanent change of common factors or factor loadings. Indeed, under suitable assumptions on the idiosyncratic uncommon components, one can still estimate the factors and factor loadings consistently by using the PCA method (Bai and Ng, 2002; Stock and Watson, 2002). However, through intensive simulations we show that the proposed P-PCA estimation procedure can outperform the PCA method in term of finite sample efficiency when estimating model parameters under a wide range of model settings. In addition, we discuss how the proposed method can be used to deal with a more general data structure, such as panel data. Throughout these works, we believe the proposed method can serve as a complementary tool for robust estimations rather than a competitive approach to those established approximate factor models.

Next we apply the proposed method to investigate whether risks from the latent factors are priced for expected returns of Fama and French 100 size and book-to-market ratio portfolios. Recently Ando and Bai (2013) analyzed possible latent common and group-specific pervasive factors of the expected stock returns in China stock market. They assumed that expected returns of stocks traded in the same exchanges are governed by the same latent within-group common factors. They found the expected returns of stocks traded in different stock exchanges are affected by different observable and latent factors. Unlike their research, we focus on how both the latent common and idiosyncratic, uncommon factors affect cross sectional expected returns. Our analysis relies on decomposing the noise term in the Fama and French three factor model into latent common and idiosyncratic factors. We find that for the 100 portfolios, risk from the common factor is priced but risks from the idiosyncratic factors are not. The latter result is consistent with Arbitrage Pricing Theory: The idiosyncratic risk of a well diversified portfolio should be negligible. It should not have an effect on expected return of the portfolio. However, we also find that model uncertainty risks of the idiosyncratic factors are priced. The result implies that with effective diversifications, only the predictable idiosyncratic risks can be reduced, but the unpredictable ones may still exist.

The rest of paper is organized as follows. In Section 2 we review the PCA method and then introduce the P-PCA estimation method. We next discuss how to select number of the latent common factors in our estimation procedure. In Section 3 we report simulation results. In Section 4 we discuss some possible extensions of our proposed method. In Section 5 we perform empirical applications. Section 6 is the conclusion.

2 Methodology

In this Section we describe a method for estimating an approximate factor model in which the candidate predictors are subject to idiosyncratic uncommon components. Specifically we assume the N dimensional time series of candidate predictors $\mathbf{X}_t$ and the variable to be forecast Y_t are subject to the following data generating process:

$$\mathbf{X}_t = \mathbf{\Lambda}\mathbf{F}_t + \mathbf{J}_t + \mathbf{e}_t, \tag{1}$$

$$Y_{t+h} = \boldsymbol{\beta}_F^{\mathbf{T}}\mathbf{F}_t + \boldsymbol{\beta}_W^{\mathbf{T}}\mathbf{W}_t + \varepsilon_{t+h}, \tag{2}$$

for $t = 1, \ldots, T$, where $\mathbf{\Lambda}$ is an $N \times r$ factor loading matrix, $\mathbf{F}_t$ is an $r \times 1$ vector of latent factors, $\mathbf{e}_t$ is an $N \times 1$ vector of measurement errors, and $\mathbf{J}_t$ is an $N \times 1$ vector of the idiosyncratic uncommon components. In addition, $\boldsymbol{\beta}_F$ is an $r \times 1$ vector of regression coefficients corresponding to latent factor $\mathbf{F}_t$, $\mathbf{W}_t$ is an $m \times 1$ vectors of observable exogenous variables, and $\boldsymbol{\beta}_W$ is an $m \times 1$ vector of regression coefficients corresponding to $\mathbf{W}_t$. The index h is the forecast horizon, and Y_{t+h} and ε_{t+h} are the variable to be forecast and error term h periods ahead, respectively. The setting is similar to the dynamic factor model considered in Stock and Watson (2002) except $\mathbf{X}_t$ has an additional idiosyncratic uncommon component $\mathbf{J}_t$. By assuming that occurrences of the uncommon components are rare, $\mathbf{J}_t$ is generically a sparse vector, i.e. some of its elements are zero, and practically it can be viewed as a jump or outlier in $\mathbf{X}_t$.

Below we review the PCA method for estimating the latent factors $\mathbf{F}_t$ and factor loadings $\mathbf{\Lambda}$. Define $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_T)^{\mathbf{T}}$, $\mathbf{F} = (\mathbf{F}_1, \ldots, \mathbf{F}_T)^{\mathbf{T}}$ and $\mathbf{J} = (\mathbf{J}_1, \ldots, \mathbf{J}_T)^{\mathbf{T}}$. Suppose $N > T$ and the number of factors r is known. Without the term $\mathbf{J}_t$, we can estimate the factor matrix $\mathbf{F}$ and factor loading matrix $\mathbf{\Lambda}$ by solving the following optimization problem:

$$\min_{\mathbf{F},\mathbf{\Lambda}} \frac{1}{TN} \left\|\mathbf{X} - \mathbf{F}\mathbf{\Lambda}^{\mathbf{T}}\right\|_F^2, \text{ subject to } \frac{\mathbf{F}^{\mathbf{T}}\mathbf{F}}{T} = \mathbf{I}_r. \tag{3}$$

Here $\|.\|_F$ is the Frobenius norm. The optimization problem (3) is closely related to the principal component analysis (PCA). The estimated factor matrix $\hat{\mathbf{F}}$ can be obtained by multiplying $\sqrt{T}$ with a matrix containing the eigenvectors corresponding to the largest r eigenvalues of the $T \times T$ matrix $\mathbf{X}\mathbf{X}^\mathbf{T}$. Given $\hat{\mathbf{F}}$, the estimated factor loading matrix $\hat{\boldsymbol{\Lambda}}$ can be obtained by using the least squares method, i.e. $\hat{\boldsymbol{\Lambda}} = \left(\left(\hat{\mathbf{F}}^\mathbf{T}\hat{\mathbf{F}}\right)^{-1}\hat{\mathbf{F}}^\mathbf{T}\mathbf{X}\right)^\mathbf{T} = \mathbf{X}^\mathbf{T}\hat{\mathbf{F}}/T$. On the other hand, when $T \geq N$, we can estimate the factor and factor loading matrices by solving the problem (3) with the constraint $\mathbf{F}^\mathbf{T}\mathbf{F}/T = \mathbf{I}_r$ replaced by $\boldsymbol{\Lambda}^\mathbf{T}\boldsymbol{\Lambda}/N = \mathbf{I}_r$. In this situation the estimated factor loading matrix $\bar{\boldsymbol{\Lambda}}$ can be obtained by multiplying $\sqrt{N}$ with a matrix containing the eigenvectors corresponding to the largest r eigenvalues of the $N \times N$ matrix $\mathbf{X}^\mathbf{T}\mathbf{X}$. Given $\bar{\boldsymbol{\Lambda}}$, the estimated factor matrix $\bar{\mathbf{F}}$ can be obtained by using the least squares method, i.e. $\bar{\mathbf{F}} = \left(\left(\bar{\boldsymbol{\Lambda}}^\mathbf{T}\bar{\boldsymbol{\Lambda}}\right)^{-1}\bar{\boldsymbol{\Lambda}}^\mathbf{T}\mathbf{X}^\mathbf{T}\right)^\mathbf{T} = \mathbf{X}\bar{\boldsymbol{\Lambda}}/N$. Now define $\mathbf{Z} = \mathbf{F}\boldsymbol{\Lambda}^\mathbf{T}$, $\hat{\mathbf{Z}} = \hat{\mathbf{F}}\hat{\boldsymbol{\Lambda}}^\mathbf{T}$ and $\bar{\mathbf{Z}} = \bar{\mathbf{F}}\bar{\boldsymbol{\Lambda}}^\mathbf{T}$. Here the matrices $\hat{\mathbf{Z}}$ and $\bar{\mathbf{Z}}$ can be viewed as low rank approximations to the matrix $\mathbf{X}$. It is known that $\hat{\mathbf{Z}} = \bar{\mathbf{Z}}$, and hence the objective function $\left\|\mathbf{X} - \mathbf{F}\boldsymbol{\Lambda}^\mathbf{T}\right\|_F^2 / (TN)$ has the same value under the two optimal solutions $\left(\hat{\mathbf{F}}, \hat{\boldsymbol{\Lambda}}\right)$ and $\left(\bar{\mathbf{F}}, \bar{\boldsymbol{\Lambda}}\right)$.

2.1 Penalized Least Squares plus PCA method

If the idiosyncratic uncommon component vector $\mathbf{J}_t$ is in the data generating process (1), estimating $\mathbf{F}$ and $\boldsymbol{\Lambda}$ may loss efficiency when the PCA method is directly applied to the matrix $\mathbf{X}\mathbf{X}^\mathbf{T}$ or $\mathbf{X}^\mathbf{T}\mathbf{X}$. In this situation if $\mathbf{J}$ is known, we can obtain better estimations of $\mathbf{F}$ and $\boldsymbol{\Lambda}$ by applying the PCA method to the matrix $\mathbf{C}\mathbf{C}^\mathbf{T}$ (or $\mathbf{C}^\mathbf{T}\mathbf{C}$), where $\mathbf{C} = \mathbf{X} - \mathbf{J}$. If $\mathbf{J}$ is unknown, we can estimate $\mathbf{F}$ and $\boldsymbol{\Lambda}$ by first obtaining an estimate $\hat{\mathbf{J}}$ for $\mathbf{J}$ and then applying the PCA method to the matrix $\hat{\mathbf{C}}\hat{\mathbf{C}}^\mathbf{T}$ (or $\hat{\mathbf{C}}^\mathbf{T}\hat{\mathbf{C}}$), where $\hat{\mathbf{C}} = \mathbf{X} - \hat{\mathbf{J}}$ is a matrix that disentangles estimate $\hat{\mathbf{J}}$ from X.

Our strategy to estimate $\mathbf{J}$ is to use the property that $\mathbf{J}$ is assumed to be a sparse matrix. To obtain an estimate for $\mathbf{J}$, a common method is to impose an l_p norm penalty on $\mathbf{J}$, where $0 \leq p \leq 1$. In the following, we focus on the l_1 norm penalty function.

Suppose $N > T$ and the number of factors r is known. In our method, we estimate $\mathbf{F}$, $\boldsymbol{\Lambda}$ and $\mathbf{J}$ by solving the following penalized l_1 norm optimization problem:

$$\min_{\mathbf{F},\boldsymbol{\Lambda},\mathbf{J}} \frac{1}{TN}\left\|\mathbf{X} - \mathbf{F}\boldsymbol{\Lambda}^\mathbf{T} - \mathbf{J}\right\|_F^2 + \frac{\delta}{TN}\|\mathbf{J}\|_1, \text{ subject to } \frac{\mathbf{F}^\mathbf{T}\mathbf{F}}{T} = \mathbf{I}_r, \tag{4}$$

where $\delta \in \mathbb{R}^+$ is a penalty parameter, and $\|\mathbf{J}\|_1$ is the sum of the absolute value of each element in $\mathbf{J}$.

The l_1 norm penalty is perhaps the most frequently used norm penalty function in sparse estimations. Examples include the lasso of Tibshirani (1996) and the robust PCA method of Candès et al. (2011). The l_1 norm penalty is a convex function of $\mathbf{J}$. This convex property makes the modified estimation problem tractable even when T and N become very large.

We call the proposed estimation method the P-PCA method (Penalized least squares plus PCA method). To solve the estimation problem (4), we propose an algorithm, which iteratively solves a principal component analysis (PCA) problem and a one dimensional shrinkage estimation problem. In Appendix A, we provide a step-by-step description on the algorithm when the number of factors r is known and the penalty parameter δ is fixed. Here we summarize the algorithm as follows.

Algorithm 1 Robust Approximate Factor Model Estimation with l_1 Norm Penalty

---

Input: The data matrix $\mathbf{X}$, the penalty penalty parameter δ, tolerance ϵ, the number of factors r and the maximum iteration k_{max}.

Output: $\hat{\mathbf{F}}$, $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{J}}$.

1: Set $\mathbf{J}^{(0)} = \mathbf{0}$ and $\delta > 0$.

2: **for** $k = 1$ to k_{max} **do**

3: Given $\mathbf{C}^{(k-1)} = \mathbf{X} - \mathbf{J}^{(k-1)}$, where $\mathbf{J}^{(k-1)} = (\mathbf{J}_1^{(k-1)}, \ldots, \mathbf{J}_T^{(k-1)})^{\mathbf{T}}$, compute $\left(\mathbf{F}^{(k)}, \mathbf{\Lambda}^{(k)}\right)$ by using the PCA method and least squares method.

4: Given $\mathbf{L}^{(k)} = \mathbf{X} - \mathbf{Z}^{(k)}$, where $\mathbf{Z}^{(k)} = \mathbf{F}^{(k)}\mathbf{\Lambda}^{(k)\mathbf{T}}$, update $J_{it}^{(k)}$ of $\mathbf{J}_t^{(k)}$ for $t = 1, \ldots, T$ as follows:

$$J_{it}^{(k)} = ST\left(L_{it}^{(k)}, \frac{\delta}{2}\right).$$

Here $ST(x, y) := sign(x)(|x| - y)_+$ is the soft thresholding function and $L_{it}^{(k)}$ is the (i, t)th entry of $\mathbf{L}^{(k)}$.

5: **if** $\left\|\mathbf{Z}^{(k)} - \mathbf{Z}^{(k-1)}\right\|_F / \left\|\mathbf{Z}^{(k-1)}\right\|_F \leq \epsilon$ **then**

6: **break**

7: **end if**

8: **end for**

9: Set output $\hat{\mathbf{F}} = \mathbf{F}^{(k)}$, $\hat{\mathbf{\Lambda}} = \mathbf{\Lambda}^{(k)}$ and $\hat{\mathbf{J}} = \mathbf{J}^{(k)}$.

To implement Algorithm 1, we assume the number of common factors r should be known. However, it is easy to incorporate methods for estimating r into step 3 of the Algorithm when using the PCA and least squares method on estimating the factors and factor loadings (see Section 2.2 for methods for estimating r). The Algorithm works well in our simulations and empirical application (see Section 3 to 5). In Appendix B, we provide some analysis on convergence of the algorithm and show that in each iteration, the Algorithm indeed decreases the objective function in (4).

2.2 Selecting the Number of Factors

So far we assumed the number of factors r is known when estimating $\mathbf{F}$ and $\mathbf{\Lambda}$, but such assumption in general does not hold in real data applications. When r is unknown, several methods have been proposed to estimate it. These methods rely on either minimizing certain loss functions (Bai and Ng, 2002; Alessi et al., 2010) or using test statistics constructed from eigenvalues of the (transformed) data matrix $\mathbf{X}$ (Onatski, 2009, 2010; Ahn and Horenstein, 2013). Since our method separately estimates $(\mathbf{F}, \mathbf{\Lambda})$ and $\mathbf{J}$, these methods can be easily incorporated into our method.

In simulations (Section 3) and empirical applications (Section 5), we use the IC_p criteria proposed by Bai and Ng (2002) to estimate the number of common factors r. Suppose $N > T$ and let $\hat{\mathbf{F}}(r)$ be the estimated $T \times r$ factor matrix (a matrix containing $\sqrt{T}$ times the first r eigenvectors of $\hat{\mathbf{C}}\hat{\mathbf{C}}^{\mathbf{T}}$). Further let

$$V\left(r, \hat{\mathbf{F}}(r)\right) = \min_{\mathbf{\Lambda}} \frac{1}{NT} \left\| \hat{\mathbf{C}} - \hat{\mathbf{F}}(r)\mathbf{\Lambda}^{\mathbf{T}} \right\|_F^2 .$$

The three IC_p criteria of Bai and Ng (2002) are defined as:

$$\begin{aligned} IC_{p^1}(r) &= \ln V\left(r, \hat{\mathbf{F}}(r)\right) + r\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right), \\ IC_{p^2}(r) &= \ln V\left(r, \hat{\mathbf{F}}(r)\right) + r\left(\frac{N+T}{NT}\right)\ln\left(\min(N,T)\right), \\ IC_{p^3}(r) &= \ln V\left(r, \hat{\mathbf{F}}(r)\right) + r\left(\frac{\ln\left(\min(N,T)\right)}{\min(N,T)}\right). \end{aligned}$$

Define $\hat{r}_i = \arg\min_r IC_{p^i}(r)$ for $i = 1, 2$ and 3. We use $\hat{r}_{BN} = \min(\hat{r}_1, \hat{r}_2, \hat{r}_3)$ as an estimate for r. In each iteration of the algorithm, we implement the method of IC_p criteria. We then use $\hat{\mathbf{F}}(\hat{r}_{BN})$, which is a matrix containing the first $\hat{r}_{BN}$ eigenvectors of $\hat{\mathbf{C}}\hat{\mathbf{C}}^{\mathbf{T}}$, and $\hat{\mathbf{\Lambda}}(\hat{r}_{BN}) = \hat{\mathbf{C}}^{\mathbf{T}}\hat{\mathbf{F}}(\hat{r}_{BN})/T$ as the estimated factor and factor loading matrices.

2.3 Setting the Penalty Parameter

The penalty parameter δ is important for carrying out the P-PCA method. However, there is no rule of thumb to specify it. Here we set the penalty parameter equal to $\delta^{naive} = \bar{\hat{\sigma}}\sqrt{8\ln T}$, where $\bar{\hat{\sigma}} = N^{-1}\sum_{i=1}^{N}\hat{\sigma}_i$ and $\hat{\sigma}_i$ is sample standard deviation of the residuals $\hat{e}_{it}$ under the PCA method. We see δ^{naive} as a "naive" choice of δ due to its simplicity. The naive setting, however, works well over entire simulations. We give some intuitions on why δ^{naive} works well. The idea comes from an observation that an ideal loss for estimating J_{it} can be attained when the soft thresholding estimator with a proper value of δ is used (Donoho and Jonhstone, 1994). Let $\omega_{it} = J_{it} + e_{it}$ denote the part of uncommon idiosyncratic component and error term in data X_{it}, and assume each e_{it} is i.i.d. normally distributed with mean zero and variance σ^2. Consider estimating J_{it} with either ω_{it} or 0. An ideal mean squared loss of an estimator over $t = 1, \ldots, T$ is given by $Loss_i^{oracle} = \sum_{t=1}^{T} \min\left(J_{it}^2, \sigma^2\right)$ when $|J_{it}| > \sigma$ is known. Without such information, however, it can be shown that the mean squared loss of $\tilde{J}_{it}$ over $t = 1, \ldots, T$ can still approach to $Loss_i^{oracle}$ if $\tilde{J}_{it} = ST\left(\omega_{it}, \sigma\sqrt{2\ln T}\right)$ is used in estimating J_{it} (Donoho and Jonhstone, 1994; Wasserman, 2006, pp.172). Let $\hat{L}_{it}$ denote the quantity calculated by subtracting X_{it} from the product of the estimated common factors and factor loadings. Note that estimating J_{it} with $\hat{J}_{it} = ST\left(\hat{L}_{it}, \sigma\sqrt{2\ln T}\right)$ is equivalent to setting $\delta = \sigma\sqrt{8\ln T}$ in the P-PCA method. Now if the common factors and factor loadings are accurately estimated, then $\hat{L}_{it} \approx \omega_{it}$ and $\hat{J}_{it} \approx \tilde{J}_{it}$. Then over $t = 1, \ldots, T$, the mean squared loss of $\hat{J}_{it}$ will approximate the mean squared loss of $\tilde{J}_{it}$, and therefore approaches to $Loss_i^{oracle}$.

Even though some technical conditions need to be satisfied in order to guarantee performance of the naive setting δ^{naive}, simulation studies show it works well with the P-PCA method under various data generating processes. In addition, due to its simplicity, using the naive setting also avoids intensive computations. Theoretically the naive setting may not be the best choice, however, it serves as an easily implemented guidance and a benchmark for further adjustments for obtaining the best results.

3 Simulation Results

We describe the data generating process as follows. First, (1) and (2) serve as the main mechanism for generating the data in our simulation studies. In addition, the following settings are adopted to generate the data:

- $T = N = 50$, 100, 200 and 400, and $r = 5$.
- $J_{it} \sim i.i.d.\ Pois\left(\nu\right) \times \mathcal{N}\left(0, \sigma_J^2\right)$, where $\nu = 0$, 0.01, 0.05 and 0.1, $\sigma_J = 5 \times \sqrt{\theta}$ and $\theta = r$.
- The number of columns of $\mathbf{X}$ that have the idiosyncratic jump components is equal to $\lfloor a \times N \rfloor$ with $a = 0$, 0.1, 0.5 and 1. The $\lfloor a \times N \rfloor$ columns are randomly chosen from the N columns of $\mathbf{X}$ without replacement.
- $\boldsymbol{\beta}_F = (1, \ldots, 1)$ with $\dim\left(\boldsymbol{\beta}_F\right) = r \times 1$ and $\boldsymbol{\beta}_W = \mathbf{0}$.

- $\varepsilon_{t+1} \sim i.i.d.\ \mathcal{N}(0,1)$ for $t = 1, \ldots, T$.

To generate $\mathbf{F}_t$, $\mathbf{\Lambda}$ and $\mathbf{e}_t$, the following five models are considered:

1. **Model 1 (i.i.d. error)**:

   - $\mathbf{F}_t \sim i.i.d.\ \mathcal{N}(\mathbf{0}, \mathbf{I}_{rr})$ and $\lambda_{ij} \sim i.i.d.\ \mathcal{N}(0,1)$ for $i = 1, \ldots, N$ and $j = 1, \ldots, r$.
   - $\mathbf{e}_t \sim i.i.d.\ \mathcal{N}(\mathbf{0}, \theta \times \mathbf{I}_{NN})$, where $\theta = r$.

2. **Model 2 (AR(1) error)**:

   - $\mathbf{F}_t \sim i.i.d.\ \mathcal{N}(\mathbf{0}, \mathbf{I}_{rr})$ and $\lambda_{ij} \sim i.i.d.\ \mathcal{N}(0,1)$ for $i = 1, \ldots, N$ and $j = 1, \ldots, r$.
   - $e_{it} = \sqrt{\theta} u_{it}$, where $u_{it} = 0.5 u_{it-1} + v_{it}$, $v_{it} \sim i.i.d.\ \mathcal{N}(0,1)$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, and $\theta = r$.

3. **Model 3 (Large break Model, Bates et al. (2013))**:

   - $\mathbf{F}_t \sim i.i.d.\ \mathcal{N}(\mathbf{0}, \mathbf{I}_{rr})$.
   - For the factor loading λ_{ij}, we first randomly select a subset B of i (without replacement) with size $4\sqrt{N}$. If $i \notin B$, we set the factor loading $\lambda_{ij} = (0.4/0.45) \times \bar{\lambda}_{ij} \tilde{\lambda}_i$, where $\bar{\lambda}_{ij} \sim i.i.d.\ \mathcal{N}(0,1)$ and $\tilde{\lambda}_i \sim i.i.d.\ U(0.1, 0.8)$. If $i \in B$, we set
   $$\lambda_{ij}^{LB} = \begin{cases} \lambda_{ij} & \text{for } t \leq \lfloor 0.5T \rfloor, \\ \lambda_{ij} + \triangle_j & \text{for } t > \lfloor 0.5T \rfloor, \end{cases}$$
   where $\triangle_j \sim i.i.d.\ \mathcal{N}(0, 0.16)$ for $j = 1, \ldots, r$.
   - $\mathbf{e}_t \sim i.i.d.\ \mathcal{N}(\mathbf{0}, \theta \times \mathbf{I}_{NN})$, where $\theta = r$.

4. **Model 4 (Cross sectionally correlated error)**:

   - $\mathbf{F}_t \sim i.i.d.\ \mathcal{N}(\mathbf{0}, \mathbf{I}_{rr})$ and $\lambda_{ij} \sim i.i.d.\ \mathcal{N}(0,1)$ for $i = 1, \ldots, N$ and $j = 1, \ldots, r$.
   - $e_{it} = \sqrt{\theta} u_{it}$, where $u_{it} = v_{it} + \sum_{l=i-L, l\neq i}^{i+L} (0.5)^{|l-i|} v_{lt}$, $v_{it} \sim i.i.d.\ \mathcal{N}(0,1)$ and $\theta = r$.
   - $L = \max(N/20, 10)$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

5. **Model 5 (AR(1) factor)**:

   - $\lambda_{ij} \sim i.i.d.\ \mathcal{N}(0,1)$ for $i = 1, \ldots, N$ and $j = 1, \ldots, r$.

- $F_{it} = 0.5F_{it-1} + v_{f,it}$, where $v_{f,it} \sim i.i.d.\ \mathcal{N}(0,1)$.
- $\mathbf{e}_t \sim i.i.d.\ \mathcal{N}(\mathbf{0}, \theta \times \mathbf{I}_{NN})$, where $\theta = r$.

Some concerns on using IC_p in the simulations should be addressed. In Lemma 2 of Amengual and Watson (2007), it was showed that when running the PCA method with the noise-contained data $\widetilde{X}_{it} = X_{it} + w_{it}$, where w_{it} is an additive error, if $(NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} w_{it}^2 = O_p\left(C_{NT}^{-2}\right)$, where $C_{NT} = \min\left(\sqrt{N}, \sqrt{T}\right)$, then information criteria IC_p proposed by Bai and Ng (2002) can still consistently estimate the number of principal components. Without the jump terms, the data generated from Model 3 can be viewed as such noise-contained data, and the noise w_{it} has the following form

$$w_{it} = \begin{cases} 0, & \text{if } i \notin B, \\ 0, & \text{if } i \in B \text{ and } t \leq \lfloor 0.5T \rfloor, \\ \sum_{j=1}^{r} F_{it} \triangle_j, & \text{if } i \in B \text{ and } t > \lfloor 0.5T \rfloor. \end{cases}$$

Note that if $i \in B$ and $t > \lfloor 0.5T \rfloor$, then $w_{it}^2 = \left(\sum_{j=1}^{r} F_{it} \triangle_j\right)^2 = O(1)$. Thus by setting $|B| = O\left(\sqrt{N}\right)$,

$$\begin{aligned} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} w_{it}^2 &= \frac{1}{NT} \sum_{t=\lfloor 0.5T \rfloor + 1}^{T} \sum_{i \in B} w_{it}^2 \\ &= O_p\left(N^{-\frac{1}{2}}\right), \end{aligned}$$

which has a rate greater than $O_p\left(C_{NT}^{-2}\right)$ when $N \asymp T$. It is still unknown whether the rate $O_p\left(C_{NT}^{-2}\right)$ can be improved (Bai and Ng, 2002). Hence for Model 3, in order to obtain a fair comparison, we will assume the number of factors is known ($\hat{r} = 5$). We will not use IC_p to estimate the number of factors.

We report the following four performance measures:

1. Distance correlation (DCOR): The performance measure is proposed by Szekely et al. (2007). It measures dependence between two sets of random vectors and has a range from zero to one. The higher (lower) the DCOR is, the higher (lower) the dependence between the two sets of random vectors will be. We apply the measure to gauge dependence between the true factors $\mathbf{F}$ and estimated factors $\hat{\mathbf{F}}$. Since PCA and P-PCA can only identify the factors up to a change of sign of the true factors, using a measure of dependence between $\mathbf{F}$ and $\hat{\mathbf{F}}$ is reasonable.

2. Squared predictive error: It is defined as $\left(\hat{y}_{T+1|T} - \widetilde{y}_{T+1|T}\right)^2$, where $\hat{y}_{T+1|T} = \hat{\beta}_F \hat{\mathbf{F}}_T$ and $\widetilde{y}_{T+1|T} = \widetilde{\beta}_F \mathbf{F}_T$. $\hat{\beta}_F$ $\left(\widetilde{\beta}_F\right)$ is obtained by regressing y_t onto $\hat{\mathbf{F}}_t$ (y_t onto $\mathbf{F}_t$) using the least squares method.

3. Trace R^2 between the true factors $\mathbf{F}$ and estimated factors $\hat{\mathbf{F}}$ (Stock and Watson, 2002): It is defined as

$$R^2_{\hat{\mathbf{F}},\mathbf{F}} = \frac{\operatorname{avg}\left(\left\|\mathbf{F}\left(\mathbf{F}^{\mathbf{T}}\mathbf{F}\right)^{-1}\mathbf{F}^{\mathbf{T}}\hat{\mathbf{F}}\right\|_F^2\right)}{\operatorname{avg}\left(\left\|\hat{\mathbf{F}}\right\|_F^2\right)}.$$

4. Proportion of better performances in the 2000 simulations: This measure is defined as the proportion that the P-PCA method has a higher DCOR (or a lower squared predictive error) than the PCA method has in the 2000 simulations.

Figure 1 to 4 show the simulation results. In each figure, plots from top to bottom correspond to Model 1 to Model 5, respectively, and plots from left to right correspond to different T and N, respectively. In Figure 1 and Figure 2, we show averages of the DCOR between the true and estimated factors and averages of the squared predictive errors together with their 99% confidence intervals obtained from the 2000 simulations. In Figure 3 and Figure 4 we show trace R-Square and proportion that the P-PCA method has a higher DCOR or a lower squared predictive error than the PCA method among the 2000 simulations. The x-axis of each plot represents (a, ν), where $a = 0, 0.1, 0.5$ and $\nu = 0, 0.01, 0.05$. The parameters (a, ν) are used to control the proportion of entries in the data matrix $\mathbf{X}$ that have nonzero idiosyncratic jump components. We sort (a, ν) in the x-axis according to the value of $a \times \nu$.

From Figure 1 to 3, we can see that the P-PCA method outperforms the PCA method in terms of the DCOR, squared predictive error and trace R-Square in almost all cases in which the nonzero idiosyncratic jump components are present (both a and ν are not zeros). Furthermore, in 4 we can see that when the uncommon components are present ($(a, \nu) \neq (0, 0)$), a large proportion of the results show that the estimated factors from the P-PCA method can have a higher DCOR with the true factors than those from PCA method. This suggests that with a high probability, the P-PCA can produce more accurate estimations for the factors. Given that the uncommon components are present in the data, when T, N and $a \times \nu$ increase, estimated factors from the P-PCA method also can have a better chance of producing a lower predictive error.

However, the performance measures vary a lot with (a, ν), T and N. Given T and N, the performance of the P-PCA method becomes worse as $a \times \nu$ increases. In contrast, given (a, ν), the performance of the P-PCA method becomes better as T and N increase. Except for some cases, the patterns of the performance measures against (a, ν) are consistent over different model settings. The performance of the P-PCA method becomes worse as the proportion of entries in $\mathbf{X}$ that have the idiosyncratic components increases. However, the performance turns better as the sample size T and the number of variables N increase.

4 Some Extensions

The proposed P-PCA method may be further applied to the data when X_{it} has a more complex generating process. For example, consider that X_{it} has a form as

$$\begin{aligned} X_{it} &= U_{it}^{\mathbf{T}}\boldsymbol{\beta}_U + \mathbf{F}_t^{\mathbf{T}}\boldsymbol{\lambda}_i + J_{it} + e_{it} \\ &= X'_{it} + J_{it}, \end{aligned} \tag{5}$$

where U_{it} is a $p \times 1$ vector of observable variables and $\boldsymbol{\beta}_U$ is a $p \times 1$ vector of coefficients. The above data generating process for X'_{it} (X_{it} without J_{it}) has been considered in Bai (2009). Suppose now $U_{it} = 1$ and $\boldsymbol{\beta}_U = \mu$, then

$$X_{it} = \mu + \mathbf{F}_t^{\mathbf{T}}\boldsymbol{\lambda}_i + J_{it} + e_{it}. \tag{6}$$

In (6), with an additional assumption that μ is constant over time and $\bar{\mathbf{F}} = T^{-1}\sum_{t=1}^{T}\mathbf{F}_t = \mathbf{0}$, we can demean the data X_{it} to eliminate the term μ. The demeaned data will not have the term μ but still will contain $\left(\mathbf{F}_t - \bar{\mathbf{F}}\right)$ as the factor term, the (demeaned) jump and the (demeaned) error terms. The demeaned data thus have a form of pure factor structure with the idiosyncratic jump component. In this situation, the P-PCA method might be applied.

4.1 Estimations

A more general strategy to estimating $\boldsymbol{\beta}_U$ when J_{it} is present can be developed with a simple modification on the proposed algorithm in Section 2. Bai (2009) shows that if X'_{it} is available, an iterative scheme which combines least squares and the PCA method can be used to estimate $\boldsymbol{\beta}_U$, factors $\mathbf{F}_t$ and factor loading $\boldsymbol{\lambda}_i$. The idea is that, given $\mathbf{F}$ and $\boldsymbol{\Lambda}$, $\boldsymbol{\beta}_U$ can be estimated by regressing $X'_{it} - \mathbf{F}_t^{\mathbf{T}}\boldsymbol{\lambda}_i$ on the observable U_{it}. While $\boldsymbol{\beta}_U$ is available, $\mathbf{F}$ and $\mathbf{\Lambda}$ can be obtained by applying the PCA method to the matrix $\mathbf{V}'\mathbf{V}'^{\mathbf{T}}$, where $\mathbf{V}' = \mathbf{X}' - \mathbf{\Gamma}_U = \mathbf{F}\mathbf{\Lambda}^{\mathbf{T}} + \mathbf{e}$ has a form of pure factor structure and $\mathbf{\Gamma}_U$ is a $T \times N$ matrix with $U_{it}^{\mathbf{T}}\boldsymbol{\beta}_U$ as its (i,j)th element. Thus, if we know $\mathbf{F}$ and $\mathbf{\Lambda}$, we can estimate $\boldsymbol{\beta}_U$ and vice versa. The estimation procedure can be done by conducting an iterative scheme, in which the least squares and the PCA method are repeatedly implemented and the procedure terminates until the estimations have converged.

If J_{it} is present, in the iterative scheme, we can replace the PCA method with the P-PCA to disentangle J_{it} from other components in X_{it}. That is, given $\boldsymbol{\beta}_U$, we can apply the P-PCA method to the matrix $\mathbf{V}\mathbf{V}^{\mathbf{T}}$, where $\mathbf{V} = \mathbf{X} - \mathbf{\Gamma}_U = \mathbf{F}\mathbf{\Lambda}^{\mathbf{T}} + \mathbf{J} + \mathbf{e}$ has a form of pure factor structure with the idiosyncratic jump component. In summary, for estimating model (5), we can construct an algorithm with two nested loops: The outer loop of the algorithm is designed for estimating $\boldsymbol{\beta}_U$, which can be done by using the observable U_{it} with the least squares, while the inner loop is designed for estimating $\mathbf{F}$, $\mathbf{\Lambda}$ and $\mathbf{J}$ with the algorithm proposed in Section 2.

4.2 More simulation results

We then conduct a simulation study to see how the P-PCA and PCA methods perform when the data have a more complex generating process like the one shown in (5). We consider two data generating processes. The first one has the following settings:

Model 6:

$$X_{it} = \sum_{k=1}^{4} U_{t,k}\beta_{ik} + \mathbf{F}_t^{\mathbf{T}}\boldsymbol{\lambda}_i + J_{it} + e_{it},$$

- $U_{t,1} = 1$ and $U_{t,k} \sim$ *i.i.d.* $U(-2.5, -2.5)$ for $k = 2, 3, 4$.
- $\beta_{i1} = 5$, $\beta_{i2} \sim$ *i.i.d.* $U(-1.5, -0.5)$, $\beta_{i3} \sim$ *i.i.d.* $U(0.3, 1.2)$ and $\beta_{i4} \sim$ *i.i.d.* $U(0.5, 1.6)$.
- $\mathbf{F}_t \sim$ *i.i.d.* $\mathcal{N}(\mathbf{0}, \mathbf{I}_{rr})$, $\lambda_{ij} \sim$ *i.i.d.* $\mathcal{N}(0, 1)$ for $i = 1, \ldots, N$ and $j = 1, \ldots, r$.
- $e_{it} \sim$ *i.i.d.* $\mathcal{N}(0, \theta)$.
- $J_{it} \sim$ *i.i.d.* *Pois* $(\nu) \times \mathcal{N}(0, \sigma_J^2)$ with $\nu = 0$, 0.01, 0.05 and 0.1, and $\sigma_J = 6 \times \sqrt{\theta}$.

We set the number of the latent common factors $r = 3$ and the scaling parameter $\theta = 4$ for Model 6. The data generating process of Model 6 says that X_{it} is governed by the observable common factors $U_{t,k}$, but coefficients of $U_{t,k}$ are cross sectionally different. In addition, X_{it} are also affected by the latent common factors, idiosyncratic jump component and noise term. The second one is a panel data model borrowed from Bai (2009):

Model 7:

$$
\begin{aligned}
X_{it} &= \sum_{k=1}^{5} U_{it,k}\beta_k + \mathbf{F}_t^{\mathbf{T}}\boldsymbol{\lambda}_i + J_{it} + e_{it}, \\
U_{it,1} &= 1, \\
U_{it,2} &= 1 + \mathbf{F}_t^{\mathbf{T}}\boldsymbol{\lambda}_i + \boldsymbol{\lambda}_i^{\mathbf{T}}\mathbf{1} + \mathbf{F}_t^{\mathbf{T}}\mathbf{1} + e_{it,1}^{U}, \\
U_{it,3} &= 1 + \mathbf{F}_t^{\mathbf{T}}\boldsymbol{\lambda}_i + \boldsymbol{\lambda}_i^{\mathbf{T}}\mathbf{1} + \mathbf{F}_t^{\mathbf{T}}\mathbf{1} + e_{it,2}^{U}, \\
U_{it,4} &= \boldsymbol{\lambda}_i^{\mathbf{T}}\mathbf{1} + e_{i,3}^{U}, \\
U_{it,5} &= \mathbf{F}_t^{\mathbf{T}}\mathbf{1} + e_{t,4}^{U},
\end{aligned}
$$

- $\beta_1 = 5$, $\beta_2 = 1$, $\beta_3 = 3$, $\beta_4 = 2$ and $\beta_5 = 4$.
- $\mathbf{F}_t \sim i.i.d.\ \mathcal{N}(\mathbf{0}, \mathbf{I}_{rr})$, $\lambda_{ij} \sim i.i.d.\ \mathcal{N}(0,1)$ for $i = 1, \ldots, N$ and $j = 1, \ldots, r$.
- $\left(e_{it,1}^{U}, e_{it,2}^{U}\right)^{\mathbf{T}} \sim i.i.d.\ \mathcal{N}(0, \mathbf{I}_{22})$, $e_{i,3}^{U} \sim i.i.d.\ \mathcal{N}(0,1)$, and $e_{t,4}^{U} \sim i.i.d.\ \mathcal{N}(0,1)$.
- $e_{it} \sim i.i.d.\ \mathcal{N}(0, \theta)$.
- $J_{it} \sim i.i.d.\ Pois(\nu) \times \mathcal{N}(0, \sigma_J^2)$ with $\nu = 0$, 0.01, 0.05 and 0.1, and $\sigma_J = 8 \times \sqrt{\theta}$.

We set the number of the latent common factors $r = 2$ and the scaling parameter $\theta = 4$ for Model 7. The data generating process of Model 7 says that X_{it} is governed by observable factors $U_{it,k}$, the latent common factors, idiosyncratic jump component and noise term. The coefficient of each $U_{it,k}$ is fixed at the same level for each X_{it}. Here each $U_{it,k}$ has quite different empirical properties: $U_{it,1}$ is a constant, $U_{it,2}$ and $U_{it,3}$ are correlated with the latent common factors, factor loadings and product of the two, $U_{it,4}$ is governed by sum of the factor loadings and is varying cross sectionally, and $U_{it,5}$ is governed by sum of the latent common factors and is varying as time changes. Finally, for controlling occurrences of the idiosyncratic jumps in Models 6 and 7, we use the same settings for parameters (a, ν) as in Section 3.

We consider two additional performance measures which are based on an oracle estimation β_{ik}^{O} (β_{k}^{O}) for β_{ik} (β_k). We define the oracle estimates β_{ik}^{O} (β_{k}^{O}) as the OLS estimation of β_{ik} (β_k) when the true latent common factors, factor loadings and idiosyncratic jump component are all known in advance. The first additional performance measure we consider is the squared estimation error of the estimated coefficient SE_{β_k}, which measures quality of each estimated coefficient of the observed factor. Since Models 6 and 7 have different specifications, the numbers of coefficients of the observable factors in these two models are quite different ($4 \times N$ for model 6 and 5 for model 7). We thus define $SE_{\beta_k} = \sum_{i=1}^{N} \left(\hat{\beta}_{ik} - \beta_{ik}^{O} \right)^2$ for Model 6 and $SE_{\beta_k} = \left(\hat{\beta}_k - \beta_k^{O} \right)^2$ for Model 7, where $\hat{\beta}_{ik}$ ($\hat{\beta}_k$) is the estimated coefficient for β_{ik} (β_k) from either using the P-PCA or PCA methods. For each k, we report the square root of average SE_{β_k} from the 2000 simulations in Tables 1 and 2. The second additional performance measure we consider is the l_2 distance ($l_2 - dis$) between the estimated coefficients and their corresponding oracle estimations, which is defined as $l_2 - dis = \sqrt{\sum_{k=1}^{K} SE_{\beta_k}}$. It measures overall quality of the estimated coefficients of the observed factors. For Model 6, $K = 4$ and for Model 7, $K = 5$. We plot average of the l_2 distance from the 2000 simulations in the upper panels of Figure 5 and Figure 6.

From Tables 1 and 2, we can see that accuracies of the estimated coefficients decline as the intensity of the idiosyncratic jump components $a \times \nu$ increases. Except when there is no idiosyncratic jump component, in general using the P-PCA method can obtain better estimated coefficients than using the PCA method[1].

Similar phenomena also occur for the l_2 distances. In Figures 5 and 6, we can see that the l_2 distances generally increase as the intensity $a \times \nu$ increases. As N and T increase, the l_2 distances get substantially improved for Model 7 but not Model 6, since the number of coefficients in the latter model increases with N in our setting. Except for the cases in which N and T are equal to 50 and 100, as the intensity $a \times \nu$ increases, the P-PCA method generates lower l_2 distances than the PCA method does.

As for other performance measures, the P-PCA method has higher values in DCOR and the trace R-square, and lower values in the squared predictive error. The P-PCA method also consistently provides a higher DCOR and a lower squared predictive error in every simulation as T, N and the intensity of the idiosyncratic jump components increase. In summary, the results shown here are consistent with those shown in Section 3: Using the P-PCA method can obtain better estimation results than using the PCA method.

5 An Empirical Application: Latent factors and Cross Sectional Expected Asset Returns

In this section, we apply the P-PCA method to investigate whether cross sectional variations of expected asset returns can be explained by latent components. In finance,

[1]Note that the numbers in Table 1 are much larger than those in Table 2. This is because in Model 6 there are much more coefficients need to be estimated than in Model 7.

a commonly used factor model for describing dynamics of expected excess asset returns is the Fama and French three-factor model (Fama and French, 1993):

$$R_{it} - R_t^f = \alpha_t + \beta_t MRT_t + \theta_t^{SMB} SMB_t + \theta_t^{HML} HML_t + u_{it}.$$

Here R_{it} and R_t^f are asset i's return and risk free rate at period t, MRT_t is the excess market portfolio return, SMB_t is a difference of returns of portfolios of small company stocks and large company stocks, and HML_t is a difference of returns of portfolios of value stocks (stocks of companies with high book to market ratio) and growth stocks (stocks of companies with low book to market ratio). The parameter α_t is a measure on asset i's abnormal return, and β_t, θ_t^{SMB} and θ_t^{HML} are factor loadings for the three factors MRT_t, SMB_t and HML_t at period t, respectively. In addition, $u_{i,t}$ is the part of the expected excess return that cannot be explained by the three factors. It is often assumed to be a pure noise, and its standard deviation is called asset i's idiosyncratic volatility ($IVOL$, Ang et al. (2006)).

However, the assumption that u_{it} is a pure noise is not realistic, since in addition to the Fama and French three factors, the expected excess returns may be associated with other factors. Recently Ando and Bai (2013) analyze possible latent common and group-specific pervasive factors of the expected stock returns in China stock market. They assume expected returns of stocks traded in the same exchanges are governed by the same latent within-group common factors. They find expected returns of stocks traded in different stock exchanges (Shanghai and Shenzhen) are indeed affected by different observable and latent factors.

Here we also assume u_{it} is governed by some latent factors and has the following form:

$$\begin{aligned} u_{it} &= \mathbf{F}_t^{\mathbf{T}} \boldsymbol{\lambda}_i + \omega_{it}, \\ &= \mathbf{F}_t^{\mathbf{T}} \boldsymbol{\lambda}_i + J_{it} + \varepsilon_{it}, \end{aligned}$$

where ω_{it} is a sum of the idiosyncratic jump J_{it} and pure noise term ε_{it}. Depending on whether including J_{it} or not, we can have two new definitions of the $IVOL$ for asset i: It can be either defined as the standard deviation of ω_{it} or the standard deviation of ε_{it}. To estimate the latent structure of u_{it}, we can first regress $R_{it} - R_t^f$ on an intercept term, MRT_t, SMB_t and HML_t and get the regression residuals and then apply the P-PCA (or PCA) to the regression residuals.

How the $IVOL$ affects the cross sectional expected returns is an important research topic in finance. From Arbitrage Pricing Theory, if an asset's expected return obeys a factor structure, then the question of whether the $IVOL$ is priced for an asset will depend on whether the asset is a well diversified portfolio or not. If the asset is a well diversified portfolio, the $IVOL$ should be negligible and thus is not (approximately) priced. If the asset is not a well diversified portfolio such as an individual stock, then as a risk factor, the $IVOL$ should have a positive effect on the asset's expected return. However, empirical evidence on whether the $IVOL$ has a positive effect on cross sectional expected stock returns is still quite mixed (Ang et al., 2006; Bali and Cakici, 2008; Ang et al., 2009; Fu, 2009).

The data used here is the Fama and French F100 size and book-to-market ratio portfolios from Jan, 1970 to Dec, 2012. We use monthly data and set the number of factors $r = 2$ to estimate the factors $\mathbf{F}_t$ and factor loadings $\boldsymbol{\lambda}_i$[2]. To estimate the $IVOL$, we follow Ang et al. (2006) to use daily data. The estimation procedures are described as follows. In each month t, we first estimate the following regression with daily data:

$$R_{i\tau} - R_\tau^f = \alpha_\tau + \beta_\tau MRT_\tau + \theta_\tau^{SMB} SMB_\tau + \theta_\tau^{HML} HML_\tau + u_{i\tau},$$

where τ denotes the τth day in month t. We then obtain estimates $\hat{\omega}_{i\tau}$, $\hat{J}_{i\tau}$ and $\hat{\varepsilon}_{i\tau}$ by applying the P-PCA (or PCA) method to the regression residuals $\hat{u}_{i\tau}$. The sample standard deviations of $\hat{u}_{i\tau}$, $\hat{\omega}_{i\tau}$ and $\hat{\varepsilon}_{i\tau}$ scaled with number of days in month t are used as different estimates for the $IVOL$ of asset i in month t. We use $IVOL^u$, $IVOL^{PCA}$ and $IVOL^{P-PCA}$ to denote such estimations, respectively. The Fama and French three-factor regression is run month-by-month, and time series of different estimations for the monthly $IVOL$ can be obtained.

In month t the expected asset return is determined by available information prior to month t. Therefore to see whether risk of the $IVOL$ is priced, it is more reasonable to use conditional expectation of the $IVOL$ rather than the realized $IVOL$. To estimate the conditional expectation of $IVOL$ of asset i in month t (denoted by $\mathbb{E}_{t-1}[IVOL_{it}]$), we first estimate the following heterogeneous autoregressive (HAR) regression (Corsi, 2009)[3]:

$$IVOL_{it} = a + b_1 IVOL_{i,t-1} + b_2 IVOL_{i,t-1}^{6m} + b_3 IVOL_{i,t-1}^{24m} + \varepsilon_{it}^{IVOL},$$

[2]For the following analysis, we also try to set $r = 3$ and 4. We find their empirical results are overall qualitatively similar as those from $r = 2$.

[3]An alternative method to obtaining $IVOL$ is by parametrically estimating conditional volatilities of the monthly regression residuals (e.g., using GARCH or stochastic volatility models). The fitted conditional volatilities of the monthly regression residuals are then used as estimates of the monthly $IVOL$ for asset i. Detail discussions on the approach can be found in Fu (2009).

where $IVOL_{i,t-1}^{6m} = 1/6\sum_{k=0}^{5} IVOL_{i,t-1-k}$ and $IVOL_{i,t-1}^{24m} = 1/24\sum_{k=0}^{23} IVOL_{i,t-1-k}$. The regression is estimated for $IVOL^{u}$, $IVOL^{PCA}$ and $IVOL^{P-PCA}$. In addition, the sample period is from January, 1968 to December, 2012. The fitted value of the $IVOL_{it}$ from the HAR regression is then used as an estimate for $\mathbb{E}_{t-1}[IVOL_{it}]$. The HAR regression is commonly used in modeling dynamics of realized variances of the asset returns, since it can well capture their highly persistent behavior. This is the reason why we use the HAR type model here, since we find the $IVOL$ of the FF100 portfolios is also highly persistent[4].

To analyze relations between the latent components and cross sectional variations of expected returns of the FF100 portfolios, we use the Fama and MacBeth (FM) regression (Fama and MacBeth, 1973). Assume portfolio i's return is a linear function of a set of control variables V_{lit}, $l = 1, \dots, L$ and a random error term ε_{it}^{cross}:

$$R_{it} = \gamma_{0t} + \sum_{l=1}^{L} \gamma_{lt} V_{lit} + \varepsilon_{it}^{cross}, \tag{7}$$

where we assume $\mathbb{E}_{t-1}[\varepsilon_{it}^{cross}] = 0$. Then portfolio i's expected return is a linear function of expectations of the control variables: $\mathbb{E}_{t-1}[R_{it}] = \gamma_0 + \sum_{l=1}^{L} \gamma_l \mathbb{E}_{t-1}[V_{lit}]$. The main control variables we are interested in are $\boldsymbol{\lambda}_i$, $IVOL_{it}$ and their expectations. As for other variables, we include the portfolio's $Beta$, logarithm of the average market value of the constituents in the portfolio ($\ln(ME)$), logarithm of the average book-to-market (BM) ratio of the constituents in the portfolio ($\ln(BE/ME)$) and the cumulative gross return from previous seven to two months ($CGR(-7,-2)$). Details on how we construct data of $Beta$, $\ln(ME)$, $\ln(BE/ME)$ and $CGR(-7,-2)$ are discussed in Appendix C. In addition, we also include information about the idiosyncratic jump component in the FM regression and see how it can affect the expected cross sectional returns. We use $ABSJ_{it} = 1/\#\{\tau \in t\}\sum_{\tau\in t}\left|\hat{J}_{i\tau}\right|$ to measure magnitude of the contemporaneous idiosyncratic jump and estimate $\mathbb{E}_{t-1}[ABSJ_{it}]$ by $1/6\sum_{k=0}^{5} ABSJ_{i,t-1-k}$.

In each month t, the FM regression (7) is estimated with the realized cross sectional portfolio returns as the left hand side variables and the specific control variables as the right hand side variables. To see whether a specific control variable has an effect on the expected cross sectional asset returns, we use t-statistic to test whether time series average of the estimated coefficients $\hat{\gamma}_{lt}$, $l = 0, \dots, L$ is significantly different from zero or not.

Tables 3 and 4 show time series average of $\hat{\gamma}_{lt}$ and the t-statistic calculated using the

[4]Note that the property of high persistence is not in $IVOL$ of individual stocks (Fu, 2009).

Newey-West standard error when different model settings are considered. Here λ_i^{PCA} (λ_i^{P-PCA}) for $i = 1, 2$ denote factor loadings of the first and second latent factors estimated from the PCA (P-PCA) method. From the two Tables, we can see the portfolio's $Beta$ and $CGR\,(-7,-2)$ on average have no significant effect on the cross sectional portfolio returns. The estimated coefficient of $\ln(ME)$ on average is negative and is also not statistically significant. The estimated coefficient of $\ln(BE/ME)$ on average is positive and has the t-statistic greater than 2 in all different model settings. The result indicates that a portfolio of stocks with high BM ratios tends to have a high expected return, which is in line with the findings of previous empirical studies.

From the two Tables, we also find the average estimated coefficient of the factor loading λ_1 is significant from zero, regardless of whether it is obtained from using the PCA or P-PCA method. The result suggests that the latent common component is priced for the expected portfolio returns. Another interesting finding is that with the P-PCA method, the significance of the factor loading is in general stronger than with PCA method. The contemporaneous $IVOL$ and $ABSJ$ on average have positive and highly significant estimated coefficients, but the expected $IVOL$ and $ABSJ$ do not. This suggests that the expected idiosyncratic volatility and jump are not priced for the expected portfolio return. Note that each of the 100 portfolios is constructed by value weighted average of individual stocks whose BM ratios and market values are within a certain range. Thus if the 100 portfolios themselves are well diversified, they should be immune from the idiosyncratic volatility risk (Merton, 1987). During our sample period, average number of stocks in the 100 portfolios increases from 36 in the period between 1970 to 1989 to 47 in the period between 1990 to 2012, which is near the number of stocks that a well diversified portfolio needs (Campbell et al., 2001)[5].

Finally, statistical significances between average estimated coefficients of the expected and contemporaneous idiosyncratic components are quite different. It suggests that forecasting errors of the portfolio returns and forecasting errors of the $IVOL$ (or $ABSJ$) may correlate with each other, or the two forecasting errors may be governed by common factors such as uncertainty of the forecasting models. To further investigate this issue, we use the absolute deviation of the estimated $\mathbb{E}_{t-1}[IVOL_{it}]$ (or $\mathbb{E}_{t-1}[ABSJ_{it}]$) and the contemporaneous $IVOL_{it}$ (or $ABSJ_{it}$) as a proxy for the forecasting errors and add them to the FM regression. Let $\left|\varepsilon_t^{IVOL^u}\right|$, $\left|\varepsilon_t^{IVOL^{PCA}}\right|$ and $\left|\varepsilon_t^{IVOL^{P-PCA}}\right|$ denote such proxies for the forecasting errors when $IVOL^u$, $IVOL^{PCA}$ and $IVOL^{P-PCA}$ are used, respectively, and let $\left|\varepsilon_t^{ABSJ}\right|$ be the same notation for $ABSJ$. From Tables 3 and 4 we

[5]Campbell et al. (2001) find that in different periods, a portfolio needs around 20 to 50 stocks to reduce its idiosyncratic risk to a small level.

can see the estimated coefficients of these proxies are on average positive and highly statistically significant, which suggests that uncertainty of the forecasting models is priced. In summary, the results shown here provide empirical evidence on the relations of portfolio diversifications, idiosyncratic risks and model uncertainty. Although investors can form a well diversified portfolio to reduce idiosyncratic risks, their efforts are limited to the extent of what information is available. With diversifications, only the predictable idiosyncratic risks can be reduced, but the unpredictable ones may still exist. Such unpredictable idiosyncratic risks can be categorized as model uncertainty risks, and as shown empirically here, they are priced risk factors.

6 Conclusion

We propose a penalized least squares estimation method, called the P-PCA method, to estimate approximate factor models in which candidate predictors are subject to idiosyncratic large uncommon components. The algorithm for the proposed method can be easily implemented and incorporated with methods for selecting the number of common factors. Simulation results indicate that the proposed method can have better finite-sample performances than the PCA method when data have uncommon components. We then use the proposed method to investigate whether the latent factors are priced for expected returns of Fama and French 100 size and book-to-market ratio portfolios. We find evidence that for the 100 portfolios, the risk from the common factor is priced but risks from the idiosyncratic factors are not. However, we also find that model with uncertainty risks of the idiosyncratic factors are priced, which suggests that with effective diversifications, only the predictable idiosyncratic risks can be reduced while the unpredictable ones may still exist.

Appendix

A. A Step-by-Step Description on The Algorithm

Below we provide a step-by-step description on the algorithm for solving problem (4) given that the number of factors r is known and the penalty parameter δ is fixed.

Step 1. Set the initial value of $\mathbf{J}$, $\mathbf{J}^{(0)} = \mathbf{0}$.

Step 2. Given $\mathbf{J} = \mathbf{J}^{(0)}$, solve (4). When $\mathbf{J}^{(0)} = \mathbf{0}$, solving (4) is equivalent to solving (3), and we can obtain an optimal solution by using the PCA method.

Let $\left(\mathbf{F}^{(1)}, \mathbf{\Lambda}^{(1)}\right)$ denote the optimal solution from using the PCA method. Define $\mathbf{Z}^{(1)} = \mathbf{F}^{(1)}\mathbf{\Lambda}^{(1)\mathbf{T}}$.

Step 3. Plug $\mathbf{Z}^{(1)}$ into (4) and solve (4) with respect to $\mathbf{J}$. This is equivalent to solving

$$\min_{\mathbf{J}} \frac{1}{TN} \left\|\mathbf{X} - \mathbf{Z}^{(1)} - \mathbf{J}\right\|_F^2 + \frac{\delta}{TN} \|\mathbf{J}\|_1 . \tag{8}$$

Define $\mathbf{L}^{(1)} = \mathbf{X} - \mathbf{Z}^{(1)}$ and let $L_{it}^{(1)}$ and J_{it} denote the (i,t)th entries of matrices $\mathbf{L}^{(1)}$ and $\mathbf{J}$, respectively. The optimization problem (8) can be reformulated as

$$\min_{J_{it}, i=1,\ldots,N, t=1,\ldots,T} -\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} J_{it}L_{it}^{(1)} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} J_{it}^2 + \frac{\delta}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} |J_{it}| . \tag{9}$$

Problem (9) is separable, meaning that each optimal J_{it} can be obtained by separately solving the following one dimensional optimization:

$$\min_{J_{it}} -2J_{it}L_{it}^{(1)} + J_{it}^2 + \delta\,|J_{it}| .$$

The optimal J_{it}, denoted by $J_{it}^{(1)}$, is given by

$$J_{it}^{(1)} = \begin{cases} L_{it}^{(1)} - \frac{\delta}{2}, & \text{if } L_{it}^{(1)} > \frac{\delta}{2}, \\ 0, & \text{if } -\frac{\delta}{2} \le L_{it}^{(1)} \le \frac{\delta}{2}, \\ L_{it}^{(1)} + \frac{\delta}{2}, & \text{if } L_{it}^{(1)} < -\frac{\delta}{2}, \end{cases}$$

or more concisely,

$$J_{it}^{(1)} = ST\left(L_{it}^{(1)}, \frac{\delta}{2}\right),$$

where $ST\left(x, y\right) := sign(x)(|x| - y)_+$ is the soft thresholding function.

Step 4. Define $\mathbf{J}_t^{(1)} = (J_{1t}^{(1)}, J_{2t}^{(1)}, \cdots, J_{Nt}^{(1)})$ and $\mathbf{J}^{(1)} = \left(\mathbf{J}_1^{(1)}, \ldots, \mathbf{J}_T^{(1)}\right)^{\mathbf{T}}$. Update $\mathbf{J}$ with $\mathbf{J}^{(1)}$, plug it into (4), and solve (4) with respect to $(\mathbf{F}, \mathbf{\Lambda})$. This is equivalent to solving

$$\min_{\mathbf{F},\mathbf{\Lambda}} \frac{1}{TN} \left\|\mathbf{C}^{(1)} - \mathbf{F}\mathbf{\Lambda}^{\mathbf{T}}\right\|_F^2, \text{ subject to } \frac{\mathbf{F}^{\mathbf{T}}\mathbf{F}}{T} = \mathbf{I}_r, \tag{10}$$

where $\mathbf{C}^{(1)} = \mathbf{X} - \mathbf{J}^{(1)}$. Again we solve (10) by using the PCA method. Let $\left(\mathbf{F}^{(2)}, \mathbf{\Lambda}^{(2)}\right)$ denote the optimal solution to problem (10). Define $\mathbf{Z}^{(2)} = \mathbf{F}^{(2)}\mathbf{\Lambda}^{(2)\mathbf{T}}$.

Step 5. Plug $\mathbf{Z}^{(2)}$ into (4) and solve (4) with respect to $\mathbf{J}$ as we do in step 3. Let

$\mathbf{J}^{(2)}$ denote the optimal solution. We use $\mathbf{J}^{(2)}$ to update $\mathbf{J}$.

Step 6. Repeat step 4 and 5 to obtain $\left(\mathbf{F}^{(k)}, \mathbf{\Lambda}^{(k)}\right)$, $\mathbf{Z}^{(k)}$ and $\mathbf{J}^{(k)}$ for $k = 1, 2, \cdots, k_{max}$ until the following convergence condition is met:

$$\frac{\left\|\mathbf{Z}^{(\bar{k})} - \mathbf{Z}^{(\bar{k}-1)}\right\|_F}{\left\|\mathbf{Z}^{(\bar{k}-1)}\right\|_F} \leq \epsilon,$$

where $\bar{k} \geq 1$ is the number of iterations for step 4 and 5. The output $\left(\mathbf{F}^{(\bar{k})}, \mathbf{\Lambda}^{(\bar{k})}\right)$ and $\mathbf{J}^{(\bar{k})}$ is used as an estimate for $(\mathbf{F}, \mathbf{\Lambda})$ and $\mathbf{J}$.

Simply to say, the above six steps can be summarized as

1. Get initial estimated $\mathbf{F}$ and $\mathbf{\Lambda}$ by using PCA on $\mathbf{X}\mathbf{X}^{\mathbf{T}}$.
2. Get residuals $\mathbf{X} - \hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{\mathbf{T}}$, and then use the softthresholding function to filter the residuals to obtain $\hat{\mathbf{J}}$.
3. Get $\hat{\mathbf{F}}$ and $\hat{\mathbf{\Lambda}}$ again by using PCA on $\hat{\mathbf{C}}\hat{\mathbf{C}}^{\mathbf{T}}$, where $\hat{\mathbf{C}} = \mathbf{X} - \hat{\mathbf{J}}$.
4. Repeat 2 and 3 until the solution has converged.

Note that in this algorithm, to identify $\mathbf{J}$, it is not necessary to know the factor $\mathbf{F}$ and factor loading matrices $\mathbf{\Lambda}$. To identify $\mathbf{J}$, knowing the product $\mathbf{Z} = \mathbf{F}\mathbf{\Lambda}^{\mathbf{T}}$ is enough.

B. Convergence of the Algorithm

Algorithm 1 iteratively chooses a low rank matrix and a sparse matrix to minimize the objective function. Below we show that the Algorithm indeed decreases the objective function in each iteration. We first show that estimating the factors and factor loadings by using the PCA and least squares methods is equivalent to solving a low rank approximation problem.

B.1 Low Rank Matrix Approximation

From the theorem of the singular value decomposition (SVD), every matrix $\mathbf{C} \in \mathbb{R}^{T \times N}$ admits a decomposition of the form:

$$\mathbf{C} = \mathbf{U}\tilde{\mathbf{L}}\mathbf{V}^{\mathbf{T}},\ \tilde{\mathbf{L}} = \begin{pmatrix} \mathbf{L} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where $\mathbf{U}$ is a $T \times T$ unitary matrix, $\mathbf{V}$ is an $N \times N$ unitary matrix, $\tilde{\mathbf{L}}$ is a $T \times N$ matrix, $\mathbf{L} = \mathbf{diag}\,(l_1, \ldots, l_q)$ is a $q \times q$ diagonal matrix, and $q \leq \min\,(T, N)$ is rank of the matrix $\mathbf{C}$. The diagonal elements $l_1, \ldots, l_q$, called the singular values of the matrix $\mathbf{C}$, are all positive and unique. The first q columns of $\mathbf{U}$ are called left singular vectors of $\mathbf{C}$, while the first q columns of $\mathbf{V}$ are called right singular vectors of $\mathbf{C}$. It can be shown that $\mathbf{CC^T} = \mathbf{U\tilde{L}\tilde{L}^TU^T}$, and $\mathbf{C^TC} = \mathbf{V\tilde{L}^T\tilde{L}V^T}$. It is known that nonzero eigenvalues of $\mathbf{CC^T}$ and $\mathbf{C^TC}$ are all the same and are given by $l_1^2, \ldots, l_q^2$. The corresponding eigenvectors of $\mathbf{CC^T}$ and $\mathbf{C^TC}$ are $\mathbf{U}$ and $\mathbf{V}$, respectively.

If the matrix $\mathbf{J}$ is known and $N > T$, solving (4) is equivalent to solving

$$\min_{\mathbf{F},\mathbf{\Lambda}} \frac{1}{TN} \left\|\mathbf{C} - \mathbf{F\Lambda^T}\right\|_F^2, \text{ subject to } \frac{\mathbf{F^TF}}{T} = \mathbf{I}_r,$$

where $\mathbf{C} = \mathbf{X} - \mathbf{J}$. Using the PCA method yields $\hat{\mathbf{F}} = \sqrt{T}\mathbf{U}_r$, where $\mathbf{U}_r$ is a matrix containing the eigenvectors corresponding to the largest r eigenvalues of the $T \times T$ matrix $\mathbf{CC^T}$. Using the least squares method yields $\hat{\mathbf{\Lambda}} = \mathbf{C^T}\hat{\mathbf{F}}/T$. By using the fact that $\mathbf{U}_r^{\mathbf{T}}\mathbf{U} = (\mathbf{I}_r, \mathbf{0})$, it can be shown that

$$\begin{aligned} \hat{\mathbf{Z}} &= \hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{\mathbf{T}} \\ &= \mathbf{U}_r\mathbf{U}_r^{\mathbf{T}}\mathbf{U}\tilde{\mathbf{L}}\mathbf{V}^{\mathbf{T}} \\ &= \mathbf{U}\tilde{\mathbf{L}}_r\mathbf{V}^{\mathbf{T}}, \end{aligned}$$

where

$$\tilde{\mathbf{L}}_r = \begin{pmatrix} \mathbf{L}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

is a $T \times N$ matrix and $\mathbf{L}_r = \mathbf{diag}\,(l_1, \ldots, l_r)$. The matrix $\hat{\mathbf{Z}}$ is in fact the solution to the following optimization problem:

$$\min_{\mathbf{Z}} \|\mathbf{C} - \mathbf{Z}\|_F, \text{ subject to } rank\,(\mathbf{Z}) = r.$$

It is the best r-rank approximation to $\mathbf{C}$. Note that if $\hat{\mathbf{Z}}$ minimizes $\|\mathbf{C} - \mathbf{Z}\|_F$, it also minimizes $\|\mathbf{C} - \mathbf{Z}\|_F^2$. Therefore the above procedures for estimating $\mathbf{F}$ and $\mathbf{\Lambda}$ is equivalent to finding an optimal r-rank matrix to approximate $\mathbf{C}$.

B.2 Descent Algorithm

Define

$$Q(\mathbf{Z},\mathbf{J}) := \frac{1}{TN}\|\mathbf{X}-\mathbf{Z}-\mathbf{J}\|_F^2 + \frac{\delta}{TN}\|\mathbf{J}\|_1.$$

We next show that under our algorithm, each iteration indeed reduces the value of $Q(\mathbf{Z},\mathbf{J})$. In our algorithm, given $\mathbf{J}^{(0)}$, we find an optimal r-rank matrix $\mathbf{Z}^{(1)}$ to minimize $Q\left(\mathbf{Z},\mathbf{J}^{(0)}\right)$, and given $\mathbf{Z}^{(1)}$ we find an optimal $\mathbf{J}^{(1)}$ to minimize $Q\left(\mathbf{Z}^{(1)},\mathbf{J}\right)$. In turn, given $\mathbf{J}^{(1)}$ we then find an optimal r-rank matrix $\mathbf{Z}^{(2)}$ to minimize $Q\left(\mathbf{Z},\mathbf{J}^{(1)}\right)$, and given $\mathbf{Z}^{(2)}$ we find an optimal $\mathbf{J}^{(2)}$ to minimize $Q\left(\mathbf{Z}^{(2)},\mathbf{J}\right)$ and so on. By induction we obtain

$$\begin{aligned}
Q\left(\mathbf{Z}^{(k)},\mathbf{J}^{(k)}\right) &\geq \min_{\mathbf{z},rank(\mathbf{z})=r} Q\left(\mathbf{Z},\mathbf{J}^{(k)}\right)\\
&= Q\left(\mathbf{Z}^{(k+1)},\mathbf{J}^{(k)}\right)\\
&\geq \min_{\mathbf{J}} Q\left(\mathbf{Z}^{(k+1)},\mathbf{J}\right)\\
&= Q\left(\mathbf{Z}^{(k+1)},\mathbf{J}^{(k+1)}\right),
\end{aligned}$$

which shows that $Q\left(\mathbf{Z}^{(k)},\mathbf{J}^{(k)}\right)$ is a decreasing function of the number of iterations k.

B.3 Convergence Condition

The convergence condition in step 6 only considers convergence of $\mathbf{Z}^{(k)} = \mathbf{F}^{(k)}\mathbf{\Lambda}^{(k)\mathbf{T}}$. The reason is that, if $\mathbf{Z}^{(k)}$ converges, $\mathbf{J}^{(k)}$ can also converges. To see this, we show that $\left\|\mathbf{J}^{(k+1)}-\mathbf{J}^{(k)}\right\|_F \leq \left\|\mathbf{Z}^{(k+1)}-\mathbf{Z}^{(k)}\right\|_F$. Note that at the kth iteration, the following condition should hold:

$$-\left(\mathbf{X}-\mathbf{Z}^{(k)}-\mathbf{J}^{(k)}\right)+\frac{\delta}{2}\mathbf{S}^{(k)}=\mathbf{0}, \tag{11}$$

where $\mathbf{S}^{(k)} = \left(\mathbf{S}_1^{(k)},\ldots,\mathbf{S}_T^{(k)}\right)^{\mathbf{T}}$ is a $T\times N$ matrix. The ith element in vector $\mathbf{S}_t^{(k)}$ is $s_{it}^{(k)} \in [-1,1]$ for $i=1,\ldots,N$. The system of equations (11) is a matrix form of the KKT conditions for problem (9). It holds for each k. One can show that

$$\left\langle \mathbf{J}^{(k+1)}-\mathbf{J}^{(k)},\mathbf{Z}^{(k+1)}-\mathbf{Z}^{(k)}\right\rangle + \left\|\mathbf{J}^{(k+1)}-\mathbf{J}^{(k)}\right\|_F^2 + \frac{\delta}{2}\left\langle \mathbf{J}^{(k+1)}-\mathbf{J}^{(k)},\mathbf{S}^{(k+1)}-\mathbf{S}^{(k)}\right\rangle = 0,$$

where $\langle\mathbf{a},\mathbf{b}\rangle$ denotes $Trace\left(\mathbf{a}^{\mathbf{T}}\mathbf{b}\right)$. The third term in the above equality is trace of an $N\times N$ matrix with the ith diagonal element

$$\frac{\delta}{2}\sum_{t=1}^{T}\left(J_{it}^{(k+1)}-J_{it}^{(k)}\right)\left(s_{it}^{(k+1)}-s_{it}^{(k)}\right).$$

It is known that if $J_{it}^{(k)} \neq 0$, $s_{it}^{(k)} = sign\left(J_{it}^{(k)}\right)$ and if $J_{it}^{(k)} = 0$, $s_{it}^{(k)} \in [-1, 1]$. Thus it can be proved that $\left(J_{it}^{(k+1)} - J_{it}^{(k)}\right)\left(s_{it}^{(k+1)} - s_{it}^{(k)}\right) \geq 0$ always holds[6], and

$$\frac{\delta}{2}\left\langle \mathbf{J}^{(k+1)} - \mathbf{J}^{(k)}, \mathbf{S}^{(k+1)} - \mathbf{S}^{(k)} \right\rangle \geq 0,$$

if δ is positive. It then follows that

$$\begin{aligned} \left\| \mathbf{J}^{(k+1)} - \mathbf{J}^{(k)} \right\|_F^2 &\leq \left\| \mathbf{J}^{(k+1)} - \mathbf{J}^{(k)} \right\|_F^2 + \frac{\delta}{2}\left\langle \mathbf{J}^{(k+1)} - \mathbf{J}^{(k)}, \mathbf{S}^{(k+1)} - \mathbf{S}^{(k)} \right\rangle \\ &= -\left\langle \mathbf{J}^{(k+1)} - \mathbf{J}^{(k)}, \mathbf{Z}^{(k+1)} - \mathbf{Z}^{(k)} \right\rangle \\ &\leq \left\| \mathbf{J}^{(k)} - \mathbf{J}^{(k+1)} \right\|_F \left\| \mathbf{Z}^{(k+1)} - \mathbf{Z}^{(k)} \right\|_F . \end{aligned}$$

Note that

$$\left\| \mathbf{J}^{(k)} - \mathbf{J}^{(k+1)} \right\|_F = \left\| \mathbf{J}^{(k+1)} - \mathbf{J}^{(k)} \right\|_F .$$

Therefore

$$\left\| \mathbf{J}^{(k+1)} - \mathbf{J}^{(k)} \right\|_F \leq \left\| \mathbf{Z}^{(k+1)} - \mathbf{Z}^{(k)} \right\|_F ,$$

which means convergence of $\mathbf{Z}^{(k)}$ implies convergence of $\mathbf{J}^{(k)}$.

C. Constructions of Financial Data

To estimate a portfolio's *Beta*, we follow Fama and French (1992) and Fu (2009) by regressing monthly returns of the portfolio on current and lagged-one-period monthly market returns with the whole sample. Sum of coefficients of the current and lagged-one-period monthly market returns in the regression is our estimate for *Beta* of the portfolio. Let ME denote a portfolio's average firm size (average market value of companies whose stocks are in the portfolio) and ME/BE denote a portfolio's book to market ratio (sum of book values to sum of market values of companies whose stocks are in the portfolio). Following Fama and French (1992), $\ln(ME)$ of a portfolio from July of this year to June next year is approximated by logarithm of the portfolio's average firm

[6] Let $JS_{it} = \left(J_{it}^{(k+1)} - J_{it}^{(k)}\right)\left(s_{it}^{(k+1)} - s_{it}^{(k)}\right)$. If $J_{it}^{(k)}, J_{it}^{(k+1)} \neq 0$ and have the same sign, $JS_{it} = 0$ by $s_{it}^{(k+1)} - s_{it}^{(k)} = 0$. If $J_{it}^{(k)}, J_{it}^{(k+1)} \neq 0$ and have different signs, $JS_{it} > 0$ since $J_{it}^{(k+1)} - J_{it}^{(k)}$ and $s_{it}^{(k+1)} - s_{it}^{(k)}$ will have the same sign. If $J_{it}^{(k+1)} = 0$ and $J_{it}^{(k)} > 0$ $\left(J_{it}^{(k)} < 0\right)$, $JS_{it} = -J_{it}^{(k)} \times ([-1, 1] - 1) \geq 0$ $\left(JS_{it} = -J_{it}^{(k)} \times ([-1, 1] + 1) \geq 0\right)$. The same logic applies to the case of $J_{it}^{(k+1)} > 0$ $\left(J_{it}^{(k+1)} < 0\right)$ and $J_{it}^{(k)} = 0$. Finally, if $J_{it}^{(k)}, J_{it}^{(k+1)} = 0$, $JS_{it} = 0$. Thus we conclude that $\left(J_{it}^{(k+1)} - J_{it}^{(k)}\right)\left(s_{it}^{(k+1)} - s_{it}^{(k)}\right) \geq 0$ holds.

size in June of this year, and $\ln(ME/BE)$ of a portfolio from July of this year to June next year is approximated by logarithm of the portfolio's book to market ratio in the last fiscal year. The raw data used to obtain $\ln(ME)$ and $\ln(BE/ME)$ can be downloaded from Professor Kenneth French's website. Now let $CGR(-7,-2)$ denotes a portfolio's monthly cumulative gross returns from month $\tau-7$ to month $\tau-2$. All of the three control variables $\ln(ME)$, $\ln(BE/ME)$, $CGR(-7,-2)$ can be obtained before the portfolio returns are realized, and hence are suitable for constructing predictions.

Table 1: The Table shows square root of average SE_{β_k} for each combination of (a,ν) over 2000 simulations. The data generating process used here is Model 6. $SE_{\beta_k} = \sum_{i=1}^{N}\left(\hat{\beta}_{ik} - \beta_{ik}^{O}\right)^2$ is the squared estimation error of the estimated coefficient, where $\hat{\beta}_{ik}$ is the estimated coefficient for β_{ik} from either using P-PCA or PCA method and β_{ik}^{O} is an oracle estimation for β_{ik}. The oracle estimation for β_{ik} is defined as the OLS estimation of β_{ik} when the true latent common factors, factor loadings and idiosyncratic jump component are all known in advance.

	$k=1$		$k=2$		$k=3$		$k=4$	
	P-PCA	PCA	P-PCA	PCA	P-PCA	PCA	P-PCA	PCA
(a,ν)	$N=50$, $T=50$							
(0,0)	0.5401	0.5361	0.3782	0.3755	0.3760	0.3737	0.3733	0.3708
(0.1,0.01)	0.5684	0.6683	0.4049	0.4712	0.3963	0.4657	0.3988	0.4677
(0.5,0.01)	0.6996	1.0345	0.4880	0.7179	0.4924	0.7201	0.4934	0.7200
(0.1,0.05)	0.7283	1.0479	0.5121	0.7365	0.5127	0.7342	0.5098	0.7341
(1,0.01)	0.8592	1.3695	0.6094	0.9702	0.5997	0.9567	0.6005	0.9595
(0.1,0.1)	0.9584	1.3874	0.6710	0.9733	0.6723	0.9763	0.6784	0.9740
(0.5,0.05)	1.3310	2.0742	0.9313	1.4449	0.9314	1.4445	0.9444	1.4673
(1,0.05)	2.0097	2.8963	1.4358	2.0462	1.4177	2.0339	1.4197	2.0465
(0.5,0.1)	2.0570	2.9192	1.4484	2.0423	1.4360	2.0282	1.4538	2.0603
(1,0.1)	3.2180	4.1247	2.2723	2.9090	2.2638	2.8938	2.2784	2.9129
	$N=100$, $T=100$							
(0,0)	0.3625	0.3602	0.2521	0.2507	0.2531	0.2517	0.2522	0.2506
(0.1,0.01)	0.4121	0.5378	0.2835	0.3672	0.2827	0.3699	0.2836	0.3737
(0.5,0.01)	0.5542	0.9375	0.3911	0.6588	0.3892	0.6536	0.3885	0.6534
(0.1,0.05)	0.5684	0.9638	0.3941	0.6633	0.3921	0.6644	0.3919	0.6633
(1,0.01)	0.7267	1.2739	0.5071	0.8858	0.5051	0.8857	0.5076	0.8903
(0.1,0.1)	0.7491	1.3440	0.5202	0.9236	0.5179	0.9272	0.5209	0.9282
(0.5,0.05)	1.1768	2.0057	0.8244	1.3965	0.8223	1.3931	0.8202	1.3888
(1,0.05)	1.8714	2.8194	1.3084	1.9623	1.3084	1.9637	1.3114	1.9656
(0.5,0.1)	1.8704	2.8993	1.3034	2.0102	1.3019	2.0084	1.3038	2.0097
(1,0.1)	3.0940	4.0689	2.1632	2.8311	2.1584	2.8290	2.1641	2.8352
	$N=200$, $T=200$							
(0,0)	0.2533	0.2521	0.1769	0.1760	0.1745	0.1737	0.1740	0.1732
(0.1,0.01)	0.3162	0.4608	0.2186	0.3220	0.2178	0.3186	0.2164	0.3194
(0.5,0.01)	0.4968	0.8894	0.3470	0.6193	0.3468	0.6189	0.3484	0.6222
(0.1,0.05)	0.5027	0.9062	0.3539	0.6382	0.3510	0.6323	0.3511	0.6359
(1,0.01)	0.6909	1.2412	0.4843	0.8647	0.4819	0.8624	0.4798	0.8608
(0.1,0.1)	0.6954	1.2900	0.4898	0.9053	0.4858	0.8957	0.4867	0.9015
(0.5,0.05)	1.1694	1.9727	0.8132	1.3687	0.8130	1.3641	0.8180	1.3750
(1,0.05)	1.8889	2.7799	1.3165	1.9321	1.3146	1.9332	1.3105	1.9254
(0.5,0.1)	1.8651	2.8477	1.2950	1.9739	1.2967	1.9744	1.2997	1.9814
(1,0.1)	3.1234	4.0259	2.1688	2.7880	2.1662	2.7881	2.1612	2.7843
	$N=400$, $T=400$							
(0,0)	0.1760	0.1752	0.1204	0.1199	0.1222	0.1217	0.1226	0.1221
(0.1,0.01)	0.2581	0.4208	0.1783	0.2905	0.1803	0.2926	0.1808	0.2928
(0.5,0.01)	0.4789	0.8729	0.3340	0.6061	0.3338	0.6065	0.3335	0.6060
(0.1,0.05)	0.4827	0.8905	0.3336	0.6150	0.3355	0.6179	0.3346	0.6165
(1,0.01)	0.6901	1.2238	0.4782	0.8485	0.4782	0.8482	0.4783	0.8483
(0.1,0.1)	0.6925	1.2772	0.4789	0.8829	0.4826	0.8866	0.4801	0.8838
(0.5,0.05)	1.1929	1.9603	0.8276	1.3572	0.8298	1.3600	0.8271	1.3562
(1,0.05)	1.9385	2.7668	1.3445	1.9185	1.3447	1.9180	1.3445	1.9177
(0.5,0.1)	1.9040	2.8335	1.3193	1.9607	1.3215	1.9625	1.3191	1.9603
(1,0.1)	3.1849	4.0042	2.2053	2.7731	2.2063	2.7730	2.2077	2.7731

Table 2: The Table shows square root of average SE_{β_k} for each combination of (a, ν) over 2000 simulations. The data generating process used here is Model 7. $SE_{\beta_k} = \left(\hat{\beta}_k - \beta_k^O\right)^2$ is the squared estimation error of the estimated coefficient, where $\hat{\beta}_k$ is the estimated coefficient for β_k from either using P-PCA or PCA method and β_k^O is an oracle estimation for β_k. The oracle estimation for β_k is defined as the OLS estimation of β_k when the true latent common factors, factor loadings and idiosyncratic jump component are all known in advance.

	$k=1$		$k=2$		$k=3$		$k=4$		$k=5$	
	P-PCA	PCA	P-PCA	PCA	P-PCA	PCA	P-PCA	PCA	P-PCA	PCA
(a, ν)					$N=50$, $T=50$					
(0,0)	0.0436	0.0435	0.0430	0.0430	0.1301	0.1300	0.0400	0.0399	0.0386	0.0386
(0.1,0.01)	0.0560	0.1007	0.0561	0.1006	0.2365	0.5154	0.0430	0.0595	0.0431	0.0605
(0.5,0.01)	0.0841	0.1737	0.0843	0.1737	0.4155	0.8274	0.0598	0.1286	0.0612	0.1300
(0.1,0.05)	0.1216	0.2113	0.1217	0.2105	0.6148	1.0017	0.0909	0.1782	0.0910	0.1782
(1,0.01)	0.1152	0.2149	0.1168	0.2153	0.5816	0.9298	0.0843	0.1906	0.0861	0.1893
(0.1,0.1)	0.1990	0.2516	0.1985	0.2507	0.9038	0.8532	0.1921	0.2856	0.1917	0.2881
(0.5,0.05)	0.2071	0.2569	0.2060	0.2550	0.8020	0.6661	0.2174	0.3148	0.2157	0.3111
(1,0.05)	0.2439	0.2618	0.2448	0.2644	0.7292	0.6040	0.2927	0.3312	0.2942	0.3351
(0.5,0.1)	0.2557	0.2639	0.2570	0.2651	0.5992	0.5527	0.3310	0.3461	0.3297	0.3448
(1,0.1)	0.2646	0.2716	0.2660	0.2738	0.5462	0.5511	0.3491	0.3537	0.3501	0.3552
					$N=100$, $T=100$					
(0,0)	0.0140	0.0140	0.0142	0.0142	0.0307	0.0306	0.0161	0.0161	0.0156	0.0156
(0.1,0.01)	0.0142	0.0207	0.0139	0.0205	0.0301	0.0733	0.0169	0.0190	0.0172	0.0190
(0.5,0.01)	0.0153	0.0412	0.0154	0.0417	0.0328	0.1811	0.0171	0.0246	0.0172	0.0255
(0.1,0.05)	0.0287	0.1138	0.0285	0.1143	0.1454	0.6265	0.0209	0.0550	0.0209	0.0550
(1,0.01)	0.0187	0.0579	0.0192	0.0593	0.0777	0.2766	0.0178	0.0300	0.0182	0.0300
(0.1,0.1)	0.0678	0.2291	0.0678	0.2294	0.3732	1.0712	0.0456	0.1878	0.0452	0.1888
(0.5,0.05)	0.0551	0.2097	0.0550	0.2094	0.2998	0.9697	0.0343	0.1524	0.0347	0.1505
(1,0.05)	0.0909	0.2322	0.0912	0.2345	0.4644	0.9598	0.0605	0.2014	0.0596	0.2012
(0.5,0.1)	0.1529	0.2563	0.1527	0.2557	0.7385	0.6764	0.1245	0.3056	0.1230	0.3030
(1,0.1)	0.2324	0.2564	0.2348	0.2594	0.8832	0.6081	0.2326	0.3170	0.2358	0.3191
					$N=200$, $T=200$					
(0,0)	0.0048	0.0048	0.0048	0.0048	0.0096	0.0096	0.0072	0.0072	0.0073	0.0073
(0.1,0.01)	0.0048	0.0054	0.0048	0.0054	0.0097	0.0106	0.0073	0.0077	0.0073	0.0077
(0.5,0.01)	0.0054	0.0079	0.0056	0.0083	0.0109	0.0156	0.0078	0.0095	0.0075	0.0093
(0.1,0.05)	0.0053	0.0133	0.0053	0.0133	0.0104	0.0579	0.0076	0.0104	0.0076	0.0100
(1,0.01)	0.0059	0.0101	0.0059	0.0104	0.0111	0.0193	0.0085	0.0115	0.0080	0.0110
(0.1,0.1)	0.0059	0.0447	0.0059	0.0446	0.0114	0.2337	0.0081	0.0191	0.0080	0.0185
(0.5,0.05)	0.0084	0.0309	0.0091	0.0313	0.0168	0.1284	0.0102	0.0172	0.0100	0.0176
(1,0.05)	0.0141	0.0441	0.0141	0.0430	0.0272	0.1735	0.0142	0.0237	0.0140	0.0236
(0.5,0.1)	0.0182	0.1205	0.0187	0.1209	0.0758	0.5747	0.0145	0.0524	0.0143	0.0531
(1,0.1)	0.0385	0.1538	0.0382	0.1535	0.1239	0.7072	0.0240	0.0780	0.0240	0.0791
					$N=400$, $T=400$					
(0,0)	0.0015	0.0015	0.0015	0.0015	0.0030	0.0030	0.0032	0.0032	0.0032	0.0032
(0.1,0.01)	0.0016	0.0019	0.0016	0.0019	0.0031	0.0035	0.0034	0.0036	0.0032	0.0033
(0.5,0.01)	0.0018	0.0029	0.0018	0.0029	0.0034	0.0051	0.0034	0.0043	0.0034	0.0043
(0.1,0.05)	0.0018	0.0030	0.0018	0.0031	0.0034	0.0054	0.0035	0.0044	0.0033	0.0042
(1,0.01)	0.0023	0.0040	0.0021	0.0038	0.0040	0.0067	0.0037	0.0053	0.0037	0.0052
(0.1,0.1)	0.0021	0.0047	0.0021	0.0047	0.0038	0.0087	0.0037	0.0056	0.0035	0.0054
(0.5,0.05)	0.0033	0.0069	0.0033	0.0067	0.0058	0.0123	0.0046	0.0073	0.0046	0.0074
(1,0.05)	0.0056	0.0099	0.0058	0.0101	0.0098	0.0187	0.0069	0.0103	0.0067	0.0100
(0.5,0.1)	0.0057	0.0143	0.0056	0.0140	0.0098	0.0290	0.0066	0.0109	0.0066	0.0111
(1,0.1)	0.0116	0.0216	0.0117	0.0214	0.0223	0.0465	0.0116	0.0155	0.0112	0.0151

Table 3: The Table shows estimation results of the Fama and MacBeth regression (5) when different model settings are considered. In the entries are time-series averages of the estimated coefficients in the Fama and MacBeth regression. In the parentheses are t-statistics for the time-series averages of the estimated coefficients, which are obtained by using the Newey West standard errors. Monthly data from January-1970 to December-2012 are used for the estimations. In the last row are the time-series averages of R^2 of the estimated Fama and MacBeth regressions.

	Model															
	bench-1	bench-2	1a	2a	3a	4a	5a	6a	7a	8a	9a	10a	11a	12a	13a	14a
$Beta$	-0.289	-0.260		0.095												
	(-0.833)	(-0.819)		(0.337)												
$\ln(ME)$	-0.050	-0.057		-0.036							-0.041	-0.025	-0.037	-0.032	-0.038	-0.023
	(-1.311)	(-1.467)		(-1.017)							(-1.027)	(-0.586)	(-0.938)	(-0.774)	(-0.964)	(-0.563)
$\ln(BE/ME)$	0.231	0.199		0.242							0.232	0.247	0.235	0.221	0.251	0.253
	(2.410)	(2.089)		(2.559)							(2.338)	(2.360)	(2.376)	(2.135)	(2.559)	(2.543)
CGR(-7,-2)		0.207		0.154												
		(0.516)		(0.413)												
λ_1^{PCA}			0.063	0.076							0.084		0.094		0.087	
			(2.072)	(2.240)							(2.590)		(2.646)		(2.189)	
λ_2^{PCA}			0.044	0.046							0.054		0.033		0.057	
			(0.792)	(1.298)							(1.471)		(0.910)		(1.592)	
$\mathbb{E}_{t-1}\left[IVOL_t^{PCA}\right]$					0.063						0.045					
					(0.770)						(0.714)					
$\mathbb{E}_{t-1}\left[IVOL_t^{u}\right]$						0.033						-0.020				
						(0.563)						(-0.374)				
$IVOL_t^{PCA}$							0.261						0.227			
							(4.629)						(4.462)			
$IVOL_t^{u}$								0.235						0.215		
								(4.964)						(4.471)		
$\left\lvert\varepsilon_t^{IVOL^{PCA}}\right\rvert$									0.201						0.177	
									(2.421)						(3.203)	
$\left\lvert\varepsilon_t^{IVOL^{u}}\right\rvert$										0.195						0.166
										(2.857)						(2.685)
AvgR^2	0.294	0.318	0.104	0.376	0.036	0.043	0.037	0.055	0.023	0.036	0.351	0.297	0.354	0.309	0.343	0.289

Table 4: The Table shows estimation results of the Fama and MacBeth regression (5) when different model settings are considered. In the entries are time-series averages of the estimated coefficients in the Fama and MacBeth regression. In the parentheses are t-statistics for the time-series averages of the estimated coefficients, which are obtained by using the Newey West standard errors. Monthly data from January-1970 to December-2012 are used for the estimations. In the last row are the time-series averages of R^2 of the estimated Fama and MacBeth regressions.

	Model													
	1b	2b	3b	4b	5b	6b	7b	8b	9b	10b	11b	12b	13b	14b
$Beta$		0.039												
		(0.137)												
$\ln(ME)$		-0.047							-0.048	-0.045	-0.053	-0.042	-0.045	-0.044
		(-1.209)							(-1.223)	(-1.155)	(-1.370)	(-1.078)	(-1.201)	(-1.170)
$\ln(BE/ME)$		0.228							0.225	0.233	0.221	0.229	0.245	0.236
		(2.418)							(2.244)	(2.422)	(2.197)	(2.332)	(2.537)	(2.389)
CGR(-7,-2)		0.088												
		(0.233)												
λ_1^{P-PCA}	0.106	0.089							0.108	0.112	0.115	0.113	0.097	0.105
	(2.918)	(3.112)							(3.645)	(3.437)	(3.920)	(3.438)	(3.033)	(3.385)
λ_2^{P-PCA}	0.015	0.033							0.037	0.021	0.037	0.020	0.043	0.039
	(0.304)	(0.921)							(0.966)	(0.571)	(1.005)	(0.540)	(1.186)	(1.087)
$\mathbb{E}_{t-1}\left[IVOL_t^{P-PCA}\right]$			0.056						0.034		0.014			
			(0.617)						(0.457)		(0.174)			
$IVOL_t^{P-PCA}$				0.267						0.200		0.150		
				(4.191)						(3.572)		(2.901)		
$\mathbb{E}_{t-1}[ABSJ_t]$					3.645						1.564			
					(0.966)						(0.502)			
$ABSJ_t$						9.467						4.237		
						(3.816)						(2.239)		
$\left\lvert\varepsilon_t^{IVOL^{P-PCA}}\right\rvert$							0.190						0.168	0.127
							(2.369)						(2.622)	(2.150)
$\left\lvert\varepsilon_t^{ABSJ}\right\rvert$								8.764						6.209
								(3.730)						(3.600)
AvgR^2	0.096	0.374	0.035	0.034	0.031	0.032	0.019	0.033	0.348	0.348	0.361	0.367	0.337	0.359

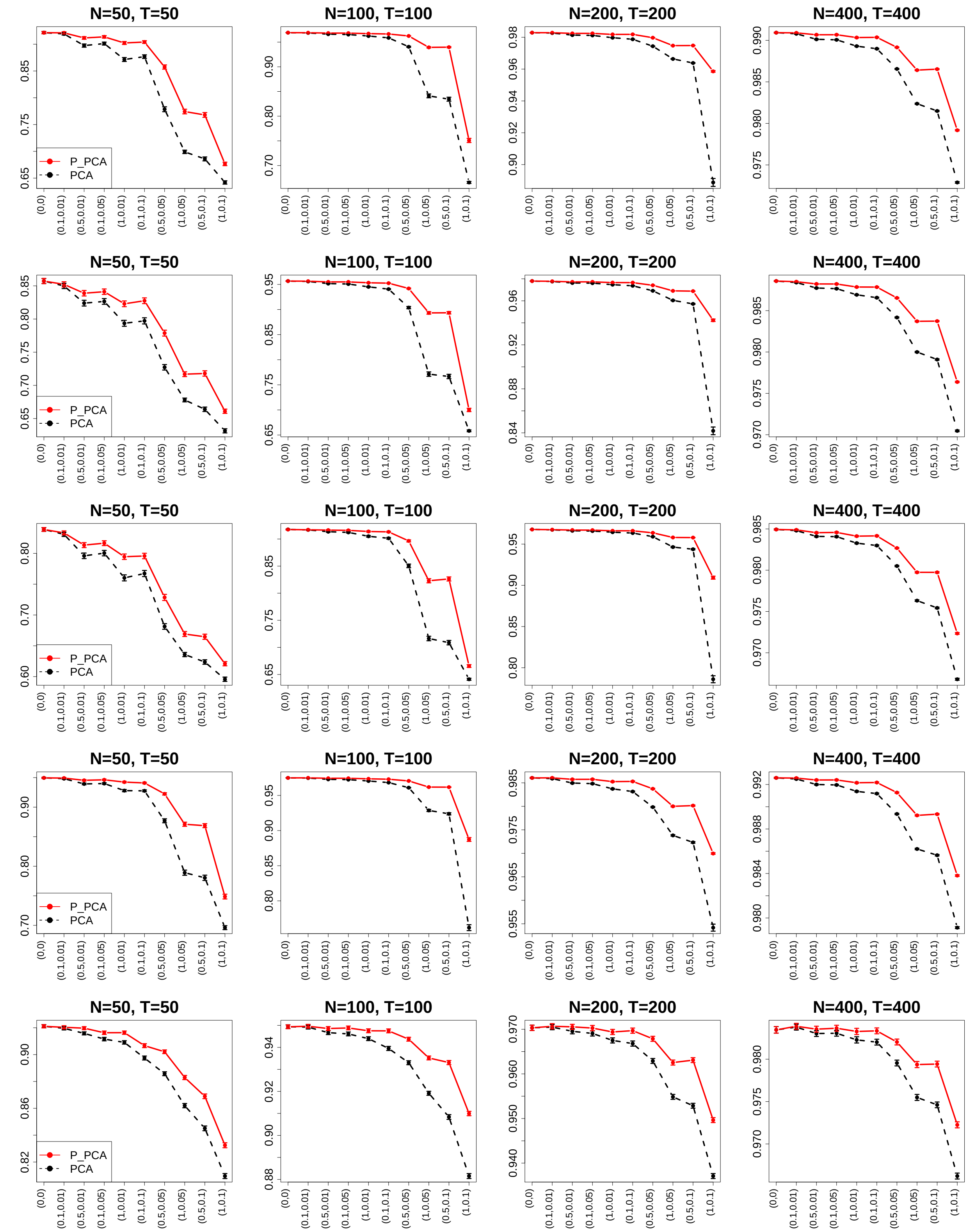

Figure 1: The Figure shows averages of Distance Correlations between the true and estimated factors (DCOR's) together with their 99% confidence intervals for each combination of (a, ν) over 2000 simulations. Plots in the first row to fifth row are corresponding to Model 1 to Model 5.

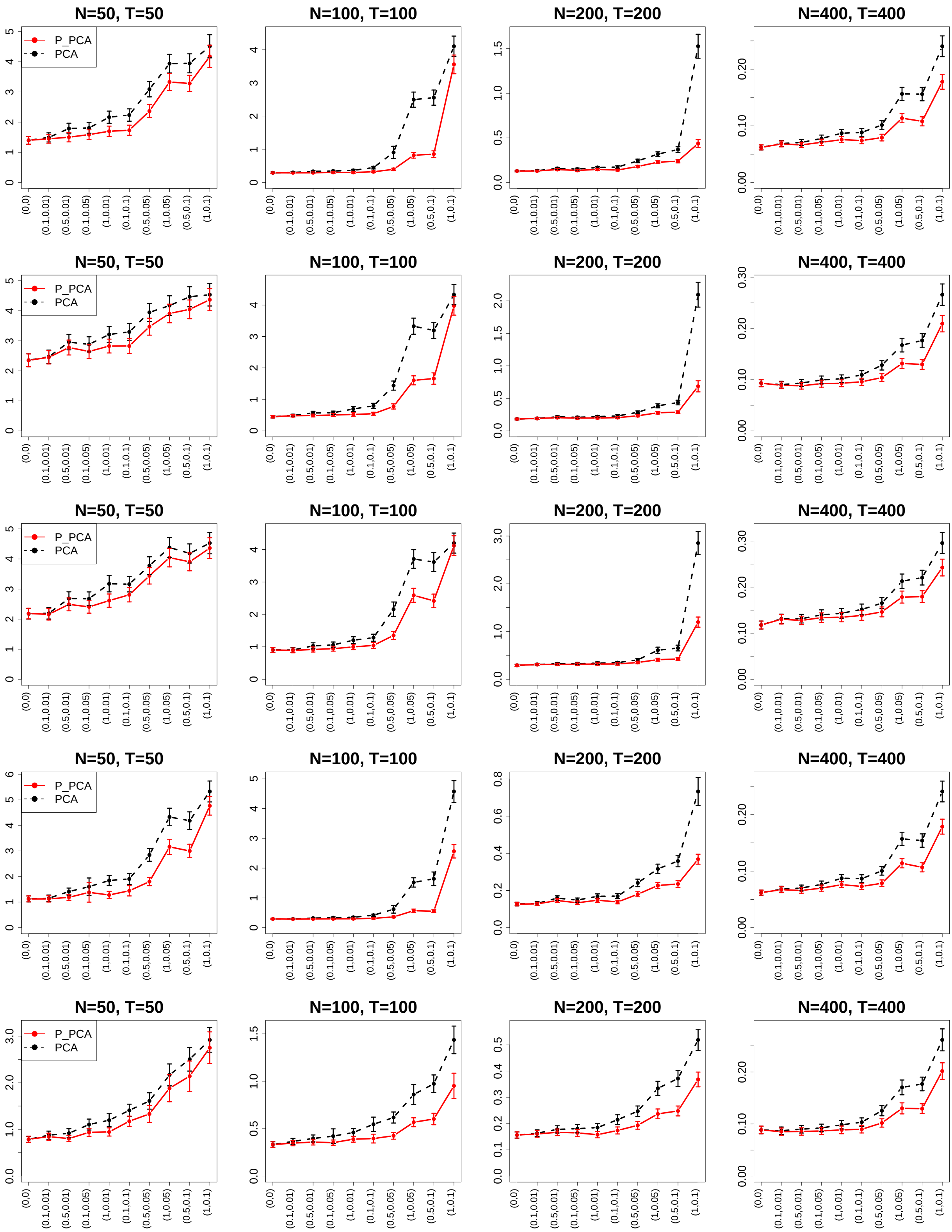

Figure 2: The Figure shows averages of squared predictive errors together with their 99% confidence intervals for each combination of (a, ν) over 2000 simulations. Plots in the first row to fifth row are corresponding to Model 1 to Model 5.

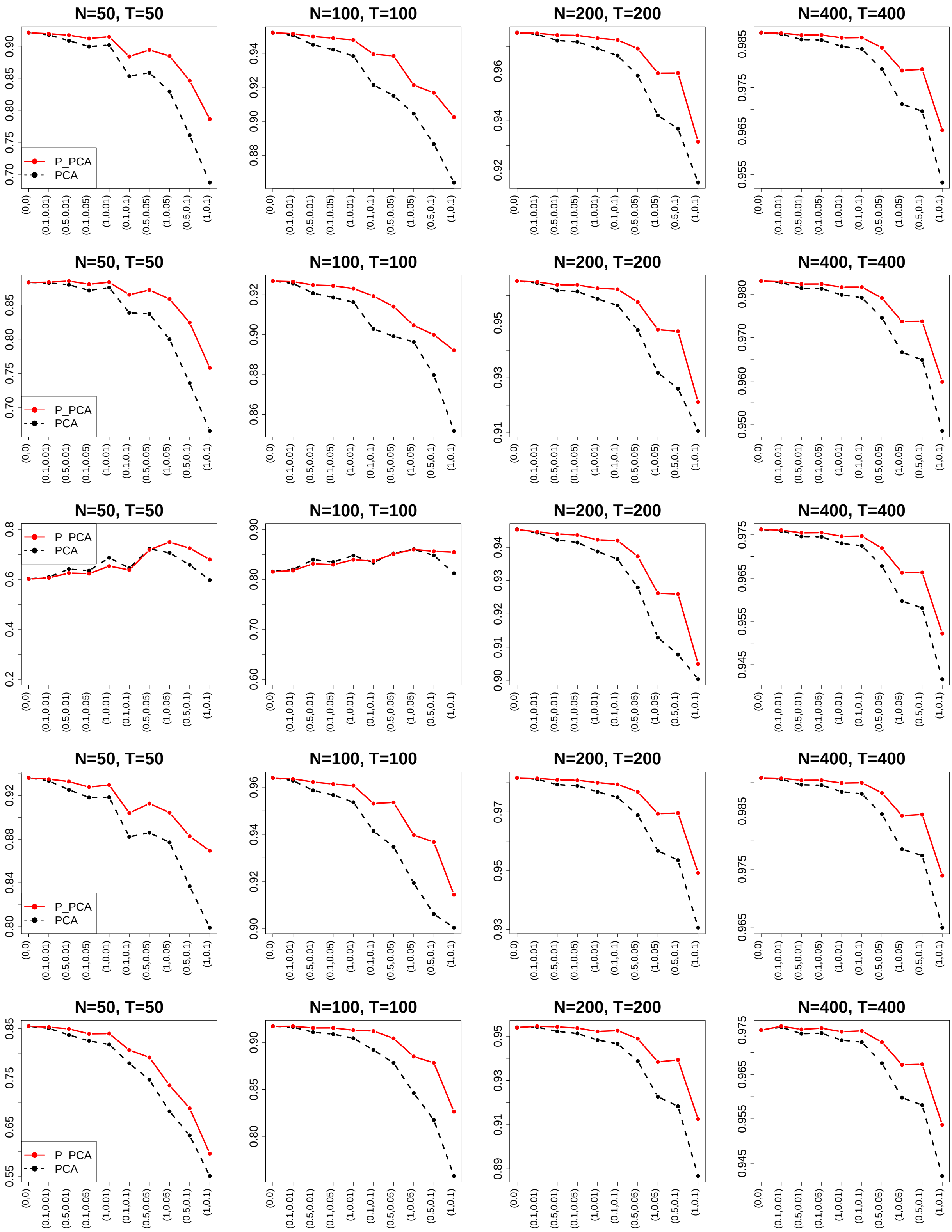

Figure 3: The Figure shows Trace R^2 between the true and estimated factors for each combination of (a, ν) from 2000 simulations. Plots in the first row to fifth row are corresponding to Model 1 to Model 5.

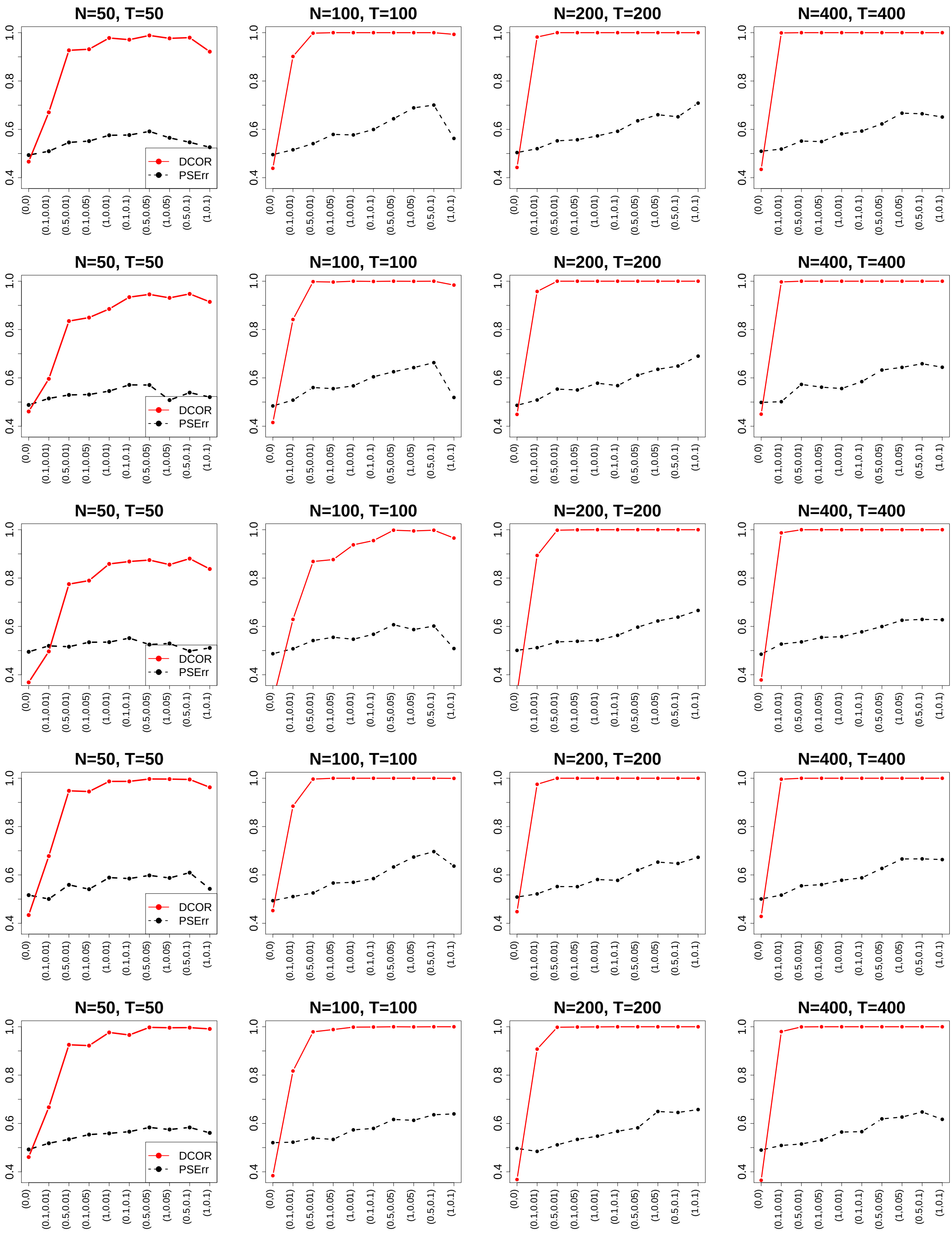

Figure 4: The Figure shows the proportion that P-PCA method has a higher Distance Correlation between the true and estimated factors (DCOR) or a lower squared predictive error than does PCA method for each combination of (a, ν) over 2000 simulations. Plots in the first row to fifth row are corresponding to Model 1 to Model 5.

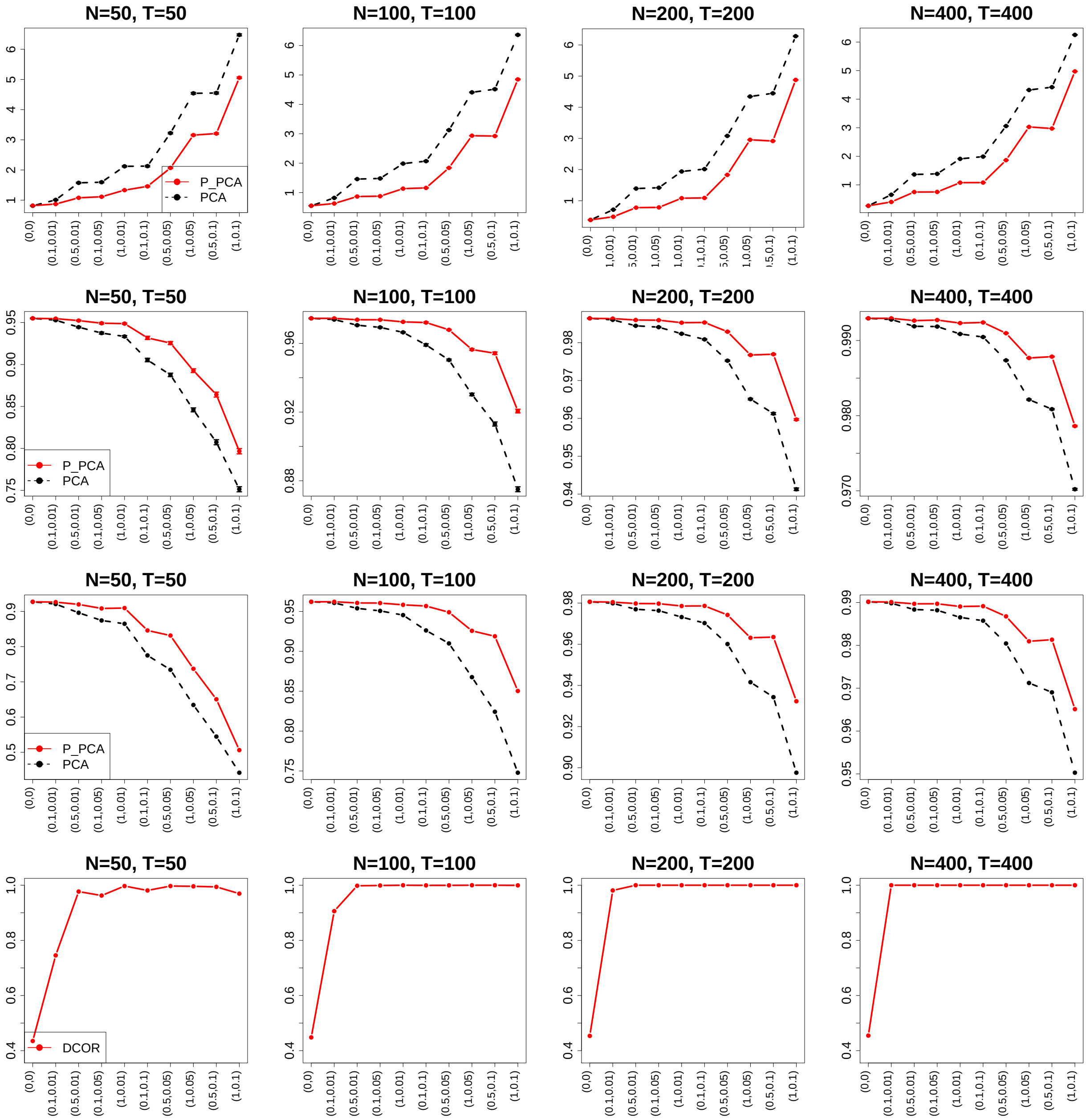

Figure 5: The Figure shows the l_2 distance between the estimated β and the oracle estimation of β (the first row), averages of Distance Correlations between the true and estimated factors (DCOR's) together with their 99% confidence intervals (the second row), Trace R^2 (the third row) and the proportion that P-PCA method has a higher DCOR than does PCA method (the fourth row) for each combination of (a, ν) over 2000 simulations. The data generating process used here is Model 6.

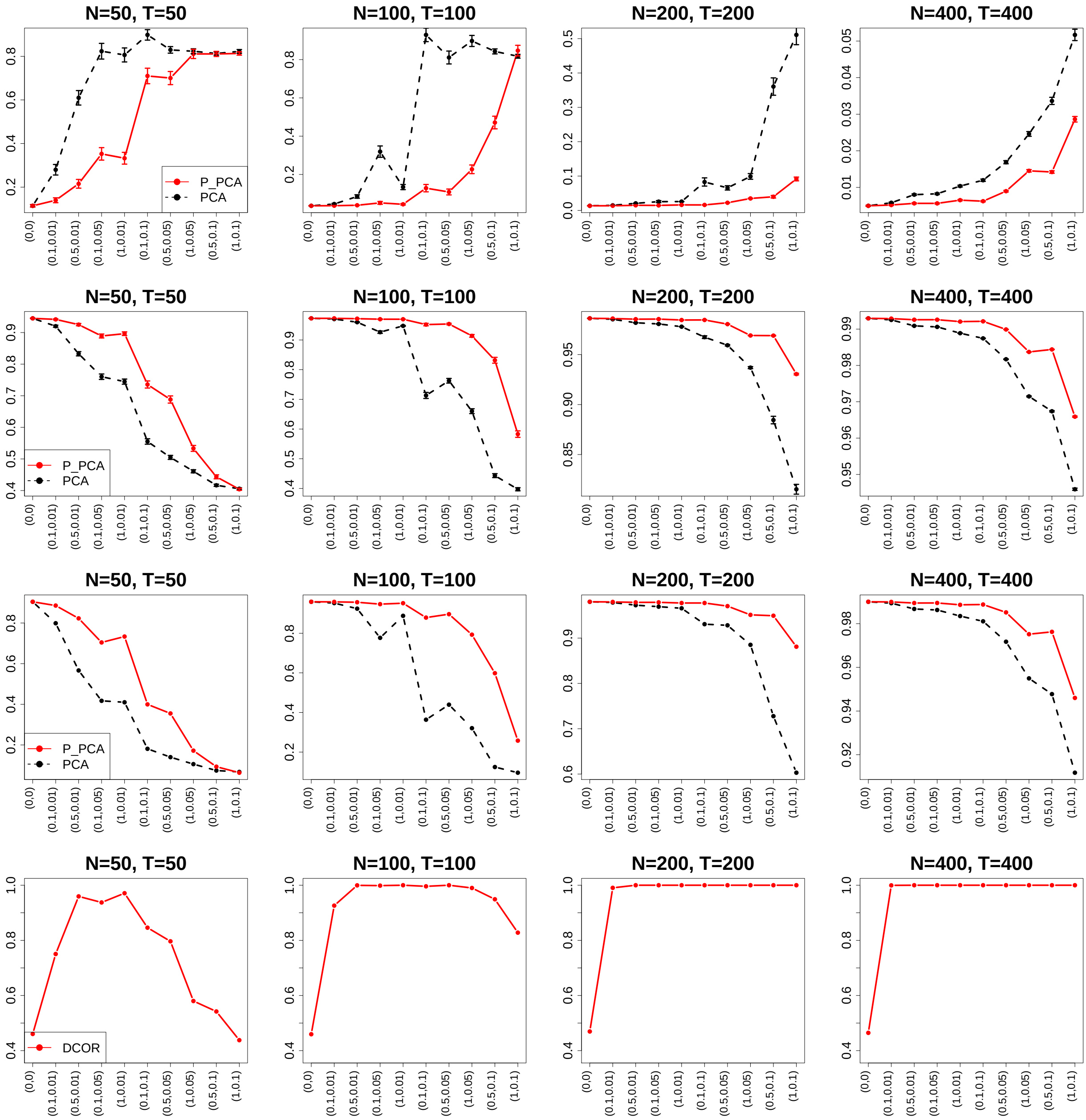

Figure 6: The Figure shows the l_2 distance between the estimated β and the oracle estimation of β (the first row), averages of Distance Correlations between the true and estimated factors (DCOR's) together with their 99% confidence intervals (the second row), Trace R^2 (the third row) and the proportion that P-PCA method has a higher DCOR than does PCA method (the fourth row) for each combination of (a, ν) over 2000 simulations. The data generating process used here is Model 7.

References

Ahn, S. C. and A. R. Horenstein (2013): "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81, 1203–1227.

Alessi, L., M. Barigozzi, and M. Capasso (2010): "Improved penalization for determining the number of factors in approximate factor models," *Statistics & Probability Letters*, 80, 1806–1813.

Amengual, D. and M. W. Watson (2007): "Consistent Estimation of the Number of Dynamic Factors in a Large N and T Panel," *Journal of Business & Economic Statistics*, 25, 91–96.

Ando, T. and J. Bai (2013): "Multifactor asset pricing with a large number of observable risk factors and unobservable common and group-specific factors," MPRA Paper 52785, University Library of Munich, Germany.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006): "The Cross-Section of Volatility and Expected Returns," *Journal of Finance*, 61, 259–299.

——— (2009): "High idiosyncratic volatility and low returns: International and further U.S. evidence," *Journal of Financial Economics*, 91, 1–23.

Bai, J. (2009): "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279.

Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

Bali, T. G. and N. Cakici (2008): "Idiosyncratic Volatility and the Cross Section of Expected Returns," *Journal of Financial and Quantitative Analysis*, 43, 29–58.

Bates, B. J., M. Plagborg-Mller, J. H. Stock, and M. W. Watson (2013): "Consistent factor estimation in dynamic factor models with structural instability," *Journal of Econometrics*, 177, 289 – 304.

Bernanke, B., J. Boivin, and P. S. Eliasz (2005): "Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach," *The Quarterly Journal of Economics*, 120, 387–422.

Campbell, J. Y., M. Lettau, B. G. Malkiel, and Y. Xu (2001): "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *The Journal of Finance*, 56, 1–43.

Candès, E. J., X. Li, Y. Ma, and J. Wright (2011): "Robust Principal Component Analysis?" *J. ACM*, 58, 11:1–11:37.

Cheng, X., Z. Liao, and F. Schorfheide (2014): "Shrinkage Estimation of High-Dimensional Factor Models with Structure Instability," NBER working paper 19792, NBER.

Corsi, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7, 174–196.

Donoho, D. L. and I. M. Jonhstone (1994): "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, 81, 425–455.

Fama, E. F. and K. R. French (1992): "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, 47, 427–465.

——— (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

Fama, E. F. and J. D. MacBeth (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–36.

Fu, F. (2009): "Idiosyncratic risk and the cross-section of expected stock returns," *Journal of Financial Economics*, 91, 24–37.

Jolliffe, I. T. (2002): *Principal Component Analysis*, Springer, second ed.

Ludvigson, S. C. and S. Ng (2009): "Macro Factors in Bond Risk Premia," *Review of Financial Studies*, 22, p. 5027 – 5067.

Merton, R. C. (1987): "A Simple Model of Capital Market Equilibrium with Incomplete Information," *Journal of Finance*, 42, 483–510.

Moench, E., S. Ng, and S. Potter (2013): "Dynamic Hierarchical Factor Model," *The Review of Economics and Statistics*, 95, 1811–1817.

Onatski, A. (2009): "Testing Hypotheses About the Number of Factors in Large Factor Models," *Econometrica*, 77, 1447–1479.

——— (2010): "Determining the Number of Factors from Empirical Distribution of Eigenvalues," *The Review of Economics and Statistics*, 92, 1004–1016.

Stock, J. H. and M. W. Watson (2002): "Forecasting using Principal components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, p. 1167 – 1179.

Szekely, G. J., M. L. Rizzo, and N. K. Bakirov (2007): "Measuring and testing dependence by correlation of distances," *The Annals of Statistics*, 35, 2769–2794.

Tibshirani, R. (1996): "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society, Series B*, 58, 267–288.

Wasserman, L. (2006): *All of Nonparametric Statistics*, Springer.

科技部補助專題研究計畫出席國際學術會議心得報告

日期：104 年 11 月 2 日

計畫編號	MOST 103-2410-H-004 -213 -		
計畫名稱	近似因子模型的有效估計-經由懲罰最小平方法		
出國人員姓名	顏佑銘	服務機構及職稱	國立政治大學國際經營與貿易學系，專任助理教授
會議時間	103 年 12 月 19 日至 103 年 12 月 20 日	會議地點	中國大陸廈門市，廈門大學
會議名稱	(中文) (英文) Conference on Financial Innovations and Bank Regulation 2014		
發表題目	(中文) (英文) Risk Evaluations with Robust Approximate Factor Models		

一、 參加會議經過

職於台灣當地時間 Dec-18-2014 早上 08:40 搭乘飛機從桃園國際中正機場出發，於中國大陸當地時間 Dec-18-2014 的早上 10:25 抵達廈門市。隨即搭乘計程車前往住宿地點，於卸下行李後、盥洗與稍做休息後，於下午兩點前往廈門大學會場(王亞南經濟研究院)進行報到和註冊，並領取大會資料和議程相關資料。

本次研討會從 Dec-19-2014 至 Dec-20-2014 為期一共兩天。大會學術交流主要為口頭報告(presentations)。本次大會兩天共排進 10 節議程，每一節次平均 3 場的口頭報告，職報告的場次為 Dec-19-2014，下午 02:00 至 03:30，第 6 節

(Dynamic risk modeling)。報告時間約為20分鐘(下午03:00-03:20)。之後大會有安排討論人(discussant) ，廈門大學王亞南經濟研究院的Haiqiang Chen教授評論此次報告。此外，這次研討會職也擔任評論人，評論的文章為”Are Market-Based Measures of Global Systemic Importance of Financial Institution Useful to Regulators and Supervisors?”報告人為University of Leeds的Qi Zhang教授。

在會議期間，職也分別遇到來自中研院的周雨田教授(報告文章作者之一)、台灣大學的沈中華教授及俞明德教授到場一起參與該次會議，感到格外的親切和榮幸。

除了既定的白天的學術交流外，藉由會議休息時間職亦與來自各國的學者能夠有機會認識彼此和交流。我們彼此交換名片，並介紹自己的研究領域和來自的學校和國家。對於職來說，能夠與各國的研究學者面對面的溝通和交流，這是一個很重要的經歷和體驗。在為期兩天會議結束後，職也於Dec-21-2014年早上08:45搭乘飛機從廈門市出發，於台灣當地時間Dec-21-2014的早上10:25抵達桃園中正國際機場。順利平安的完成本次的出國報告。

二、 與會心得

職這次除了報告論文外，也擔任評論人。評論的文章為“Are Market-Based Measures of Global Systemic Importance of Financial Institution Useful to Regulators and Supervisors?” 報告人為University of Leeds的Qi Zhang教授。這篇文章主要是探討四個常用的衡量系統風險指標是否在金融危機時能有效地提供預警的作用。這四個指標為ΔCoVar、ΔA_CoVar、SRISK指標及

EXSHORT。這篇文章的實證期間橫跨1998年的金融危機、亞洲金融危機及2007年的全球金融危機。作者發現這四個指標僅有ΔCoVar比較可以提供一致性的預警，但是它的功效在1998年金融危機及亞洲金融危機時還是有疑義。作者另外發現其他常用的財務指標，僅有公司市值較為有一致性的預警效果。職給這篇文章的建議是，是否可以將這些指標給組合起來，像是使用主成分分析法(PCA)及模型平均等等計量方法，截長補短，來建構出更具一致性預警能力的綜合指標。此外在實證分析方面，職也建議在回歸式中加入一些常用的總體經濟變數，可使得分析更為完善。

這次的研討會職也依據自己的喜好去旁聽其他學者的研究，如俞明德老師團隊應用尖端的小波分析法(Wavelet)在風險管理上面。另外還有沈中華老師對於影子銀行(Shadow Bank)對於公司資本及流動性的研究，都使職獲益匪淺。

三、發表論文全文或摘要

論文摘要(英文)

Approximate factor models and their extensions are widely used in risk evaluations due to their ability to extracting useful information from a large number of relevant variables. In these models, candidate predictors are typically subject to some common components. In this paper we evaluated risks by proposing a new method for robustly estimating the approximate factor models. We considered a class of approximate factor models in which the candidate predictors are additionally subject to idiosyncratic large uncommon components such as jumps or outliers. By assuming that occurrences of the uncommon components are rare, we developed an estimation procedure to simultaneously disentangle and estimate the common and uncommon components. Through intensive simulations, we compared the proposed method and traditional PCA method in terms of their finite sample efficiency. We then used the proposed method to investigate whether risks from the latent factors are priced for expected returns of Fama and French 100 size and book-to-market ratio portfolios. We found that while the risk from the common factor is priced for the 100 portfolios, the risks from the idiosyncratic factors are not. However, we also found that model uncertainty risks of the idiosyncratic factors are priced, suggesting that with effective diversifications, only the predictable idiosyncratic risks can be reduced, but the unpredictable ones may still exist.

四、建議

首先感謝科技部計畫 MOST 103-2410-H-004 -213 對於本次差旅費及其他經費的補助。另外政大國貿系同仁在行政程序上的支持，在此也一併感謝。對於職來說，參加本次研討會，除了觀摩各地優秀財金學者的最新研究外，對於研究人脈的建構也頗有助益。

五、攜回資料名稱及內容

研討會論文時程表(如附件)。

六、其他

來回機票單據、研討會邀請信及發表論文全文。

科技部補助計畫衍生研發成果推廣資料表

日期:2016/02/15

科技部補助計畫	計畫名稱: 近似因子模型的有效估計-經由懲罰最小平方法	
	計畫主持人: 顏佑銘	
	計畫編號: 103-2410-H-004-213-	學門領域: 數理與數量方法

無研發成果推廣資料

103年度專題研究計畫研究成果彙整表

計畫主持人：顏佑銘			計畫編號：103-2410-H-004-213-				
計畫名稱：近似因子模型的有效估計-經由懲罰最小平方法							
成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數（含實際已達成數）	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	1	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	1	0	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
其他成果（無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。）		無					

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

科技部補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估 ■達成目標 □未達成目標（請說明，以100字為限） □實驗失敗 □因故實驗中斷 □其他原因 說明：
2. 研究成果在學術期刊發表或申請專利等情形： 論文：□已發表 ■未發表之文稿 □撰寫中 □無 專利：□已獲得 □申請中 ■無 技轉：□已技轉 □洽談中 ■無 其他：（以100字為限）
3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以500字為限）