3-2019

Scaled PCA: A new approach to dimension reduction

Dashan HUANG
Singapore Management University, DASHANHUANG@smu.edu.sg

Fuwei JIANG
Central University of Finance and Economics

Guoshi TONG
Renmin University of China

Guofu ZHOU
Washington University in St. Louis

Scaled PCA: A New Approach to Dimension Reduction*

Dashan Huang, Fuwei Jiang, Guoshi Tong, Guofu Zhou

First draft: February 2019

This version: March 2019

*Huang: Lee Kong Chian School of Business, Singapore Management University, Singapore; Email: dashanhuang@smu.edu.sg. Jiang: School of Finance, Central University of Finance and Economics, China; Email: jfuwei@gmail.com. Tong: Hanqing Advanced Institute of Economics and Finance, Renmin University of China, China; Email: gstong@ruc.edu.cn. Zhou: Olin School of Business, Washington University in St. Louis, USA; Email: zhou@wustl.edu. We are grateful to Hui Chen, Zhanhui Chen, Anna Cieslak, Zhi Da, Jonas N. Eriksen, James D. Hamilton, Wenxin Huang, Ben Jacobsen, Scott Joslin, Toshio Kimura, Junye Li, Laura Xiaolei Liu, Roger Loh, Ruichang Lu, Sydney C. Ludvigson, Neil Pearson, Christopher Polk, David Rapach, Allan Timmermann, Jun Tu, Baolian Wang, Dacheng Xiu, Jiangmin Xu, Hong Zhang, Xiaoyan Zhang, Hao Zhou, Ning Zhu, and seminar and conference participants at the Central University of Finance and Economics, Peking University, Tsinghua PBC, and 2018 SMU Finance Summer Camp for valuable comments and suggestions.

Scaled PCA: A New Approach to Dimension Reduction

Abstract

We propose a novel modification to the popular principal component analysis (PCA) by scaling each predictor according to its predictive power on the target to be forecasted. Unlike the PCA that maximizes the variations of predictors, our scaled PCA, s-PCA, identifies factors that are particularly useful for forecasting the target. Asymptotically, the s-PCA factors converge to true latent factors that are important for the target. Empirically, we find that the s-PCA outperforms the popular PCA substantially in forecasting market return with a variety of investor sentiment proxies and forecasting inflation with a large panel of macro variables.

JEL codes: C22, C23, C53

Keywords: Forecasting, PCA, Big Data, Machine Learning, Supervised Learning

1 Introduction

The principal component analysis (PCA) is the oldest, dated back to Pearson (1901), and the most widely used dimension reduction method (Trevor, Robert, and Jerome, 2009). It transforms a large number of variables into orthogonal components so that the original data can be replaced by a few principal components. It has wide applications in all areas of science, including in particular management, finance, and economics. Gu, Kelly, and Xiu (2018), Giglio and Xiu (2018), Mullainathan and Spiess (2017), and Belloni, Chernozhukov, Fernández-Val, and Hansen (2017) are examples of its recent applications. Nowadays, in the age of big data, it is important in dealing with the "curse of dimensionality". Without dimension reduction, prediction by conventional multivariate regressions can be subject to over-fitting and suffer from poor out-of-sample performance. While the PCA is useful in reducing a large number of predictors to just a few combinations of them, one recognized weakness is that it ignores the target information completely.

In this paper, we propose a modification to the PCA by scaling each predictor according to its predictive power on the target to be forecasted. By design, our scaled PCA, s-PCA, puts more weights on predictors that are more important in forecasting the target. In contrast, the PCA puts equal weights on all predictors. While the PCA helps to summarize information from a large number of predictors into a few factors and filter out idiosyncratic noises, it ignores the target and is an unsupervised learning technique. If a predictor is noisier than others, it inevitably affects the weights of factors disproportionately. The s-PCA exactly corrects this deficiency by putting less weight on the noisier predictor. In a certain sense, our s-PCA method is designed to let the target guide dimension reduction.

The s-PCA extracts target-specific factors in two steps. First, it runs a predictive regression of the target on lagged values of each predictor to assess its predictive power. Instead of treating all the standardized predictors equally as in the PCA, we scale each standardized predictor by its predictive regression slope. Second, we extract factors from these scaled predictors by using the PCA method. Intuitively, the s-PCA tends to underweight those variables with weak predictive power, while overweight those with strong predictive power. The resulting s-PCA factors are thus more likely to outperform the PCA factors for forecasting, because they take into account the target in the dimension reduction procedure.

Theoretically, we derive the asymptotic properties of the s-PCA by imposing a factor structure on the set of predictors, where each predictor loads on both a latent factor that is important to the target and the latent factors irrelevant to the target. To gain analytical insights, we first consider a simple case of two predictors with different predictive powers. We solve explicitly for the s-PCA factor and show that it outperforms the PCA factor in terms of predicting the target. We then generalize it to the case of a large number of predictors and show the asymptotic consistency of the s-PCA factors under mild regularity conditions.

We next explore the forecasting performance of the s-PCA with two applications. First, we forecast the US stock market return using six sentiment proxies in Baker and Wurgler (2006). It is widely believed that investor sentiment is a driver of bubbles and crashes and hence supposed to predict the market return. However, a sentiment index measured by the first principal component of the six sentiment proxies fails to significantly predict the market return (see, e.g., Baker and Wurgler, 2007; Huang, Jiang, Tu, and Zhou, 2015). Second, we forecast the US inflation with 123 macro variables from FRED-MD, a widely used database in macroeconomic forecasting (McCracken and Ng, 2016). Although Stock and Watson (2002) find that information in a large number of macro variables can be effectively summarized by a small number of principal components, Boivin and Ng (2005) show that using macro PCA factors to forecast inflation remains challenging.

We apply the s-PCA method to these two applications and find that it substantially raises the predictive power in- and out-of-sample across 1- to 12-month horizons. Specifically, the in-sample predictive regression coefficients turn from statistically insignificant with the PCA factors to significant with the s-PCA factors in terms of the Newey-West t-statistics. The out-of-sample forecast performances as measured by the Campbell and Thompson (2008) out-of-sample R^2_{OS} also turn from negative or insignificant with the PCA factors to significantly positive with the s-PCA factors.

Our s-PCA method is related to the target PCA (t-PCA) of Bai and Ng (2008). In their pioneering study, Bai and Ng apply the PCA to a subset of predictors that are tested to have predictive power to the target under either soft or hard threshold rules. The presumption is that not every predictor is relevant to the target. Hence, they assign a binary (0-1) weight to each predictor to rule out the influence of uninformative ones. In contrast, the s-PCA approach assigns a continuous weight to each predictor, assuming that all are target-relevant but differ quantitatively in their predictive powers. The s-PCA methods also bypasses a need to pre-specify a threshold in selecting predictors and thus is a dense modelling technique in the sense of Chernozhukov, Hansen, and Liao (2017) and Giannone, Lenza, and Primiceri (2018). As pointed out by Bai and Ng (2008), the threshold selection can be sensitive to small changes in the data due to discreteness of the decision rule. In contrast, our s-PCA accommodates all predictors without selecting a cut-off level.

Our study is also related to existing approaches of dimension reduction with supervised learning, such as the partial least squares (PLS) regression, initially proposed by Wold (1966) and further developed by Kelly and Pruitt (2015). The PLS extracts factors from predictors with a three-pass regression filter to reduce common noise. Kelly and Pruitt (2013), Huang, Jiang, Tu, and Zhou (2015), and Light, Maslov, and Rytchkov (2017), among others, show that it outperforms the PCA in various applications. Our s-PCA method shares similar insights to increase predictive efficiency in dimension reduction through supervised learning. In their extensively empirical applications of all these methods to forecast bond returns with real-time macro data, Huang, Jiang, Tong, and Zhou (2019) find that the s-PCA performs better than both the t-PCA and the PLS.

The rest of the article is organized as follows. Section 2 introduces the s-PCA method and discusses its analytical properties. Section 3 explores two empirical applications. Section 4 concludes.

2 Methodology

In this section, we introduce the s-PCA in details. We show that the s-PCA analytically outperforms the PCA for forecasting in a simplified setting and the s-PCA factors are asymptotically consistent in a general setting.

2.1 s-PCA

Suppose there are N predictors, denoted by $X_t = (x_{1,t}, \cdots, x_{N,t})'$ for $i = 1, \cdots, N$ and $t = 1, \cdots, T$. We are interested in forecasting the target y_{t+h}, with a forecast horizon of h. Each individual predictor $x_{i,t}$ is a relevant but imperfect predictor of the target. Hence, relying on a few predictors is unlikely to capture well the dynamics of the target. However, including all the predictors in a standard multivariate regression suffers from the curse of dimensionality, which often leads to in-sample over fitting and poor out-of-sample forecasting performance, especially when N is large (see, e.g., Welch and Goyal, 2008; Ng, 2013).

To address the challenge of curse of dimensionality, a commonly adopted approach is to use the PCA to reduce the dimension. Mathematically, the PCA extracts common factors F_t as linear combinations of $(x_{1,t}, \cdots, x_{N,t})'$ via the following linear factor model,

$$x_{i,t} = \lambda_i' F_t + e_{i,t}, \tag{1}$$

where F_t are K-dimensional linear factors to be estimated, λ_i are K-dimensional parameters to be estimated, and $e_{i,t}$ is the idiosyncratic noise term. An intuitive way to understand F_t is that it provides a natural ranking of N mutually orthogonal linear combinations of X_t, which span the space of X_t. If the bulk of information in X_t can be summarized by a small number of linear factors, i.e. $K \ll N$, dimension reduction can be achieved.

With the PCA factors F_t, one can forecast the target as:

$$y_{t+h} = \alpha + \beta F_t + \varepsilon_{t+h}. \tag{2}$$

In practice before extracting F_t, the N predictors are typically standardized to have the same variance.

While the PCA via Equation (1) maximally represents the total variations of the N predictors, it ignores the target, and therefore is an un-supervised learning technique for dimension reduction. The resulting F_t are not necessarily the most relevant for prediction, especially when the individual predictors consist of common noises that are irrelevant to the target.

To overcome the PCA deficiency, we propose a more efficient method, s-PCA, which is designed to use the target information to guide dimension reduction and filter out both the idiosyncratic and common noise terms. Specifically, we estimate the s-PCA factor in two steps. In the first step, we form a panel of scaled predictors, $(\beta_1 x_{1,t}, \cdots, \beta_N x_{N,t})$, where the scaled coefficient β_i is the slope from the predictive regression of the target on the i-th (standardized) predictor:

$$y_{t+h} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{t+h}, \quad i = 1, \cdots, N. \tag{3}$$

In the second step, we perform the PCA on $(\beta_1 x_1, \cdots, \beta_N x_N)$ to extract target-specific factors, denoted by f_t,

$$\beta_i x_{i,t} = \lambda_i' f_t + e_{i,t}. \tag{4}$$

Intuitively, the scaled series $\beta_i x_{i,t}$ reflects the i-th predictor's predictive power on the target. A predictor with strong forecasting power receives a larger weight (i.e., higher absolute value of β_i), whereas a predictor with weak forecasting power receives a smaller weight. In short, the s-PCA performs the PCA on the scaled predictors, rather than on the raw predictors.

Finally, we predict the target with the s-PCA factors as:

$$y_{t+h} = \alpha + \beta f_t + \varepsilon_{t+h}. \tag{5}$$

It should be noted that, in our out-of-sample predictions, all the data standardization and model estimation are done recursively, so that the forecast at time t uses information available only up to time t.

2.2 Analytical Comparison with PCA

In this subsection, we provide some insights as to why the s-PCA is supposed to outperform the PCA for forecasting in a simplified setting, which permits analytical solutions to both the s-PCA and PCA.

We assume that there are only two predictors, $x_{1,t}$ and $x_{2,t}$, which have the following latent factor structure:

$$x_{1,t} = Z_t + E_t + u_{1,t}, \tag{6}$$

$$x_{2,t} = \eta_1 Z_t + \eta_2 E_t + u_{2,t}, \tag{7}$$

where Z_t is a latent predictor related to the target y_{t+h} through:

$$y_{t+h} = \alpha + \beta Z_t + e_{t+h}, \tag{8}$$

where E_t is the common noise component unrelated to the target, $\eta_1, \eta_2 \in [0,1]$ are the sensitivity parameters of $x_{2,t}$ to Z_t and E_t, and $u_{i,t}$ $(i = 1,2)$ are the idiosyncratic noises terms. Without loss of generality, we assume that Z_t, E_t, and $u_{i,t}$ $(i = 1,2)$ are independent of each other and have means zero and variances σ_Z^2, σ_E^2, and σ_u^2, where the idiosyncratic noises $u_{1,t}$ and $u_{2,t}$ have the same variance.

For notational simplicity, we ignore below the time subscript t. The covariance matrix of x_1 and x_2 is then

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_Z^2+\sigma_E^2+\sigma_u^2 & \eta_1\sigma_Z^2+\eta_2\sigma_E^2 \\ \eta_1\sigma_Z^2+\eta_2\sigma_E^2 & \eta_1^2\sigma_Z^2+\eta_2^2\sigma_E^2+\sigma_u^2 \end{pmatrix}. \tag{9}$$

Let $\sigma_1 = s\sigma_2$, , where $s \geq 1$ with our assumption $\eta_1, \eta_2 \in [0,1]$. Σ can be rewritten as:

$$\Sigma = \begin{pmatrix} s^2 & \rho h \\ \rho s & 1 \end{pmatrix} \sigma_2^2. \tag{10}$$

Hence, the first PCA factor is

$$F = w'x = w_1x_1 + w_2x_2,$$

where the weights, as a 2-dimensional vector, are proportional to the eigenvector of the larger eigenvalue of Σ as

$$w \quad \propto \quad \begin{pmatrix} 1 \\ \frac{1-s^2+\sqrt{(1-s^2)^2+4\rho^2s^2}}{2\rho s} \end{pmatrix}, \tag{11}$$

where $w_2 \leq w_1 = 1$, with equality when $s = 1$, i.e. x_1 and x_2 have the same variance.

When regressing the target y at time $t+h$ on x_1 or x_2 at time t, we have

$$\begin{aligned} \beta_1 &= \frac{\text{Cov}(y,x_1)}{\text{Var}(x_1)} = \frac{\text{Cov}(y,Z)}{\text{Var}(x_1)}, \\ \beta_2 &= \frac{\text{Cov}(y,x_2)}{\text{Var}(x_2)} = \eta_1\frac{\text{Cov}(y,Z)}{\text{Var}(x_2)}, \end{aligned}$$

which implies that $\beta_1 = \beta_2/(\eta_1 s^2)$. The covariance matrix of $\beta_1 x_1$ and $\beta_2 x_2$ is then given by

$$\tilde{\Sigma} = \begin{pmatrix} \beta_1^2\sigma_1^2 & \beta_1\beta_2\rho\sigma_1\sigma_2 \\ \beta_1\beta_2\rho\sigma_1\sigma_2 & \beta_2^2\sigma_2^2 \end{pmatrix} = \begin{pmatrix} 1 & \rho\eta_1 s \\ \rho\eta_1 s & \eta_1^2 h^2 \end{pmatrix} \beta_1^2\sigma_1^2. \tag{12}$$

Hence, the first s-PCA factor is $f = \tilde{w}_1x_1 + \tilde{w}_2x_2$, with weights proportional to the eigenvector of the larger eigenvalue of $\tilde{\Sigma}$ as

$$\tilde{w} \quad \propto \quad \begin{pmatrix} 1 \\ \frac{\eta_1^2s^2-1+\sqrt{(\eta_1^2s^2-1)^2+4\rho^2\eta_1^2s^2}}{2\rho\eta_1 s} \end{pmatrix}, \tag{13}$$

which suggests that $\tilde{w}_2 > \tilde{w}_1 = 1$ if η_1 is relatively large, i.e., the predictor x_2 is relatively more informative.

Comparing (11) and (13), the first PCA factor always allocates less weight on x_2 no matter if it is more informative or not when $\eta_1 < 1$ or $\eta_2 < 1$ or both. In contrast, the x_2 weight of the first s-PCA factor does depend on its predictive power. For example, when $\eta_1 = 1$ and $\eta_2 = 0$ (i.e., x_2 is more informative), (11) and (13) suggest that $w_2 < 1 < \tilde{w}_2$ (i.e., the s-PCA factor allocates a larger weight on x_2), and therefore, the s-PCA factor should have more power in predicting the target. When $\eta_1 = 0$ and $\eta_2 = 1$ (i.e., x_2 is a pure noise), (11) and (13) suggest that $\tilde{w}_2 = 0 < w_2$, and therefore, the PCA factor assigns too much weight on x_2 and will underperform the s-PCA factor. Since the s-PCA weights the more important predictor properly, its predictive power is generally higher than the PCA.

2.3 Asymptotic Consistency of s-PCA

In this subsection, we show that the s-PCA method is asymptotically consistent in extracting the target-specific factors. Suppose the target and predictors follow the following structure:

$$y_{t+h} = Z_t + e_{t+h}, \quad t = 1, \cdots, T, \tag{14}$$

$$X_{i,t} = \lambda_i' G_t + \varepsilon_{i,t} = m_i Z_t + j_i' E_t + \varepsilon_{i,t}, \quad i = 1, \cdots, N. \tag{15}$$

For simplicity, Z_t is assumed to be a one-dimensional latent factor that is important to the target (it can be also a combination of multiple factors), and E_t is an $(r-1)$-dimensional vector of latent common factors irrelevant to target. $G_t = (Z_t, E_t')'$ is an r-dimensional vector that collects both sets of the latent factors, and $\lambda_i = (m_i, j_i')'$ is the associated vector of factor loadings. e_{t+1} and $\varepsilon_{i,t}$ are the innovations of the target at $t+h$ and the idiosyncratic part of the predictor $x_{i,t}$, respectively.

We make the following assumptions on the above factor model. For ease of exposition, denote $\Lambda = (\lambda_1, \cdots, \lambda_N)'$, $M = (m_1, \ldots, m_N)'$, $G = (G_1, \ldots, G_T)'$, $Z = (Z_1, \ldots, Z_T)'$, and $E = (E_1, \ldots, E_T)'$.

Assumption 2.1. *(i) y_{t+h} and $x_{i,t}$ for each $i = 1, \cdots, N$ are jointly normally distributed.*

(ii) $\mathbb{E}(Z_t e_{t+h}) = 0$, $\mathbb{E}(E_t e_{t+h}) = 0$, $\mathbb{E}(e_{t+h}\varepsilon_{i,t}) = 0$, and $\mathbb{E}(Z_t \varepsilon_{i,t}) = 0$.

Assumption 2.2. *There exists a positive constant c, such that for all N and T,*

(i) As $N \to \infty$, $\frac{1}{N}\Lambda'\Lambda \xrightarrow{p} \Sigma_\lambda > 0$, $\Sigma_\lambda = \begin{pmatrix} \sigma_{\lambda,Z} & 0_{1\times(r-1)} \\ 0_{(r-1)\times 1} & \Sigma_{\lambda,E} \end{pmatrix}$, and $\max_{1\le i\le N} \mathbb{E}\|\lambda_i\|^{2q} \le c$ for some $q \ge 4$.

(ii) $\mathbb{E}\|G_t\|^{2q+\varepsilon} \leq c$ *for some* $\varepsilon > 0$, $q > 4$, *and all* t*; as* $T \to \infty$, $\frac{1}{T}\sum_{t=1}^{T} G_t G_t' \xrightarrow{p} \Sigma_G > 0$, *where* Σ_G *is positive definite.*

(iii) Let $\gamma_N(s,t) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}(\varepsilon_{it}\varepsilon_{is})$ *and* $\xi_{st} = \frac{1}{N}\sum_{i=1}^{N}[\varepsilon_{it}\varepsilon_{is} - \mathbb{E}(\varepsilon_{it}\varepsilon_{is})]$, $\max_{1\leq s,t\leq T} N^2\mathbb{E}|\xi_{st}|^4 \leq c$ *and* $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\gamma_N(s,t)\|^2 \leq c$.

Assumption 2.1 on normal distribution is commonly stated in the factor analysis literature, while Assumption 2.2 specifies regularity conditions on the time-series and cross-sectional dependance of factors and loadings akin to Bai and Ng (2002).

Proposition 2.1. *Under Assumptions 2.1 and 2.2, and denote* $C_{NT} = \min(\sqrt{N}, \sqrt{T})$*, there exists an asymptotically non-singular* $H = O_p(1)$*, such that*

$$\frac{1}{\sqrt{T}}\|\hat{f}H^{-1} - Z\| = O_p(C_{NT}^{-1}).$$

The proof is provided in the Appendix. The proposition says that the true factor is asymptotically identified, establishing the consistency of the s-PCA factors.

3 Empirical Applications

In this section, we show that the s-PCA outperforms the PCA for prediction with two empirical applications.

The first application is predicting the stock market return with investor sentiment. Baker and Wurgler (2006) propose a top-down sentiment index, which is the first PCA factor of six individual measures of investor sentiment. The PCA captures their common component and filters out idiosyncratic noises in the six measures, consisting of 1) closed end fund discount rate (CEFD), 2) share turnover de-trended by past 5 years's average (TURN), 3) number of IPOs (NIPO), 4) first day return of IPO (RIPO), 5) dividend premium measured as difference in market-to-book ratios of dividend payers and nonpayers (PDND), and 6) equity share in new issuance (S). Baker and Wurgler (2006) and subsequent studies, such as Stambaugh, Yu, and Yuan (2012), document a strong effect of investor sentiment on smaller, hard-to-value, and difficult-to-arbitrage firms. However, when predicting the market return, the evidence is weak or insignificant (Huang, Jiang, Tu, and Zhou, 2015).

The second application is predicting inflation with a large panel of macro variables. Boivin and Ng (2005) find that the PCA factor extracting from macro variables has much weaker power in predicting inflation than in predicting industrial production. In this paper, we consider monthly US economic indictors from the FRED-MD database, which is maintained by St. Louis Fed.[1] As described in more detail in McCracken and Ng (2016), this database represents a recent effort by the authors and Fed staffs to compile a standard macroeconomic database to facilitate "big-data" macro research. It extends the widely used Stock-Watson dataset (Stock and Watson, 2006) and covers broad economic categories such as output and income, labor force and unemployment, consumption expenditure and housing indicators, money stock and credit, and price indices. We collect a total of 123 indicators with no missing data during the span of January 1960 to December 2017. The detailed list and transformation codes to ensure stationarity of each macro variable is provided in the data appendix of McCracken and Ng (2016).

The market return is measured by the S&P 500 index excess return, and inflation is calculated as the log change in the US CPI all items. The forecast horizon varies from 1 to 12 months.

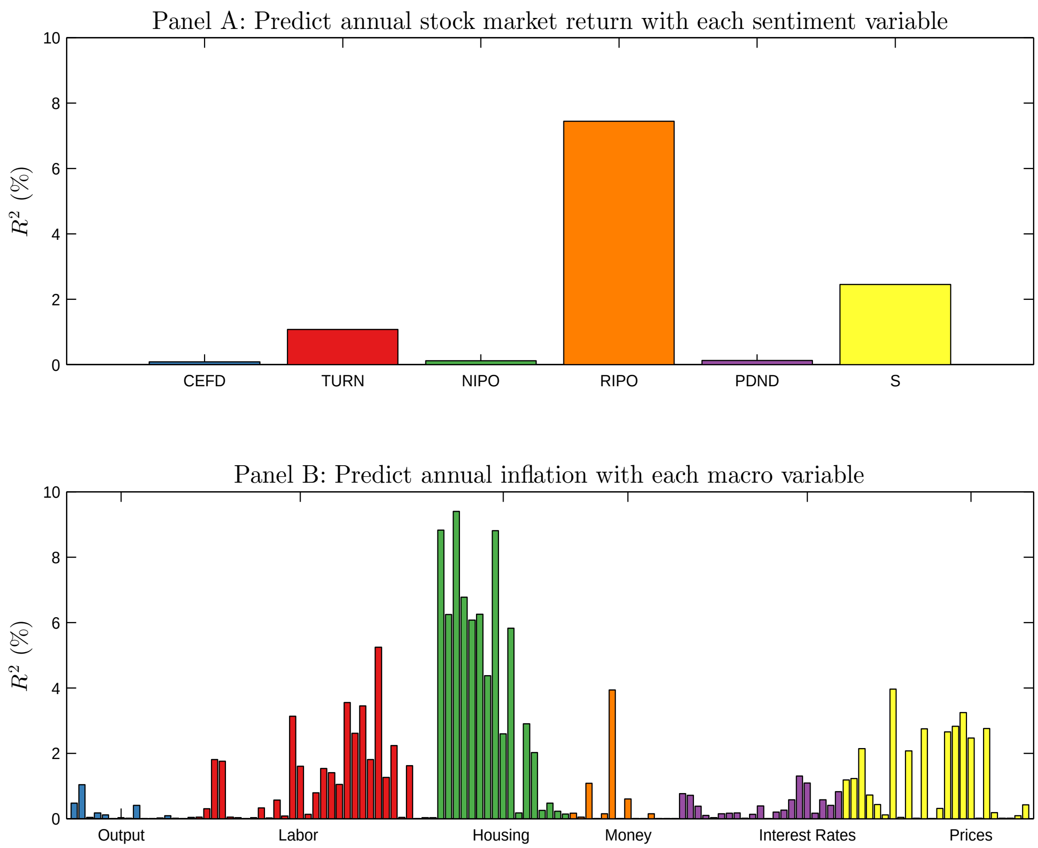
Figure 1 presents the predictive power of each variable in predicting the target. Specifically, Panel A plots the R^2s of regressing 12-month ahead market return on each individual sentiment proxy separately. We observe that the first day return of IPO (RIPO) displays the highest R^2 of 7%, followed by equity share in new issuance (S) and turnover (TURN) (about 2% and 1%). The predictive power of other sentiment proxies are rather weak, with close to zero R^2. Analogously, Panel B plots the R^2s of predicting 12-month ahead inflation with each macro variable. To highlight the incremental predictive power of macro variables, we control for lagged values of inflation with number of lags determined by the BIC and use the residual of inflation as the target. It is evident that among different categories, housing variables on average have the highest predictive power, which are followed by labor market condition, prices and one of the money variables. The other categories, such as the output and interest rates, have marginal predictive power.

3.1 In-sample Results

In this subsection, we explore the in-sample forecasting performance of the PCA and s-PCA factors. The sample period for predicting market return is 1965:07–2016:12, and for predicting inflation is 1960:01–2017:12. We focus on the first order factors of the two approaches because they capture the majority of predictive power.

[1] The dataset is updated in a timely manner and can be downloaded for free from the website http://research.stlouisfed.org/econ/mccracken/sel/.

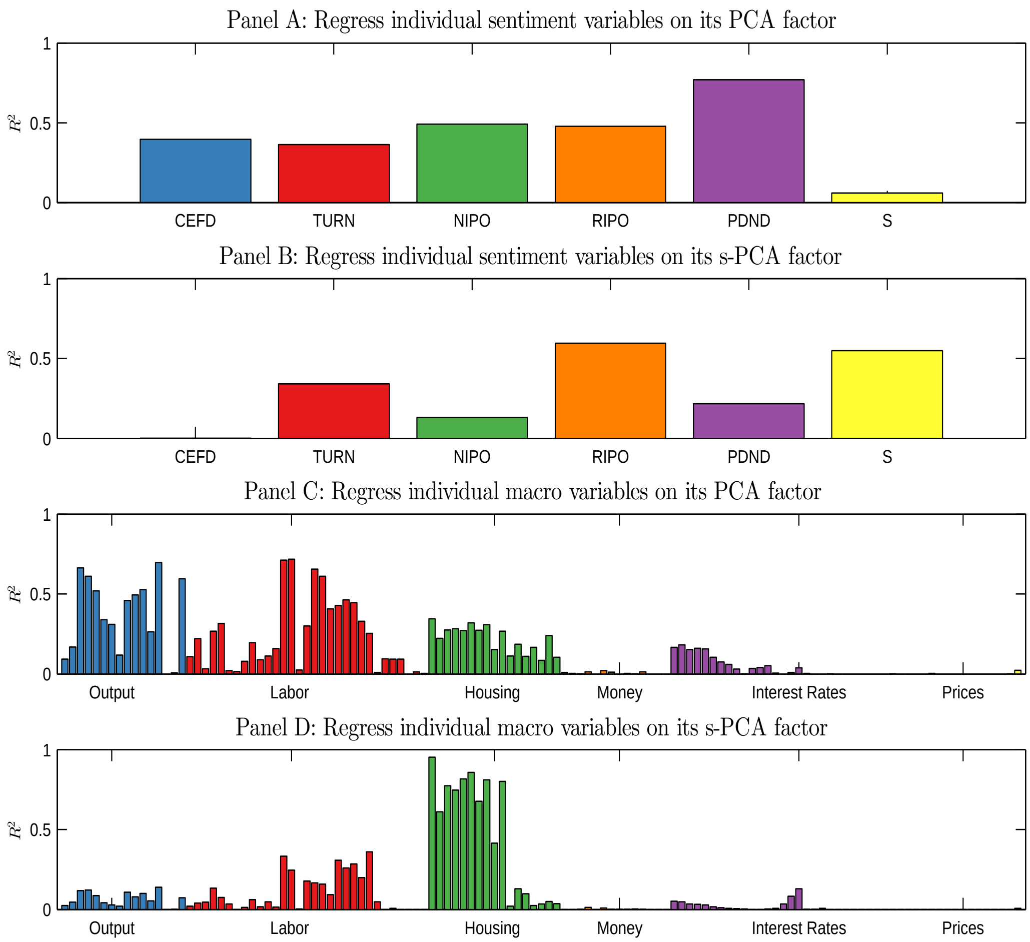
Figure 1: In-Sample R^2s of Predicting Market Return and Inflation with Individual Predictors
Panel A plots the in-sample R^2s (in percentage) of predicting 12-month ahead market return using each of the six individual sentiment proxies from Baker and Wurgler (2006), consisting of closed end fund discount rate (CEFD), share turnover (TURN), number of IPOs (NIPO), first day return of IPO (RIPO), dividend premium (PDND), and equity share in new issuance (S). Panel B plots the in-sample R^2s (in percentage) of predicting the one year ahead inflation using each of the 123 individual macro variables from FRED-MD data set of McCracken and Ng (2016), consisting of output and income (No. 1-16), labor market (No. 17-47), consumption and housing (No. 48-64), money and credit (No. 65-78), interest and exchange rate (No. 79-99), and prices (No. 100-123). The sample period is 1965:07–2016:12 in Panel A, and 1960:01–2017:12 in Panel B.



To zoom in the composition of these factors, Figure 2 plots the R^2s of regressing each predictor on the PCA and s-PCA factors, respectively. A higher R^2 suggests that the factor captures more variation and therefore loads more heavily on the predictor. Panel A shows that the PCA factor loads most heavily on dividend premium (PDND) with R^2 above 60%, which is followed by number of IPO (NIPO), first day IPO return (RIPO), closed end fund discount (CEFD) and turnover (TURN) with R^2s of around 40%, and has the

Figure 2: R^2s of Regressing Individual Predictors on PCA and s-PCA Factors

This figure plots the R^2s of regressing individual sentiment proxies and macro variables on the PCA and s-PCA factors, respectively. The sample period for sentiment proxies is 1965:07–2016:12 and for macro variables is 1960:01–2017:12. Sentiment proxies include closed end fund discount rate (CEFD), turnover (TURN), number of IPOs (NIPO), first day return of IPO (RIPO), dividend premium (PDND), and equity share in new issuance (S). Macro variables include output and income (No. 1-16), labor market (No. 17-47), consumption and housing (No. 48-64), money and credit (No. 65-78), interest and exchange rate (No. 79-99), and prices (No. 100-123).



least loading on equity share of new issuance (S). In contrast, Panel B displays a distinct pattern of loadings for the s-PCA factor. First, the R^2s are on average lower than those of the PCA factor, indicating that the s-PCA factor does not maximally explain the total variation of the sentiment proxies. A comparison with Panel A of Figure 1 further reveals that the pattern of loadings for the s-PCA factor is tilting towards those variables with higher predictive powers. In particular, the equity share of new issuance, due to its relatively high predictive power, is having the highest loading in constructing the s-PCA factor.

Panel C suggests that the PCA factor loads most heavily on output and labor market variables with an average R^2 of around 40%. Panel D reveals that the s-PCA factor loads most heavily on the housing variables. The average R^2s are again lower than that of the PCA factor, implying that the s-PCA factor does not explain the majority of the total variation of macro variables. The R^2 pattern in Panel D also echoes the pattern in Panel B of Figure 1, as housing variables display higher predictive power.

We next conduct in-sample predictive regression tests for both the PCA and s-PCA factors. Table 1 reports the regression slopes, Newey-West t-statistics, and R^2s of the PCA and s-PCA factors in predicting market return and inflation. The factors are extracted from 6 individual sentiment proxies in Panel A and from 123 macro variables in Panel B. For ease of interpretation, we normalize all the factors to have a zero mean and a standard deviation of one.

Table 1: In-Sample Results of Forecasting Market Return and Inflation

This table reports the slopes, Newey-West t-statistics, and R^2s of the PCA and s-PCA factors in predicting market return and inflation, respectively. The factors are extracted from 6 sentiment proxies over 1965:07–2016:12 in Panel A, and from 123 macro variables over 1960:01–2017:12 in Panel B. Since inflation is persistent, we control for lagged inflation to highlight the incremental forecasting power of macro variables. Forecast horizon ranges from 1 to 12 months. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	PCA			s-PCA		
horizon	β	t-stat	R^2	β	t-stat	R^2
Panel A: Forecast market return						
1 month	-0.24	-1.23	0.00	-0.60^{***}	-3.54	0.02
3 months	-0.64	-1.23	0.01	-1.61^{***}	-3.87	0.05
6 months	-1.14	-1.11	0.01	-2.84^{***}	-3.18	0.07
12 months	-1.56	-0.82	0.01	-4.27^{**}	-2.13	0.07
Panel B: Forecast inflation						
1 month	0.01	1.16	0.00	0.03^{***}	2.66	0.02
3 months	0.03	0.69	0.00	0.10^{***}	2.40	0.03
6 months	0.07	0.77	0.00	0.20^{**}	2.18	0.04
12 months	0.30	1.62	0.03	0.56^{***}	2.74	0.09

Panel A shows that the PCA sentiment factor fails to significantly predict market return across horizons, consistent with Huang, Jiang, Tu, and Zhou (2015). In contrast, the s-PCA factor significantly forecasts market returns with t-statistic ranging from -3.54 to -2.13 and R^2s from 2% to 7% across horizons. For instance, a one standard deviation increase in the s-PCA sentiment factor is associated with a 0.60% decrease in the next month expected return. In contrast, a one standard deviation increase in the PCA sentiment factor is associated with only a 0.24% decrease in the next month expected return.

Panel B indicates that the macro PCA factor exhibits weak predictive power for inflation at the 1- to 12-month horizons. The s-PCA factor manages to raise the predictive power with t-statistic ranging from 2.66 to 2.74 and R^2 from 2% to 9% across horizons. In particular, a one standard deviation increase in the s-PCA factor is associated with a 0.56% increase of one year ahead inflation. In contrast, a one standard deviation increase in the PCA factor is associated with a 0.30% increase of one year ahead inflation.

3.2 Out-of-sample Results

This subsection examines the out-of-sample predictive power of the PCA and s-PCA factors in the above two applications. Although the in-sample tests provide more efficient estimation using all available data, the out-of-sample exercise helps to avoid in-sample over-fitting and mimics the real time forecasting situation faced by the forecasters.

We use a recursive estimation procedure with expanding windows to obtain the out-of-sample forecasts. In particular, we first divide the full sample into an in-sample training period consisting of the first m observations, and an out-of-sample evaluation period consisting of the last q observations. Take the s-PCA factor as an example. At the initial forecasting time of month m, we are restricted to use $x_{i,t}$, with $i = 1, \cdots, N$ and $t = 1, \cdots, m$, to estimate the s-PCA factor, denoted by f_t for $t = 1, \cdots, m$, through equations (3) and (4). The initial out-of-sample forecast based on f_m is then given by $\hat{y}_{m+h} = \hat{\alpha}_m + \hat{\beta}_m f_m$, where $\hat{\alpha}_m$ and $\hat{\beta}_m$ are estimated through the factor forecasting equation (5) using target and estimated factor up to month m, i.e. $\{y_t\}_{t=h+1}^{m}$ and $\{f_t\}_{t=1}^{m-h}$. Proceeding in this manner through the end of the out-of-sample evaluation period, we generate a series of $q-h$ out-of-sample forecasts of the target, $\{\hat{y}_{t+h}\}_{t=m}^{T-h}$. For the purpose of comparison, we also generate a series of $q-h$ out-of-sample forecasts by using the PCA factor.

We calculate the Campbell and Thompson (2008) R^2_{OS} statistic to evaluate the out-of-sample performance, which is defined as

$$R^2_{OS} = 1 - \frac{\sum_{t=m}^{T-h} (y_{m+h} - \hat{y}_{m+h})^2}{\sum_{t=m}^{T-h} (y_{m+h} - \bar{y}_{m+h})^2}, \tag{16}$$

where $\{\bar{y}_{t+h}\}_{t=m}^{T-h}$ in the numerator is the benchmark out-of-sample forecast by historical average or autoregressive forecasting model. Thus, the R^2_{OS} statistic can be interpreted as the percentage reduction in mean squared prediction error (MSPE) for the forecasts generated by latent factors relative to the benchmark. If a factor carries out-of-sample predictive information, we should observe a positive R^2_{OS} as an indication of higher forecast accuracy. We use the Clark and West (2007) statistic to test whether the percentage reduction in MSPE by the s-PCA factor against the benchmark forecast is statistically significant. This amounts to a test on the null hypothesis of $R^2_{OS} \leq 0$ against the one-sided alternative of $R^2_{OS} > 0$.

Table 2: Out-of-Sample R^2_{OS}s of of Forecasting Market Return and Inflation

This table reports out-of-sample R^2_{OS}s (in percentage) of the PCA and s-PCA factors in predicting market return and inflation, respectively. The factors are extracted from 6 sentiment proxies in Panel A and from 123 macro variables in Panel B. All the factors and predictive regressions are recursively estimated with an expanding window scheme. R^2_{OS} is computed against the historical average as benchmark in Panel A and against an autoregressive model with lagged inflation as benchmark in Panel B. Statistical significance for R^2_{OS} is based on the p-value of the Clark and West (2007) MSPE-adjusted statistic for testing $H_0 : R^2_{OS} \leq 0$ against $H_A : R^2_{OS} > 0$. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Forecast horizon ranges from 1 to 12 months. The out-of-sample period is 1985:01–2016:12 in Panel A and 1985:01–2017:12 in Panel B.

	1 month		3 months		6 months		12 months	
	R^2_{OS} (%)	p-value	R^2_{OS} (%)	p-value	R^2_{OS} (%)	p-value	R^2_{OS} (%)	p-value
Panel A: Forecast market return								
PCA	0.21	0.11	0.43*	0.08	0.36	0.22	-1.87	0.85
s-PCA	1.21***	0.01	3.31***	0.00	5.21***	0.00	2.97*	0.07
Panel B: Forecast inflation								
PCA	-1.30	0.61	-4.92	0.59	-9.01	0.62	-10.30	0.48
s-PCA	4.99***	0.00	8.82***	0.00	10.42***	0.00	7.48***	0.00

Panel A of Table 2 presents the out-of-sample R^2_{OS} of predicting market return with the PCA and s-PCA factors. We use the data over 1965:07 to 1984:12 as the initial estimation period so that the forecast evaluation period spans from 1985:01 to 2016:12. The R^2_{OS}s are computed against the benchmark forecasts based on historical average, which is a very stringent out-of-sample benchmark for stock return predictability according to Welch and Goyal (2008). We observe that the PCA sentiment factor generates positive R^2_{OS}s from 0.21% to 0.36% for a horizon of 1 to 6 months, but delivers a negative R^2_{OS} when the forecast horizon is 1 year. The positive R^2_{OS}s are not statistically significant based on the Clark and West (2007) tests except for the forecast horizon of 3 months. In contrast, the R^2_{OS}s with the s-PCA factor increase substantially to a range of 1.21% to 5.21%, suggesting that it delivers a lower MSPE than the the PCA factor. Moreover, this outperformance is statistically significant.

Panel B of Table 2 presents the out-of-sample R^2_{OS}s of predicting inflation with the macro PCA and s-PCA factors. The initial estimation period is set to be 1960:01 to 1984:12, so that the forecast evaluation period spans from 1985:01 to 2017:12. We compute R^2_{OS}s generated by a factor forecasting model augmented with lagged inflation against the benchmark inflation forecasts based only on autoregressive models with lagged inflation. The number of lags in both the competing and benchmark models are determined by the BIC. We observe that the PCA factor generates negative R^2_{OS}s across the forecast horizons. This weak out-of-sample predictive performance is consistent with our previous in-sample results in Panel B of Table 1. Instead, the s-PCA factor manages to substantially raise the R^2_{OS}s to a range from 4.99% to 7.48% across horizons, all of which are statistically significant.

In sum, this section empirically shows that the s-PCA is a better method for dimension reduction than the PCA in terms of predictability.

4 Conclusion

In this paper, we propose a novel dimension reduction technique, s-PCA, to extract factors for predicting a target with many predictors. In comparison with the PCA, the s-PCA is a supervised learning technique that uses the target to guide the factor extraction. It is also a simple modification to the standard PCA: it extracts principal components from predictors scaled by their predictive powers instead of from the raw values of predictors.

We derive analytical properties of the s-PCA factors by assuming a linear latent factor structure on the predictors. In a simple two predictors case, we show that compared with the PCA, the s-PCA factor tilts towards the more informative predictor. In a more general case with a large number of predictors, we show that the s-PCA factors are asymptotically consistent.

We then apply the s-PCA method to two empirical applications: forecasting market return with sentiment proxies and forecasting inflation with a large number of macro variables. While the PCA factors generally fail to display significant predictive power, the s-PCA factors exhibit significant predictive ability both in-and out-of-sample across 1- to 12-month forecast horizons.

References

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 135–172.

Bai, J., Ng, S., 2008. Forecasting economic time series using targeted predictors. Journal of Econometrics 146, 304–317.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. Journal of Finance 61, 1645–1680.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. Journal of Economic Perspectives 21, 129–152.

Belloni, A., Chernozhukov, V., Fernández-Val, I., Hansen, C., 2017. Program evaluation and causal inference with high-dimensional data. Econometrica 85, 233–298.

Boivin, J., Ng, S., 2005. Understanding and comparing factor-based forecasts. International Journal of Central Banking 1, 117–152.

Campbell, J. Y., Thompson, S. B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies 21, 1509–1531.

Chernozhukov, V., Hansen, C., Liao, Y., 2017. A lava attack on the recovery of sums of dense and sparse signals. Annals of Statistics 45, 39–76.

Clark, T. E., West, K. D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291–311.

Giannone, D., Lenza, M., Primiceri, G. E., 2018. Economic predictions with big data: The illusion of sparsity. Working Paper.

Giglio, S., Xiu, D., 2018. Inference on risk premia in the presence of omitted factors. Working Paper.

Gu, S., Kelly, B. T., Xiu, D., 2018. Empirical asset pricing via machine learning. Working paper.

Huang, D., Jiang, F., Tong, G., Zhou, G., 2019. Are bond returns predictable with real-time macro data? Working paper.

Huang, D., Jiang, F., Tu, J., Zhou, G., 2015. Investor sentiment aligned: A powerful predictor of stock returns. Review of Financial Studies 28, 791–837.

Kelly, B., Pruitt, S., 2013. Market expectations in the cross-section of present values. Journal of Finance 68, 1721–1756.

Kelly, B., Pruitt, S., 2015. The three-pass regression filter: A new approach to forecasting using many predictors. Journal of Econometrics 186, 294–316.

Light, N., Maslov, D., Rytchkov, O., 2017. Aggregation of information about the cross section of stock returns: A latent variable approach. Review of Financial Studies 30, 1339–1381.

McCracken, M. W., Ng, S., 2016. Fred-md: A monthly database for macroeconomic research. Journal of Business & Economic Statistics 34, 574–589.

Mullainathan, S., Spiess, J., 2017. Machine learning: an applied econometric approach. Journal of Economic Perspectives 31, 87–106.

Ng, S., 2013. Chapter 14 - variable selection in predictive regressions. In: Elliott, G., Timmermann, A. (eds.), *Handbook of Economic Forecasting*, vol. 2 of *Handbook of Economic Forecasting*, pp. 752–789.

Pearson, K., 1901. Liii. on lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2, 559–572.

Stambaugh, R. F., Yu, J., Yuan, Y., 2012. The short of it: Investor sentiment and anomalies. Journal of Financial Economics 104, 288–302.

Stock, J. H., Watson, M. W., 2002. Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics 20, 147–162.

Stock, J. H., Watson, M. W., 2006. Forecasting with many predictors. In: Elliot, G., Granger, C. W., Timmermann, A. (eds.), *Handbook of Economic Forecasting*, Elsevier, vol. 1, chap. 10, pp. 515–554.

Trevor, H., Robert, T., Jerome, F., 2009. The elements of statistical learning: data mining, inference, and prediction.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455–1508.

Wold, H., 1966. Estimation of principal components and related models by iterative least squares. Multivariate Analysis. New York: Academic Press.

Appendix: Proof of Proposition 2.1

We first rewrite the scaled predictors as

$$\begin{aligned} x^*_{i,t} &= \hat{\beta}_i x_{it} = \left(\frac{\frac{1}{T}\sum_{t=1}^{T} y_{t+h} x_{i,t}}{\frac{1}{T}\sum_{t=1}^{T} x_{i,t}^2} \right) x_{i,t} \\ &= \left(\frac{\frac{1}{T}\sum_{t=1}^{T} (Z_t + e_{t+h})(m_i Z_t + j_i' E_t + \varepsilon_{i,t})}{\frac{1}{T}\sum_{t=1}^{T} x_{i,t}^2} \right) (m_i Z_t + j_i' E_t + \varepsilon_{i,t}). \end{aligned}$$

Consider the numerator

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}(Z_t+e_{t+h})(m_iZ_t+j_i'E_t+\varepsilon_{i,t})\\
=&m_i\frac{1}{T}\sum_{t=1}^{T}Z_t^2+\underbrace{m_i\frac{1}{T}\sum_{t=1}^{T}Z_te_{t+h}}_{\mathbb{E}(Z_te_{t+h})=0}+\underbrace{j_i'\frac{1}{T}\sum_{t=1}^{T}E_tZ_t}_{\text{Under normal assumption, PCA components are asymptotically independent}}\\
&+\underbrace{j_i'\frac{1}{T}\sum_{t=1}^{T}E_te_{t+h}}_{\mathbb{E}(E_te_{t+h})=0}+\underbrace{\frac{1}{T}\sum_{t=1}^{T}Z_t\varepsilon_{i,t}}_{\mathbb{E}(Z_t\varepsilon_{i,t})=0}+\underbrace{\frac{1}{T}\sum_{t=1}^{T}e_{t+h}\varepsilon_{i,t}}_{\mathbb{E}(e_{t+h}\varepsilon_{i,t})=0}\\
\xrightarrow{p}&m_i\frac{1}{T}\sum_{t=1}^{T}Z_t^2 \quad \text{as } T\to\infty.
\end{aligned}
$$

Now we have the scaled predictors as

$$
x_{i,t}^* = \hat{\beta}_i x_{i,t} = \left(\frac{\frac{1}{T}\sum_{t=1}^{T}Z_t^2}{\frac{1}{T}\sum_{t=1}^{T}x_{i,t}^2}\right)\left(m_i^2Z_t+m_ij_i'E_t+m_i\varepsilon_{i,t}\right)+o_p(1)=m_i^*Z_t+\varepsilon_{i,t}^*+o_p(1),
$$

where $m_i^*=m_i^2\times\left(\frac{\frac{1}{T}\sum_{t=1}^{T}Z_t^2}{\frac{1}{T}\sum_{t=1}^{T}x_{i,t}^2}\right)$ and $\varepsilon_{i,t}^*=\left(\frac{\frac{1}{T}\sum_{t=1}^{T}Z_t^2}{\frac{1}{T}\sum_{t=1}^{T}x_{i,t}^2}\right)(m_ij_i'E_t+m_i\varepsilon_{i,t})$.

Rewriting the above equation in vector form, we have

$$
x_i^* = Zm_i^*+\varepsilon_i^*+o_p(1), \tag{17}
$$

where $x_i^*=\left(x_{i,1}^*,\ldots,x_{i,T}^*\right)'$ and Z, ε_i^* are defined in an analogous manner.

We then estimate the s-PCA factor $\hat{f}$ through the eigenvalue decomposition as,

$$
\left[\frac{1}{NT}\sum_{i=1}^{N}x_i^*x_i^{*\prime}\right]\hat{f}=\hat{f}V_{NT}, \tag{18}
$$

where $\frac{1}{T}\hat{f}'\hat{f}=I_1$ and V_{NT} is the largest eigenvalue of the matrix inside the square brackets.

Combing the above two equations, we have

$$
\begin{aligned}
\hat{f}V_{NT}=&\frac{1}{NT}Zm_i^*\varepsilon_i^{*\prime}\hat{f}+\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i^*m_i^*Z'\hat{f}+\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i^*\varepsilon_i^{*\prime}\hat{f}+\frac{1}{NT}\sum_{i=1}^{N}Zm_i^*m_i^*Z'\hat{f}+o_p(1)\\
=&I_1+I_2+I_3+\frac{1}{NT}\sum_{i=1}^{N}Zm_i^*m_i^*Z'\hat{f}+o_p(1).
\end{aligned}
$$

It follows that $\hat{f}V_{NT}-Z\left(\frac{1}{N}\sum_{i=1}^{N}m_i^*m_i^*\right)\frac{Z'\hat{f}}{T}=I_1+I_2+I_3+o_p(1)$.

Let $H=\left(\frac{1}{N}\sum_{i=1}^{N}m_i^*m_i^*\right)\frac{Z'\hat{f}}{T}V_{NT}^{-1}$. It is easy to show that $H=O_p(1)$ and is asymptotically nonsingular.

Then we have $\hat{f}H^{-1} - Z = [I_1 + I_2 + I_3]\left(\frac{Z'\hat{f}}{T}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} m_i^* m_i^*\right)^{-1} + o_p(1)$ and

$$\frac{1}{\sqrt{T}}\left\|\hat{f}H^{-1} - Z\right\| \leq \frac{1}{\sqrt{T}}\left(\|I_1\| + \|I_2\| + \|I_3\|\right)\left\|\left(\frac{Z'\hat{f}}{T}\right)^{-1}\right\|\left\|\left(\frac{1}{N}\sum_{i=1}^{N} m_i^* m_i^*\right)^{-1}\right\| + o_p(T^{-1/2}).$$

It remains to analyze the $\|I_l\|$ for $l = 1,2,3$. For I_1, we have

$$\frac{1}{\sqrt{T}}\|I_1\| = \left\|\frac{1}{\sqrt{T}}\frac{1}{NT}ZM^{*\prime}\varepsilon^*\hat{f}\right\| \leq \frac{1}{\sqrt{N}}\left\|\frac{f}{\sqrt{T}}\right\|\left\|\frac{\hat{f}}{\sqrt{T}}\right\|\frac{1}{\sqrt{NT}}\left\|M^{*\prime}\varepsilon^*\right\| = O_p(N^{-1/2}),$$

where $M^* = (m_1, ..., m_N)'$ and $\varepsilon^* = (\varepsilon_1^*, ..., \varepsilon_N^*)$ and we use the fact that $\frac{1}{NT}\|M^{*\prime}\varepsilon^*\|^2 = O_p(1)$ by Assumption 2(iii). Similarly, we can show that $\frac{1}{\sqrt{T}}\|I_2\| = O_p(N^{-1/2})$.

For I_3, we have

$$\begin{aligned}\frac{1}{T}\|I_3\|^2 =& \frac{1}{T}\left\|\frac{1}{NT}\varepsilon^{*\prime}\varepsilon^*\hat{f}\right\|^2 \leq 2\sum_{t=1}^{T}\left\|T^{-3/2}\sum_{s=1}^{T}\gamma_N(s,t)\hat{f}_s'\right\|^2 + 2\sum_{t=1}^{T}\left\|T^{-3/2}\sum_{s=1}^{T}\xi_{st}\hat{f}_s'\right\|^2\\ =& 2\left\|I_3(a)\right\| + 2\left\|I_3(b)\right\|,\end{aligned}$$

where $\gamma_N(s,t)$ and ξ_{st} are defined in Assumption 2.2(iii). Note that

$$\|I_3(a)\| = T^{-1}\left(T^{-1}\sum_{s=1}^{T}\left\|\hat{f}_s\right\|^2\right)\left(T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\gamma_N(s,t)\|^2\right) = O_p(T^{-1})$$

and

$$\|I_3(b)\| = N^{-1}\left(T^{-1}\sum_{s=1}^{T}\left\|\hat{f}_s\right\|^2\right)\left(T^{-2}N\sum_{s=1}^{T}\sum_{t=1}^{N}\|\xi_{s,t}\|^2\right) = O_p(N^{-1})$$

by the fact that $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\gamma_N(s,t)\|^2 \leq c$ and $E(\|\xi_{st}\|^2) \leq N^{-1}c$ under Assumption 2(iii). Then $\frac{1}{\sqrt{T}}\|I_3\| = O_p(T^{-1/2} + N^{-1/2})$. Denote $C_{NT} = \min(\sqrt{N}, \sqrt{T})$, it follows that

$$\frac{1}{\sqrt{T}}\|\hat{f}H^{-1} - Z\| = O_p(C_{NT}^{-1}).$$

This completes the proof.