Manuscript submitted to *The Econometrics Journal*, pp. 1–20.

Feasible weighted projected principal component analysis for semiparametric factor models

Sung Hoon Choi[†]

[†]*Department of Economics, University of Connecticut, Storrs, CT 06269, USA.*
E-mail: sung_hoon.choi@uconn.edu

Summary Various factor estimation procedures have been developed, based on the latent factor model. They often consider general conditions that allow for correlations and heteroskedasticity. However, the conventional principal components (PC) method does not efficiently estimate the parameters. It also does not accommodate additional covariates, which explain the unknown factors, even if they are available. In particular, a few aggregated macroeconomic variables can be used as covariates in diffusion index forecasts. To account for these features, I propose the feasible weighted projected principal component (WPPC) analysis, based on semi-parametric factor models, and also establish its asymptotic properties. In addition, I apply the WPPC method to the diffusion index forecasting model. Finally, I investigate the performance of the WPPC estimator in forecasting excess bond returns using U.S. bond market and macroeconomic data.

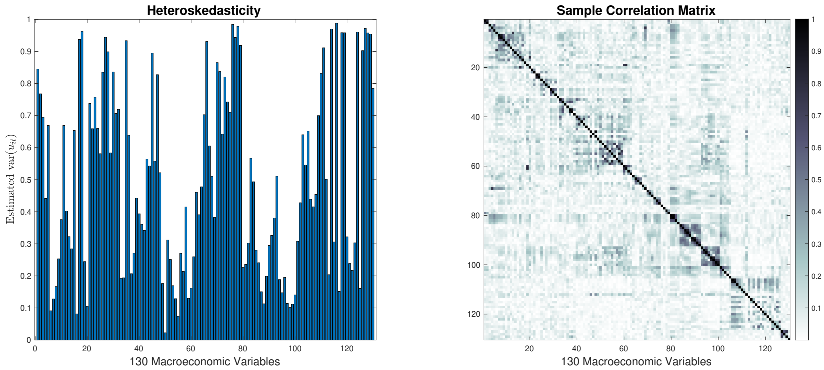
Keywords: *Conditional sparsity, Cross-sectional correlation, Heteroskedasticity, Semi-parametric factor models, Unknown factors.*

1. INTRODUCTION

Factor analysis and principal component (PC) analysis are powerful tools for dimension reduction, and they have been extensively studied (Ait-Sahalia and Xiu, 2017; Bai and Ng, 2002; Bernanke et al., 2005; Fan et al., 2013; Stock and Watson, 2002; Su and Wang, 2017). In many economic and financial applications, it is crucial to accurately estimate latent factors. For example, one would like to understand precisely how individual stock depends on latent factors to examine its relative performance and risks. Also, extracting more accurate factors improves forecasting performance with large datasets. In the literature, high-dimensional factor models often require the idiosyncratic error components to be cross-sectionally heteroskedastic and correlated. In other words, the large error covariance matrix is a non-diagonal matrix, and the diagonal entries may vary widely. For example, Figure 1 shows that cross-sectional heteroskedasticity and correlations exist in a large dataset of 130 macroeconomic variables. However, the conventional PC method, such as Bai (2003) and Stock and Watson (2002), essentially treats the error term to be homoskedastic and uncorrelated across the cross-sectional individuals. Hence, it may lead to inefficient estimators under cross-sectional heteroskedasticity with unknown dependence structures.

Based on high-dimensional factor models, several latent factor estimation procedures have been developed to improve estimation efficiency. For example, Choi (2012) and Breitung and Tenhofen (2011) considered the generalized PC method using weighted least squares for static and dynamic factor models, respectively. When analyzing macroeconomics or the financial market, we often observe that a few observable proxies (e.g., Fama-French factors, firm-specific characteristics, etc.) explain the factor structure. Then, we

Figure 1. Cross-sectional heteroskedasticity and correlations in macroeconomic variables



Note: The first panel shows the estimated error variance for each variable using the estimated residuals by the conventional PC method using 130 macroeconomic variables. The second panel displays an image of the sample error correlation matrix (in elementwise absolute value) with scaled colors.

can employ a "supervised" version of the PC method using the extra information. Specifically, the eigenvectors (i.e., the principal components) are supervised by the additionally observed covariates. This is very suitable in the context of asset pricing and economic forecasts. For example, Fan et al. (2016) proposed a projected principal component (PPC) analysis, which employs the PC method to the projected data matrix onto a given linear space spanned by the covariates. Because the projection approach removes noise components, it helps to estimate the factors more accurately than the conventional PC method. Kelly et al. (2020) considered a generalized factor model structure that allows time-varying loadings, which are explained by instrumental variables. In addition, Pelger and Xiong (2022) used an observed state process as the additional information to estimate the latent factors and state-varying loadings. Recent studies, such as Chen et al. (2020); Fan et al. (2021); Kelly et al. (2019), have also been conducted in this direction.

This paper proposes a novel factor estimation procedure that incorporates cross-sectional heterogeneity and additional covariates. Specifically, I consider the latent factor models that observed time-specific covariates (e.g., aggregated macroeconomic variables) partially explain the unknown factors. Such a model specification is appropriate for the factor-augmented regression (Bai and Ng, 2006; Ludvigson and Ng, 2009; Stock and Watson, 2002), which is one of the most popular applications of the factor models in the econometric literature. Also, it is obvious that more accurate factor estimators can substantially improve out-of-sample forecasts (Bai and Liao, 2016; Choi, 2012). Note that the additional information could be unit-specific (e.g., individual characteristics for health studies) rather than time-specific, and the proposed approach in this paper can be applicable to the setup of unit-specific information. To account for cross-sectional heteroskedasticity and dependences, I assume that the error covariance matrix is sparse. Under this factor model structure, I first consistently estimate the error covariance matrix by using an adaptive thresholding method. Then, I adapt the regular PC method to the projected data combined with the inverse error covariance estimator, which I call this the feasible weighted projected principal component (WPPC). I then drive convergence rates for WPPC estimators and discuss the benefit of the proposed WPPC compared to the regular PC. Furthermore, I suggest the WPPC-based diffusion index forecasting model.

The empirical study supports the merit of the WPPC method. Specifically, I employ the proposed method to forecast excess bond returns using four aggregated macroeconomic variables as the covariates. Interestingly, the proposed method remarkably outperforms other existing PC methods during the recession periods.

The rest of the paper is organized as follows. Section 2 sets up the model and proposes the WPPC factor estimation procedure. Section 3 presents an asymptotic analysis of the proposed WPPC estimators. Moreover, I study the WPPC-based DI model. I present simulation studies in Section 4 and a real data application on forecasting bond excess returns in Section 5. In Section 6, I conclude the study. All proofs are presented in Appendix A and Online supplement.

2. MODEL SETUP AND ESTIMATION PROCEDURE

Throughout the paper, let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues of a matrix A, respectively. I also let $\|A\|_F = \sqrt{\mathrm{tr}(A'A)}$, $\|A\|_2 = \sqrt{\lambda_{\max}(A'A)}$, and $\|A\|_1 = \max_i \sum_j |A_{ij}|$ denote the Frobenius norm, the spectral norm (also called the operator norm) and the L_1-norm of a matrix A, respectively. Note that if A is a vector, both $\|A\|$ and $\|A\|_F$ are equal to the Euclidean norm.

2.1. Semiparametric factor model

This paper considers the following semiparametric factor model:

$$\mathbf{y}_t = \mathbf{\Lambda}\mathbf{f}_t + \mathbf{u}_t, \qquad \mathbf{f}_t = \mathbf{g}(\mathbf{X}_t) + \boldsymbol{\gamma}_t, \qquad t \leq T, \tag{2.1}$$

where $\mathbf{y}_t$ is an $N \times 1$ vector of observed data at time t, $\mathbf{f}_t$ is a $K \times 1$ vector of latent factors, $\mathbf{\Lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N)'$ indicates an $N \times K$ matrix of factor loadings, $\mathbf{u}_t$ is an $N \times 1$ vector of idiosyncratic components, $\mathbf{X}_t$ is a $d \times 1$ vector of observable covariates that partially explain the latent factors, $\mathbf{g}(\cdot)$ is the unknown nonparametric function, and $\boldsymbol{\gamma}_t$ is the component of common factors that cannot be explained by the covariates $\mathbf{X}_t$.

The model (2.1) can be written in matrix form as follows:

$$\mathbf{Y} = \mathbf{\Lambda}\{\mathbf{G}(\mathbf{X}) + \mathbf{\Gamma}\}' + \mathbf{U}, \tag{2.2}$$

where $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_T)$ is an $N \times T$ matrix of observed data, $\mathbf{G}(\mathbf{X}) = (\mathbf{g}(\mathbf{X}_1), \ldots, \mathbf{g}(\mathbf{X}_T))'$ is a $T \times K$ matrix of factors explained by the covariates, $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_T)'$ is a $T \times K$ matrix of factor components that cannot be explained by the covariates, and $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_T)$ is an $N \times T$ matrix of idiosyncratic errors. Throughout the paper, I assume that $\mathbf{\Gamma}$ and $\mathbf{U}$ are orthogonal to the function space spanned by the covariates $\mathbf{X}$. I also assume $d = \dim(\mathbf{X}_t)$ and $K = \dim(\mathbf{F}_t)$ to be constant. The number of factors, K, is assumed to be known. In practice, it can be consistently estimated by eigenvalue ratio methods (Ahn and Horenstein, 2013; Lam and Yao, 2012) or information criteria (Bai and Ng, 2002).

To estimate $\mathbf{G}(\mathbf{X})$, the nonparametric function is assumed to be additive and the sieve method can be applied (Fan et al., 2016). Specifically,

$$\mathbf{G}(\mathbf{X}) = \mathbf{\Phi}(\mathbf{X})\mathbf{B}' + \mathbf{R}(\mathbf{X}),$$

where $\mathbf{\Phi}(\mathbf{X}) = (\phi(\mathbf{X}_1), \cdots, \phi(\mathbf{X}_T))'$ is a $T \times (Jd)$ matrix of basis functions, $\mathbf{B} = (\mathbf{b}_1, \cdots, \mathbf{b}_K)'$ is a $K \times (Jd)$ matrix of sieve coefficients, and $\mathbf{R}(\mathbf{X})$ is a $T \times K$ matrix of approximation error term. Here, J is the number of sieve terms, and it grows slowly as $T \to \infty$. Then the model (2.2) can be rewritten as

$$\mathbf{Y} = \mathbf{\Lambda}\{\mathbf{\Phi}(\mathbf{X})\mathbf{B}' + \mathbf{\Gamma}\}' + \mathbf{\Lambda}\mathbf{R}(\mathbf{X})' + \mathbf{U}. \tag{2.3}$$

In high-dimensional factor analysis, the projection using additional covariates removes noise components. Specifically, let $\mathcal{X}$ be a sieve space spanned by the basis functions of $\mathbf{X}$. Define the $T \times T$ projection matrix onto $\mathcal{X}$:

$$\mathbf{P} = \mathbf{\Phi}(\mathbf{X})(\mathbf{\Phi}(\mathbf{X})'\mathbf{\Phi}(\mathbf{X}))^{-1}\mathbf{\Phi}(\mathbf{X})'.$$

The projected data by multiplying $\mathbf{P}$ on both sides of (2.3) has the following representation:

$$\mathbf{YP} = \mathbf{\Lambda}\mathbf{B}\mathbf{\Phi}(\mathbf{X})' + \widetilde{\mathbf{E}},$$

where $\widetilde{\mathbf{E}} = \mathbf{\Lambda}\mathbf{\Gamma}'\mathbf{P} + \mathbf{\Lambda}\mathbf{R}(\mathbf{X})'\mathbf{P} + \mathbf{UP}$. Then, $\widetilde{\mathbf{E}} \approx \mathbf{0}$ because $\mathcal{X}$ is orthogonal to $\mathbf{\Gamma}$ and $\mathbf{U}$, and $\mathbf{R}(\mathbf{X})$ is the sieve approximation error. Therefore, analyzing the projected data is an approximately noiseless problem, and it helps to obtain more accurate estimators.

Fan et al. (2016) proposed the projected principal component (PPC) method, based on the semi-parametric factor model that the factor loading has the semiparametric structure.[1] The idiosyncratic components are often cross-sectionally heteroskedastic and correlated in the approximate factor structure (Bai, 2003; Chamberlain and Rothschild, 1983; Fan et al., 2013). However, the PPC method does not require estimating the $N \times N$ covariance matrix, $\mathbf{\Sigma}_u = \text{cov}(\mathbf{u}_t)$, hence it essentially treats u_{it} to be homoskedastic and uncorrelated over i. As a result, it is inefficient. Therefore, this paper considers the following a weighted least squares problem to efficiently estimate the approximate semiparametric factor models:

$$\min_{\mathbf{\Lambda},\mathbf{B}} \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{\Lambda}\mathbf{B}\phi(\mathbf{X}_t))'W(\mathbf{y}_t - \mathbf{\Lambda}\mathbf{B}\phi(\mathbf{X}_t))$$

subject to the normalization of $\frac{1}{N}\mathbf{\Lambda}'W\mathbf{\Lambda} = \mathbf{I}_K$. Here, W is an $N \times N$ positive definite weighted matrix. This paper considers $W = \mathbf{\Sigma}_u^{-1}$ to take into account both cross-sectional heteroskedasticity and correlations. Bai and Liao (2017) also used $\mathbf{\Sigma}_u^{-1}$ as the first-order asymptotic optimal weight matrix in the panel data model with interactive fixed effects. See also Boivin and Ng (2006); Choi (2012); Forni et al. (2005); Stock and Watson (2006) for related articles. Since the weight matrix is not feasible in practice, the proposed estimation procedure requires a consistent estimator $\widehat{\mathbf{\Sigma}}_u$ under the sparsity assumption on $\mathbf{\Sigma}_u = (\Sigma_{u,ij})_{N \times N}$ as follows:

$$m_N = \max_{i \le N} \sum_{j=1}^{N} |\Sigma_{u,ij}|^q,$$

for some $q \in [0, 1)$, where m_N diverges slowly, such as $\log N$. This special structure is known as the "conditional sparsity" given the common factors in an approximate factor model (Fan et al., 2011). To accommodate additional covariates and improve estimation efficiency, I propose a novel factor estimation procedure in the following section.

[1]The conditions including normalization and the PPC estimation procedure considered in this paper are symmetric to those of Fan et al. (2016), because this paper incorporates the time-specific covariates, which have explanatory power on the common factors.

Table 1. Four different principal component methods.

	Objective function	Eigenvectors of
PC	$\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{\Lambda f}_t)'(\mathbf{y}_t - \mathbf{\Lambda f}_t)$	$\mathbf{YY}'$
WPC	$\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{\Lambda f}_t)'\widehat{\mathbf{\Sigma}}_u^{-1}(\mathbf{y}_t - \mathbf{\Lambda f}_t)$	$\widehat{\mathbf{\Sigma}}_u^{-1/2}\mathbf{YY}'\widehat{\mathbf{\Sigma}}_u^{-1/2}$
PPC	$\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{\Lambda B}\phi(\mathbf{X}_t))'(\mathbf{y}_t - \mathbf{\Lambda B}\phi(\mathbf{X}_t))$	$\mathbf{YPY}'$
WPPC	$\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{\Lambda B}\phi(\mathbf{X}_t))'\widehat{\mathbf{\Sigma}}_u^{-1}(\mathbf{y}_t - \mathbf{\Lambda B}\phi(\mathbf{X}_t))$	$\widehat{\mathbf{\Sigma}}_u^{-1/2}\mathbf{YPY}'\widehat{\mathbf{\Sigma}}_u^{-1/2}$

Note: The estimated factor loading matrix is $\sqrt{N}$ times the eigenvectors of the leading K eigenvectors of each covariance matrix.

2.2. Implementation of WPPC

The adaptive thresholding method is applied to estimate $\mathbf{\Sigma}_u^{-1}$ following Bickel and Levina (2008) and Fan et al. (2013). Specifically, let $\widetilde{R}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{it}\widehat{u}_{jt}$, where $\widehat{u}_{it}$ is the estimated residuals using the PPC method defined in Table 1. Then, define $\widehat{\mathbf{\Sigma}}_u$ as the thresholded error covariance matrix estimator:

$$\widehat{\mathbf{\Sigma}}_u = (\widehat{\Sigma}_{u,ij})_{N\times N}, \quad \widehat{\Sigma}_{u,ij} = \begin{cases} \widetilde{R}_{ij}, & i = j \\ s_{ij}(\widetilde{R}_{ij})I(|\widetilde{R}_{ij}| \geq \tau_{ij}), & i \neq j \end{cases},$$

where $s_{ij}(\cdot)$ is a generalized shrinkage function such as hard or soft thresholding (Cai and Liu, 2011; Rothman et al., 2008) and an entry-dependent threshold

$$\tau_{ij} = M\omega_{N,T}\sqrt{\widetilde{R}_{ii}\widetilde{R}_{jj}}, \text{ where } \omega_{N,T} = \sqrt{\frac{\log N}{T}} + \frac{1}{\sqrt{N}},$$

for some pre-determined threshold constant $M > 0$. In practice, the tuning parameter M can be chosen by multifold cross-validation, which is discussed in Section 2.2.1.

The WPPC estimators are then obtained as follows. Let $\widetilde{\mathbf{Y}} = \widehat{\mathbf{\Sigma}}_u^{-\frac{1}{2}}\mathbf{Y}$, $\widetilde{\mathbf{\Lambda}} = \widehat{\mathbf{\Sigma}}_u^{-\frac{1}{2}}\mathbf{\Lambda}$, and $\widetilde{\mathbf{U}} = \widehat{\mathbf{\Sigma}}_u^{-\frac{1}{2}}\mathbf{U}$. Then, the estimated feasible weighted loading matrix for $\mathbf{\Sigma}_u^{-\frac{1}{2}}\mathbf{\Lambda}$, denoted by $\widehat{\widetilde{\mathbf{\Lambda}}}$, is $\sqrt{N}$ times the eigenvectors corresponding to the K largest eigenvalues of the $N \times N$ matrix $\widetilde{\mathbf{Y}}\mathbf{P}\widetilde{\mathbf{Y}}' = \widehat{\mathbf{\Sigma}}_u^{-\frac{1}{2}}\mathbf{YPY}'\widehat{\mathbf{\Sigma}}_u^{-\frac{1}{2}}$. Note that the estimator of $\mathbf{\Lambda}$ is $\widehat{\mathbf{\Lambda}} = \widehat{\mathbf{\Sigma}}_u^{\frac{1}{2}}\widehat{\widetilde{\mathbf{\Lambda}}}$. Given $\widehat{\widetilde{\mathbf{\Lambda}}}$, the estimator of common factor is

$$\widehat{\mathbf{F}} = \frac{1}{N}\widetilde{\mathbf{Y}}'\widehat{\widetilde{\mathbf{\Lambda}}} = \frac{1}{N}\mathbf{Y}'\widehat{\mathbf{\Sigma}}_u^{-1}\widehat{\mathbf{\Lambda}}.$$

Moreover, given $\widehat{\widetilde{\mathbf{\Lambda}}}$,

$$\widehat{\mathbf{G}}(\mathbf{X}) = \frac{1}{N}\mathbf{P}\widetilde{\mathbf{Y}}'\widehat{\widetilde{\mathbf{\Lambda}}}, \quad \widehat{\mathbf{\Gamma}} = \frac{1}{N}(\mathbf{I} - \mathbf{P})\widetilde{\mathbf{Y}}'\widehat{\widetilde{\mathbf{\Lambda}}}$$

are estimators of $\mathbf{G}(\mathbf{X})$ and $\mathbf{\Gamma}$, respectively.

In summary, given the observed covariates, we first project $\mathbf{Y}$ onto the sieve space spanned by $\{\mathbf{X}_t\}_{t\leq T}$, then employ the regular PC method to the projected data (i.e., the PPC method). Next, using the estimated residuals from the first step, the consistent estimator $\widehat{\mathbf{\Sigma}}_u^{-1}$ can be obtained by thresholding approach under the conditional sparsity assumption. Section 3 presents asymptotic theory for the proposed WPPC estimators. Table 1 presents the main differences of the estimators. Four different methods minimize different objective functions, depending on the model specification and the weight matrix. Then, the loading estimators are obtained from different covariance matrices. Note that the feasible weighted principal component (WPC) is similar to the estimator suggested in Choi (2012), using the same thresholding error covariance matrix estimation procedure discussed above.

2.2.1. Choice of thresholding parameter The suggested covariance matrix estimator, $\widehat{\mathbf{\Sigma}}_u$, requires the choice of tuning parameters M, which is the threshold constant. Define $\widehat{\mathbf{\Sigma}}_u(M) = \widehat{\mathbf{\Sigma}}_u$, where the covariance estimator depends on M.

The thresholding constant, M, can be chosen through multifold cross-validation (Bickel and Levina, 2008; Fan et al., 2013). First, we obtain the estimated $N \times 1$ vector residuals $\widehat{\mathbf{u}}_t$ by PPC, then divide the data into $P = \log(T)$ blocks $J_1, ..., J_P$ with block length $T/\log(T)$. Here, we take one of the P blocks as the validation set. At the pth split, let $\widehat{\mathbf{\Sigma}}_u^p$ be the sample covariance matrix based on the validation set, defined by $\widehat{\mathbf{\Sigma}}_u^p = J_p^{-1} \sum_{t \in J_p} \widehat{\mathbf{u}}_t \widehat{\mathbf{u}}_t'$. Let $\widehat{\mathbf{\Sigma}}_u^{S,p}(M)$ be the soft-thresholding estimator with threshold constant M using the training data set $\{\widehat{\mathbf{u}}\}_{t \notin J_p}$. Then, we choose the constant M^* by minimizing a cross-validation objective function

$$M^* = \arg \min_{c_{\min} < M < c_{\max}} \frac{1}{P} \sum_{j=1}^{P} \|\widehat{\mathbf{\Sigma}}_u^{S,p}(M) - \widehat{\mathbf{\Sigma}}_u^p\|_F^2,$$

where $c_{\min}$ is the minimum constant that $\widehat{\mathbf{\Sigma}}_u(M)$ is positive definite for $M > c_{\min}$: $c_{\min} = \inf[C > 0 : \lambda_{\min}\{\widehat{\mathbf{\Sigma}}_u(M)\} > 0, \forall M > C]$, and $c_{\max}$ is a large constant such that $\widehat{\mathbf{\Sigma}}_u(c_{\max})$ is a diagonal matrix. Then, the resulting estimator of $\mathbf{\Sigma}_u$ is $\widehat{\mathbf{\Sigma}}_u(M^*)$.

3. ASYMPTOTIC ANALYSIS

In this section, I establish the asymptotic properties of the WPPC estimators. To do this, I impose the following technical conditions.

Assumption 3.1.

(i) There are constants $c_1, c_2 > 0$ such that $\lambda_{\min}(\mathbf{\Sigma}_u) > c_1$ and $\max_{i \le N} \sum_{j=1}^N |\Sigma_{u,ij}| < c_2$.

(ii) There is $q \in [0, 1)$ such that,

$$m_N \omega_{N,T}^{1-q} = o(1), \text{ where } \omega_{N,T} = \sqrt{\frac{\log N}{T}} + \frac{1}{\sqrt{N}}. \tag{3.1}$$

Condition (i) requires that $\mathbf{\Sigma}_u$ be well conditioned. This is a standard assumption of idiosyncratic term in the approximate factor model literature (Bai, 2003; Fan et al., 2013). Condition (ii) is needed for the $\|\cdot\|_1$-convergence of estimating $\mathbf{\Sigma}_u$ and its inverse. If (3.1) holds, we have

$$\|\widehat{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1}\|_1 = O_P(m_N \omega_{N,T}^{1-q}). \tag{3.2}$$

When m_N grows slowly with N, $\widehat{\mathbf{\Sigma}}_u^{-1}$ is consistent estimator with a nice convergence rate. In addition, when $m_N = O(1)$, $q = 0$ and $N > T$, the rate would be $O_P(\sqrt{\frac{\log N}{T}})$, which is minimax optimal rate as proved by Cai and Zhou (2012).

ASSUMPTION 3.2. *There are two positive constants c_1 and c_2 so that with probability approaching one as $T \to \infty$,*

$$c_1 < \lambda_{\min}(T^{-1}\mathbf{G}(\mathbf{X})'\mathbf{G}(\mathbf{X})) < \lambda_{\max}(T^{-1}\mathbf{G}(\mathbf{X})'\mathbf{G}(\mathbf{X})) < c_2.$$

Since $\mathbf{F} = \mathbf{G}(\mathbf{X}) + \mathbf{\Gamma}$, $\mathbf{G}(\mathbf{X})$ can be regarded as the projection of $\mathbf{F}$ onto the sieve space spanned by $\mathbf{X}$. Therefore, this assumption is a sufficient condition for pervasiveness on the factors in the semiparametric factor model.

ASSUMPTION 3.3.

(i) There are d_1, $d_2 > 0$ so that with probability approaching one as $T \to \infty$,

$$d_1 < \lambda_{\min}(T^{-1}\mathbf{\Phi}(\mathbf{X})'\mathbf{\Phi}(\mathbf{X})) < \lambda_{\max}(T^{-1}\mathbf{\Phi}(\mathbf{X})'\mathbf{\Phi}(\mathbf{X})) < d_2.$$

(ii) $\max_{j\leq J, t\leq T, l\leq d} E\phi_j(X_{tl})^2 \leq \infty$.

Since $T^{-1}\mathbf{\Phi}(\mathbf{X})'\mathbf{\Phi}(\mathbf{X}) = T^{-1}\sum_{t=1}^{T}\phi(\mathbf{X}_t)\phi(\mathbf{X}_t)'$ and $\phi(\mathbf{X}_t)$ is a $Jd \times 1$ vector, where $Jd \ll T$, the strong law of large numbers implies condition (i). This condition can be satisfied over normalizations of commonly used basis functions (e.g., Fourier basis, B-splines, and polinomial basis).

ASSUMPTION 3.4.

(i) $\{\mathbf{f}_t, \mathbf{u}_t\}_{t\leq T}$ *is strictly stationary.* $E\boldsymbol{\gamma}_t = 0$ *and* $E\mathbf{u}_t = 0$ *for all* $t \leq T$; $\{\mathbf{u}_t\}_{t\leq T}$ *is independent of* $\{\mathbf{f}_t\}_{t\leq T}$; $\{\mathbf{X}_t\}_{t\leq T}$ *is independent of* $\{\boldsymbol{\gamma}_t\}_{t\leq T}$.

(ii) Define $\boldsymbol{\gamma}_t = (\gamma_{t1}, \cdots, \gamma_{tK})'$. $\max_{k\leq K, t\leq T} Eg_k(\mathbf{X}_t)^2 \leq \infty, \nu_T < \infty$ *and*

$$\max_{k\leq K, s\leq T}\sum_{t\leq T}|E\gamma_{tk}\gamma_{sk}| = O(\nu_T), \text{ where } \nu_T = \max_{k\leq K}\frac{1}{T}\sum_{t\leq T}\mathrm{var}(\gamma_{tk}).$$

(iii) Exponential tail: There exist $r_2, r_3 > 0$ *satisfying* $r_1^{-1} + r_2^{-1} + r_3^{-1} > 1$ *and* $b_1, b_2 > 0$, *such that for any* $s > 0, i \leq N$ *and* $k \leq K$,

$$P(|u_{it}| > s) \leq \exp(-(s/b_1)^{r_2}), \quad P(|f_{kt}| > s) \leq \exp(-(s/b_2)^{r_3}).$$

(iv) Weak dependence: There exists a positive constant $C < \infty$ so that

$$\max_{s\leq T}\sum_{t=1}^{T}|Eu_{it}u_{is}| < C,$$

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|Eu_{it}u_{js}| < C,$$

$$\max_{t\leq T}\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{q=1}^{T}|\mathrm{cov}(u_{it}u_{is}, u_{jt}u_{jq})| < C.$$

Condition (ii) allows serial weak dependence for $\{\boldsymbol{\gamma}_t\}_{t\leq T}$. Condition (iii) ensures the Bernstein-type inequality for weakly dependent data. Note that the underlying distributions are assumed to be thin-tailed. Allowing for heavy-tailed distributions is also an important issue, but it would require a very different estimation method (Fan et al., 2021). Condition (iv) is commonly imposed in high-dimensional factor analysis (Bai, 2003; Stock and Watson, 2002).

Assumption 3.5. *Define, for each $k \leq K$ and for each $t \leq T$, $g_k(\mathbf{X}_t) = \sum_{l=1}^d g_{kl}(X_{tl})$. For all $l \leq d, k \leq K$,*
(i) the factor component $g_{kl}(\cdot)$ belongs to a Hölder class $\mathcal{G}$ defined by, for some $L > 0$,

$$\mathcal{G} = \{g : |g^{(r)}(s) - g^{(r)}(t)| \leq L|s-t|^{\alpha}\}.$$

(ii) the sieve coefficients of $g_{kl}(X_{tl})$, $\{b_{j,kl}\}_{j\leq J}$, satisfy for $\kappa = 2(r+\alpha) \geq 4$, as the sieve dimension $J \to \infty$,

$$\sup_{x\in\mathcal{X}_l} |g_{kl}(x) - \sum_{j=1}^J b_{j,kl}\phi_j(x)|^2 = O(J^{-\kappa}),$$

where $\mathcal{X}_l$ is the support of the lth element of $\mathbf{X}_t$. In addition, $\max_{k,j,l} b_{j,kl}^2 < \infty$.

The above assumption is needed for the accuracy of the sieve approximation. When $\{\phi_j\}$ is B-splines or polynomial basis, condition (i) implies condition (ii) as discussed in Chen (2007). The following theorem provides the convergence rates of WPPC estimators.

Theorem 3.1. *Suppose $J = o(\sqrt{T})$ and Assumptions 3.1–3.5 hold. There is an invertible matrix $\mathbf{H}$, as $N, T, J \to \infty$, we have, for $\omega_{N,T} = \sqrt{\frac{\log N}{T}} + \frac{1}{\sqrt{N}}$,*

$$\frac{1}{N}\|\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\mathbf{H}\|_F^2 = O_P\left(\frac{1}{T}\right), \tag{3.3}$$

$$\frac{1}{T}\|\widehat{\mathbf{G}}(\mathbf{X}) - \mathbf{G}(\mathbf{X})\mathbf{H}\|_F^2 = O_P\left(\frac{1}{J^\kappa} + \frac{J\nu_T}{T} + \frac{J}{T}m_N^2\omega_{N,T}^{2-2q}\right), \tag{3.4}$$

$$\frac{1}{T}\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\mathbf{H}\|_F^2 = O_P\left(\frac{1}{N} + \frac{1}{J^\kappa} + \frac{J}{T^2} + \frac{J\nu_T}{T} + \frac{1}{T}m_N^2\omega_{N,T}^{2-2q}\right). \tag{3.5}$$

In addition, for any $t \leq T$,

$$\|\widehat{\mathbf{f}}_t - \mathbf{H}^{-1}\mathbf{f}_t\| = O_P\left(m_N\omega_{N,T}^{1-q}\right). \tag{3.6}$$

Remark 3.1. The additional term $1/\sqrt{N}$ in $\omega_{N,T}$ results from the estimation of unknown factor components through the PPC procedure. Moreover, the estimation error of the inverse error covariance matrix in (3.2) carries over to the estimation of $\mathbf{G}(\mathbf{X})$ and $\boldsymbol{\Gamma}$ in the WPPC procedure, and it affects their rates of convergence. However, they are negligible when the dimensional N and T grow sufficiently large.

Since $\widehat{\mathbf{F}} = \widehat{\mathbf{G}}(\mathbf{X}) + \widehat{\boldsymbol{\Gamma}}$, the convergence rate for the estimated common factor can be obtained by two convergences. I have the following remark about the rates of convergence above compared with those using the conventional PC method.

REMARK 3.2. Denote $\widehat{\boldsymbol{\Lambda}} = (\widehat{\boldsymbol{\lambda}}_1, ..., \widehat{\boldsymbol{\lambda}}_N)'$, $\widehat{\mathbf{G}}(\mathbf{X}) = (\widehat{\mathbf{g}}(\mathbf{X}_1), ..., \widehat{\mathbf{g}}(\mathbf{X}_T))'$, and $\widehat{\boldsymbol{\Gamma}} = (\widehat{\boldsymbol{\gamma}}_1, ..., \widehat{\boldsymbol{\gamma}}_T)'$. For the factor loadings, we have

$$\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\boldsymbol{\lambda}}_i - \mathbf{H}'\boldsymbol{\lambda}_i\|^2 = O_P\left(\frac{1}{T}\right).$$

For the factor components, consider $m_N = O(1)$ and $q = 0$ as a simple case. Define the optimal $J^* = (T\min\{N, T/\log N, \nu_T^{-1}\})^{1/(\kappa+1)}$. With $J = J^*$, we have

$$\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{g}}(\mathbf{X}_t) - \mathbf{H}^{-1}\mathbf{g}(\mathbf{X}_t)\|^2 = O_P\left(\frac{1}{(T\min\{N, T/\log N, \nu_T^{-1}\})^{1-1/(\kappa+1)}}\right),$$
$$\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{\gamma}}_t - \mathbf{H}^{-1}\boldsymbol{\gamma}_t\|^2 = O_P\left(\frac{1}{N} + \frac{1}{(T\min\{N, T/\log N, \nu_T^{-1}\})^{1-1/(\kappa+1)}}\right).$$

In contrast, for some rotation matrix $\widetilde{\mathbf{H}}$, the rates of the regular PC method estimators $(\widetilde{\boldsymbol{\lambda}}_i, \widetilde{\mathbf{f}}_t)$ (Bai, 2003; Stock and Watson, 2002) are

$$\frac{1}{N}\sum_{i=1}^{N}\|\widetilde{\boldsymbol{\lambda}}_i - \widetilde{\mathbf{H}}'\boldsymbol{\lambda}_i\|^2 = O_P\left(\frac{1}{N} + \frac{1}{T}\right), \quad \frac{1}{T}\sum_{t=1}^{T}\|\widetilde{\mathbf{f}}_t - \widetilde{\mathbf{H}}^{-1}\mathbf{f}_t\|^2 = O_P\left(\frac{1}{N} + \frac{1}{T}\right).$$

The WPPC procedure can have faster rates of convergence than the regular PC method. Specifically, the rate of convergence for the estimated loadings is improved for small N. When both N and T are large, the rate of convergence for nonparametric functions is $O_P(1/(NT)^{\kappa/(\kappa+1)} + (\log N/T^2)^{\kappa/(\kappa+1)})$. This implies that, when $\mathbf{f}_t = \mathbf{g}(\mathbf{X}_t)$, the rate for the estimated factors is always faster than that of the regular PC as long as $\log N = o(T^{1-1/\kappa})$. When $\mathbf{f}_t \neq \mathbf{g}(\mathbf{X}_t)$, the rate for the estimated factors is still faster than the regular PC when N is larger than T, while they are the same when T is larger than N.

3.1. *Diffusion index forecasting models*

In this subsection, I apply the WPPC factor estimates to the diffusion index (DI) forecasting model (Bai and Ng, 2006; Stock and Watson, 2002).

Consider the following DI forecasting model:

$$z_{t+h} = \alpha'\mathbf{f}_t + \beta'\mathbf{v}_t + \epsilon_{t+h}, \tag{3.7}$$

where h is a forecasting horizon, $\mathbf{f}_t$ is a vector of unobservable factors and $\mathbf{v}_t$ is a vector of observable variables. Because $\mathbf{f}_t$ is latent, we often use the estimated factors $\widehat{\mathbf{f}}_t$ using PC methods, based on the factor model.

Define $L_t = (\mathbf{f}_t', \mathbf{v}_t')'$. To predict the conditional mean, $z_{T+h|T} = E(z_{T+h}|L_T, L_{T-1}, ...)$, we first obtain the least squares estimates $\widehat{\alpha}$ and $\widehat{\beta}$ from regression z_{t+h} on $\widehat{L}_t = (\widehat{\mathbf{f}}_t', \mathbf{v}_t')'$, for $t = 1, ..., T-h$, and define $\widehat{\delta} = (\widehat{\alpha}', \widehat{\beta}')'$. Then, the feasible prediction can be obtained by $\widehat{z}_{T+h|T} = \widehat{\alpha}'\widehat{\mathbf{f}}_T + \widehat{\beta}'\mathbf{v}_T = \widehat{\delta}'\widehat{L}_T$. Stock and Watson (2002) showed consistency of $\widehat{z}_{T+h|T}$ for $z_{T+h|T}$, and Bai and Ng (2006) established the limiting distributions of the least squares estimates and forecast errors. They used the regular PC estimation method under the conventional factor model. In this paper, I consider the WPPC-based DI forecasting model and its asymptotic properties. The following assumption is standard for forecasting regression analysis.

Assumption 3.6. *Let $L_t = (\mathbf{f}_t', \mathbf{v}_t')'$. $E\|L_t\|^4$ is bounded for every t.*
(i) $E(\epsilon_{t+h}|z_t, L_t, z_{t-1}, L_{t-1}, ...) = 0$ for any $h > 0$, and L_t and ϵ_t are independent of the idiosyncratic errors u_{is} for all i and s.
(ii) $\frac{1}{T}\sum_{t=1}^{T} L_t L_t' \xrightarrow{p} \Sigma_L$, which is a positive definite matrix.
(iii) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} L_t\epsilon_{t+h} \xrightarrow{d} N(0, \Sigma_{L,\epsilon})$, where $\Sigma_{L,\epsilon} = \operatorname{plim} \frac{1}{T}\sum_{t=1}^{T} \epsilon_{t+h}^2 L_t L_t'$.

Condition (i) implies that the idiosyncratic errors from the factor model and all the random variables in the forecasting model are independent. Conditions (ii)-(iii) are needed for regression analysis.

Theorem 3.2. *Let $\widehat{\delta} = (\widehat{\alpha}', \widehat{\beta}')'$ and $\delta = (\alpha'\mathbf{H}, \beta')'$. Suppose the assumptions of Theorem 3.1 and Assumption 3.6 hold. For q, m_N, and $\omega_{N,T}$ defined in (3.1), if $\sqrt{T}m_N^2\omega_{N,T}^{2-2q} = o(1)$,*

$$\sqrt{T}(\widehat{\delta} - \delta) \xrightarrow{d} N(0, \Sigma_\delta),$$

where $\Sigma_\delta = \Pi'^{-1}\Sigma_L^{-1}\Sigma_{L,\epsilon}\Sigma_L^{-1}\Pi'$ with $\Pi = \operatorname{diag}(\mathbf{H}', \mathbf{I})$. A heteroskedasticity consistent estimator for Σ_δ is

$$\widehat{\Sigma}_\delta = \left(\frac{1}{T}\sum_{t=1}^{T-h}\widehat{L}_t\widehat{L}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\epsilon}_{t+h}^2\widehat{L}_t\widehat{L}_t'\right)\left(\frac{1}{T}\sum_{t=1}^{T-h}\widehat{L}_t\widehat{L}_t'\right)^{-1}.$$

Remark 3.3. Consider a special case where $m_N = O(1)$ and $q = 0$ (i.e., a strictly sparse case), which means the number of nonzero elements in each row of $\boldsymbol{\Sigma}_u$ is bounded. Then the condition $\sqrt{T}m_N^2\omega_{N,T}^{2-2q} = o(1)$ becomes $\frac{\log N}{\sqrt{T}} + \frac{\sqrt{T}}{N} = o(1)$, which holds if $\sqrt{T} = o(N)$ and $\log N = o(\sqrt{T})$. Implicitly, requiring $\sqrt{T}/N \to 0$ is needed for the asymptotic normality of $\widehat{\delta}$ as Bai and Ng (2006) imposed.

The following theorem presents the convergence rate of the estimated conditional mean.

Theorem 3.3. *Let $\widehat{z}_{T+h|T} = \widehat{\delta}'\widehat{L}_T$. Suppose that the assumptions of Theorem 3.2 hold. Then, for $\omega_{N,T} = \sqrt{\frac{\log N}{T}} + \frac{1}{\sqrt{N}}$,*

$$\widehat{z}_{T+h|T} - z_{T+h|T} = O_P(m_N\omega_{N,T}^{1-q}).$$

4. MONTE CARLO SIMULATIONS

In this section, I conducted simulations to examine the finite sample performance of the proposed WPPC method. I considered the following semiparametric factor model:

$$\mathbf{y}_t = \boldsymbol{\Lambda}\mathbf{f}_t + \mathbf{u}_t, \text{ and } \mathbf{f}_t = \sigma_g\mathbf{g}(\mathbf{X}_t) + \sigma_\gamma\boldsymbol{\gamma}_t, \text{ for } t = 1, \cdots T,$$

where $\boldsymbol{\Lambda}$ is drawn from i.i.d. Uniform$(0, 1)$. I chose $\dim(\mathbf{X}_t) = 3$ and three factors (i.e., $K = 3$), which is assumed to be known. I introduced serial dependences on $\mathbf{X}_t$ and $\boldsymbol{\gamma}_t$ as follows:

$$\mathbf{X}_t = \Psi\mathbf{X}_{t-1} + \boldsymbol{\xi}_t, \text{ and } \boldsymbol{\gamma}_t = \Psi\boldsymbol{\gamma}_{t-1} + \boldsymbol{\nu}_t, \text{ for } t = 1, \cdots T,$$

with $\mathbf{X}_0 = \mathbf{0}$, $\boldsymbol{\gamma}_0 = \mathbf{0}$ and a 3×3 diagonal matrix Ψ. Each diagonal element of Ψ is generated from Uniform$(0.3, 0.7)$. In addition, $\boldsymbol{\xi}_t$ and $\boldsymbol{\nu}_t$ are drawn from i.i.d. $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

Table 2. Canonical correlations of loadings/factors.

		Strong($w=3$)				Weak($w=1/3$)			
N	T	PC	WPC	PPC	WPPC	PC	WPC	PPC	WPPC
		Loadings							
50	120	0.209	0.253	0.592	0.719	0.249	0.293	0.320	0.457
	240	0.257	0.268	0.823	0.853	0.294	0.289	0.490	0.647
100	120	0.134	0.574	0.643	0.736	0.211	0.718	0.350	0.537
	240	0.135	0.548	0.832	0.850	0.220	0.704	0.556	0.684
		Factors							
50	120	0.135	0.199	0.486	0.852	0.179	0.262	0.302	0.679
	240	0.149	0.174	0.622	0.897	0.188	0.213	0.410	0.826
100	120	0.143	0.729	0.644	0.949	0.223	0.852	0.400	0.890
	240	0.130	0.632	0.754	0.959	0.212	0.765	0.570	0.940

Note: Entries in this table denote averaged minimum canonical correlations of the estimated loadings and factors: the larger the better.

To address different correlations between $\mathbf{f}_t$ and $\mathbf{g}(\mathbf{X}_t)$, define $\sigma_g^2 = \frac{w}{1+w}$ and $\sigma_\gamma^2 = \frac{1}{1+w}$. In the simulations, I considered $w = 3$ and $1/3$, where the larger w represents the stronger explanatory power of the observed covariates. The unknown function $\mathbf{g}(\cdot)$ has the following model: $\mathbf{g}(\mathbf{X}_t) = (g_1(\mathbf{X}_t), \cdots, g_K(\mathbf{X}_t))'$, where $g_k(\mathbf{X}_t) = \sum_{l=1}^{3} g_{kl}(X_{tl})$. The three characteristic functions are $g_{1l}(x) = x, g_{2l}(x) = x^2 - 1$, and $g_{3l}(x) = x^3 - 2x$, for all $l \leq d$.

Next, the idiosyncratic errors are generated using a $N \times N$ banded covariance matrix Σ_u as follows: let $\{\varepsilon_{it}\}_{i\leq N, t\leq T}$ be i.i.d. $\mathcal{N}(0,1)$. Let

$$\eta_{1t} = \varepsilon_{1t}, \eta_{2t} = \varepsilon_{2t} + a_1\varepsilon_{1t}, \eta_{3t} = \varepsilon_{3t} + a_2\varepsilon_{2t} + b_1\varepsilon_{1t},$$
$$\eta_{i+1,t} = \varepsilon_{i+1,t} + a_i\varepsilon_{it} + b_{i-1}\varepsilon_{i-1,t} + c_{i-2}\varepsilon_{i-2,t},$$

where the constants $\{a_i, b_i, c_i\}_{i=1}^N$ are generated from i.i.d. $\mathcal{N}(0, \sqrt{5})$. Denote the correlation matrix of $\eta_t = (\eta_{1t}, \cdots, \eta_{Nt})'$ by R_η, which is a banded matrix. Then the cross-sectional heteroskedasticity is introduced as follows: let $D = \mathrm{diag}(d_i)$, where $\{d_i\}_{i\leq N}$ is drawn from i.i.d. Uniform$(0, \sqrt{7})$. Finally, define $\Sigma_u = DR_\eta D$, and generate $\{u_t\}_{t\leq T}$ as i.i.d. $\mathcal{N}(0, \Sigma_u)$.

For comparison, I employed WPPC, PPC, WPC, and PC to estimate the factors loadings and common factors. The estimation accuracies are measured using the canonical correlation between the estimators and parameters because the factors and loading may be estimated up to a rotation matrix (Bai and Liao, 2016). For WPPC and PPC, the additive polynomial basis with $J = 5$ is used for the sieve basis. For WPPC and WPC, the thresholding tuning parameter M is chosen by the 5-fold cross-validation, as discussed in Section 2.2.1. In this simulation study, I considered $N = 50, 100$ and $T = 120, 240$. The simulation is replicated 1000 times for each scenario.

Table 2 presents sample means of the smallest canonical correlations. In addition, averaged mean squared errors (MSE) of the estimated common components, $(\frac{1}{NT}\sum_{i,t}(\widehat{\lambda}_i'\widehat{f}_t - \lambda_i' f_t)^2)^{1/2}$, are also compared in Table 3. From Tables 2–3, I find that the WPPC method performs the best when the observed covariates have strong explanatory power. On the

Table 3. Mean squared errors of $\mathbf{\Lambda F}'$.

N	T	Strong($w=3$) PC	WPC	PPC	WPPC	Weak($w=1/3$) PC	WPC	PPC	WPPC
50	120	0.801	0.613	0.554	0.345	0.807	0.616	0.677	0.517
	240	0.769	0.586	0.460	0.258	0.779	0.595	0.591	0.400
100	120	0.659	0.389	0.447	0.313	0.670	0.360	0.576	0.464
	240	0.616	0.361	0.361	0.233	0.628	0.331	0.482	0.364

Note: Entries in this table denote averaged mean squared errors of the estimated common components: the smaller the better.

Table 4. Out-of-sample relative mean squared forecasting errors.

N	T	Strong($w=3$) PC	WPC	PPC	WPPC	Weak($w=1/3$) PC	WPC	PPC	WPPC
					$h=1$				
50	120	1.000	0.867	0.944	0.823	1.000	0.856	0.982	0.831
	240	1.000	0.873	0.939	0.831	1.000	0.871	0.956	0.826
100	120	1.000	0.869	0.954	0.853	1.000	0.858	0.979	0.852
	240	1.000	0.874	0.944	0.852	1.000	0.864	0.967	0.849
					$h=12$				
50	120	1.000	0.870	0.950	0.836	1.000	0.864	0.985	0.837
	240	1.000	0.864	0.933	0.815	1.000	0.864	0.953	0.822
100	120	1.000	0.869	0.952	0.850	1.000	0.855	0.985	0.852
	240	1.000	0.873	0.942	0.852	1.000	0.860	0.963	0.845

Note: Entries in this table denote out-of-sample averaged relative mean squared forecast errors (with PC as the benchmark): the smaller the better.

other hand, when $w = 1/3$ and $N = 100$, WPC performs better than PPC and WPPC for loadings as well as common components. This is because the observed covariates are not informative enough, while the magnitude of cross-sectional heteroskedasticity and correlation gets larger as N increases. For factors, however, the WPPC always outperforms others even in the weak explanatory case.

I now demonstrate the out-of-sample forecasting performance based on the WPPC method. Consider a linear forecasting model as follows:

$$z_{t+h} = \alpha'\mathbf{f}_t + \beta'\mathbf{v}_t + \epsilon_{t+h},$$

where $\mathbf{v}_t = 1$, $\beta = 1$, and ϵ_{t+1} is drawn from i.i.d. $\mathcal{N}(0,1)$. To cover a variety of model settings, unknown coefficients, α, are generated from Uniform$(0.5, 1)$ for each simulation. I generated the data for the factor model as above.

I conducted h-periods ahead out-of-sample forecasting 50 times using rolling window scheme. In this experiment, I considered $h = 1$ and 12. The moving window size is fixed as T. In each simulation, the total $T + h + 49$ observations are generated. To forecast z_{T+m+h} for $m = 0, \cdots, 49$, the observations from $m + 1$ to $m + T$ are used. Specifically, the unknown factors are estimated by either PC, WPC, PPC or

Table 5. Components of observed covariates, $\mathbf{X}_t$.

	Series
$X_{1,t}$	Linear combination of five forward rates
$X_{2,t}$	Real gross domestic product
$X_{3,t}$	Consumption price index
$X_{4,t}$	Non-agriculture employment

WPPC and are denoted by $\{\widehat{\mathbf{f}}_{m+1}, \cdots, \widehat{\mathbf{f}}_{m+T}\}$. Then, $\widehat{\alpha}$ and $\widehat{\beta}$ are obtained by regressing $\{z_{m+1+h}, \cdots, z_{m+T}\}$ on $\{(\widehat{\mathbf{f}}'_{m+1}, \mathbf{v}_{m+1})', \cdots, (\widehat{\mathbf{f}}'_{m+T-h}, \mathbf{v}_{m+T-h})'\}$. Finally, forecasts are $\widehat{z}_{T+m+h|T+m} = \widehat{\alpha}'\widehat{\mathbf{f}}_{m+T} + \widehat{\beta}\mathbf{v}_{m+T}$. This procedure continues for $m = 0, \cdots, 49$. I calculated the relative mean squared forecasting errors by setting PC as a benchmark:

$$\text{RMSFE} = \frac{\sum_{m=0}^{49}(z_{T+m+h} - \widehat{z}_{T+m+h|T+m})^2}{\sum_{m=0}^{49}(z_{T+m+h} - \widehat{z}^{PC}_{T+m+h|T+m})^2}.$$

Table 4 presents the averages of RMSFEs for several methods. Table 4 shows that WPPC performs the best. In the previous in-sample estimation results, WPPC outperforms others for factor estimation even in the weak explanatory power case. This implies that more efficient estimations of the factors results in better forecasting performances.

5. APPLICATION: US BOND RISK PREMIA

In this section, I applied the proposed WPPC method to forecast the excess return of U.S. government bonds. I collected monthly bond return data spanning from the period 1964:1–2016:4 ($T = 628$), which is an updated version of the Ludvigson and Ng (2009) dataset. The bond return data are obtained from the Fama-Bliss dataset from the Center for Research in Securities Prices (CRSP). The factors are estimated by the PC, WPC, PPC, and WPPC methods from a monthly balanced panel of disaggregated 130 macroeconomic time series (Bai and Ng, 2008; Ludvigson and Ng, 2009; Stock and Watson, 2002). A specific description and transformation code of panel data is provided in McCracken and Ng (2016). For WPPC and PPC methods, I used four covariates listed in Table 5.[2] These aggregate series are widely used to describe the co-movement of the macroeconomic activities (Stock and Watson, 2014).

In this study, I set the number of factors $K = 8$ for all methods, which is determined by the information criteria suggested in Bai and Ng (2002). In addition, the additive polynomial basis with $J = 5$ is used for the sieve basis of PPC and WPPC. The thresholding constants of WPC and WPPC are chosen as suggested in Section 2.2.1 for each estimation period.

Following Ludvigson and Ng (2009) and Cochrane and Piazzesi (2005), the excess return with maturity of n-years is defined by

$$rx^{(n)}_{t+12} = r^{(n)}_{t+12} - y^{(1)}_t, \text{ for } t = 1, ..., T,$$

[2] I interpolated gross domestic product to a monthly frequency following Chow and Lin (1971).

Table 6. In-sample adjusted R^2 of U.S. bonds excess return forecasting.

Maturity	PC	WPC	PPC	WPPC
2 year	15.5	17.1	17.4	19.5
3 year	15.9	17.2	16.8	18.3
4 year	16.1	17.4	17.0	18.5
5 year	17.3	18.5	17.7	19.2

Note: Entries in this table denote in-sample adjusted $R^2(\%)$ of predictive regressions for excess bond returns: the larger the better. Sample period: 1964:1-2016:4.

where $r_{t+12}^{(n)} = p_{t+12}^{(n-1)} - p_t^{(n)}$ is the log holding period return for the log price of n-year discount bond $p_t^{(n)}$, and $y_t^{(1)}$ is the log yield on the one-year bond.

First, I conducted an in-sample regression analysis, based on the following simple linear forecasting model:

$$rx_{t+12}^{(n)} = \alpha + \beta'\widehat{F}_t + \epsilon_{t+12},$$

where $\widehat{F}_t$ is one of the factor estimators from WPPC, PPC, WPC, and PC. Table 6 reports the adjusted R^2 statistics for 2- to 5-year log excess bond returns. From Table 6, I find that WPPC outperforms others overall. For example, the factors estimated by WPPC explain 19.5% of the variation one year ahead in the 2-year return, while the factors estimated by PC only explain about 15.5%. I confirmed that to predict the excess bond returns, the estimated factors using WPPC outperform the factors using other methods. Thus, we can conjecture that incorporating the cross-sectional correlations and additional covariates improves the accuracy of factor estimation.

I also conducted the one-year-ahead out-of-sample forecast of the excess bond returns. I considered the following diffusion index forecasting model (Ludvigson and Ng, 2009):

$$rx_{t+12}^{(n)} = \alpha + \delta'\widehat{L}_t + \epsilon_{t+12},$$

where $\widehat{L}_t = (\widehat{F}_t', CP_t)'$, where CP_t is the single forward rate factor suggested by Cochrane and Piazzesi (2005), and $\widehat{F}_t$ is one of the factor estimators from WPPC, PPC, WPC, and PC. Forecasts are constructed based on rolling estimation windows using the data from the past 240 months. I considered three out-of-sample periods: recessions, expansions, and the whole period (January 1984–April 2016).

To evaluate the forecast performance, I use the out-of-sample R^2 suggested by Campbell and Thompson (2007), defined as

$$\text{Out-of-sample } R^2 = 1 - \frac{\sum_{t=Q}^{T-12}(rx_{t+12}^{(n)} - \widehat{rx}_{t+12|t}^{(n)})^2}{\sum_{t=Q}^{T-12}(rx_{t+12}^{(n)} - \overline{rx}_{t+12}^{(n)})^2},$$

where, for each maturity n, $\widehat{rx}_{t+12|t}^{(n)}$ is the forecast of bond excess returns using each model and $\overline{rx}_{t+12}^{(n)}$ is the historical average of bond excess return. Note that the out-of-sample R^2 values can be negative, indicating that the forecasting performance of the particular model is even worse than the historical averages.

Table 7 reports the out-of-sample R^2 of various factor estimation methods. WPPC outperforms others in out-of-sample predictive accuracy by using additional covariates

Table 7. Out-of-sample R^2 of U.S. bonds excess return forecasting.

	Recessions				Expansions				1984:1-2016:4			
	PC	WPC	PPC	WPPC	PC	WPC	PPC	WPPC	PC	WPC	PPC	WPPC
2 year	6.6	7.6	11.8	11.7	41.1	41.1	43.2	42.9	36.9	37.0	39.3	39.1
3 year	-4.3	-1.7	8.2	11.6	40.4	40.8	42.8	42.8	35.2	35.8	38.7	39.1
4 year	-10.1	-6.2	6.0	11.8	41.8	42.4	45.4	46.0	36.6	37.6	41.5	42.6
5 year	-9.8	-4.4	5.7	14.2	39.7	40.4	42.5	43.1	35.2	36.3	39.2	40.4

Note: Entries in this table denote out-of-sample R^2(%) of U.S. bonds excess return forecasting: the larger the better. Out-of-sample period: 1984:1-2016:4.

and the error covariance matrix estimator. For the whole period, WPPC increases the out-of-sample R^2 by 6.5%-16.4% compared to PC. Interestingly, during the recession periods, WPPC and PPC yield remarkably larger out-of-sample R^2 than PC and WPC. Overall, these results indicate that WPPC can accurately estimate the latent factors by harnessing the additional covariates and the error covariance matrix estimator.

6. CONCLUSIONS

This paper proposes a novel factor estimation procedure based on the semiparametric factor models. The proposed WPPC method estimates the unknown factors and loadings efficiently by incorporating cross-sectional heteroskedasticity and correlations of the error terms. It also removes the noise components by using the projection approach with a few observed covariates. I then apply the proposed estimation method to the diffusion index forecasting model. I show the asymptotic behaviors of the WPPC method compared to the regular PC method.

In the empirical study, in order to forecast excess bond returns, the proposed estimator shows the best performance. It confirms the presence of both cross-sectional dependence and heteroskedasticity, which provides the theoretical basis for employing the error covariance matrix estimation. In addition, using aggregated macroeconomic time series as covariates yields a substantial gain in out-of-sample forecasting.

In this paper, I assume the dimension of additional covariates to be fixed while the covariates investigate the nonlinearlity in the common factors. On the other hand, in the literature on asset pricing, there are some recent papers discussing linear factor models and model selection from a large set of characteristics, the so-called factor zoo (Cochrane, 2011; Feng et al., 2020; Harvey et al., 2016). However, the factor zoo may contain many redundant or useless characteristics that lead to efficiency loss. Thus, it is an interesting research topic to incorporate a high-dimensional set of covariates in the semiparametric factor models. I leave this for a future study.

ACKNOWLEDGEMENTS

I am grateful to two anonymous referees and the Co-Editor for their insightful comments, which lead to a better presentation of the current paper.

REFERENCES

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

Ait-Sahalia, Y. and D. Xiu (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics 201*(2), 384–399.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J. and Y. Liao (2016). Efficient estimation of approximate factor models via penalized maximum likelihood. *Journal of econometrics 191*(1), 1–18.

Bai, J. and Y. Liao (2017). Inferences in panel data with interactive effects using large covariance matrices. *Journal of econometrics 200*(1), 59–78.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica 74*(4), 1133–1150.

Bai, J. and S. Ng (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics 146*(2), 304–317.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. *The Quarterly journal of economics 120*(1), 387–422.

Bickel, P. J. and E. Levina (2008). Covariance regularization by thresholding. *The Annals of Statistics 36*(6), 2577–2604.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132*(1), 169–194.

Breitung, J. and J. Tenhofen (2011). Gls estimation of dynamic factor models. *Journal of the American Statistical Association 106*(495), 1150–1166.

Cai, T. and W. Liu (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association 106*(494), 672–684.

Cai, T. T. and H. H. Zhou (2012). Optimal rates of convergence for sparse covariance matrix estimation. *The Annals of Statistics 40*(5), 2389–2420.

Campbell, J. Y. and S. B. Thompson (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies 21*(4), 1509–1531.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica 51*(5), 1281–1304.

Chen, E. Y., D. Xia, C. Cai, and J. Fan (2020). Semiparametric tensor factor analysis by iteratively projected svd. *arXiv preprint arXiv:2007.02404*.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of econometrics 6*, 5549–5632.

Choi, I. (2012). Efficient estimation of factor models. *Econometric Theory 28*(2), 274–308.

Chow, G. and A.-l. Lin (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics 53*(4), 372–75.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of finance 66*(4), 1047–1108.

Cochrane, J. H. and M. Piazzesi (2005). Bond risk premia. *American Economic Review 95*(1), 138–160.

Fan, J., Y. Ke, and Y. Liao (2021). Augmented factor models with applications to validating market risk factors and forecasting bond risk premia. *Journal of Econometrics 222*(1), 269–294.

Fan, J., Y. Liao, and M. Mincheva (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics 39*(6), 3320.

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(4), 603–680.

Fan, J., Y. Liao, and W. Wang (2016). Projected principal component analysis in factor models. *Annals of statistics 44*(1), 219.

Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance 75*(3), 1327–1370.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American statistical association 100*(471), 830–840.

Harvey, C. R., Y. Liu, and H. Zhu (2016). … and the cross-section of expected returns. *The Review of Financial Studies 29*(1), 5–68.

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 134*(3), 501–524.

Kelly, B. T., S. Pruitt, and Y. Su (2020). Instrumented principal component analysis. *Available at SSRN 2983919*.

Lam, C. and Q. Yao (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics 40*(2), 694–726.

Ludvigson, S. C. and S. Ng (2009). Macro factors in bond risk premia. *The Review of Financial Studies 22*(12), 5027–5067.

McCracken, M. W. and S. Ng (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Pelger, M. and R. Xiong (2022). State-varying factor models of large dimensions. *Journal of Business & Economic Statistics 40*(3), 1315–1333.

Rothman, A. J., P. J. Bickel, E. Levina, and J. Zhu (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics 2*, 494–515.

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*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2006). Forecasting with many predictors. *Handbook of economic forecasting 1*, 515–554.

Stock, J. H. and M. W. Watson (2014). Estimating turning points using large data sets. *Journal of Econometrics 178*, 368–381.

Su, L. and X. Wang (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics 198*(1), 84–101.

APPENDIX A: PROOFS OF RESULTS

In this Appendix, I present main proofs for the theoretical results in Section 3. All other proofs and and technical lemmas can be found in the online supplement.

Proof of Theorem 3.1: In the semiparametric factor model, let $\mathbf{K}$ denote a $K \times K$ diagonal matrix of the first K eigenvalues of $\frac{1}{NT}\widetilde{\mathbf{Y}}\mathbf{P}\widetilde{\mathbf{Y}}'$ in descending order. By definition of eigenvalues, we have

$$\frac{1}{NT}(\widetilde{\mathbf{Y}}\mathbf{P}\widetilde{\mathbf{Y}}')\widehat{\widetilde{\mathbf{\Lambda}}} = \widehat{\widetilde{\mathbf{\Lambda}}}\mathbf{K}.$$

Let $\mathbf{H} = \frac{1}{NT}(\mathbf{Q}\mathbf{Q}'\widetilde{\mathbf{\Lambda}}' + \mathbf{Q}\widetilde{\mathbf{U}}')\widehat{\widetilde{\mathbf{\Lambda}}}\mathbf{K}^{-1}$, where $\mathbf{Q} = \mathbf{B}\mathbf{\Phi}(\mathbf{X})' + \mathbf{\Gamma}'\mathbf{P} + \mathbf{R}(\mathbf{X})'\mathbf{P}$. Note that $\widetilde{\mathbf{\Lambda}} = \widehat{\mathbf{\Sigma}}_u^{-\frac{1}{2}}\mathbf{\Lambda}$, and $\widetilde{\mathbf{U}} = \widehat{\mathbf{\Sigma}}_u^{-\frac{1}{2}}\mathbf{U}$. Substituting $\widetilde{\mathbf{Y}} = \widetilde{\mathbf{\Lambda}}\mathbf{B}\mathbf{\Phi}(\mathbf{X})' + \widetilde{\mathbf{\Lambda}}\mathbf{R}(\mathbf{X})' + \widetilde{\mathbf{\Lambda}}\mathbf{\Gamma}' + \widetilde{\mathbf{U}}$, we have

$$\widehat{\widetilde{\mathbf{\Lambda}}} - \widetilde{\mathbf{\Lambda}}\mathbf{H} = \left(\sum_{i=1}^{4}\mathbf{A}_i\right)\mathbf{K}^{-1}, \tag{A.1}$$

where

$$\mathbf{A}_1 = \frac{1}{NT}\widetilde{\mathbf{U}}\mathbf{\Phi}(\mathbf{X})\mathbf{B}'\widetilde{\mathbf{\Lambda}}'\widehat{\widetilde{\mathbf{\Lambda}}}, \quad \mathbf{A}_2 = \frac{1}{NT}\widetilde{\mathbf{U}}\mathbf{P}\mathbf{R}(\mathbf{X})\widetilde{\mathbf{\Lambda}}'\widehat{\widetilde{\mathbf{\Lambda}}},$$
$$\mathbf{A}_3 = \frac{1}{NT}\widetilde{\mathbf{U}}\mathbf{P}\mathbf{\Gamma}\widetilde{\mathbf{\Lambda}}'\widehat{\widetilde{\mathbf{\Lambda}}}, \quad \mathbf{A}_4 = \frac{1}{NT}\widetilde{\mathbf{U}}\mathbf{P}\widetilde{\mathbf{U}}'\widehat{\widetilde{\mathbf{\Lambda}}}.$$

We consider (3.3). By Lemmas S.4 and S.5 in the online supplement and $\|\widehat{\mathbf{\Sigma}}_u\|_2 < \infty$, we have

$$\frac{1}{N}\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}\|_F^2 \leq O_P\left(\frac{1}{N}\right)\|\widehat{\widetilde{\mathbf{\Lambda}}} - \widetilde{\mathbf{\Lambda}}\mathbf{H}\|_F^2 \leq O_P\left(\frac{1}{N}\|\mathbf{K}^{-1}\|_2^2\right)\sum_{i=1}^{4}\|\mathbf{A}_i\|_F^2 = O_P(1/T).$$

Consider (3.4). By Lemma S.9 in the online supplement, we have

$$\begin{aligned}\frac{1}{T}\|\widehat{\mathbf{G}}(\mathbf{X}) - \mathbf{G}(\mathbf{X})\mathbf{H}\|_F^2 &\leq \frac{2}{T}\|\mathbf{\Phi}(\mathbf{X})(\widehat{\mathbf{B}}' - \mathbf{B}'\mathbf{H})\|_F^2 + \frac{2}{T}\|\mathbf{R}(\mathbf{X})\mathbf{H}\|_F^2 \\ &= O_P(\|\widehat{\mathbf{B}}' - \mathbf{B}'\mathbf{H}\|_F^2 + J^{-\kappa}) = O_P\left(\frac{1}{J^\kappa} + \frac{J\nu_T}{T} + \frac{J}{T}m_N^2\omega_{N,T}^{2-2q}\right).\end{aligned}$$

Consider (3.5). Since $\widehat{\mathbf{\Gamma}} = \frac{1}{N}(\mathbf{I} - \mathbf{P})\widetilde{\mathbf{Y}}'\widehat{\widetilde{\mathbf{\Lambda}}}$,

$$\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\mathbf{H} = \sum_{i=1}^{6}\mathbf{D}_i,$$

where

$$\mathbf{D}_1 = \frac{1}{N}(\mathbf{I} - \mathbf{P})\mathbf{\Gamma}\widetilde{\mathbf{\Lambda}}'(\widehat{\widetilde{\mathbf{\Lambda}}} - \widetilde{\mathbf{\Lambda}}\mathbf{H}), \quad \mathbf{D}_2 = \frac{1}{N}\widetilde{\mathbf{U}}'(\widehat{\widetilde{\mathbf{\Lambda}}} - \widetilde{\mathbf{\Lambda}}\mathbf{H}),$$
$$\mathbf{D}_3 = -\mathbf{P}\mathbf{\Gamma}\mathbf{H}, \quad \mathbf{D}_4 = (\mathbf{I} - \mathbf{P})\mathbf{R}(\mathbf{X})(\mathbf{H} + \frac{1}{N}\widetilde{\mathbf{\Lambda}}'(\widehat{\widetilde{\mathbf{\Lambda}}} - \widetilde{\mathbf{\Lambda}}\mathbf{H})),$$
$$\mathbf{D}_5 = -\frac{1}{N}\mathbf{P}\widetilde{\mathbf{U}}'(\widehat{\widetilde{\mathbf{\Lambda}}} - \widetilde{\mathbf{\Lambda}}\mathbf{H}), \quad \mathbf{D}_6 = \frac{1}{N}(\mathbf{I} - \mathbf{P})\widetilde{\mathbf{U}}'\widetilde{\mathbf{\Lambda}}\mathbf{H}.$$

Then, by Lemma S.12 in the online supplement, we have

$$\frac{1}{T}\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\mathbf{H}\|_F^2 \leq O(\frac{1}{T})\sum_{i=1}^{6}\|\mathbf{D}_i\|_F^2 = O_P(\frac{1}{N} + \frac{J}{T^2} + \frac{1}{J^\kappa} + \frac{J\nu_T}{T} + \frac{1}{T}m_N^2\omega_{N,T}^{2-2q}).$$

Consider (3.6). Since $\widehat{\mathbf{F}} = \frac{1}{N}\widetilde{\mathbf{Y}}'\widehat{\widetilde{\mathbf{\Lambda}}}$, $\widehat{\mathbf{f}}_t = \frac{1}{N}\widehat{\widetilde{\mathbf{\Lambda}}}'\widetilde{\mathbf{\Lambda}}\mathbf{f}_t + \frac{1}{N}\widehat{\widetilde{\mathbf{\Lambda}}}'\widetilde{\mathbf{u}}_t$. Using $\widetilde{\mathbf{\Lambda}} = \widetilde{\mathbf{\Lambda}} - \widehat{\widetilde{\mathbf{\Lambda}}}\mathbf{H}^{-1} + \widehat{\widetilde{\mathbf{\Lambda}}}\mathbf{H}^{-1}$ and $\frac{1}{N}\widehat{\widetilde{\mathbf{\Lambda}}}'\widehat{\widetilde{\mathbf{\Lambda}}} = \mathbf{I}_K$, we have

$$\widehat{\mathbf{f}}_t - \mathbf{H}^{-1}\mathbf{f}_t = \sum_{i=1}^{3}\mathbf{M}_i,$$

where

$$\mathbf{M}_1 = \frac{1}{N}\widehat{\widetilde{\mathbf{\Lambda}}}'(\widetilde{\mathbf{\Lambda}}\mathbf{H} - \widehat{\widetilde{\mathbf{\Lambda}}})\mathbf{H}^{-1}\mathbf{f}_t, \quad \mathbf{M}_2 = \frac{1}{N}(\widehat{\widetilde{\mathbf{\Lambda}}} - \widetilde{\mathbf{\Lambda}}\mathbf{H})'\widetilde{\mathbf{u}}_t, \quad \mathbf{M}_3 = \frac{1}{N}\mathbf{H}'\widetilde{\mathbf{\Lambda}}'\widetilde{\mathbf{u}}_t.$$

Then, by Lemmas S.8 and S.13 in the online supplement,

$$\|\mathbf{M}_1\| \leq \|\frac{1}{N}\widehat{\widetilde{\mathbf{\Lambda}}}'(\widehat{\widetilde{\mathbf{\Lambda}}} - \widetilde{\mathbf{\Lambda}}\mathbf{H})\|_F\|\mathbf{H}^{-1}\|_2\|\mathbf{f}_t\| = O_P\left(\frac{1}{\sqrt{T}}m_N\omega_{N,T}^{1-q}\right).$$

Note that $\|\mathbf{\Lambda}'\mathbf{u}_t\|^2 = O_P(N)$ since $|\lambda_k'\mathbf{u}_t|^2 = O_P(1)E|\lambda_k'\mathbf{u}_t|^2 = O_P(1)\text{var}(\lambda_k'\mathbf{u}_t) =$

$O_P(1)\lambda_k' \text{var}(\mathbf{u}_t)\lambda_k \le O_P(1)\|\lambda_k\|^2\|\text{var}(\mathbf{u}_t)\| = O_P(N)$. Let $\mathbf{\Lambda}^* = \mathbf{\Sigma}_u^{-\frac{1}{2}}\mathbf{\Lambda}$ and $\mathbf{u}_t^* = \mathbf{\Sigma}_u^{-\frac{1}{2}}\mathbf{u}_t$. Then $\|N^{-1}\mathbf{\Lambda}^{*'}\mathbf{u}_t^*\| = O_P(1/\sqrt{N})$. By Lemma S.5, we have

$$\begin{aligned}\|\mathbf{M}_3\| &\le \frac{1}{N}\|\mathbf{H}\|_2(\|\mathbf{\Lambda}'(\widehat{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1})\mathbf{u}_t\| + \|\mathbf{\Lambda}^{*'}\mathbf{u}_t^*\|) \\ &\le O_P(\frac{1}{N})(\|\mathbf{\Lambda}\|_F\|\widehat{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1}\|_1\|\mathbf{u}_t\| + \|\mathbf{\Lambda}^{*'}\mathbf{u}_t^*\|) \\ &\le O_P(m_N\omega_{N,T}^{1-q} + \frac{1}{\sqrt{N}}) = O_P(m_N\omega_{N,T}^{1-q}).\end{aligned}$$

For each fixed t, it follows from $\frac{1}{N}\|\widehat{\widetilde{\mathbf{\Lambda}}} - \widetilde{\mathbf{\Lambda}}\mathbf{H}\|^2 = O_P(T^{-1})$ and $\frac{1}{N}\sum_{i=1}^N u_{it}^2 = O_P(1)$ that $\|\mathbf{M}_2\| = O_P(T^{-1/2})$. Hence, for each $t \le T$, we have $\|\widehat{\mathbf{f}}_t - \mathbf{H}^{-1}\mathbf{f}_t\| = O_P(m_N\omega_{N,T}^{1-q})$. □

Proof of Theorem 3.2: The forecasting model (3.7) can be written as

$$\begin{aligned}z_{t+h} &= \alpha'\mathbf{H}^{-1'}\widehat{\mathbf{f}}_t + \beta'\mathbf{v}_t + \epsilon_{t+h} + \alpha'\mathbf{H}^{-1'}(\mathbf{H}'\mathbf{f}_t - \widehat{\mathbf{f}}_t) \\ &= \widehat{L}_t'\delta + \epsilon_{t+h} + \alpha'\mathbf{H}^{-1'}(\mathbf{H}'\mathbf{f}_t - \widehat{\mathbf{f}}_t).\end{aligned}$$

In matrix notation, the model can be rewritten as

$$Z = \widehat{L}\delta + \epsilon + (\mathbf{F}\mathbf{H} - \widehat{\mathbf{F}})\mathbf{H}^{-1}\alpha,$$

where $Z = (z_{h+1}, ..., z_T)'$, $\epsilon = (\epsilon_{h+1}, ..., \epsilon_T)'$, and $\widehat{L} = (\widehat{L}_1, ..., \widehat{L}_{T-h})'$. Then, the ordinary least squares estimator of δ is $\widehat{\delta} = (\widehat{L}'\widehat{L})^{-1}\widehat{L}'Z$. By Lemma S.14 in the online supplement, $T^{-1/2}\widehat{L}'(\mathbf{F}\mathbf{H} - \widehat{\mathbf{F}}) = O_P(T^{1/2}m_N^2\omega_{N,T}^{2-2q})$ and $T^{-1/2}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})'\epsilon = O_P(T^{1/2}m_N^2\omega_{N,T}^{2-2q})$. Then, if $T^{1/2}m_N^2\omega_{N,T}^{2-2q} \to 0$,

$$\begin{aligned}\sqrt{T}(\widehat{\delta} - \delta) &= (T^{-1}\widehat{L}'\widehat{L})^{-1}T^{-1/2}\widehat{L}'\epsilon + (T^{-1}\widehat{L}'\widehat{L})^{-1}T^{-1/2}\widehat{L}'(\mathbf{F}\mathbf{H} - \widehat{\mathbf{F}})\mathbf{H}^{-1}\alpha \\ &= (T^{-1}\widehat{L}'\widehat{L})^{-1}T^{-1/2}\widehat{L}'\epsilon + o_P(1),\end{aligned}$$

and

$$\frac{\widehat{L}'\epsilon}{\sqrt{T}} = \begin{bmatrix} \frac{(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})'\epsilon}{\sqrt{T}} + \frac{\mathbf{H}'\mathbf{F}'\epsilon}{\sqrt{T}} \\ \frac{V'\epsilon}{\sqrt{T}} \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{H}'\mathbf{F}'\epsilon}{\sqrt{T}} \\ \frac{V'\epsilon}{\sqrt{T}} \end{bmatrix} + o_P(1).$$

Then, for a block diagonal matrix $\Pi = \text{diag}(\mathbf{H}', I)$,

$$\sqrt{T}(\widehat{\delta} - \delta) = (T^{-1}\widehat{L}'\widehat{L})^{-1}\Pi(T^{-1/2}L'\epsilon) + o_P(1).$$

Since $T^{-1/2}L'\epsilon \xrightarrow{d} N(0, \Sigma_{L,\epsilon})$ by Assumption 3.6, we can show the statement. □

Proof of Theorem 3.3: Since $\|\widehat{\mathbf{f}}_t - \mathbf{H}^{-1}\mathbf{f}_t\| = O_P(m_N\omega_{N,T}^{1-q})$ by (3.6) and $\sqrt{T}(\widehat{\delta} - \delta)$ is asymptotically normal, we have

$$\begin{aligned}\widehat{z}_{T+h|T} - z_{T+h|T} &= \widehat{\alpha}'\widehat{\mathbf{f}}_T + \widehat{\beta}'\mathbf{v}_T - \alpha'\mathbf{f}_T - \beta'\mathbf{v}_T \\ &= (\widehat{\alpha} - \mathbf{H}'\alpha)'\widehat{\mathbf{f}}_T + \alpha'\mathbf{H}(\widehat{\mathbf{f}}_T - \mathbf{H}^{-1}\mathbf{f}_T) + (\widehat{\beta} - \beta)'\mathbf{v}_T \\ &= \frac{1}{\sqrt{T}}\widehat{L}_T'\sqrt{T}(\widehat{\delta} - \delta) + \alpha'\mathbf{H}(\widehat{\mathbf{f}}_T - \mathbf{H}^{-1}\mathbf{f}_T) = O_P(m_N\omega_{N,T}^{1-q}).\end{aligned}$$

□