12-2023

Are bond returns predictable with real-time macro data?

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

Fuwei JIANG

Kunpeng LI

Guoshi TONG

Guofu ZHOU

Citation

HUANG, Dashan; JIANG, Fuwei; LI, Kunpeng; TONG, Guoshi; and ZHOU, Guofu. Are bond returns predictable with real-time macro data?. (2023). *Journal of Econometrics*. 237, (2), 1-20.
Available at: https://ink.library.smu.edu.sg/lkcsb_research/7368

Are Bond Returns Predictable with Real-Time Macro Data?*

Dashan Huang,[†] Fuwei Jiang,[‡] Kunpeng Li,[§] Guoshi Tong,[¶] Guofu Zhou[‖]

This version: September 2022

Abstract: We investigate the predictability of bond returns using real-time macro variables and consider the possibility of a nonlinear predictive relationship and the presence of weak factors. To address these issues, we propose a scaled sufficient forecasting (sSUFF) method and analyze its asymptotic properties. Using both the existing and the new method, we find empirically that real-time macro variables have significant forecasting power both in-sample and out-of-sample. Moreover, they generate sizable economic values, and their predictability is not spanned by the yield curve. We also observe that the forecasted bond returns are countercyclical, and the magnitude of predictability is stronger during economic recessions, which lends empirical support to well-known macro finance theories.

JEL codes: C22, C53, G11, G12, G17

Keywords: Bond Return Predictability, Real-Time Macro Data, Scaled Sufficient Forecasting, Machine Learning

*Corresponding authors. F. Jiang: jfuwei@gmail.com; G. Tong: gstong@fudan.edu.cn

†Lee Kong Chian School of Business, Singapore Management University, 178899, Singapore.

‡School of Finance, Central University of Finance and Economics, 102206, China.

§International School of Economics and Management, Capital University of Economics and Business, 100070, China.

¶Fanhai International School of Finance, Fudan University, 200001, China.

‖Olin School of Business, Washington University in St. Louis, Missouri, 63130, USA.

1. Introduction

One of the central themes in modern term structure models is that bond risk premia are time-varying with macroeconomic condition (see, e.g., Joslin, Priebsch, and Singleton, 2014, and references therein). Consistent with the theory, Ludvigson and Ng (2009, 2011) show empirically that bond returns in the US Treasury market can be significantly predicted by macro variables. From an investment perspective, Gargano, Pettenuzzo, and Timmermann (2019) and Bianchi, Büchner, and Tamoni (2021) find that forecasting bond returns can generate sizeable economic values. However, Ghysels, Horan, and Moench (2018) and Fulop, Li, and Wan (2021) show that macro variables are unable to significantly predict bond returns once real-time data are used, thereby raising an important question on the conclusion of many empirical and theoretical studies.

In this paper, we re-examine bond return predictability with real-time macro variables. We allow for both a nonlinear predictive relation and the presence of weak factors. There are two reasons for this assumption. First, traditional studies typically assume that the forecast target is linearly linked with some latent strong factors that summarize the predictive information from a panel of macro variables. While this modelling strategy is intuitive and appealing, it is restrictive in the real world. Perhaps because of this, Ludvigson and Ng (2009) include partial squared factors to improve the forecasting ability. Since there is no firm evidence to support such a squared form, it is of interest to permit a general nonlinear setup. Second, strong factors are largely imposed in existing studies, but it is often the case in applications that the factors with the strongest predictive power may not be the largest principal components, and, for this reason, they may not be strong factors. Recently, there is also a growing interest to incorporate weak factors in empirical studies, e.g., Lettau and Pelger (2020a,b), Giglio, Xiu, and Zhang (2021), among others. This line of research emphasizes the necessity of using weak factors for modeling financial data.

We propose a new forecasting method, called scaled sufficient forecasting (hereafter abbreviated by sSUFF), to predict bond returns, inspired by and based on the sufficient forecasting (SUFF) method originally proposed by Fan, Xue, and Yao (2017).[1] SUFF is suitable for a nonlinear predictive relation with a large set of predictors, but we show that the presence of weak factors can deteriorate the performance of SUFF, due to a term representing the noise-to-signal ratio that arises when the factors are weak.

The theoretical contribution of this paper is twofold. First, we extend SUFF to sSUFF by allowing for weak factors, which is new in the econometric literature and has potentially wide applications. We establish the asymptotic properties for sSUFF, and show that it outperforms SUFF in the presence of weak factors. As a byproduct, we find that the convergence rates found in Fan et al. (2017) can be slightly sharpened. Second, we run simulations to assess the finite sample performance of sSUFF. The experiment results lend support to sSUFF in small samples. It is worthwhile to mention that, although sSUFF is designed to address issues related to weak factors, its forecasting performance is robust when all the factors are strong.

In the empirical application, we consider the same forecast targets and real-time macro variables as Ghysels et al. (2018). That is, we use annual holding period returns on 2- to 5-year government bonds as the forecast targets, which are computed based on the Fama-Bliss zero-coupon bond yields.[2] We collect real-time vintages of 60 macro variables from March 1982 to December 2019, where each vintage has observations from January 1968. The real-time data are different from the final revised data used in Ludvigson and Ng (2009, 2011), because the value of a variable in one month could be revised multiple times after its first release, due to estimation errors or measurement errors. On the other hand, although the final revised data are of high quality, they are not available at the time of forecasting, and only the real-time data are.

There are four main findings when we apply sSUFF to forecasting bond returns

[1] sSUFF and the scaled principal component analysis (sPCA) in Huang, Jiang, Li, Tong, and Zhou (2022) both scale predictors, but they are different as sSUFF is nonlinear.

[2] For robustness, in Section 3.4 we also consider forecasting monthly holding period bond returns based on Gürkaynak, Sack, and Wright's (2007) dataset and find similar results.

in real time. First, and foremost, we show that bond returns can be significantly predicted by the noisier real-time macro variables although the results are stronger with the final revised data. For example, over the March 1982 to December 2019 period, the in- and out-of-sample R^2s are 22.99% and 11.21% in forecasting the 5-year bond returns, and they are 31.81% and 7.00% in forecasting 2-year bond returns,[3] where the out-of-sample R^2 is computed against the historical mean forecast. This result is robust to different model specifications. In contrast, SUFF displays weaker predictive power, especially for out-of-sample forecasting. When comparing with other forecast methods such as PCA, sPCA, PLS, and SUFF,[4] sSUFF performs the best across all bond maturities. We attribute the success to its ability to capture factor weakness and nonlinearity.

Second, we find that forecasted bond returns with sSUFF generate sizeable economic values. Consider a mean-variance investor who allocates her wealth between an n-year bond ($n > 1$) and a 1-year risk-free bond. We show that the investor is able to improve her portfolio performance if she estimates the n-year bond expected returns by using sSUFF relative to other forecast benchmarks, such as historical mean forecast. For example, when investing in the 5-year bond with a risk aversion of 5, the investor can obtain 1.31% more annualized certainty equivalent returns than the case if she uses the historical mean forecast. This finding provides a support to the positive economic values of bond return predictability documented by Gargano et al. (2019) with final revised macro variables. In contrast, the forecasts based on SUFF and linear forecasts of PCA and PLS generate negative certainty equivalent returns and, therefore, underperform the historical mean forecast if investing in the 2-year bond. Thus, our results help to reconcile Thornton and Valente (2012) and Sarno, Schneider, and Wagner (2016) who find it difficult to translate statistical predictability into economic value.

Third, we explore the sources of bond return predictability and find that real-

[3]It should be noted that, as shown by Inoue and Kilian (2005), there is no theoretical relation between the in- and out-of-sample R^2s and the out-of-sample R^2 can be larger than the in-sample one. Also, the out-of-sample R^2 depends on the choice of the forecast benchmark.

[4]See, e.g., Huang, Jiang, Tu, and Zhou (2015); Jiang, Lee, Martin, and Zhou (2019); Huang et al. (2022).

time macro factors strongly predict future macroeconomic conditions as measured by the Chicago Fed National Activity Index (CFNAI), consumption, employment, recession probability, macro uncertainty, and yield spread. This result is important as Cochrane (2007) argues that return predictability is more economically compelling if the predictors are able to predict future macroeconomic conditions. Hence, putting together our results on statistical tests, economic value, and future macroeconomic conditions, there is strong evidence that bond risk premia are time-varying and can be predicted by the real-time macro variables.

Finally, we examine the relation between the degree of bond return predictability and macroeconomic conditions, and find that the return predictability is more pronounced in recessions, and that the bond term premia are countercyclical. Specifically, we show that the forecasting accuracy and forecasted bond returns generated by sSUFF are both stronger in periods with low economic activity, high recession probability and high uncertainty. This finding is consistent with Gargano et al. (2019) and Bianchi et al. (2021), though their forecasts are based on the final revised macro variables. Moreover, in line with Ludvigson and Ng (2009), the term premium implied by sSUFF is negatively correlated with macroeconomic conditions such as the employment rate.

Our paper contributes to five strands of literature. First, it helps to resolve the debate on real-time macro variable predictability on bond returns. Second, it provides evidence supporting recent growing studies on nonlinearity of bond factors (see, e.g., Feldhütter, Heyerdahl-Larsen, and Illeditsch, 2018; Breach, D'Amico, and Orphanides, 2020; Giacoletti, Laursen, and Singleton, 2021). Third, it generalizes the weak factor studies on equity markets into bond markets. Fourth, it provides evidence to support bond pricing beyond the spanning hypothesis that restricts all pricing information to yields, extending Zhao, Zhou, and Zhu (2021) without non-US macros. Fifth, it adds to the large predictability literature by providing a new method, sSUFF, which can be widely used for forecasting stocks, foreign exchanges, and other asset classes.

The rest of the paper is organized as follows. Section 2 proposes sSUFF and explores its asymptotic properties. Section 3 applies sSUFF and other methods to forecasting bond returns with real-time macro variables. Section 4 concludes.

2. Econometric Framework

2.1. *Nonlinear factor-based forecasting framework*

We consider a predictive framework in which the forecast target is nonlinearly related to latent factor-related indices, where the latent factors are contained in a large number of predictors. Specifically, the forecast target y_{t+1} and the predictor x_{it} admit the following expression for $i = 1, \ldots, N$ and $t = 1, \ldots, T$:

$$y_{t+1} = h(\phi_1' \boldsymbol{f}_t, \phi_2' \boldsymbol{f}_t, \ldots, \phi_k' \boldsymbol{f}_t) + \epsilon_{t+1}, \tag{1}$$

$$x_{it} = \boldsymbol{b}_i' \boldsymbol{f}_t + u_{it}, \tag{2}$$

where x_{it} is the i-th observable predictor at time t, $\boldsymbol{b}_i$ is an r-dimensional vector of factor loadings, and $\boldsymbol{f}_t$ is an r-dimensional vector of factors, which carries the forecasting information contained in x_{it}'s. Both $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$ are unobservable. Throughout the paper, we assume $r > k$, i.e., only partial linear combinations of the factors enter into the predictive equation (1). u_{it} is the idiosyncratic error that cannot be explained by $\boldsymbol{f}_t$, and ϵ_{t+1} is the forecasting error. $h(\cdot)$ is some unknown nonlinear link function. Leaving $h(\cdot)$ unspecified has an advantage of the model's flexibility and generality. However, we emphasize that although (1) does not specify the link function, it does specify that given the predictive indices $\phi_1' \boldsymbol{f}_t, \ldots, \phi_k' \boldsymbol{f}_t$, the factors $\boldsymbol{f}_t$ or the predictors x_{it}'s do not provide any incremental forecasting power. Fan et al. (2017) call these indices *sufficient predictive indices*, and their forecasting method *sufficient forecasting*. In retrospect, the idea of sufficient forecasting dates back to the sliced inverse regression of the seminal work by Li (1991), where a salient result is that under some linear conditions, the sufficient predictive indices, which are also known as central subspace in the sufficient dimension reduction literature (Cook and Ni, 2005; Wang and Xia, 2008; Cook, 2009), can be recovered by the eigenvectors' space of the covariance matrix of the predictors conditional on the forecast target.

Fan et al. (2017) extend the sufficient dimension reduction technique of Li (1991) to a factor model, which on the one hand makes the forecasting framework consistent with the popular economic and financial theories, such as factor asset pricing models (e.g., Lettau and Pelger, 2020a,b), and on the other hand maintains the flexibility of specification, particularly on the relationship between the forecast target and the latent factors. This paper further extends Fan et al. (2017) to make the sufficient forecasting robust to weak factors. To this end, we first give a brief review on the sufficient forecasting

2.2. *Sufficient forecasting (SUFF)*

The sufficient forecasting proposed by Fan et al. (2017) is based on Li (1991), which shows that if the linear projection condition

$$\mathbb{E}(\boldsymbol{\psi}'\boldsymbol{f}_t|\boldsymbol{\phi}_1'\boldsymbol{f}_t,\ldots,\boldsymbol{\phi}_k'\boldsymbol{f}_t) = \alpha_1\boldsymbol{\phi}_1'\boldsymbol{f}_t + \alpha_2\boldsymbol{\phi}_2'\boldsymbol{f}_t + \cdots + \alpha_k\boldsymbol{\phi}_k'\boldsymbol{f}_t \tag{3}$$

holds for any $\boldsymbol{\psi} \in \mathbb{R}^r$ with $\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_k$ being the indices given in (1), then

$$\mathbb{E}(\boldsymbol{f}_t|y_{t+1}) = \Phi\boldsymbol{a}(y_{t+1}), \tag{4}$$

where $\Phi = [\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\ldots,\boldsymbol{\phi}_k]$ is an $r \times k$ indices matrix and $\boldsymbol{a}(y_{t+1})$ is a k-dimensional vector.[5] With this result, one can estimate the covariance of $\boldsymbol{f}_t$ conditional on y_{t+1} as

$$\mathrm{cov}(\boldsymbol{f}_t|y_{t+1}) = \Phi\mathbb{E}\Big[\boldsymbol{a}(y_{t+1})\boldsymbol{a}(y_{t+1})'\Big]\Phi'. \tag{5}$$

Equation (5) is important to subsequent analyses. It says that the unknown indices matrix Φ falls in the k largest eigenvectors subspace of the conditional

[5]Readers are referred to Yu, Yao, and Xue (2022) for a detailed proof or Theorem 3.1 of Li (1991) for a general treatment.

covariance matrix $\mathrm{cov}(\boldsymbol{f}_t|y_{t+1})$. So the remaining job is to consistently estimate this covariance matrix and extract its k largest eigenvectors. Unfortunately, the estimation of $\mathbb{E}(\boldsymbol{f}_t|y_{t+1})$ or $\mathrm{cov}(\boldsymbol{f}_t|y_{t+1})$ is a non-trivial task. Li (1991) proposes the sliced inverse regression, which has been widely adopted in the literature. To see the intuition, we partition the support of y_{t+1} into H intervals and denote the h-th interval by I_h. If the length of I_h is small, $\mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h)$ is approximate to $\mathbb{E}(\boldsymbol{f}_t|y_{t+1} = x)$ for $x \in I_h$. Based on this observation, a feasible estimator for $\mathrm{cov}(\boldsymbol{f}_t|y_{t+1})$ is

$$\Sigma_{\boldsymbol{f}|\boldsymbol{y}} = \frac{1}{H}\sum_{h=1}^{H} \mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h)\mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h)'. \tag{6}$$

Then, we replace $\mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h)$ by its sample analog. One natural method is to replace $\boldsymbol{f}_t$ with its PCA estimate $\widehat{\boldsymbol{f}}_t$. For a given I_h, let $\widehat{\boldsymbol{f}}_{h,l}$ denote the l-th sample point of $\widehat{\boldsymbol{f}}$. That is, for those $\widehat{\boldsymbol{f}}_t$ satisfying $y_{t+1} \in I_h$, $\widehat{\boldsymbol{f}}_{h,l}$ is the l-th sample point in this subset. Hence, $\mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h)$ can be estimated by $\frac{1}{c_h}\sum_{l=1}^{c_h} \widehat{\boldsymbol{f}}_{h,l}$, where c_h is the number of sample points in I_h. Fan et al. (2017) suggest choosing I_h to make c_h identical across the intervals, which gives

$$\widehat{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}} = \frac{1}{H}\sum_{h=1}^{H}\left[\frac{1}{c}\sum_{l=1}^{c}\widehat{\boldsymbol{f}}_{h,l}\right]\left[\frac{1}{c}\sum_{l=1}^{c}\widehat{\boldsymbol{f}}_{h,l}\right]'. \tag{7}$$

Based on (7), Fan et al. (2017) proposes the following sufficient forecasting algorithm.

Algorithm 1 Sufficient forecasting with strong factors

1. Estimate the PCA factors $\widehat{\boldsymbol{f}}_t$;
2. Construct $\widehat{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}}$ as (7);
3. Obtain $\widehat{\phi}_1, \widehat{\phi}_2, \ldots, \widehat{\phi}_k$ by extracting the k largest eigenvectors of $\widehat{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}}$;
4. Construct the predictive indices $\widehat{\phi}_1'\widehat{\boldsymbol{f}}_t, \widehat{\phi}_2'\widehat{\boldsymbol{f}}_t, \ldots, \widehat{\phi}_k'\widehat{\boldsymbol{f}}_t$;
5. Using local linear regression (Fan and Gijbels, 1996) to estimate $h(\cdot)$ with indices from Step 4, and forecast y_{t+1}.[6]

[6]We use the R package *np* to implement the local linear regression. Specifically, we use the *npregbw* function with parameters *regtype="ll"* and *bwmethod="cv.aic"* to estimate the bandwidth, and the *npreg* function to obtain the fitted values.

Algorithm 1 involves three tuning parameters, the number of slices (H), the number of factors (r), and the number of predictive indices (k). Because it is found in practice that the value of H has little effect on the final result, we follow Fan et al. (2017) and assume that H is larger than k but a finite value. Under this assumption, there are $O(T)$ sample points in each interval I_h that leads to a faster convergence rate (Remark 1 below has further discussion). The number of factors r can be determined by the method of Onatski (2010). We note that since the factors to be identified can be weak, some existing methods on strong factors such as Bai and Ng (2002) are inapplicable. Other methods, such as the eigenvalue-ratio method of Ahn and Horenstein (2013), seem working in a weak factor setup, but rigorous justifications are still deficient in the literature. As regard to determining the value of k, one can use the χ^2 test of Li (1991). For details, see Li (1991) and Fan et al. (2017).

2.3. *Scaled sufficient forecasting (sSUFF)*

We extend Fan et al. (2017) to a weak factor framework and propose the following algorithm.

Algorithm 2 Scaled sufficient forecasting with weak factors

1-3. The same as SUFF in Algorithm 1

4. For each i, regress x_{it} on $\widehat{\phi}_1'\widehat{\boldsymbol{f}}_t, \cdots, \widehat{\phi}_k'\widehat{\boldsymbol{f}}_t$ to obtain $\widehat{\gamma}_i$, the R-squared of the regression.
5. Apply PCA to the scaled panel $\{\widehat{\gamma}_i x_{it}\}$ to obtain new factor $\widetilde{\boldsymbol{f}}_t$.
6. Construct $\widetilde{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}}$ with $\widetilde{\boldsymbol{f}}_t$ by the same way as in Step 2.
7. Obtain $\widetilde{\phi}_1, \cdots, \widetilde{\phi}_k$ from the k largest eigenvectors of $\widetilde{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}}$ and construct predictive indices $\widetilde{\phi}_1'\widetilde{\boldsymbol{f}}_t, \cdots, \widetilde{\phi}_k'\widetilde{\boldsymbol{f}}_t$.
8. Use local linear regression to estimate $h(\cdot)$ with $\widetilde{\phi}_1'\widetilde{\boldsymbol{f}}_t, \cdots, \widetilde{\phi}_k'\widetilde{\boldsymbol{f}}_t$ and forecast y_{t+1}.

The intuition of sSUFF can be interpreted as follows. When the factors are weak, as long as the weakness is not extreme, SUFF can still deliver a consistent estimate of the predictive indices. We then use these estimated predictive indices as the base to detect which predictors contain useful information and which do not. For a predictor that contains more useful information, the regression of this predictor on the estimated predictive indices will have a larger R^2, and vice versa. As shown below, for those irrelevant predictors, the corresponding R^2s have an order of $O_p(\frac{1}{N^\nu}) + O_p(\frac{1}{T})$ with $\nu \in (0,1]$, which shrinks to zero as N and T approach infinite. Hence, by using R^2 as the scaling parameter, we strengthen the useful predictors and downplay the irrelevant ones, thereby improving the forecasting power relative to SUFF. The tuning parameters H, r, and k can be chosen the same as Algorithm 1.

2.4. *Asymptotic results*

We make the following assumptions for the asymptotic analysis. Hereafter, C denotes a generic constant that is large enough and may be different at each appearance.

Assumption A: The factors satisfy

(i) $\boldsymbol{f}_t$ is independent and identically distributed over t with $E(\boldsymbol{f}_t) = 0$, $\sup_t E\|\boldsymbol{f}_t\|^8 \leq C$ and $\frac{1}{T}\boldsymbol{F}'\boldsymbol{F} = I_r$, where $\boldsymbol{F} = (\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_T)'$.

(ii) $\mathbb{E}(\psi'\boldsymbol{f}_t|\phi_1'\boldsymbol{f}_t, \ldots, \phi_k'\boldsymbol{f}_t)$ is a linear function of $\phi_1'\boldsymbol{f}_t, \ldots, \phi_k'\boldsymbol{f}_t$ for any $\psi \in \mathbb{R}^r$, where $\phi_1'\boldsymbol{f}_t, \ldots, \phi_k'\boldsymbol{f}_t$ are the indices in model (1).

Assumption B: The factor loadings $\boldsymbol{b}_i$ are nonrandom and satisfy $\max_{1\leq i\leq N} \|\boldsymbol{b}_i\| \leq C$. Let $\mathcal{I}_o = \{i : \|P_\Phi \boldsymbol{b}_i\| > \underline{c}\}$ and $\mathcal{I}_b = \{i : \|P_\Phi \boldsymbol{b}_i\| = 0; \boldsymbol{b}_i \neq 0\}$ denote the sets of predictors that contain the relevant and irrelevant factors, respectively. $\underline{c}$ is a small positive constant and $P_\Phi = \Phi(\Phi'\Phi)^{-1}\Phi'$ with $\Phi = [\phi_1, \phi_2, \ldots, \phi_k]$. We assume $\mathrm{Card}(\mathcal{I}_o) \asymp N^\nu$ and $\mathrm{Card}(\mathcal{I}_b) \asymp N^\nu$ for some $\nu \in (0,1]$, where $\mathrm{Card}(\cdot)$ denotes the cardinality of the input set, i.e., the number of elements of the input set, and $a \asymp b$ means that there exist two constants c and C such that $cb \leq a \leq Cb$. In addition, $N^{-\nu}\boldsymbol{B}'\boldsymbol{B} \to \Sigma_b$ for some $r \times r$ positive definite matrix Σ_b, where $\boldsymbol{B} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_N)'$.

Assumption C: $u_{it} = \sigma_i e_{it}$, where e_{it} is independent and identically distributed over i and t with $\mathbb{E}(e_{it}^{16}) \leq C$. In addition, $C^{-1} \leq \sigma_i \leq C$.

Assumption D: ϵ_{t+1} is independent over t. Furthermore, $\{\boldsymbol{f}_t\}_{t=1,\ldots,T}$, $\{\epsilon_{t+1}\}_{t=1,\ldots,T}$ and $\{u_{it}\}_{t=1,\ldots,T}$ are mutually independent for each i.

Assumption A is about factors. Assumption A(i) is often imposed in the studies on large dimensional factor analysis, see, e.g., Bai and Ng (2002) and Bai (2003). The eighth moment condition is more restrictive than the usual fourth moment one, but it is necessary for our theoretical analysis of sSUFF. Assumption A(ii) is crucial to the validity of sliced inverse regression, and is also imposed in Fan et al. (2017).

Assumption B is new and imposed to characterize weak factors. Let $\mathcal{I}$ be the whole set of predictors. In this paper, we divide $\mathcal{I}$ into three sets. The predictors (i) have a factor structure and are relevant to the forecast target in the first set $\mathcal{I}_o$, (ii) have a factor structure but are irrelevant to the forecast target in the second set $\mathcal{I}_b$, and (iii) are pure noises in the third set $\mathcal{I} \backslash (\mathcal{I}_o \cup \mathcal{I}_b)$. Assumption B specifies that the weakness of factors is due to the inclusion of purely noisy predictors in the third set. This specification, which is different from the existing literature, proposes an alternative and plausible way to define weak factors.

To see the insight of Assumption B, let us take the typical predictive regression as an example. When there are a large number of predictors, one basic thought in modern statistics is that only a small number of predictors are relevant, and the remaining are irrelevant. It is this thought that motivates the machine learning techniques to simultaneously select and estimate the model, such as LASSO, smoothly clipped absolute deviation (SCAD), and Dantzig selector. In this paper, we assume that only a small fraction of the predictors, belonging to $\mathcal{I}_o \cup \mathcal{I}_b$ (of order N^ν), contain the factors, and the remaining predictors are pure noises. Our assumption on weak factors is consistent with the above basic thought. It is in the similar spirit of Huang et al. (2022) but different from the usual definition in the literature. Consider $x_{it} = \boldsymbol{b}_i' \boldsymbol{f}_t + u_{it}$. Existing studies typically assume $\|\boldsymbol{b}_i\| = O(N^{\nu-1})$ for each i and some $\nu < 1$ to specify the weakness of factors (see, e.g., Lettau and Pelger, 2020a; Onatski, 2010). This assumption has an undesirable implication. The size of signals contained in each predictor, the variance of $\boldsymbol{b}_i' \boldsymbol{f}_t$, is $O(N^{2(\nu-1)})$, depends on the number of predictors N and shrinks to zero as $N \to \infty$. But asset pricing theories assert that the signals contained in each predictor are independent of the number of predictors. To preclude this inconsistency, we instead assume that each predictor contains a positive amount of, or zero, signals. We also allow for the presence of seemingly important predictors, which we classify to set $\mathcal{I}_b$. Different from the pure noise predictors in $\mathcal{I}\backslash(\mathcal{I}_o \cup \mathcal{I}_b)$, the predictors in $\mathcal{I}_b$ have high correlations with those in $\mathcal{I}_o$ due to the irrelevant factors, which may cause a false impression that these predictors in $\mathcal{I}_b$ are also important for prediction.

Assumption B also specifies that the loadings are fixed values and treated as parameters, which facilitates our theoretical analysis. For example, it allows us to bound $\max_i \|\boldsymbol{b}_i\|$ by a large constant. If the loadings are random, our analysis can be viewed as conditional on their realizations. This treatment is widely used in factor analysis, see Bai and Ng (2002) and Bai (2003). Following the literature (see, e.g., Lettau and Pelger, 2020a; Onatski, 2010), Assumption B also specifies that all the factors have the same degree of weakness. In real data, it is possible that partial factors are strong and the remaining ones are weak. Then, the current analysis can be viewed as exploring the transformed data that are obtained by subtracting the estimated strong factors from the original data.

Assumption C precludes the correlations in time series and cross section. It can be relaxed to allow more general cross sectional and temporal correlations and heteroskedasticity, at the cost of imposing higher moment conditions. This type of extension has been elaborated in many studies, see, e.g., Bai and Li (2016), so we do not pursue it in this paper.

The following theorems provide the asymptotic results for the estimators of SUFF and sSUFF in a weak factor framework.

Proposition 1. *Let $\widehat{\boldsymbol{D}}$ be the diagonal matrix whose diagonal elements are the largest r eigenvalues of $\frac{1}{N^\nu T}\boldsymbol{X}'\boldsymbol{X}$ with $\boldsymbol{X} = (x_{it})_{N\times T}$. Under Assumptions A-D, as $N, T \to \infty, N^{1-\nu}/T \to 0$,*

$$\tau_i(\widehat{\boldsymbol{D}}) - \tau_i\Big(\frac{1}{N^\nu}\boldsymbol{B}'\boldsymbol{B}\Big) = o_p(1),$$

where $\tau_i(\cdot)$ denotes the i-th largest eigenvalue of its input.

Theorem 1. *Let $\widehat{\boldsymbol{F}}$ be the PCA estimator of $\boldsymbol{F}$ for model (2), and $\widehat{\Phi}$ be the estimator of Φ given in Algorithm 1. Under Assumptions A–D, as $N \to \infty, T \to \infty$, and $N^{1-\nu}/T \to 0$, we* *have*

$$(i)\ \ \frac{1}{\sqrt{T}}\|\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R}\| = O_p(\frac{1}{\sqrt{N^{\nu}}}) + O_p(\frac{N^{1-\nu}}{T}),$$
$$(ii)\ \ \|\widehat{\Phi}-\boldsymbol{R}'\Phi\boldsymbol{R}^*\| = O_p(\frac{N^{1-\nu}}{T}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{N^{\nu}}),$$

where $\boldsymbol{R} = \frac{1}{N^{\nu}T}\boldsymbol{B}'\boldsymbol{B}\boldsymbol{F}'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1}$ *is the rotational matrix among factors* $\boldsymbol{f}_t$ *and* $\boldsymbol{R}^* = (\Phi'\Phi)^{-1/2}\boldsymbol{R}_{\Omega}$ *is the rotational matrix among indices* $\phi_1,\ldots,\phi_k$*, with* $\boldsymbol{R}_{\Omega}$ *being the eigenvector matrix of*

$$(\Phi'\Phi)^{1/2}\Big[\frac{1}{H}\sum_{h=1}^{H}\mathbb{E}\Big(\boldsymbol{a}(y_{t+1})|y_{t+1}\in I_h\Big)\mathbb{E}\Big(\boldsymbol{a}(y_{t+1})|y_{t+1}\in I_h\Big)'\Big](\Phi'\Phi)^{1/2}.$$

We have two remarks on Theorem 1.

Remark 1. When $\nu = 1$, factors in model (2) are strong. Theorem 1 indicates that the convergence rate of the estimated indices $\widehat{\Phi}$ is $O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{T}})$, which is slightly sharper than $O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{\sqrt{T}})$ in Fan et al. (2017). The reason is as follows. In the weak factor setup, the $O_p(\frac{1}{\sqrt{N^{\nu}}})$ term comes from the leading term of $\widehat{\boldsymbol{f}}_t - \boldsymbol{R}'\boldsymbol{f}_t$, which admits

$$\begin{aligned}\widehat{\boldsymbol{f}}_t - \boldsymbol{R}'\boldsymbol{f}_t &= \boldsymbol{R}'(\boldsymbol{B}'\boldsymbol{B})^{-1}\sum_{i\in\mathcal{I}_o\cup\mathcal{I}_b}\boldsymbol{b}_i u_{it} + \kappa_t \\ &= \frac{1}{\sqrt{N^{\nu}}}\boldsymbol{R}'\Big[\frac{1}{N^{\nu}}\boldsymbol{B}'\boldsymbol{B}\Big]^{-1}\frac{1}{\sqrt{N^{\nu}}}\sum_{i\in\mathcal{I}_o\cup\mathcal{I}_b}\boldsymbol{b}_i u_{it} + \kappa_t, \end{aligned} \tag{8}$$

where κ_t is the remaining term. Note that a key step to construct the conditional covariance matrix is to take average over $\widehat{\boldsymbol{f}}_t$ in each sliced interval I_h for $h = 1, 2, \ldots, H$. For those f_t's whose corresponding y_{t+1}'s fall in the same sliced interval I_h, we have

$$\frac{1}{c}\sum_{l=1}^{c}\widehat{\boldsymbol{f}}_{h,l} - \boldsymbol{R}'\frac{1}{c}\sum_{l=1}^{c}\boldsymbol{f}_{h,l} = \frac{1}{\sqrt{N^{\nu}}}\boldsymbol{R}'\Big[\frac{1}{N^{\nu}}\boldsymbol{B}'\boldsymbol{B}\Big]^{-1}\frac{1}{\sqrt{N^{\nu}}c}\sum_{l=1}^{c}\sum_{i\in\mathcal{I}_o\cup\mathcal{I}_b}\boldsymbol{b}_i u_{i,hl} + \frac{1}{c}\sum_{l=1}^{c}\kappa_{h,l}. \tag{9}$$

Given the fact that $\mathbb{E}(\|\sum_{l=1}^{c}\sum_{i\in\mathcal{I}_o\cup\mathcal{I}_b}\boldsymbol{b}_i u_{i,hl}\|^2) = O(N^{\nu}T)$ because there are $O(T)$ sample points falling in this interval, the first term becomes $O_p(\frac{1}{\sqrt{N^{\nu}T}})$, which is negligible in that it is dominated by $O_p(\frac{1}{\sqrt{T}})$. Due to this reason, the $O_p(\frac{1}{\sqrt{N^{\nu}}})$ term is gone in Theorem 1. $\square$

Remark 2. In Theorem 1, the term $O_p(\frac{N^{1-\nu}}{T})$ is of our interest. Intuitively, since we have N predictors and T periods, the magnitude of noise is of $O_p(N) + O_p(T)$. In the presence of weak factors, the magnitude of information is $O_p(N^\nu T)$. So the noise-to-signal ratio is of $O_p(\frac{N^{1-\nu}}{T}) + O_p(N^{-\nu})$. In practical applications, it is not uncommon that the number of predictors N is larger than the number of periods T, so the noise-to-signal ratio further reduces to $O_p(\frac{N^{1-\nu}}{T})$. Note that when N is much larger than T, the term $O_p(\frac{N^{1-\nu}}{T})$ is likely to dominate the term $O_p(\frac{1}{\sqrt{T}})$. In this case, the estimation error is unnecessarily large and unacceptable since if we discard the irrelevant predictors, this term would disappear and the estimation precision is improved. This fact motivates us to propose the new procedure of sSUFF. $\square$

Theorem 2. *Let $\widetilde{\boldsymbol{F}}$ is the PCA estimator of $\boldsymbol{F}$ for the scaled data $(\widehat{\gamma}_i x_{it})_{N\times T}$, and $\widetilde{\Phi}$ be the estimator of Φ given in Algorithm 2. Under Assumptions A–D, as $N \to \infty, T \to \infty$, and $N^{1-\nu}/T \to 0$, we have*

$$
\begin{aligned}
&(i) \quad \frac{1}{\sqrt{T}}\|\widetilde{\boldsymbol{F}} - \boldsymbol{F}\widetilde{\boldsymbol{R}}\| = O_p(\frac{1}{\sqrt{N^\nu}}) + O_p(\frac{1}{T}) + O_p(\frac{N^{1-\nu}}{T\sqrt{T}}),\\
&(ii) \quad \|\widetilde{\Phi} - \widetilde{\boldsymbol{R}}'\Phi\boldsymbol{R}^*\| = O_p(\frac{1}{N^\nu}) + O_p(\frac{1}{\sqrt{T}}),
\end{aligned}
$$

where $\widetilde{\boldsymbol{R}} = \frac{1}{N^\nu T}\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{b}_i\boldsymbol{b}_i'\boldsymbol{F}'\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{D}}^{-1}$ is the rotational matrix for the scaled data and $\boldsymbol{R}^*$ is defined in Theorem 1

Theorem 2 shows that Algorithm 2 kills the $O_p(\frac{N^{1-\nu}}{T})$ term in the estimation error of $\widetilde{\boldsymbol{F}}$ and $\widetilde{\Phi}$, thereby improving the forecasting power.

Given the above two theorems, we have the following corollary.

Corollary 1. *Under Assumptions A–D, $N^{1-\nu}/T \to 0$, $N^{1-\nu}/\sqrt{T} \to \infty$, and $N/T \to \infty$, sSUFF outperforms SUFF.*

Corollary 1 is obtained based on the condition that $\frac{N^{1-\nu}}{T}$ dominates $\frac{1}{\sqrt{T}}$ and $\frac{1}{N^\nu}$, so the estimation precisions of $\boldsymbol{F}$ and Φ are substantially improved. The comparison of forecasting performance here is indirect in the sense that we resort to a comparison of the key part $\Phi \boldsymbol{f}_t$, instead of the direction comparison of the forecasting error of y_{t+1}. Indeed, we can implement the latter comparison, but the related theoretical analysis involves the smoothness condition on the link function $h(\cdot)$, as well as other regularity conditions, which are not closely related to Assumptions A–D. For this reason, we do not pursue this direction. We note that recently Yu, Yao, and Xue (2022) provide a rigorous analysis in this direction with strong, but diverging number of, factors. For readers who are interested in this direct comparison, we refer to their paper.

It should be noted that if one has the correct prior or knows which predictors are useful and which are not, an application of SUFF method will deliver better asymptotic convergence rates than the sSUFF. Moreover, if we can find consistently the useful units by screening on the initial PCA estimates, using the sSUFF may not be necessary. Nevertheless, to the extent that it is often difficult to have prior knowledge on the true predictor and there is no availability of highly accurate screening methods, our proposed sSUFF is useful and applicable generally.

2.5. *Simulation evidence*

In this section, we conduct Monte Carlo simulations to explore the finite sample forecasting performance of sSUFF with either weak or strong factors.

To make the comparison convincing, we consider two factor models. The first is from Huang et al. (2022), where the predictors are governed by only two latent factors, f_{1t} and f_{2t}, with f_{1t} being target relevant. f_{1t} and f_{2t} are independent and normally distributed with mean zero and unit variance, i.e., $f_{jt} \sim N(0,1)$ for $j = 1,2$. The forecast target is $y_{t+1} = f_{1t} + \epsilon_{t+1}$ when the link function $h(\cdot)$ is linear, and $y_{t+1} = f_{1t} + f_{1t}^2 + \epsilon_{t+1}$ when the link function $h(\cdot)$ is nonlinear. In both cases, we assume $\epsilon_{t+1} \sim N(0,1)$, so that the infeasible best forecast has a lowest mean squared forecast error (MSFE) of 1. There are N observable predictors $X_{i,t}$ ($i = 1,...,N$), which load on both the relevant factor f_{1t} and the irrelevant factor f_{2t}. The factors are weak if only $n < N$ predictors have non-zero loadings on them, and strong otherwise. Each predictor X_i also contains an idiosyncratic noise term u_{it}, which is generated from a normal distribution with zero mean and standard deviation of σ_i.

The second factor model is from Fan et al. (2017). When the link function is linear, we assume a total of five factors, $f_{1t},...,f_{5t}$, and generate the forecast target as $y_{t+1} = 0.8f_{1t} + 0.5f_{2t} + 0.3f_{3t} + \epsilon_{t+1}$. When the link function is nonlinear, we instead assume a total of seven factors, $f_{1t},...,f_{7t}$, and generate the forecast target as $y_{t+1} = f_{1t}(f_{2t} + f_{3t} + 1) + \epsilon_{t+1}$. Similar as the first factor model, the factors are independently and normally distributed, and either $n < N$ predictors or all N predictors have non-zero loadings on the latent factors. The idiosyncratic noise term u_{it} of X_i is normally distributed with zero mean and standard deviation σ_i.

Table 1 reports the simulated results of sSUFF under varying degrees of factor weakness. For comparison, we also report the results using PCA, sPCA, and SUFF. Thus, we consider four forecasting methods. For each method, the simulation relies on 1,000 repetitions with $N = 500$ predictors and $T = 550$ observations. Among the 500 predictors, we assume $n = 10$, 20, 30, 40, 50, and 500 predictors having non-zero loadings on the latent factors, respectively. The non-zero loadings are drawn independently from a standard uniform distribution with support $[0,1]$, denoted by $U[0,1]$. The standard deviations of idiosyncratic noises, $\sigma_1, \cdots, \sigma_N$, are also drawn independently from $U[0,1]$. For each simulated sample, we use the first 500 observations for parameter training and the remaining 50 observations for out-of-sample evaluation. We then report the median MSFE of the 1,000 repetitions.

Table 1 has four panels, corresponding to the two factor models with and without a nonlinear link function. According to the DGPs, the true number of predictive indices, k, is known and equal to one in the first three panels and two in the last panel in implementing SUFF and sSUFF. When k is unknown in practical applications, one can estimate it with the χ^2 test in Li (1991). Table 1 considers both known and unknown k. In Panels A and C when the link function is linear, we extract $r = 2$ and 5 factors, respectively. In this case, with correct identification, SUFF performs similarly as PCA, in that the single predictive index of SUFF is a linear combination of the PCA factors and they both are linearly related to the forecast target. With the

Table 1: Simulated forecasting performance with weak and strong factors

This table reports the simulated mean squared forecast errors (MSFEs) of PCA, sPCA, SUFF, and sSUFF with either weak or strong factors. Panels A and B assume two latent factors f_{1t} and f_{2t} driving the predictors, and the forecast targets are $y_{t+1} = f_{1t} + \epsilon_{t+1}$ and $y_{t+1} = f_{1t} + f_{1t}^2 + \epsilon_{t+1}$, respectively. Panels C and D assume five and seven latent factors driving the predictors, and the forecast targets are $y_{t+1} = 0.8f_{1t} + 0.5f_{2t} + 0.3f_{3t} + \epsilon_{t+1}$ and $y_{t+1} = f_{1t}(f_{2t} + f_{3t} + 1) + \epsilon_{t+1}$, respectively. For each simulated sample, we assume $(T, N) = (550, 500)$, and use the first 500 observations for parameter training and the rest 50 for out-of-sample evaluation. All factors are i.i.d normally distributed. In the weak factor setup, $n = 10$, 20, 30, 40, or 50 out of 500 predictors are loading on the latent factors. In the strong factor setup, all $n = 500$ predictors have non-zero loadings, which are drawn independently from an uniform distribution with support $[0, 1]$. The idiosyncratic errors are cross-sectionally heterogenous with standard deviations following an uniform distribution with support $[0, 1]$. Reported are the median MSFEs of each forecasting method on the basis of 1,000 simulations. k is the number of predictive indices used in SUFF and sSUFF.

			Known k		Unknown k				Known k		Unknown k	
n	PCA	sPCA	SUFF	sSUFF	SUFF	sSUFF	PCA	sPCA	SUFF	sSUFF	SUFF	sSUFF
	A: Linear model with 2 factors						B: Nonlinear model with 2 factors					
10	1.55	1.21	1.55	1.15	1.55	1.15	3.15	2.40	3.17	1.74	3.17	1.74
20	1.38	1.10	1.38	1.07	1.38	1.07	3.02	2.23	2.58	1.37	2.58	1.36
30	1.19	1.07	1.19	1.07	1.19	1.07	3.04	2.23	1.86	1.34	1.86	1.33
40	1.10	1.05	1.10	1.05	1.10	1.05	3.08	2.23	1.50	1.26	1.50	1.25
50	1.07	1.04	1.07	1.04	1.07	1.04	3.00	2.16	1.33	1.21	1.33	1.20
500	1.01	1.01	1.01	1.01	1.01	1.01	3.08	2.18	1.06	1.06	1.06	1.06
	C: Linear model with 5 factors						D: Nonlinear model with 7 factors					
10	1.51	1.25	1.51	1.22	1.51	1.22	3.90	3.95	3.88	2.94	3.88	3.42
20	1.49	1.14	1.49	1.10	1.49	1.10	3.92	3.74	3.90	2.42	3.91	3.24
30	1.43	1.09	1.43	1.07	1.43	1.07	3.88	3.58	3.88	2.15	3.91	3.04
40	1.34	1.07	1.35	1.06	1.35	1.05	3.72	3.40	3.71	2.00	3.74	2.82
50	1.27	1.05	1.27	1.05	1.27	1.05	3.64	3.32	3.59	1.97	3.65	2.59
500	1.01	1.01	1.01	1.01	1.01	1.01	2.99	3.01	1.75	1.75	1.78	1.78

same reason, the sSUFF forecast has similar performance as the sPCA forecast; they differ in some cases because sSUFF uses the regression R^2 as scaling (see Step 4 of Algorithm 2) while sPCA uses the slope of regressing the predictor on the forecast target. When factors are weak ($n \leq 50$), all the forecasts are inconsistent with MSFEs being unanimously larger than 1, especially in the extremely weak case ($n = 10$). Comparing the four methods, because sPCA and sSUFF are proposed to deal with factor weakness, they outperform their counterparts, PCA and SUFF, with smaller MSFEs. When factors are strong ($n = 500$), all the forecasts are consistent and their

MSFEs are equal to 1.01, almost the same as that with the infeasible best forecast.

Panels B and D of Table 1 are more interesting. In addition to weak factors, the link function is nonlinear in these two panels. Because SUFF and sSUFF are able to accommodate nonlinearity, their forecasts outperform the PCA and sPCA forecasts with smaller MSFEs. This result remains true no matter the factors are weak or strong. An interesting observation is that there is a tradeoff between factor weakness and nonlinearity. sPCA can deal with factor weakness but not nonlinearity, whereas SUFF can deal with nonlinearity but not factor weakness. As a result, in Panel B, sPCA outperforms SUFF when the factors are extremely weak ($n = 10$ and 20), but it underperforms SUFF in other cases. In all the cases, sSUFF performs the best as expected, because it is proposed to deal with both factor weakness and nonlinearity.

3. Bond Return Predictability with Real-Time Macro Data

In this section, we apply sSUFF to examine bond return predictability with real-time macro variables. We compare its performance with SUFF, as well as the linear forecasts by PCA, sPCA, and PLS.

3.1. *Data and key variables*

We obtain monthly US government bond prices from the Fama-Bliss dataset, which is available at the Center for Research in Security Prices (CRSP). In line with the bond return predictability literature (e.g., Fama and Bliss, 1987; Cochrane and Piazzesi, 2005), we focus on US government bonds with a remaining time to maturity of one through five years.

We closely follow Cochrane and Piazzesi (2005) and Ludvigson and Ng (2009) on the construction of bond returns. $r^{(n)}_{t+12} = p^{(n-1)}_{t+12} - p^{(n)}_t$ is the log annual holding period return from buying an n-year bond at time t and selling it as an $(n-1)$-year bond one year later, where $p_t^{(n)}$ denotes the log price of an n-year zero-coupon bond at time t. The log bond yield is defined as $y_t^{(n)} = -(1/n)p_t^{(n)}$, and $y_t^{(1)}$ is the log yield of a 1-year bond and is set to be the risk-free rate known at time t.[7] In this paper, we focus on bond excess returns, $rx_{t+12}^{(n)} = r_{t+12}^{(n)} - y_t^{(1)}$, which refer to the continuously compounded log excess returns of an n-year bond. The average bond excess return across maturities is defined as $\overline{rx}_{t+12} = \frac{1}{4}\sum_{n=2}^{5} rx_{t+12}^{(n)}$. For brevity, bond returns in the sequel always refer to bond excess returns, unless otherwise stated.

We obtain monthly real-time macro variables from the Archival Federal Reserve Economic Data (ALFRED) database, which is maintained by the Federal Reserve Bank of St. Louis. For a large panel of macro variables, the ALFRED database keeps track of their historical monthly vintages with the exact time stamp of information release pertaining to the data date, because macro variables are often released with delay and subject to subsequent revisions. Throughout the paper, we denote $X_{i,t}$ as the value of variable i in fiscal month $t-1$ but is collected and released in month t, which may be further revised later. In reality, however, some macro variables are even released with a two-month lag. For ease of exposition, we uniformly assume a one-month release lag and refer to the last reading of variable i released in month t as $X_{i,t}$.

Following Ghysels et al. (2018), we collect 60 macro variables with a balance between the number of variables included and the length of the time-series observations. These 60 variables cover broad economic categories such as output and income, labor force and unemployment, consumption expenditure and housing indicators, money stock and credit, and price indices, largely mirroring the set of macro variables considered in Ludvigson and Ng (2009). Since the vintage data are sparse before March 1982, we choose the initial vintage release date as March 1982. Also, all the 60 selected variables have a long series of monthly observations starting no later than January 1968 for each vintage. Although most of the macro variables' vintages are updated at a monthly frequency, there are a few cases where one series is updated twice in a month and no vintage is released in the following month. In these cases, we keep track of the information based on its release date and follow a general rule

[7] The log forward rate between time $t+12(n-1)$ and $t+12n$ is defined as $fwd_t^{(n)} = p_t^{(n-1)} - p_t^{(n)}$.

Table 2: Snapshot of vintages of total nonfarm payroll from ALFRED
This table presents a snapshot of the vintage data on total nonfarm payroll from ALFRED, where Column "PAYEMS_20170203" refers to the vintage released on February 3, 2017, pertaining to the monthly observations with data date January 2017, with a one-month publication lag, and Column "PAYEMS_20170310" to the vintage released on March 10, 2017, pertaining to the monthly observations with data date February 2017, etc.

Data date	PAYEMS_20170203	PAYEMS_20170310	PAYEMS_20170407	PAYEMS_20170505
2016-07	144,457	144,457	144,457	144,457
2016-08	144,633	144,633	144,633	144,633
2016-09	144,882	144,882	144,882	144,882
2016-10	145,006	145,006	145,006	145,006
2016-11	145,170	145,170	145,170	145,170
2016-12	145,327	145,325	145,325	145,325
2017-01	145,554	145,563	145,541	145,541
2017-02		145,798	145,760	145,773
2017-03			145,858	145,852
2017-04				146,063

of picking the most recently observed vintage by the end of each month. Besides, for each vintage, we also identify the outliers following Ludvigson and Ng (2009) and replace them with the previous month values to avoid look-ahead bias.

The online appendix provides a detailed description of the 60 macro variables, along with the transformation codes applied to each variable to ensure its stationarity.[8] In order to visualize the structure of real-time macro data in ALFRED, Table 2 presents a snapshot of the vintage data on *All Employee: Total Nonfarm Payroll Employment*, a widely followed macro variable by market participants. The last reading in the second column, 145,554, refers to the data date of January 2017, and is released in the February 3, 2017 vintage due to one month publication lag. This value was subsequently revised to 145,563 and to 145,541 in the March 10, 2017 and April 7, 2017 vintages, respectively. Apparently, the first release on February 3, 2017 and its subsequent two revisions are not available to investors at the end of January 2017, and cannot be used to predict bond returns in February 2017.

[8] We use the same codes for variable transformation as the FRED-MD database, which are discussed in detail in McCracken and Ng (2016).

3.2. *In-sample forecasting performance*

We start by examining the in-sample performance of predicting annual holding period excess returns of n-year bonds, $rx_{t+12}^{(n)}$. While our data dates start from January 1968, the real-time vintage observations start from March 1982 and data before March 1982 may contain publication lags and revisions. As such, we estimate real-time macro PCA, sPCA, and PLS factors, as well as the SUFF and sSUFF predictive indices, with vintage data from March 1982 to December 2019, and compare their forecasting performances over the March 1982 to December 2019 period. For comparison, we also report the forecasting results with the final revised macro data.

Following Ludvigson and Ng (2009), with each forecasting method, we construct a single predictor factor to predict bond returns at all maturities. Specifically, by using the average bond excess return across maturity, $\overline{rx}_{t+12} = \frac{1}{4}\sum_{n=2}^{5} rx_{t+12}^{(n)}$, as the forecast target, we extract $r = 6$ latent factors with PCA, sPCA and PLS, and $k = 2$ predictive indices with SUFF and sSUFF. The number of latent factors is determined by the eigenvalue difference test of Onatski (2010), and the number of indices is determined by the χ^2 test of Li (1991). We then regress the target $\overline{rx}_{t+12}$ on the six latent factors or the two predictive indices, and define the fitted value as the single predictor factor, which is in turn used to predict bond returns with different maturities. While SUFF and sSUFF use local linear regressions to estimate the nonlinear link function $h(\cdot)$ for forecasting, in this subsection we also consider directly using linear forecasts with the two extracted predictive indices.

Table 3 reports the regression slopes, Newey-West t-statistics, and R^2s in predicting $rx_{t+12}^{(n)}$ with $n = 2, ..., 5$. For ease of interpretation, we normalize each single predictor factor to have mean zero and standard deviation one, and negate it in the predictive regression to make the slope align with the economic theory-implied relationship between macro variables and bond risk premia. The regression intercepts hence measure the average annual holding period bond returns, from 0.54% to 1.52% for 2- to 5-year bonds, and are untabulated for brevity.

Table 3: In-sample performance of forecasting bond returns with macro factors This table reports the slopes, Newey-West t-statistics, and R^2s from the predictive regressions of 2- to 5-year bond returns on a single predictor factor, which is extracted from 60 final revised or real-time macro variables. The single predictor factor is estimated as the fitted values of regressing average bond excess return $\overline{rx}_{t+12}$ on the first PCA factor in row PCA_1, the first six PCA, sPCA, and PLS factors in the next three rows, and two SUFF and sSUFF predictive indices with a linear link function in rows SUFF_l and sSUFF_l and a nonlinear link function $h(\cdot)$ estimated by local linear regression in the last two rows, respectively. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is 1982:03–2019:12.

	$rx^{(2)}_{t+12}$			$rx^{(3)}_{t+12}$			$rx^{(4)}_{t+12}$			$rx^{(5)}_{t+12}$		
	β	t-stat	R^2	β	t-stat	R^2	β	t-stat	R^2	β	t-stat	R^2
Panel A: Revised data												
PCA_1	0.32*	1.71	5.17	0.65**	2.04	5.98	0.78*	1.80	4.26	0.89*	1.83	3.54
PCA	0.42***	2.91	8.90	0.77***	3.25	8.34	0.89***	2.80	5.54	0.95***	2.67	4.04
sPCA	0.71***	4.70	25.17	1.23***	4.58	21.22	1.56***	3.91	17.15	1.83***	3.78	15.02
PLS	0.65***	4.61	21.06	1.15***	4.63	18.50	1.49***	4.18	15.50	1.77***	4.12	14.16
SUFF_l	0.42***	2.99	8.65	0.76***	3.34	8.10	0.88***	2.90	5.43	0.94***	2.77	3.97
sSUFF_l	0.62***	3.92	19.08	1.05***	3.85	15.69	1.32***	3.33	12.18	1.51***	3.22	10.26
SUFF	0.61***	6.49	18.56	1.08***	6.94	16.42	1.40***	6.88	13.83	1.56***	6.01	11.02
sSUFF	0.86***	8.48	37.33	1.53***	7.74	32.92	2.08***	6.87	30.34	2.46***	6.64	27.36
Panel B: Real-time data												
PCA_1	0.18	1.08	1.70	0.41	1.38	2.35	0.47	1.15	1.54	0.55	1.16	1.38
PCA	0.33***	3.06	5.63	0.61***	3.24	5.33	0.75***	3.00	3.95	0.85***	2.88	3.25
sPCA	0.63***	4.82	19.98	1.08***	4.69	16.42	1.39***	4.23	13.53	1.62***	4.08	11.84
PLS	0.57***	4.75	16.51	1.01***	5.02	14.38	1.32***	4.73	12.24	1.58***	4.70	11.22
SUFF_l	0.34***	3.04	5.80	0.62***	3.18	5.35	0.73***	2.84	3.77	0.82***	2.68	2.99
sSUFF_l	0.57***	3.98	16.15	0.95***	3.73	12.76	1.19***	3.20	9.96	1.38***	3.08	8.63
SUFF	0.50***	5.54	12.71	0.92***	5.73	11.98	1.21***	5.40	10.30	1.40***	4.99	8.87
sSUFF	0.80***	6.86	31.81	1.39***	6.53	27.37	1.89***	6.08	25.19	2.26***	5.84	22.99

Panel A uses final revised macro data.[9] Consistent with the literature like Ludvigson and Ng (2009), the first PCA factor can predict future bond returns and its power is marginally significant at the 10% level. Moreover, including six PCA factors makes the the predictability significant at the 1% level, with R^2s ranging from 4.04% to 8.90%. Economically, a one-standard deviation increase in the single predictor factor leads to a decrease in expected bond return across maturities by 0.42% to 0.95%, more than half of the unconditional average bond returns. This pattern holds for sPCA and PLS. By assuming a linear link function, SUFF and sSUFF perform

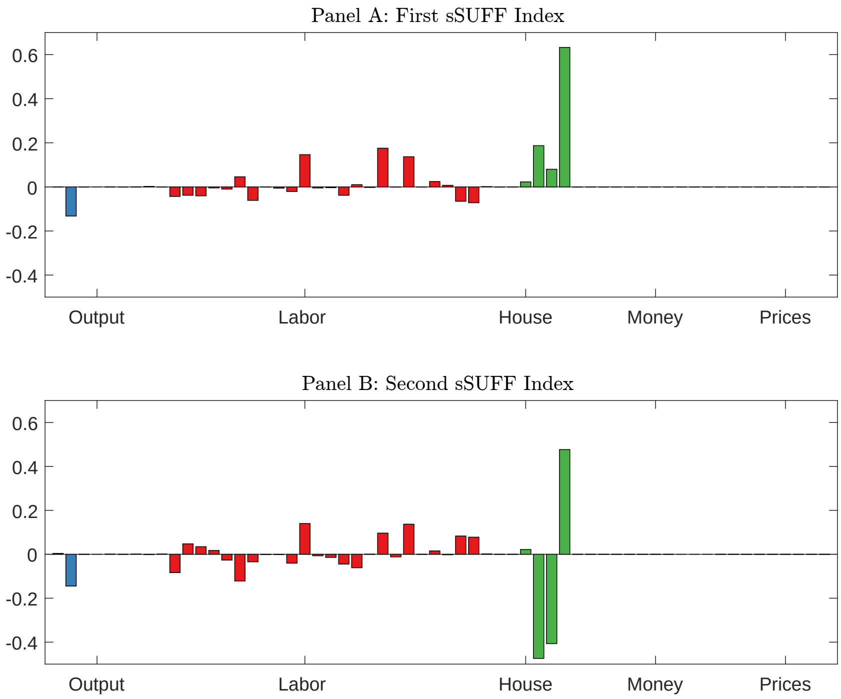
[9] As our focus is on real-time bond return predictability, we use final revised data only in this panel and real-time data in other analyses by default.

similarly as PCA and sPCA, because the two predictive indices of SUFF and sSUFF are linear combinations of the six PCA and sPCA factors, respectively. When allowing for the presence of nonlinearity and factor weakness, sSUFF outperforms SUFF and all other methods, with R^2s ranging from 27.36% to 37.33%. This result suggests that the relationship between bond returns and macro variables is more complex than what existing methods assume.

When using the noisier real-time macro data, Panel B confirms Ghysels et al. (2018) that the forecasting power of the first PCA factor becomes insignificant. However, when more PCA factors are included, the predictability can be improved substantially, although the magnitude is still smaller than that with final revised data. For example, the R^2 in forecasting the 2-year bond returns is 8.90% with final revised data-based PCA factors and 5.63% with real-time data-based PCA factors. An interesting feature is that while the R^2 declines substantially when using real-time macro data, the predictive slope does not decease very much. This is reasonable because the real-time predictor factor is noisier but unbiased. Compared with PCA, the predictive power of the sPCA and PLS return factors do not decrease very much. The reason is that both sPCA and PLS use the forecast target information in extracting the factors. Finally, sSUFF performs the best regardless of whether the final revised data or real-time data are used. Assuming the presence of weak factors and nonlinearity significantly improves the forecasting performance.

The key difference of our method from the existing ones lies in two ingredients, factor weakness and nonlinearity. We now explore whether these two ingredients are supported by real-time macro data. Regarding factor weakness, Figure 1 plots the loadings of two predictive indices on each macro variable. It shows that the two indices load on industrial production, labor and housing variables. While both indices load negatively on industrial production, their loadings on the labor and housing variables have opposite signs. Moreover, there are a large number of macro variables having zero loadings. This figure has two implications. First, the two predictive indices are not as diffusive as the strong factors in the sense of Stock and Watson (1998) and, therefore, are better captured by our sparsity-induced weak

Figure 1: Loadings of sSUFF indices on individual macro variables
This figure plots the loadings of two sSUFF predictive indices on each individual macro variables. Macro variables include output and income (No. 1-7), labor market (No. 8-33), consumption and housing (No. 34-40), money and credit (No. 41-52), and prices (No. 52-60). The sample period is 1982:03–2019:12.

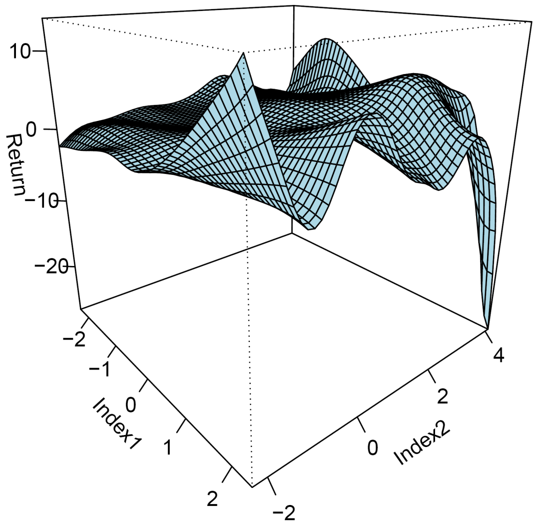


factors. Second, the typically used estimation methods of assuming strong factors would underestimate the forecasting power of macro variables.

Regarding nonlinearity, Figure 2 plots the surface of the fitted nonlinear link function. Clearly, the relationship between bond expected returns and macro factors is highly nonlinear. This result is not new but implies the necessity for using SUFF or sSUFF. In the literature, Breach et al. (2020) and Giacoletti et al. (2021) find that dispersion in inflation, output growth, and future yields forecasts contain incremental information for future bond returns, and suggest incorporating these second moments into the term structure models. Bianchi et al. (2021) show that nonlinear models, such as random forest and neural networks, outperform other linear models in forecasting bond returns with final revised macro variables.

Figure 2: Nonlinear relationship between bond returns and macro factors

This figure plots the surface of the fitted nonlinear link function $\hat{h}(\cdot)$ from the model $\overline{rx}_{t+12} = h(\phi_1' \boldsymbol{f}_t, \phi_2' \boldsymbol{f}_t) + \varepsilon_{t+12}$, where the two predictive indices are $\phi_1' \boldsymbol{f}_t$ and $\phi_2' \boldsymbol{f}_t$, and $\boldsymbol{f}_t$ are the first six factors extracted by using Steps 1–5 of Algorithm 2 from 60 real-time macro variables. The sample period is 1982:03–2019:12.



3.3. Incremental power of macro variables relative to yield curve predictors

In the literature, bond pricing models are usually based on a small set of factors that are linear combinations of bond yields (see, e.g., Ang and Piazzesi, 2003; Bernanke and Kuttner, 2005; Duffee, 2013; Joslin, Le, and Singleton, 2013), which leads to the famous spanning hypothesis as to whether the yield curve contains all available information about future yields. However, recent studies, such as Joslin et al. (2014), argue that previous macro-finance term structure models impose "counterfactual restrictions on the joint distribution of bond yields and the macroeconomy." To examine whether the forecasting power of macro variables in Table 3 is incremental relative to the yield curve, we consider controlling for two yield curve predictors. One is Cochrane and Piazzesi's (2005) single hump shaped combination of the 1- to 5-year forward rates, and the other is the first three principal components of the 1- to 5-year yields (Fama and French, 1988).

Let f be the single predictor factor in Table 3 and CP be the Cochrane and Piazzesi

Table 4: Incremental power of macro factors relative to yield curve predictors
This table reports the slopes, Newey-West t-statistics, and incremental R^2 (ΔR^2s) of macro factors relative to yield curve predictors in predicting bond returns, where macro factors are single predictor factors extracted from 60 real-time macro variables, and yield curve predictors are the Cochrane and Piazzesi (2005) return predictor (CP) or the first three principal components of the 1- to 5-year yields. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is 1982:03–2019:12.

	$rx_{t+12}^{(2)}$			$rx_{t+12}^{(3)}$			$rx_{t+12}^{(4)}$			$rx_{t+12}^{(5)}$		
	β	t-stat	ΔR^2	β	t-stat	ΔR^2	β	t-stat	ΔR^2	β	t-stat	ΔR^2
Panel A: CP factor as the yield curve predictor												
PCA	0.26***	2.55	3.25	0.47***	2.72	3.07	0.53***	2.35	1.90	0.60**	2.17	1.56
sPCA	0.55***	4.92	14.34	0.92***	4.66	11.41	1.13***	4.11	8.60	1.33***	3.95	7.64
PLS	0.50***	4.92	12.08	0.87***	5.11	10.35	1.10***	4.62	8.22	1.33***	4.46	7.71
SUFF	0.39***	3.92	6.96	0.71***	4.25	6.54	0.87***	4.02	4.88	1.02***	3.75	4.28
sSUFF	0.71***	7.02	21.87	1.22***	6.93	18.25	1.59***	6.69	15.52	1.93***	6.23	14.64
Panel B: First three principal components of yields as the yield curve predictors												
PCA	0.28***	2.78	3.82	0.51***	2.96	3.64	0.58***	2.66	2.35	0.65***	2.45	1.90
sPCA	0.55***	5.00	14.65	0.93***	4.77	11.76	1.14***	4.22	8.75	1.33***	4.05	7.69
PLS	0.50***	4.90	12.30	0.88***	5.14	10.61	1.11***	4.73	8.32	1.33***	4.59	7.74
SUFF	0.40***	4.05	7.64	0.74***	4.41	7.25	0.91***	4.20	5.39	1.05***	3.9	4.64
sSUFF	0.71***	7.06	21.86	1.23***	7.01	18.41	1.58***	6.72	15.19	1.90***	6.27	14.16

(2005) factor. We run the following regression,

$$rx_{t+12}^{(n)} = \alpha + \beta f_t + \gamma \mathrm{CP}_t + \epsilon_{t+12}^{(n)}, \tag{10}$$

and report the estimate of β, t-statistic, and the incremental R^2, ΔR^2, in Panel A of Table 4, where ΔR^2 is the difference in R^2 between including and excluding f_t in (10). The results show that the PCA predictor factor remains significant after controlling for the CP factor, with incremental R^2s ranging from 1.56% to 3.25% across maturities. The incremental R^2s by using the sPCA, PLS, SUFF, and sSUFF predictor factors are much larger. For example, the ΔR^2s with PCA, SUFF, and sSUFF in forecasting the 5-year bond returns are 1,56%, 4.28%, and 14.64%, respectively.

Panel B of Table 4 replaces CP in (10) with the first three principal components of the 1- to 5-year yields, and presents quantitatively similar results as Panel A, in terms of the incremental R^2 and regression slope on f. Thus, real-time macro factors

Table 5: Forecasting bond returns with macro factors: Robustness tests
Panels A and B report Hodrick (1992) t-statistics and bootstrapped p-values of forecasting 12-month holding period bond returns with real-time macro factors. Panel C reports the regression slopes of forecasting 1-month holding period bond returns, where monthly bond returns are backed out from the yield curve dataset of Gürkaynak et al. (2007).

	$rx^{(2)}_{t+12}$	$rx^{(3)}_{t+12}$	$rx^{(4)}_{t+12}$	$rx^{(5)}_{t+12}$
Panel A: Hodrick (1992) t-statistics				
PCA	2.88	3.10	2.92	2.88
sPCA	4.47	4.37	3.94	3.85
PLS	4.40	4.72	4.51	4.56
SUFF	5.52	5.88	5.67	5.33
sSUFF	6.17	5.87	5.43	5.25
Panel B: Bootstrapped p-values				
PCA	0.01	0.00	0.01	0.01
sPCA	0.02	0.03	0.04	0.05
PLS	0.01	0.01	0.02	0.02
SUFF	0.08	0.08	0.11	0.15
sSUFF	0.03	0.04	0.06	0.07
Panel C: Slopes of forecasting 1-month ahead bond returns				
PCA	0.09***	0.11***	0.12**	0.13**
sPCA	0.18***	0.26***	0.33***	0.39***
PLS	0.19***	0.28***	0.36***	0.43***
SUFF	0.16***	0.24***	0.32***	0.39***
sSUFF	0.21***	0.27***	0.33***	0.39***

contain incremental information relative to the well-known yield curve predictors in predicting bond returns. This result echoes the concurrent finding in Huang and Shi (2022), who propose a supervised adaptive group LASSO method to forecast bond returns with macro variables. We differ from them by accounting for both factor weakness and nonlinearity.

3.4. *Alleviating the concern of overlapping bond returns*

This paper follows Ludvigson and Ng (2009) and uses annual holding period bond returns as the forecast targets. A concern with such returns is that the overlapping feature could inflate the Newey-West t-statistic in the predictive regression (Ang and Bekaert, 2007). To alleviate this concern, we perform three tests.

The first test is to reexamine the sSUFF forecast with Hodrick (1992) t-statistic. Hodrick (1992) shows that forecasting h-period ahead returns with x_t is equivalent to forecasting 1-period ahead returns with h-period average of x_t. With this insight, he proposes a method to remove the overlapping nature in calculating the t-statistic. Panel A of Table 5 shows that the forecasting power of macro predictor factors remains significant and all the t-statistics are larger than 2, with most of them being larger than 3.

The second test is to perform bootstrapped simulations. Following Bauer and Hamilton (2018), we reconstruct bond returns by bootstrapping the yields with four steps. First, based on the observed 1- to 5-year yields, we extract three principal components and obtain the loadings of the observed yields on the principal components. Second, we model the three principal components as a VAR(1) process and estimate its coefficients. Third, based on the estimated VAR(1) coefficients, we bootstrap the yield principal components by setting the initial value equal to that in March 1982. Finally, with the yield loadings and the simulated principal components, we construct a yield sample by assuming its residuals to be normally distributed, and then calculate the artificial bond returns accordingly. With the same procedure, we bootstrap macro variables by assuming a six-factor structure. For each bootstrapped bond returns and macro variables sample, we reexamine Table 3 and report the p-values in Panel B of Table 5. Overall, the results indicate that bond returns can be significantly predicted by macro variables, except for forecasting 4- and 5-year bond returns with SUFF.

The third and last test is to consider forecasting *monthly* holding period bond returns directly. In addition to the Fama-Bliss dataset, Gürkaynak et al. (2007) construct an alternative dataset and can back out bond returns at a monthly frequency without overlapping. We replace the 12-month holding period returns with 1-month holding period returns and report the forecasting slopes of macro factors in Panel C of Table 5. The results continue to support our conclusion that real-time macro variables can predict bond returns, especially when the forecasting method accommodates factor weakness and nonlinearity. However, as noted by Liu and Wu (2021), this dataset extrapolates the short end of the yield curve and has large pricing errors, thereby raising a question whether it is appropriate in our framework. For this reason, we focus on the more recognized and widely used Fama-Bliss dataset, while acknowledging its potential drawbacks.

3.5. *Alternative specifications of sSUFF*

In Section 3.2, the benchmark specification for sSUFF is constructing $k = 2$ predictive indices based on $r = 6$ latent factors. This section considers alternative number of latent factors and predictive indices. First, we consider extracting seven or eight latent factors from macro variables while keeping $k = 2$ unchanged. The result in Panel A of Table 6 shows that increasing the number of latent factors does not increase the forecasting performance significantly. In comparison with Table 3, the R^2 is 31.81% with $r = 7$ and 31.73% with $r = 8$ in forecasting the 2-year bond returns. Second, we assume $k = 3$ or 4 predictive indices while keeping $r = 6$ unchanged. Interestingly, this increase in the number of indices slightly weakens the forecasting power. For example, the R^2 of forecasting the 2-year bond returns reduces to 26.01%. Hence, sSUFF is relatively robust to alternative number of latent factors and predictive indices, and the benchmark specification is reasonable for the empirical application.

To construct the predictive indices, Steps 1–5 of Algorithm 2 are proposed to extract the latent factors. For robustness, we explore three alternative methods: sPCA in Huang et al. (2022), target PCA (tPCA) in Bai and Ng (2008),[10] and weighted PCA (wPCA) in Bailey (2012). With each method, we extract six factors from the real-time macro variables and then repeat Steps 6–8 of Algorithm 2 to construct two predictive indices for forecasting. The results are reported in Panel B of Table 6 and show that the forecasts with these alternative methods underperform the benchmark forecasts based on Algorithm 2. For example, the R^2 of forecasting the 2-year bond returns decreases from 31.81% by using Algorithm 2 to 15.17% by using the soft threshold

[10] In the implementation, we use the hard threshold to choose relevant predictors. The results with soft threshold are quantitatively similar. The supervised PCA in Bair, Hastie, Paul, and Tibshirani (2006) is in the same spirit of tPCA and differs by choosing predictors with cross-validation.

Table 6: Alternative specifications of sSUFF

This table reports the performances of predicting 2- to 5-year bond returns with macro variables by alternative specifications of sSUFF. In contrast to the benchmark $(r,k) = (6,2)$ in Table 3, Panel A considers alternative number of factors ($r = 7$ and 8) and alternative number of predictive indices ($k = 3$ and 4), respectively. Panel B considers replacing the latent factors in Step 5 of Algorithm 2 with wPCA, sPCA or tPCA factors to construct the sSUFF predictive indices, where the factors are extracted from real-time macro variables by using the sPCA of Huang et al. (2022), target PCA (tPCA) of Bai and Ng (2008), and weighted PCA (wPCA) of Bailey (2012). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is 1982:03–2019:12.

	$rx^{(2)}_{t+12}$			$rx^{(3)}_{t+12}$			$rx^{(4)}_{t+12}$			$rx^{(5)}_{t+12}$		
	β	t-stat	R^2	β	t-stat	R^2	β	t-stat	R^2	β	t-stat	R^2
Panel A: Alternative number of latent factors or predictive indices												
$r=7$	0.77***	6.87	29.59	1.39***	6.94	27.14	1.92***	6.79	25.76	2.33***	6.56	24.37
$r=8$	0.80***	7.68	31.73	1.45***	7.65	29.73	2.02***	7.19	28.58	2.45***	6.95	26.98
$k=3$	0.71***	9.42	25.07	1.29***	10.04	23.37	1.75***	9.74	21.52	2.07***	8.65	19.36
$k=4$	0.72***	8.56	26.01	1.34***	9.20	25.41	1.86***	9.11	24.24	2.21***	8.26	22.01
Panel B: Extracting latent factors $\boldsymbol{f}_t$ with alternative methods												
sPCA	0.67***	7.70	22.30	1.18***	7.13	19.67	1.58***	6.52	17.57	1.87***	5.96	15.69
tPCA	0.55***	6.59	14.95	1.08***	8.23	16.52	1.48***	8.66	15.41	1.76***	8.35	13.94
wPCA	0.53***	5.88	14.07	0.95***	6.10	12.80	1.24***	5.78	10.85	1.47***	5.52	9.72

tPCA factors.

3.6. *Out-of-sample forecasting performance*

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

$$R^2_{OS} = 1 - \frac{\sum_{t=m}^{T-12} (rx^{(n)}_{t+12} - \widehat{rx}^{(n)}_{t+12})^2}{\sum_{t=m}^{T-12} (rx^{(n)}_{t+12} - \widetilde{rx}^{(n)}_{t+12})^2}, \tag{11}$$

where $\widehat{rx}^{(n)}_{t+12}$ is the forecast based on PCA, sPCA, PLS, SUFF or sSUFF, and $\widetilde{rx}^{(n)}_{t+12}$ is a benchmark forecast, say the historical mean. Both $\widehat{rx}^{(n)}_{t+12}$ and $\widetilde{rx}^{(n)}_{t+12}$ are recursively estimated with all information up to time t. The out-of-sample forecast starts from March 1982 in that it is the initial vintage of our real-time data set. As such, the out-of-sample period is the same as the in-sample period in Table 3, from March 1982

Table 7: Out-of-sample performance of forecasting bond returns with macro factors
This table reports the out-of-sample R^2_{OS}s of predicting 2- to 5-year bond returns with real time macro variables. We consider six factors when employing PCA, sPCA and PLS and two predictive indices when employing SUFF and sSUFF. The factors, predictive indices, and return forecasts are recursively estimated with data up to time t. In calculating R^2_{OS}, we use three benchmark forecasts: historical mean forecast (Panel A), yield curve (measured by the first three principal components of 1- to 5-year bond yields) forecast (Panel B), and yield curve and PCA factor-based forecast (Panel C). 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. The out-of-sample period is 1982:03–2019:12.

	$rx^{(2)}_{t+12}$		$rx^{(3)}_{t+12}$		$rx^{(4)}_{t+12}$		$rx^{(5)}_{t+12}$	
	R^2_{OS} (%)	p-value	R^2_{OS} (%)	p-value	R^2_{OS} (%)	p-value	R^2_{OS} (%)	p-value
Panel A: Historical mean forecast as benchmark								
PCA	2.45	0.12	4.39**	0.03	4.96***	0.01	6.02***	0.01
sPCA	2.30**	0.04	4.88**	0.02	6.74***	0.01	8.56***	0.01
PLS	2.18**	0.03	4.51**	0.02	5.87***	0.01	7.51***	0.01
SUFF	1.88	0.18	3.71*	0.06	4.11**	0.03	5.25***	0.01
sSUFF	7.00**	0.03	9.35***	0.01	9.94***	0.01	11.21***	0.00
Panel B: Yield curve forecast as benchmark								
PCA	−0.91	0.16	1.88	0.12	1.55	0.12	2.95	0.11
sPCA	−3.84	0.09	0.36*	0.08	1.58*	0.07	3.96*	0.07
PLS	−3.82	0.08	0.12*	0.07	0.73*	0.07	2.92*	0.07
SUFF	−1.01	0.16	1.61	0.12	1.11	0.13	2.52	0.12
sSUFF	3.39*	0.06	6.78**	0.04	6.73**	0.04	8.28**	0.04
Panel C: Yield curve and macro PCA factors forecast as benchmark								
sPCA	−2.91	0.19	−1.55	0.18	0.03	0.15	1.04	0.13
PLS	−2.89	0.14	−1.80	0.13	−0.84	0.13	−0.03	0.12
SUFF	−0.10	0.40	−0.28	0.59	−0.45	0.84	−0.45	0.86
sSUFF	4.27**	0.04	5.00**	0.02	5.26**	0.01	5.49**	0.01

to December 2019. Since the data date for each vintage of time-series observations starts from January 1968, there are about 15 years of historical data used in the first forecast when using the March 1982 vintage. In line with the in-sample forecasting, we employ six PCA, sPCA and PLS factors and two SUFF and sSUFF predictive indices.

In calculating R^2_{OS} in (11), we consider three benchmark forecasts: 1) historical mean forecast, 2) yield curve (measured by the first three principal components of 1- to 5-year bond yields) forecast,[11] and 3) yield curve and PCA factor-based forecast. Table 7 reports the results and makes two observations. First, sSUFF performs the best across the benchmarks and its R^2_{OS} is the largest one in all the specifications. Second, the forecasting performance depends on the benchmark one uses. When assuming there is no predictability by benchmarking on the historical mean, all the forecasting methods we consider deliver positive R^2_{OS}s. When assuming there is some predictability by benchmarking on yield curve and PCA macro factors, the R^2_{OS}s decrease dramatically. In this case, sSUFF is the only one consistently beating the benchmark.

We note that our PCA results are better than Ghysels et al. (2018). There are two possible reasons. First, we focus on different sample periods. Second, we use the vintage data differently. That is, we take the latest vintage time series as information set one can use, while Ghysels et al. (2018) do not consider further revisions; their focus is to compare whether the first release or the revision contains more information about future bond returns. Take the total nonfarm payroll in Table 2 as an example. To predict bond returns in June 2017, we use the vintage released on May 5, 2017 (last column of Table 2), while Ghysels et al. (2018) use the diagonal values in Table 2. As a result, the data used by Ghysels et al. (2018) are noisier than ours, because the vintages before May 2017 were further revised on May 5, 2017 to reflect the true total nonfarm payroll.

3.7. *Economic value of macro variables in forecasting bond returns*

We investigate the economic value of real-time macro factors in making asset allocation decisions. Following Eriksen (2017) and Gargano et al. (2019), we compute the certainty equivalent return (CER) gain for an investor who has access to real-time macro factors when allocating her wealth between an n-year bond and a 1-year bond over an annual horizon. For simplicity, we assume that the investor has a

[11] The result by using the Cochrane and Piazzesi (2005) return factor is quantitatively similar and untabulated for brevity.

Table 8: Economic value of forecasting bond returns with real-time macro factors
This table reports the portfolios gains of forecasting bond returns with real-time macro factors with transaction costs of 20 basis points. We consider two portfolio gain measures: certainty equivalent return (CER) gain and Goetzmann, Ingersoll, Spiegel, and Welch (2007) manipulation-proof performance (MPP) improvement, which are calculated as the incremental values of using the PCA, sPCA, PLS, SUFF or sSUFF forecasts relative to the benchmark forecasts. Three benchmark forecasts are historical mean forecast (Panel A), yield curve (measured by the first three principal components of 1- to 5-year bond yields) forecast (Panel B), and yield curve and PCA factor-based forecast (Panel C). We assume a mean-variance investor, with risk aversion $\gamma = 5$ and investment horizon of one year, decides to allocate his wealth between a 1-year (risk free) bond and an n-year bond ($n = 2, \cdots, 5$), and estimates the expected n-year bond return by using the real-time macro factors. All the macro factors and expected bond returns are recursively estimated with expanding windows. The investment period is 1982:03–2019:12.

	CER gain (%)				MPP improvement			
	$n=2$	$n=3$	$n=4$	$n=5$	$n=2$	$n=3$	$n=4$	$n=5$
Panel A: Historical mean forecast as benchmark								
PCA	−0.15	0.34	0.48	0.64	−0.12	0.43	0.57	0.73
sPCA	−0.15	0.43	0.80	0.97	0.04	0.89	1.29	1.50
PLS	−0.32	0.13	0.40	0.67	−0.14	0.65	1.02	1.26
SUFF	−0.15	0.33	0.42	0.58	−0.11	0.41	0.49	0.66
sSUFF	0.19	0.91	1.17	1.31	0.37	1.31	1.53	1.69
Panel B: Yield curve forecast as benchmark								
PCA	0.17	0.81	1.45	2.12	0.00	0.24	0.10	0.00
sPCA	0.42	1.11	1.89	2.50	0.30	0.46	0.49	0.47
PLS	0.21	0.80	1.53	2.17	0.07	0.27	0.24	0.22
SUFF	0.13	0.78	1.43	2.10	−0.03	0.18	0.04	−0.05
sSUFF	0.61	1.31	1.96	2.63	0.49	0.73	0.71	0.70
Panel C: Yield curve and macro PCA factors forecast as benchmark								
sPCA	0.25	0.30	0.44	0.38	0.30	0.22	0.40	0.47
PLS	0.04	−0.02	0.07	0.05	0.07	0.03	0.14	0.22
SUFF	−0.03	−0.03	−0.02	−0.02	−0.03	−0.06	−0.06	−0.05
sSUFF	0.44	0.50	0.50	0.52	0.49	0.49	0.61	0.69

mean-variance preference with risk aversion $\gamma = 5$, and repeatedly makes allocation decisions at the end of each month over the March 1982 to December 2019 period.

Specifically, at time t, the investor optimally allocates a proportion of $w_t^* = \mathrm{E}_t[R_{t+12}^{(n)}]/(\gamma \mathrm{Var}_t[R_{t+12}^{(n)}])$ of her wealth to the n-year bond, where $\mathrm{E}_t[R_{t+12}^{(n)}]$ is the return forecast on n-year bond, and $\mathrm{Var}_t[R_{t+12}^{(n)}]$ is the conditional variance. To highlight the effect of return forecasts, we fix the variance of bond return at its unconditional mean over the full sample, and entertain varying real-time macro factors in return forecasts to judge their economic values. The investor then allocates $1 - w_t^*$ of her wealth to the 1-year bond, assuming a risk free asset, and the portfolio return realized at $t+12$ is $R_{t+12}^p = w_t^* R_{t+12}^{(n)} + R_t^{(1)}$. To prevent the investor from taking extreme positions, we restrict the portfolio weight w_t^* to lie in $[-1,5]$, which amounts to a maximum short-sale of 100% and a maximum leverage of 400%.

The CER of the portfolio is $\text{CER} = \hat{\mu}_p - 0.5\gamma\hat{\sigma}_p^2$, where $\hat{\mu}_p$ and $\hat{\sigma}_p^2$ are the sample mean and variance of the portfolio returns. We calculate the CER gain as the difference between the CER for the investor who uses the real-time macro factor based forecast and the CER for an investor who uses a benchmark forecast such as the historical mean. Intuitively, this CER gain can be interpreted as a portfolio management fee that an investor would be willing to pay in order to have access to this real-time macro factor based forecast.

In addition to the CER gain, we also consider an alternative measure, the manipulation-proof performance (MPP) measure (Goetzmann et al., 2007). Following Thornton and Valente (2012) and Eriksen (2017), we define MPP as

$$\text{MPP} = \frac{1}{1-\gamma}\ln\left(\frac{1}{m}\sum_{t=1}^{m}\left[\frac{\hat{R}_{t+12}^p}{R_t^{(1)}}\right]^{1-\gamma}\right) - \frac{1}{1-\gamma}\ln\left(\frac{1}{m}\sum_{t=1}^{m}\left[\frac{\bar{R}_{t+12}^p}{R_t^{(1)}}\right]^{1-\gamma}\right),$$

where $\hat{R}_{t+12}^p$ and $\bar{R}_{t+12}^p$ denote portfolio returns by using the real-time macro factor forecast and a benchmark forecast, respectively.

Table 8 reports the CER gain and MPP, assuming a proportional transaction cost of 20 basis points (bps) following Gargano et al. (2019). Following Table 7, we consider three benchmark forecasts, historical mean forecast, yield curve forecast, and yield curve and PCA factor-based forecast. The results indicate that forecasting bond returns with real-time macro variables are economically profitable, but investing in 5-year bonds is more profitable than investing in 2-year bonds. For example, when using historical mean forecast as the benchmark in Panel A, the investor obtains a CER gain of 19 bps for investing in 2-year bonds and 131 bps for investing in 5-year bonds, if she employs sSUFF.

Thornton and Valente (2012) and Sarno et al. (2016), among others, find it difficult to translate statistical predictability into economic value for investors who trade on government bonds. Table 8 provides two interpretations to such a finding. First, the predictability of bond returns is heterogeneous. The portfolio gain is likely to be negative for investing in 2-year bonds if the investor employs PCA, PLS, or SUFF. Instead, it is high likely to be positive for investing in 3- to 5-year bonds. Second, the portfolio gains are dependent on the benchmark forecasts. The gains for benchmarking on yield curve forecast are larger than that benchmarking on historical mean forecast. This result suggests that yield curve-based forecasts underperform the historical mean forecasts for real-time investments.

In sum, we conclude that real-time macro factors can generate economic values in predicting bond returns, so long as the econometric method takes into account nonlinearity and factor weakness.

3.8. Forecasting macroeconomic condition

Cochrane (2007) argues that return predictability is more plausibly related to macroeconomic risk if the return predictor also predicts macroeconomic condition. This section examines whether Cochrane's (2007) argument applies to real-time macro factors. To this end, we run the following predictive regression,

$$\Delta Y_{t+12} = \alpha + \beta f_t + \epsilon_{t+12}, \tag{12}$$

where f_t is the estimated single return predictor factor of PCA, SUFF, sPCA, or sSUFF, and $\Delta Y_{t+12} = Y_{t+12} - Y_t$, in which Y_t measures the macroeconomic condition, including the CFNAI, CFNAI sub-index on consumption and housing (consumption), CFNAI sub index on employment and hours (employment), St. Louis Fed smooth recession probability, Jurado et al. (2015) macroeconomic uncertainty index, and spread between the BAA and AAA yields.

Table 9: Forecasting macroeconomic condition with real-time macro factors
This table reports the regression slope, t-statistic, and R^2 of predicting annual changes of macroeconomic condition with real-time return predictor factors, which are extracted from six PCA, sPCA or PLS factors, or two SUFF or sSUFF predictive indices. Macroeconomic condition (Y_t) is measured by six proxies: the Chicago Fed National Activity Index (CFNAI), CFNAI sub index on consumption and housing (Consumption), CFNAI sub index on employment and hours (Employment), St Louis Fed smooth recession probability, Jurado, Ludvgison, and Ng's (2015) macroeconomic uncertainty index, and yield spread between the BAA and AAA yields. The annual change is computed as $\Delta Y_{t+12} = Y_{t+12} - Y_t$. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The evaluation period is 1982:03–2019:12.

	CFNAI	Consumption	Employment	Recession probability	Macro uncertainty	Yield spread
Panel A: $\Delta Y_{t+12} = \alpha + \beta f_t^{\text{PCA}} + \epsilon_{t+12}$						
β	0.14^{***}	0.12^{*}	0.20^{***}	-0.16^{**}	-0.18^{**}	-0.11^{**}
t-stat	2.79	1.95	3.06	-1.96	-2.14	-2.29
R^2	0.03	0.02	0.07	0.04	0.05	0.02
Panel B: $\Delta Y_{t+12} = \alpha + \beta f_t^{\text{sPCA}} + \epsilon_{t+12}$						
β	0.15^{***}	0.11^{***}	0.21^{***}	-0.17^{***}	-0.21^{***}	-0.14^{***}
t-stat	3.50	2.89	4.26	-2.52	-3.17	-2.51
R^2	0.07	0.04	0.13	0.09	0.13	0.06
Panel C: $\Delta Y_{t+12} = \alpha + \beta f_t^{\text{PLS}} + \epsilon_{t+12}$						
β	0.13^{***}	0.08^{**}	0.17^{***}	-0.14^{**}	-0.18^{***}	-0.12^{**}
t-stat	2.96	2.16	3.31	-2.29	-2.71	-2.15
R^2	0.05	0.02	0.09	0.06	0.09	0.04
Panel D: $\Delta Y_{t+12} = \alpha + \beta f_t^{\text{SUFF}} + \epsilon_{t+12}$						
β	0.13^{***}	0.12^{**}	0.20^{***}	-0.16^{**}	-0.18^{**}	-0.11^{**}
t-stat	2.66	1.99	3.08	-2.00	-2.19	-2.31
R^2	0.03	0.02	0.06	0.04	0.05	0.02
Panel E: $\Delta Y_{t+12} = \alpha + \beta f_t^{\text{sSUFF}} + \epsilon_{t+12}$						
β	0.16^{***}	0.13^{***}	0.21^{***}	-0.18^{***}	-0.21^{***}	-0.13^{**}
t-stat	3.28	2.86	3.83	-2.39	-2.75	-2.29
R^2	0.06	0.04	0.11	0.08	0.11	0.04

Table 9 reports the results. For ease of exposition, we standardize both the left and right-hand side variables to have mean zero and unit standard deviation. We find the sPCA, PLS and sSUFF return predictor factors exhibit stronger predictive power in forecasting future one-year ahead changes in macroeconomic condition. In particular, all the factors positively predict future increase in the CFNAI and their sub-index on consumption and employment, and negatively predict the annual changes in recession probability, macro uncertainty and credit spread. Overall, Table 9 provides empirical evidence in support of Cochrane (2007) that real-time macro factors predict future bond returns because they are able to predict the changes of future macroeconomic condition.

3.9. Counter-cyclical term premia

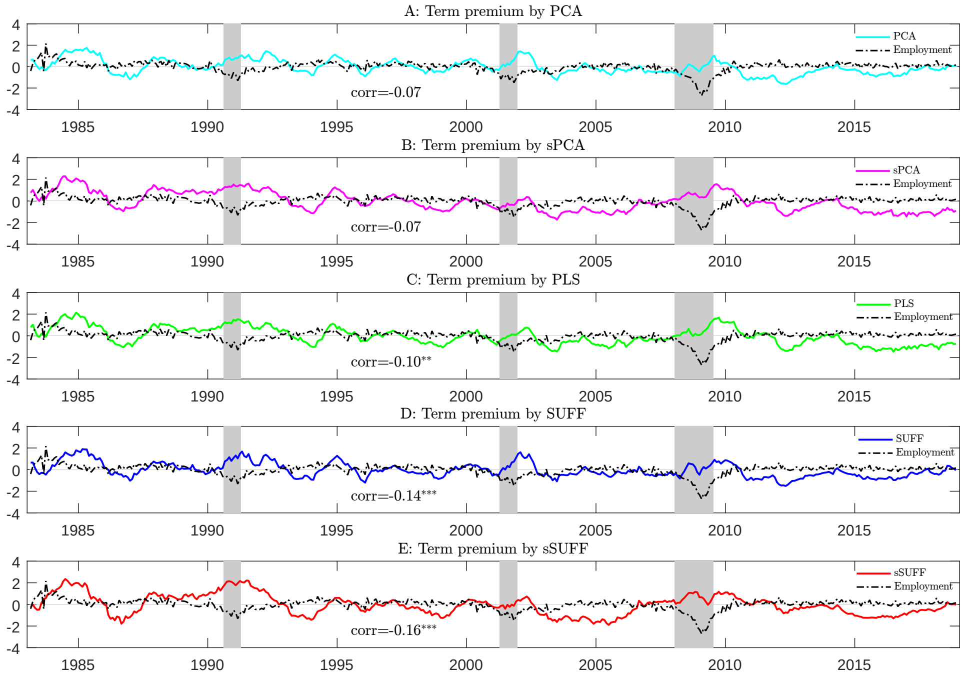
Many macro finance models predict time-varying term premia due to time-varying macroeconomic risk or time-varying risk bearing ability, high in recessions and low in expansions (see, e.g., Ludvigson and Ng, 2009; Wright, 2011; Joslin, Priebsch, and Singleton, 2014). With this insight, Creal and Wu (2020) extend consumption-based models to incorporate time variations in both the prices and quantities of growth and inflation risks and show that time-varying inflation risk is an important driver of bond risk premia. However, Ghysels et al. (2018) show that real-time macro data substantially reduce the countercyclicality of the implied term premia, and argue that the previously documented countercyclical term premia may be largely driven by the macro data revisions.

We re-examine the countercyclical behavior of term premia using the real-time macro factors. Following Ludvigson and Ng (2009) and Ghysels et al. (2018), we estimate term premia using a VAR model that includes the annual returns of 2- to 5-year bonds. Specifically, the term premium of an n-year bond is estimated as

$$tp_t^{(n)} = \frac{1}{n}\left[\mathrm{E}_t(rx_{t+12}^{(n)}) + \mathrm{E}_t(rx_{t+24}^{(n-1)}) + \cdots + \mathrm{E}_t(rx_{t+(n-1)*12}^{(2)})\right],$$

where $\mathrm{E}_t(rx_{t+h*12}^{(n-h+1)})$ denotes the time t forecast for the h-year ahead annual bond return on an $(n-h+1)$-year bond return from $t+(h-1)*12$ to $t+h*12$, where h spans from 1 to $n-1$. The h-year ahead annual bond return forecast is estimated using the monthly VAR model with 12 lags. The above definition for $tp_t^{(n)}$ is equivalent to the difference between the yield of an n-year bond and the average expected short rate over the life of the bond. The expectation hypothesis assumes

Figure 3: Term premium and macroeconomic condition
This figure plots the time-series of term premium (i.e., bond risk premium in yields), which is estimated with a VAR model that includes the annual return of 2- to 5-year government bonds. The yield curve predictors are the first three principal components of 1- to 5-year bond yields. Macro factors are the single return predictor factors of PCA, sPCA, PLS, SUFFand sSUFF, respectively. Each panel also plots the Chicago Fed Employment Index (Employment) as a proxy for the state of macroeconomic condition. The vertical bars indicate NBER-dated recessions. The sample period is 1982:03–2019:12.



that the term premium is constant, while the macro finance models suggest that the term premium is countercyclical.

Figure 3 plots the time-series dynamics of the 5-year bond term premium $tp_t^{(5)}$ implied by the VAR model. Across all panels, the term premia implied by the real-time macro factor appear to exhibit countercyclical patterns, rising around NBER recessions and falling during expansions, consistent with Ludvigson and Ng (2009). To quantify the degree of countercyclicality of the implied term premia, we further plot in each panel the time-series of the monthly CFNAI employment index and calculate its correlation with the term premia. We find that the correlation between the term premium implied by the PCA, sPCA and PLS forecasts and the CFNAI employment index are around -0.07 to -0.10, while the correlations with the SUFF and sSUFF forecasts are -0.14 and -0.16, with the latter two being statistically significant at the 1% level. Thus, once the forecasts are improved, the term premia would display a more countercyclical feature, thereby lending support to leading macro-finance term structure models.

3.10. Time-varying forecasting performance

Finally, we investigate the time-varying forecasting performance of the real-time macro factors, in order to further explore their link with macroeconomic condition. In the stock market, Rapach, Strauss, and Zhou (2010), Dangl and Halling (2012), and Henkel, Martin, and Nardari (2011), among others, show that return predictability is stronger in economic recessions. In the bond market, Sarno et al. (2016) and Gargano et al. (2019) find a similar pattern when the final revised macro factors are incorporated into the predictive regression framework.

Following Eriksen (2017), we examine the correlations of the relative forecast accuracy (RFA) and relative realized portfolio return (RPR) with macroeconomic condition.[12] RFA and RPR are respectively defined as the difference in squared forecast error and the difference in realized portfolio return generated by the real-time macro factors against those generated by the historical return mean. For ease of exposition, we focus on the 5-year bond returns below.

Panel A of Table 10 shows that the RFA generally exhibits negative correlations with the CFNAI, consumption, and employment, and positive correlations with the recession probability, macro uncertainty, and yield spread. Panel B focuses on the RPR and documents evident countercyclical patterns across all the real-time macro factors. The magnitudes of correlations with sSUFF forecasts are larger in magnitude than those with other forecasts. These results indicate that the real-time macro factors have stronger forecasting power statistically and economically in periods of economic

[12] A popular way to assess the time-varying monthly return predictability is to evaluate the forecast errors separately during the NBER-dated expansions and recessions (e.g., Gargano, Pettenuzzo, and Timmermann, 2019). However, such a sample split test seems infeasible and inaccurate for us, since we focus on annual holding period returns, which may cover multiple economic states.

Table 10: Macro factors' forecasting performance and macroeconomic condition This table reports the correlation between the forecasting performances of real-time macro factors and macroeconomic condition. We extract six factors when using PCA, sPCA and PLS and two predictive indices when using SUFF and sSUFF from 60 real-time macro variables. We use two proxies to measure the forecasting performance: relative forecast accuracy ($RFA_t^{(5)}$) and relative realized portfolio return ($RPR_t^{(5)}$), calculated as the difference in squared forecast error and the difference in realized portfolio return for a mean-variance investor who forecasts 5-year bond returns with a macro factor against the historical return mean. Macroeconomic condition (Y_t) is measured by seven proxies: the Chicago Fed National Activity Index (CFNAI), CFNAI sub index on consumption and housing (Consumption), CFNAI sub index on employment and hours (Employment), St Louis Fed smooth recession probability, Jurado et al.'s (2015) macroeconomic uncertainty index, and yield spread between the BAA and AAA yields. All the macro factors and expected bond returns are recursively estimated with an expanding window approach. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The evaluation period is 1982:03–2019:12.

	Y_t					
	CFNAI	Consumption	Employment	Recession probability	Macro uncertainty	Yield spread
Panel A: corr($RFA_t^{(5)}, Y_t$)						
PCA	-0.05	-0.17^{***}	-0.08^{*}	0.02	0.04	0.15^{***}
sPCA	-0.05	-0.18^{***}	-0.06^{*}	0.07	0.06	0.15^{***}
PLS	-0.11^{**}	-0.20^{***}	-0.12^{**}	0.13^{**}	0.13^{**}	0.21^{***}
SUFF	-0.06	-0.15^{***}	-0.11	0.09	0.08	0.20^{***}
sSUFF	-0.13^{***}	-0.12^{***}	-0.17^{***}	0.16^{***}	0.11^{**}	0.18^{***}
Panel B: corr($RPR_t^{(5)}, Y_t$)						
PCA	-0.04	-0.26^{***}	-0.07	0.00	0.09^{*}	0.23^{***}
sPCA	-0.03	-0.36^{***}	-0.02	0.05	0.08^{*}	0.22^{***}
PLS	-0.09^{*}	-0.29^{***}	-0.09^{*}	0.11^{**}	0.19^{***}	0.29^{***}
SUFF	-0.08	-0.22^{***}	-0.14^{***}	0.09^{*}	0.16^{***}	0.27^{***}
sSUFF	-0.14^{***}	-0.20^{***}	-0.18^{***}	0.15^{***}	0.17^{***}	0.25^{***}

recessions. It also suggests that macroeconomic risk can play an important role in driving the predictability of bond returns.

4. Conclusion

We study an important and debatable question of whether bond returns can be predicted by factors extracted from real-time macro variables. We argue that macro factors are likely to be nonlinearly related to future bond returns, and some of them are weak, in the sense that they are loaded by a partial set of macro variables. To improve the forecasting power, we extend Fan et al. (2017) sufficient forecasting to a weak and partially relevant factor framework that is suitable for both nonlinearity and weak factors. Similar to the scaled PCA in Huang et al. (2022), the idea is to overweight factors with high forecasting power while downweight those with low power. Econometrically, we study the properties of the scaled sufficient forecasting method, shedding light on why it can extract relevant factors more efficiently than the sufficient forecasting. We also conduct simulations to gain additional econometric insights.

Empirically, we apply the scaled sufficient forecasting approach to predict 2- to 5-year government bond returns with 60 real-time macro variables. We find that bond return forecasts are significant in- and out-of-sample at all maturities, and they outperform those based on other alternative methods, such as the sufficient forecasting, and linear algorithms of PCA, scaled PCA, and PLS. We also find that the bond risk premia based on our new method are countercycilical, consistent with well-known macro finance theories. Overall, the success of our method is due to the fact that it can accommodate both nonlinearity and factor weakness.

There are a number of topics that seem important for future research. First, while this paper focuses on bond return predictability, it is of interest to examine how the scaled sufficient forecasting improves return predictability in a big data environment for other markets such as the stock and currency markets. Second, from an investment perspective, one may also include real-time macro survey data or high frequency data to improve the economic value of return predictability even further. Third, with a more efficient return forecast based on real-time macro data, one may better decompose the yield curve to assess how the bond market's responses to monetary policy changes, helping policy makers to better judge in real time the separate impacts of monetary policies on the market's expectation of future inflation and the market wide risk appetite for government bonds.

Acknowledgement

We are extremely grateful to the Editor (Torben G. Andersen), an anonymous associate editor, and three anonymous referees for their very insightful comments that improved the paper substantially. We thank Hui Chen, Zhanhui Chen, Anna Cieslak, Zhi Da, Jonas Eriksen, Jianqing Fan, Eric Ghysels, James Hamilton, Wenxin Huang, Ben Jacobsen, Scott Joslin, Toshio Kimura, Junye Li, Laura Xiaolei Liu, Roger Loh, Ruichang Lu, Sydney C. Ludvigson, Neil Pearson, Markus Pelger, Christopher Polk, David Rapach, Allan Timmermann, Jun Tu, Baolian Wang, Dacheng Xiu, Jiangmin Xu, Hong Zhang, Xiaoyan Zhang, Hao Zhou, Ning Zhu, and participants from a number of seminars and conferences. This paper won the *WRDS Best Paper Award* at the 2018 AsianFA Annual Meeting under the title "Forecasting Bond Returns with Real Time Macroeconomic Data: A Predictive Principal Component Approach". Authors are listed alphabetically and contributed equally to this paper. F. Jiang acknowledges financial support from the National Social Science Fund of China [22&ZD063], National Natural Science Foundation of China [72072193, 71872195]. K. Li gratefully acknowledges financial support from the National Natural Science Foundation of China [72225006].

References

Ahn, S., Horenstein, A., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81, 1203–1227.

Ang, A., Bekaert, G., 2007. Stock return predictability: Is it there? Review of Financial Studies 20, 651–707.

Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of the term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 50, 745–787.

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

Bai, J., Li, K., 2016. Maximum likelihood estimation and inference for approximate factor models of high dimension. Review of Economics and Statistics 98, 298–309.

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.

Bailey, S., 2012. Principal component analysis with noisy and/or missing data. Astronomical Society of the Pacific 124, 1015.

Bair, E., Hastie, T., Paul, D., Tibshirani, R., 2006. Prediction by supervised principal components. Review of Financial Studies 101, 119–137.

Bauer, M. D., Hamilton, J. D., 2018. Robust bond risk premia. Review of Financial Studies 31, 399–448.

Bernanke, B. S., Kuttner, K. N., 2005. What explains the stock market's reaction to federal reserve policy? Journal of Finance 60, 1221–1257.

Bianchi, D., Büchner, M., Tamoni, A., 2021. Bond risk premia with machine learning. Review of Financial Studies 34, 1046–1089.

Breach, T., D'Amico, S., Orphanides, A., 2020. The term structure and inflation uncertainty. Journal of Financial Economics 138, 388–414.

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.

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

Cochrane, J. H., 2007. Financial markets and the real economy. In: Mehra, R., Prescott, E. C. (eds.), *Handbook of the Equity Premium*, Elsevier: Amsterdam, pp. 237–330.

Cochrane, J. H., Piazzesi, M., 2005. Bond risk premia. American Economic Review 95, 138–160.

Cook, R., Ni, L., 2005. Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. Journal of the American Statistical Association 100, 410–428.

Cook, R. D., 2009. Regression graphics: Ideas for studying regressions through graphics. John Wiley & Sons.

Creal, D. D., Wu, J. C., 2020. Bond risk premia in consumption-based models. Quantitative Economics 11, 1461–1484.

Dangl, T., Halling, M., 2012. Predictive regressions with time-varying coefficients. Journal of Financial Economics 106, 157–181.

Duffee, G. R., 2013. Forecasting interest rates. In: Elliott, G., Timmermann, A. (eds.), *Handbook of Economic Forecasting*, Elsevier, vol. 2, Part A of *Handbook of Economic Forecasting*, chap. 7, pp. 385–426.

Eriksen, J. N., 2017. Expected business conditions and bond risk premia. Journal of Financial and Quantitative Analysis 52, 1667–1703.

Fama, E., French, K., 1988. Dividend yields and expected stock returns. Journal of Financial Economics 19, 3–29.

Fama, E. F., Bliss, R. R., 1987. The information in long-maturity forward rates. American Economic Review 77, 680–692.

Fan, J., Gijbels, I., 1996. Local polynomial modelling and its applications. CRC Press.

Fan, J., Xue, L., Yao, J., 2017. Sufficient forecasting using factor models. Journal of econometrics 201, 292–306.

Feldhütter, P., Heyerdahl-Larsen, C., Illeditsch, P., 2018. Risk premia and volatilities in a nonlinear term structure model. Review of Finance 22, 337–380.

Fulop, A., Li, J., Wan, R., 2021. Real-time bayeisan learning and bond return predictability. Journal of Econometrics 230, 114–130.

Gargano, A., Pettenuzzo, D., Timmermann, A., 2019. Bond return predictability: Economic value and links to the macroeconomy. Management Science 65, 459–954.

Ghysels, E., Horan, C., Moench, E., 2018. Forecasting through the rear-view mirror: Data revisions and bond return predictability. Review of Financial Studies 31, 678–714.

Giacoletti, M., Laursen, K., Singleton, K. J., 2021. Learning from disagreement in the u.s. treasury bond market. Journal of Finance 76, 395–441.

Giglio, S., Xiu, D., Zhang, D., 2021. Test assets and weak factors. Chicago Booth Research Paper.

Goetzmann, W., Ingersoll, J., Spiegel, M., Welch, I., 2007. Portfolio performance manipulation and manipulation-proof performance measures. Review of Financial Studies 20, 1503–1546.

Gürkaynak, R. S., Sack, B., Wright, J. H., 2007. The U.S. treasury yield curve: 1961 to the present. Journal of Monetary Economics 54, 2291–2304.

Henkel, S. J., Martin, J. S., Nardari, F., 2011. Time-varying short-horizon predictability. Journal of Financial Economics 99, 560–580.

Hodrick, R. J., 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. Review of Financial Studies 5, 357–386.

Huang, D., Jiang, F., Li, K., Tong, G., Zhou, G., 2022. Scaled PCA: A new approach to dimension reduction. Management Science 3, 1678–1695.

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.

Huang, J.-Z., Shi, Z., 2022. Machine-learning-based return predictors and the spanning controversy in macro-finance. Management Science, forthcoming.

Inoue, A., Kilian, L., 2005. In-sample or out-of-sample tests of predictability: Which one should we use? Econometric Review 23, 371–402.

Jiang, F., Lee, J., Martin, X., Zhou, G., 2019. Manager sentiment and stock returns. Journal of Financial Economics 132, 126–149.

Joslin, S., Le, A., Singleton, K. J., 2013. Why gaussian macro-finance term structure models are (nearly) unconstrained factor-VARs. Journal of Financial Economics 109, 604–622.

Joslin, S., Priebsch, M., Singleton, K. J., 2014. Risk premiums in dynamic term structure models with unspanned macro risks. Journal of Finance 69, 1197–1233.

Jurado, K., Ludvgison, S. C., Ng, S., 2015. Measuring uncertainty. American Economic Review 105, 1177–1216.

Lettau, M., Pelger, M., 2020a. Estimating latent asset-pricing factors. Journal of Econometrics 218, 1–31.

Lettau, M., Pelger, M., 2020b. Factors that fit the time series and cross-section of stock returns. Review of Financial Studies 33, 2274–2325.

Li, K.-C., 1991. Sliced inverse regression for dimension reduction. Journal of the American Statistical Association 86, 316–327.

Liu, Y., Wu, J. C., 2021. Reconstructing the yield curve. Journal of Financial Economics 142, 1395–1425.

Ludvigson, S. C., Ng, S., 2009. Macro factors in bond risk premia. Review of Financial Studies 22, 5027–5067.

Ludvigson, S. C., Ng, S., 2011. A factor analysis of bond risk premia. In: Ullah, A., Giles, D. E. A. (eds.), *Handbook of Empirical Economics and Finance*, Chapman and Hall, pp. 313–372.

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

Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. Review of Economics and Statistics 92, 1004–1016.

Rapach, D., Strauss, J. K., Zhou, G., 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. Review of Financial Studies 23, 821–862.

Sarno, L., Schneider, P., Wagner, C., 2016. The economic value of predicting bond risk premia. Journal of Empirical Finance 37, 247–267.

Stock, J. H., Watson, M. W., 1998. Diffusion indexes. Working paper.

Thornton, D. L., Valente, G., 2012. Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. Review of Financial Studies 25, 3141–3168.

Wang, H., Xia, Y., 2008. Sliced regression for dimension reduction. Journal of the American Statistical Association 103, 811–821.

Wright, J. H., 2011. Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. American Economic Review 101, 1514–1534.

Yu, X., Yao, J., Xue, L., 2022. Nonparametric estimation and conformal inference of the sufficient forecasting with adiverging number of factors. Journal of Business & Economic Statistics 40, 342–354.

Zhao, F., Zhou, G., Zhu, X., 2021. Unspanned global macro risks in bond returns. Management Science 67, 7825–7843.

Online Appendix

Are Bond Returns Predictable with Real-Time Macro Data?

A. Proofs for the theoretical results

This appendix provides the detailed proofs of the two theorems in the main text, which are based on the following six propositions. Specifically, Propositions A.1– A.4 jointly prove Theorem 1, and Propositions A.6–A.7 jointly prove Theorem 2.

We first introduce some notations. Let $\boldsymbol{X}$ be $N \times T$ matrix with its (i,t)th element being x_{it}, and $\boldsymbol{X}_i$ be the transpose of its ith row. With this definition,

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{F}' + \boldsymbol{U}, \qquad \text{or} \qquad \boldsymbol{X}_i = \boldsymbol{F}\boldsymbol{b}_i + \boldsymbol{U}_i,$$

where $\boldsymbol{U}$ and $\boldsymbol{U}_i$ are defined in the same way as $\boldsymbol{X}$ and $\boldsymbol{X}_i$. We have the following proposition on the eigenvalues.

Proposition A.1. *Let $\widehat{\boldsymbol{D}}$ be the diagonal matrix whose diagonal elements are the largest r eigenvalues of $\frac{1}{N^\nu T}\boldsymbol{X}'\boldsymbol{X}$ with $\boldsymbol{X} = (x_{it})_{N\times T}$. Under Assumptions A-D, as $N, T \to \infty, N^{1-\nu}/T \to 0$,*

$$\tau_i(\widehat{\boldsymbol{D}}) - \tau_i\Big(\frac{1}{N^\nu}\boldsymbol{B}'\boldsymbol{B}\Big) = o_p(1),$$

where $\tau_i(\cdot)$ denotes the i-th largest eigenvalue of its input.

Proposition A.1 can be proved by the same arguments in the proof of Proposition 1 of Bai (2003).

Corollary A.1. *Define $\boldsymbol{R} = \frac{1}{N^\nu T}\boldsymbol{B}'\boldsymbol{B}\boldsymbol{F}'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1}$, where $\widehat{\boldsymbol{F}}$ is the PC estimator for the matrix $\boldsymbol{X}$, we have $\boldsymbol{R} = O_p(1)$. In addition, if $\boldsymbol{R}'\boldsymbol{R} = I_r + o_p(1)$ as asserted in result (e) of Proposition A.2, we have $\boldsymbol{R}^{-1} = O_p(1)$.*

Proof of Corollary A.1. By Proposition A.1, we see

$$\|\boldsymbol{R}\| \le \frac{\|\boldsymbol{B}'\boldsymbol{B}\|}{N^\nu}\frac{\|\boldsymbol{F}\|}{\sqrt{T}}\frac{\|\widehat{\boldsymbol{F}}\|}{\sqrt{T}}\|\widehat{\boldsymbol{D}}^{-1}\| = O_p(1).$$

For the second result, we note that, by condition, $\boldsymbol{R}^{-1} = \boldsymbol{R}' + o_p(1)$. So $\boldsymbol{R}^{-1} = O_p(1)$ because $\boldsymbol{R} = O_p(1)$. □

Proposition A.2. *Under the assumptions of Proposition A.1, we have*

$$(a)\ \ \frac{1}{\sqrt{T}}\|\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R}\| = O_p(\frac{1}{\sqrt{N^\nu}}) + O_p(\frac{N^{1-\nu}}{T}),$$

$$(b)\ \ \frac{1}{T}\|\boldsymbol{F}'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\| = O_p(\frac{1}{\sqrt{N^\nu T}}) + O_p(\frac{1}{N^\nu}) + O_p(\frac{N^{1-\nu}}{T}),$$

$$(c)\ \ \frac{1}{T}\|\boldsymbol{U}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\| = O_p(\frac{1}{\sqrt{N^\nu T}}) + O_p(\frac{1}{N^\nu}) + O_p(\frac{N^{1-\nu}}{T^{3/2}}) = O_p(\frac{1}{N^\nu}) + o_p(\frac{1}{\sqrt{T}}),$$

$$(d)\ \ \frac{1}{T}\|\boldsymbol{X}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\| = O_p(\frac{1}{N^\nu}) + O_p(\frac{N^{1-\nu}}{T}),$$

$$(e)\ \ \boldsymbol{R}'\boldsymbol{R} - I_r = O_p(\frac{1}{N^\nu}) + O_p(\frac{N^{1-\nu}}{T}).$$

Proof of Proposition A.2. Consider (a). Note that

$$\frac{1}{N^\nu T}\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{F}} = \widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}.$$

Substitute $\boldsymbol{X} = \boldsymbol{B}\boldsymbol{F}' + \boldsymbol{U}$ into the above expression,

$$\widehat{\boldsymbol{F}} = \frac{1}{N^\nu T}\boldsymbol{F}\boldsymbol{B}'\boldsymbol{B}\boldsymbol{F}'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1} + \frac{1}{N^\nu T}\boldsymbol{F}\boldsymbol{B}'\boldsymbol{U}\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1} + \frac{1}{N^\nu T}\boldsymbol{U}'\boldsymbol{B}\boldsymbol{F}'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1} + \frac{1}{N^\nu T}\boldsymbol{U}'\boldsymbol{U}\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1}.$$

By the definition of $\boldsymbol{R} = \frac{1}{N^\nu T}\boldsymbol{B}'\boldsymbol{B}\boldsymbol{F}'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1}$,

$$\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R} = \frac{1}{N^\nu T}\boldsymbol{F}\boldsymbol{B}'\boldsymbol{U}\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1} + \frac{1}{N^\nu T}\boldsymbol{U}'\boldsymbol{B}\boldsymbol{F}'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1} + \frac{1}{N^\nu T}\boldsymbol{U}'\boldsymbol{U}\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1}. \tag{A.1}$$

Given the above expression,

$$\begin{aligned}\frac{1}{\sqrt{T}}\|\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R}\| &\le \frac{1}{N^\nu T\sqrt{T}}\|\boldsymbol{F}\boldsymbol{B}'\boldsymbol{U}\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1}\| + \frac{1}{N^\nu T\sqrt{T}}\|\boldsymbol{U}'\boldsymbol{B}\boldsymbol{F}'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1}\| \\ &\quad + \frac{1}{N^\nu T\sqrt{T}}\|\boldsymbol{U}'\boldsymbol{U}\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1}\| \\ &= I_1 + I_2 + I_3, \qquad \text{say.}\end{aligned}$$

We consider the above three terms one by one. For I_1, we have

$$I_1 \le \frac{1}{\sqrt{N^\nu}}\frac{\|\boldsymbol{F}\|}{\sqrt{T}}\frac{\|\boldsymbol{B}'\boldsymbol{U}\|}{\sqrt{N^\nu T}}\frac{\|\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R}\|}{\sqrt{T}}\|\widehat{\boldsymbol{D}}^{-1}\| + \frac{1}{\sqrt{N^\nu T}}\frac{\|\boldsymbol{F}\|}{\sqrt{T}}\frac{\|\boldsymbol{B}'\boldsymbol{U}\boldsymbol{F}\|}{\sqrt{N^\nu T}}\|\boldsymbol{R}\|\|\widehat{\boldsymbol{D}}^{-1}\|.$$

The first term is $O_p(\frac{1}{\sqrt{N^\nu}})\|\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R}\|/\sqrt{T}$ which is a term of smaller order relative to $\|\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R}\|/\sqrt{T}$ and therefore is negligible. The second term is $O_p(\frac{1}{\sqrt{N^\nu T}})$. Given this, we have $I_1 = O_p(\frac{1}{\sqrt{N^\nu T}})$. For I_2,

$$I_2 \leq \frac{1}{\sqrt{N^\nu}} \frac{\|\boldsymbol{U}'\boldsymbol{B}\|}{\sqrt{N^\nu T}} \frac{\|\boldsymbol{F}'\widehat{\boldsymbol{F}}\|}{T} \|\widehat{\boldsymbol{D}}^{-1}\| = O_p(\frac{1}{\sqrt{N^\nu}}).$$

For I_3, we have

$$I_3 \leq \frac{1}{N^\nu T} \|\boldsymbol{U}'\boldsymbol{U}\|_2 \frac{\|\widehat{\boldsymbol{F}}\|}{\sqrt{T}} \|\widehat{\boldsymbol{D}}^{-1}\|.$$

By Theorem 5.8 of Bai and Silverstein (2010), we have that $\|\frac{1}{\max(N,T)}\boldsymbol{U}'\boldsymbol{U}\|_2 = O_p(1)$. This implies that $\|\boldsymbol{U}'\boldsymbol{U}\|_2 = O_p(\max(N,T)) = O_p(N) + O_p(T)$. With this result, we have that $I_3 = O_p(\frac{N^{1-\nu}}{T}) + O_p(N^{-\nu})$. Summarizing the results on I_1, I_2 and I_3, we conclude (a).

Consider (b). By (A.1),

$$\begin{aligned}
\frac{1}{T}\boldsymbol{F}'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R}) &= \frac{1}{N^\nu T}\boldsymbol{B}'\boldsymbol{U}\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1} + \frac{1}{N^\nu T^2}\boldsymbol{F}'\boldsymbol{U}'\boldsymbol{B}\boldsymbol{F}'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1} + \frac{1}{N^\nu T^2}\boldsymbol{F}'\boldsymbol{U}'\boldsymbol{U}\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1} \\
&= II_1 + II_2 + II_3, \qquad \text{say}
\end{aligned}$$

We consider the three terms one by one. For II_1, it can be decomposed into

$$II_1 = \frac{1}{N^\nu T}\boldsymbol{B}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-1} + \frac{1}{N^\nu T}\boldsymbol{B}'\boldsymbol{U}\boldsymbol{F}\boldsymbol{R}\widehat{\boldsymbol{D}}^{-1}$$

The second term is bounded in norm by

$$\frac{1}{\sqrt{N^\nu T}} \frac{\|\boldsymbol{B}'\boldsymbol{U}\boldsymbol{F}\|}{\sqrt{N^\nu T}} \|\boldsymbol{R}\| \|\widehat{\boldsymbol{D}}^{-1}\| = O_p(\frac{1}{\sqrt{N^\nu T}}).$$

We next consider the first term. By (A.1),

$$\begin{aligned}
\frac{1}{N^\nu T}\boldsymbol{B}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-1} &= \frac{1}{N^{2\nu}T^2}\boldsymbol{B}'\boldsymbol{U}\boldsymbol{F}\boldsymbol{B}'\boldsymbol{U}\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-2} + \frac{1}{N^{2\nu}T^2}\boldsymbol{B}'\boldsymbol{U}\boldsymbol{U}'\boldsymbol{B}\boldsymbol{F}'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-2} \\
&\quad + \frac{1}{N^{2\nu}T^2}\boldsymbol{B}'\boldsymbol{U}\boldsymbol{U}'\boldsymbol{U}\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-2}
\end{aligned}$$

The first term is bounded in norm by

$$\frac{1}{N^\nu\sqrt{T}} \frac{\|\boldsymbol{B}'\boldsymbol{U}\boldsymbol{F}\|}{\sqrt{N^\nu T}} \frac{\|\boldsymbol{B}'\boldsymbol{U}\|}{\sqrt{N^\nu T}} \frac{\|\widehat{\boldsymbol{F}}\|}{\sqrt{T}} \|\widehat{\boldsymbol{D}}^{-2}\| = O_p(\frac{1}{N^\nu\sqrt{T}}).$$

The second term is bounded in norm by

$$\frac{1}{N^\nu\sqrt{T}} \frac{\|\boldsymbol{B}'[\boldsymbol{U}\boldsymbol{U}' - \mathbb{E}(\boldsymbol{U}\boldsymbol{U}')]\boldsymbol{B}\|}{N^\nu\sqrt{T}} \frac{\|\boldsymbol{F}\|}{\sqrt{T}} \frac{\|\widehat{\boldsymbol{F}}\|}{\sqrt{T}} \|\widehat{\boldsymbol{D}}^{-2}\| + \frac{1}{N^\nu} \frac{\|\boldsymbol{B}'\mathbb{E}(\boldsymbol{U}\boldsymbol{U}')\boldsymbol{B}\|}{N^\nu T} \frac{\|\boldsymbol{F}\|}{\sqrt{T}} \frac{\|\widehat{\boldsymbol{F}}\|}{\sqrt{T}} \|\widehat{\boldsymbol{D}}^{-2}\|.$$

which is $O_p(\frac{1}{N^\nu}) + O_p(\frac{1}{N^\nu\sqrt{T}})$. The third term can be further decomposed into

$$\frac{1}{N^{2\nu}T^2}\boldsymbol{B}'\boldsymbol{U}\boldsymbol{U}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-2} + \frac{1}{N^{2\nu}T^2}\boldsymbol{B}'\boldsymbol{U}\boldsymbol{U}'\boldsymbol{U}\boldsymbol{F}\boldsymbol{R}\widehat{\boldsymbol{D}}^{-2}$$

Note that

$$\begin{aligned}
\|\boldsymbol{B}'\boldsymbol{U}\boldsymbol{U}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-2}\|^2 &= \mathrm{tr}\Big[\boldsymbol{B}'\boldsymbol{U}\boldsymbol{U}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-4}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}'\boldsymbol{U}\boldsymbol{U}'\boldsymbol{B}\Big] \\
&\leq \|\widehat{\boldsymbol{D}}^{-1}\|_2^2 \cdot \|\boldsymbol{U}\boldsymbol{U}'\|_2^2 \cdot \mathrm{tr}\Big[\boldsymbol{B}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-2}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}'\boldsymbol{B}\Big] \\
&= \|\widehat{\boldsymbol{D}}^{-2}\|_2^2 \cdot \|\boldsymbol{U}\boldsymbol{U}'\|_2^2 \cdot \|\boldsymbol{B}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-1}\|^2.
\end{aligned}$$

With the above result, we have

$$\begin{aligned}
\frac{1}{N^{2\nu}T^2}\|\boldsymbol{B}'\boldsymbol{U}\boldsymbol{U}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-2}\| &\leq \|\widehat{\boldsymbol{D}}^{-1}\|_2\Big[\frac{1}{N^\nu T}\|\boldsymbol{U}\boldsymbol{U}'\|_2\Big]\Big[\frac{1}{N^\nu T}\|\boldsymbol{B}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\|^2\widehat{\boldsymbol{D}}^{-1}\Big] \\
&= o_p(1)\Big[\frac{1}{N^\nu T}\|\boldsymbol{B}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-1}\|\Big]
\end{aligned}$$

So the first expression is of smaller order term relative to $\frac{1}{N^\nu T}\|\boldsymbol{B}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-1}\|$ and is therefore negligible. For the second expression, it is bounded in norm by

$$\frac{N^{1/2-3\nu/2}}{\sqrt{T}}\frac{\|\boldsymbol{B}'\boldsymbol{U}[\boldsymbol{U}'\boldsymbol{U} - \mathbb{E}(\boldsymbol{U}'\boldsymbol{U})]\|}{N^{1/2+\nu/2}T}\frac{\|\boldsymbol{F}\|}{\sqrt{T}}\|\boldsymbol{R}\|\|\widehat{\boldsymbol{D}}^{-2}\| + \frac{N^{1-3\nu/2}}{T\sqrt{T}}\frac{\|\boldsymbol{B}'\boldsymbol{U}\mathbb{E}(\boldsymbol{U}'\boldsymbol{U})\boldsymbol{F}\|}{N^{1+\nu/2}\sqrt{T}}\|\boldsymbol{R}\|\|\widehat{\boldsymbol{D}}^{-2}\|$$

which is $o_p(N^{-\nu}) + o_p(\frac{1}{\sqrt{N^\nu T}})$ under $N^{1-\nu}/T \to 0$. With this result, we have

$$II_1 = O_p(\frac{1}{N^\nu}) + O_p(\frac{1}{\sqrt{N^\nu T}}).$$

For II_2, we have

$$\|II_2\| \leq \frac{1}{\sqrt{N^\nu T}}\frac{\|\boldsymbol{F}'\boldsymbol{U}'\boldsymbol{B}\|}{\sqrt{N^\nu T}}\frac{\|\boldsymbol{F}'\widehat{\boldsymbol{F}}\|}{T}\|\widehat{\boldsymbol{D}}^{-1}\| = O_p(\frac{1}{\sqrt{N^\nu T}}).$$

For II_3, it can be decomposed into

$$\frac{1}{N^\nu T^2}\boldsymbol{F}'\boldsymbol{U}'\boldsymbol{U}\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1} = \frac{1}{N^\nu T^2}\boldsymbol{F}'\boldsymbol{U}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-1} + \frac{1}{N^\nu T^2}\boldsymbol{F}'\boldsymbol{U}'\boldsymbol{U}\boldsymbol{F}\boldsymbol{R}\widehat{\boldsymbol{D}}^{-1}.$$

Note that

$$\begin{aligned}
\|\boldsymbol{F}'\boldsymbol{U}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-1}\|^2 &= \mathrm{tr}\Big[\boldsymbol{F}'\boldsymbol{U}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-2}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}'\boldsymbol{U}\boldsymbol{F}\Big] \\
&\leq \|\widehat{\boldsymbol{D}}^{-1}\|_2^2 \cdot \|\boldsymbol{U}'\boldsymbol{U}\|_2^2\mathrm{tr}\Big[\boldsymbol{F}'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\Big] \\
&\leq \|\widehat{\boldsymbol{D}}^{-1}\|_2^2 \cdot \|\boldsymbol{U}'\boldsymbol{U}\|_2^2 \cdot \|\boldsymbol{F}'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\|^2
\end{aligned}$$

With the above result, we have

$$\frac{1}{N^\nu T^2}\|\boldsymbol{F}'\boldsymbol{U}'\boldsymbol{U}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-1}\| \le \|\widehat{\boldsymbol{D}}^{-1}\|_2\Big[\frac{1}{N^\nu T}\|\boldsymbol{U}'\boldsymbol{U}\|_2\Big]\Big[\frac{1}{T}\|\boldsymbol{F}'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\|\Big]$$
$$= o_p(1)\Big[\frac{1}{T}\|\boldsymbol{F}'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\|\Big].$$

So the first term is of smaller order relative to $\|\boldsymbol{F}'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\|/T$ and is therefore negligible. The second term is bounded in norm by

$$\frac{N^{1/2-\nu}}{T}\frac{\|\boldsymbol{F}'[\boldsymbol{U}'\boldsymbol{U}-\mathbb{E}(\boldsymbol{U}'\boldsymbol{U})]\boldsymbol{F}\|}{\sqrt{N}T}\|\boldsymbol{R}\|\|\widehat{\boldsymbol{D}}^{-1}\| + \frac{N^{1-\nu}}{T}\frac{\|\boldsymbol{F}\|^2}{T}\frac{\|\mathbb{E}(\boldsymbol{U}'\boldsymbol{U})\|_2}{N}\|\boldsymbol{R}\|\|\widehat{\boldsymbol{D}}^{-1}\|,$$

which is $O_p(\frac{N^{1/2-\nu}}{T}) + O_p(\frac{N^{1-\nu}}{T})$. Note that under $N^{1-\nu}/T \to 0$, $O_p(\frac{N^{1/2-\nu}}{T})$ is of smaller order relative to $O_p(\frac{1}{\sqrt{N^\nu T}})$. Given this result, $II_3 = o_p(\frac{1}{\sqrt{N^\nu T}}) + O_p(\frac{N^{1-\nu}}{T})$.

With the results on II_1, II_2 and II_3, we therefore conclude that

$$\frac{1}{T}\|\boldsymbol{F}'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\| = O_p(\frac{1}{\sqrt{N^\nu T}}) + O_p(\frac{1}{N^\nu}) + O_p(\frac{N^{1-\nu}}{T}).$$

Consider (c). By (A.1),

$$\frac{1}{T}\boldsymbol{U}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R}) = \frac{1}{N^\nu T^2}\boldsymbol{U}_i'\boldsymbol{F}\boldsymbol{B}'\boldsymbol{U}\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1} + \frac{1}{N^\nu T^2}\boldsymbol{U}_i'\boldsymbol{U}'\boldsymbol{B}\boldsymbol{F}'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1} + \frac{1}{N^\nu T^2}\boldsymbol{U}_i'\boldsymbol{U}'\boldsymbol{U}\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}}^{-1}$$
$$= III_1 + III_2 + III_3, \qquad \text{say}$$

For III_1, we have

$$\|III_1\| \le \frac{1}{\sqrt{N^\nu T}}\frac{\|\boldsymbol{U}_i'\boldsymbol{F}\|}{\sqrt{T}}\frac{\|\boldsymbol{B}'\boldsymbol{U}\|}{\sqrt{N^\nu T}}\frac{\|\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R}\|}{\sqrt{T}}\|\widehat{\boldsymbol{D}}^{-1}\| + \frac{1}{\sqrt{N^\nu T}}\frac{\|\boldsymbol{U}_i'\boldsymbol{F}\|}{\sqrt{T}}\frac{\|\boldsymbol{B}'\boldsymbol{U}\boldsymbol{F}\|}{\sqrt{N^\nu T}}\|\boldsymbol{R}\|\|\widehat{\boldsymbol{D}}^{-1}\|,$$

which is $o_p(\frac{1}{\sqrt{N^\nu T}})$. For III_2, we have

$$\|III_2\| \le \frac{1}{\sqrt{N^\nu T}}\frac{\|[\boldsymbol{U}_i'\boldsymbol{U}'-\mathbb{E}(\boldsymbol{U}_i'\boldsymbol{U}')]\boldsymbol{B}\|}{\sqrt{N^\nu T}}\frac{\|\boldsymbol{F}'\widehat{\boldsymbol{F}}\|}{T}\|\widehat{\boldsymbol{D}}^{-1}\| + \frac{1}{N^\nu}\frac{\|\mathbb{E}(\boldsymbol{U}_i'\boldsymbol{U}')\boldsymbol{B}\|}{T}\frac{\|\boldsymbol{F}'\widehat{\boldsymbol{F}}\|}{T}\|\widehat{\boldsymbol{D}}^{-1}\|,$$

which is $O_p(\frac{1}{\sqrt{N^\nu T}}) + O_p(N^{-\nu})$. For III_3, we have

$$\|III_3\| = \frac{1}{N^\nu T^2}\|\boldsymbol{U}_i'\boldsymbol{U}'\boldsymbol{U}\boldsymbol{F}\boldsymbol{R}\widehat{\boldsymbol{D}}^{-1}\| + \frac{1}{N^\nu T^2}\|\boldsymbol{U}_i'\boldsymbol{U}'\boldsymbol{U}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-1}\| \tag{A.2}$$

Note that

$$\|\boldsymbol{U}_i'\boldsymbol{U}'\boldsymbol{U}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-1}\|^2 = \mathrm{tr}\Big[\boldsymbol{U}_i'\boldsymbol{U}'\boldsymbol{U}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-2}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}'\boldsymbol{U}\boldsymbol{U}_i\Big]$$

$$\leq \|\widehat{\boldsymbol{D}}^{-1}\|_2^2 \|\boldsymbol{U}'\boldsymbol{U}\|_2^2 \mathrm{tr}\Big[\boldsymbol{U}_i'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\Big]$$
$$= \|\widehat{\boldsymbol{D}}^{-1}\|_2^2 \|\boldsymbol{U}'\boldsymbol{U}\|_2^2 \|\boldsymbol{U}_i'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\|^2$$

With the above result, we have

$$\begin{aligned}\frac{1}{N^\nu T^2}\|\boldsymbol{U}_i'\boldsymbol{U}'\boldsymbol{U}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\widehat{\boldsymbol{D}}^{-1}\| &\leq \frac{1}{N^\nu T^2}\|\widehat{\boldsymbol{D}}^{-1}\|_2 \|\boldsymbol{U}'\boldsymbol{U}\|_2 \|\boldsymbol{U}_i'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\| \\ &= \|\widehat{\boldsymbol{D}}^{-1}\|_2 \Big[\frac{1}{N^\nu T}\|\boldsymbol{U}'\boldsymbol{U}\|_2\Big]\Big[\frac{1}{T}\|\boldsymbol{U}_i'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\|\Big] \\ &= o_p(1)\Big[\frac{1}{T}\|\boldsymbol{U}_i'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\|\Big]\end{aligned}$$

So the second term of (A.2) is negligible since it is of smaller order relative to $\|\boldsymbol{U}_i'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\|/\sqrt{T}$. For the first term, it is bounded in norm by

$$\frac{N^{1/2-\nu}}{T}\frac{\|\boldsymbol{U}_i'[\boldsymbol{U}'\boldsymbol{U} - \mathbb{E}(\boldsymbol{U}'\boldsymbol{U})]\boldsymbol{F}\|}{\sqrt{N}T}\|\boldsymbol{R}\|\|\widehat{\boldsymbol{D}}^{-1}\| + \frac{N^{1-\nu}}{T\sqrt{T}}\frac{\|\boldsymbol{U}_i'\mathbb{E}(\boldsymbol{U}'\boldsymbol{U})\boldsymbol{F}\|}{N\sqrt{T}}\|\|\boldsymbol{R}\|\|\widehat{\boldsymbol{D}}^{-1}\|$$

which is $O_p(\frac{N^{1/2-\nu}}{T}) + O_p(\frac{N^{1-\nu}}{T\sqrt{T}})$. We therefore have $III_3 = O_p(\frac{N^{1/2-\nu}}{T}) + O_p(\frac{N^{1-\nu}}{T\sqrt{T}})$. Given the results on III_1, III_2 and III_3, together with the fact that $O_p(\frac{N^{1/2-\nu}}{T}) = o_p(\frac{1}{\sqrt{N^\nu T}})$ by $\frac{N^{1-\nu}}{T} \to 0$, we have (c).

Consider (d). Note that $\boldsymbol{X}_i' = \boldsymbol{b}_i'\boldsymbol{F}' + \boldsymbol{U}_i'$. So result (d) is a direct consequence of results (b) and (c).

Consider (e). By the fact that $\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}/T = I_r$ and $\boldsymbol{F}'\boldsymbol{F}/T = I_r$, we have

$$I_r = \frac{1}{T}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R}) + \frac{1}{T}\boldsymbol{R}'\boldsymbol{F}'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R}) + \frac{1}{T}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{R} + \boldsymbol{R}'\boldsymbol{R}.$$

The first term is $O_p(\frac{1}{N^\nu}) + O_p(\frac{N^{2-2\nu}}{T^2})$ due to result (a). The second term is given in result (b). We therefore have (e).

This completes the proof. □

Proposition A.3. *Under the assumptions of Proposition A.1, we have that for each h,*

$$\frac{1}{c}\sum_{l=1}^{c}(\widehat{\boldsymbol{f}}_{h,l} - \boldsymbol{R}'\boldsymbol{f}_{h,l}) = \frac{1}{\sqrt{N^\nu T}} + O_p(\frac{1}{N^\nu}) + O_p(\frac{N^{1-\nu}}{T}).$$

Proof of Proposition A.3. We can alternatively write (A.1) as, for each t,

$$\widehat{\boldsymbol{f}}_t - \boldsymbol{R}'\boldsymbol{f}_t = \boldsymbol{R}'(\boldsymbol{B}'\boldsymbol{B})^{-1}\sum_{i=1}^{N}\boldsymbol{b}_i u_{it} + \widehat{\boldsymbol{D}}^{-1}\Big[\frac{1}{N^\nu T}\sum_{i=1}^{N}\sum_{s=1}^{T}\widehat{\boldsymbol{f}}_s\boldsymbol{b}_i' u_{is}\Big]\boldsymbol{f}_t + \widehat{\boldsymbol{D}}^{-1}\frac{1}{N^\nu T}\sum_{i=1}^{N}\sum_{s=1}^{T}\widehat{\boldsymbol{f}}_s u_{is} u_{it}.$$

With the above result, we therefore have that for each h,

$$
\begin{aligned}
\frac{1}{c}\sum_{l=1}^{c}(\widehat{\boldsymbol{f}}_{h,l}-\boldsymbol{R}'\boldsymbol{f}_{h,l}) &= \boldsymbol{R}'(\boldsymbol{B}'\boldsymbol{B})^{-1}\frac{1}{c}\sum_{l=1}^{c}\sum_{i=1}^{N}\boldsymbol{b}_i u_{i,hl} + \widehat{\boldsymbol{D}}^{-1}\Big[\frac{1}{N^\nu T}\sum_{i=1}^{N}\sum_{s=1}^{T}\widehat{\boldsymbol{f}}_s'\boldsymbol{b}_i' u_{is}\Big]\Big(\frac{1}{c}\sum_{l=1}^{c}\boldsymbol{f}_{hl}\Big)\\
&\quad + \widehat{\boldsymbol{D}}^{-1}\frac{1}{N^\nu Tc}\sum_{l=1}^{c}\sum_{i=1}^{N}\sum_{s=1}^{T}\widehat{\boldsymbol{f}}_s u_{is}u_{i,hl}\\
&= I_{h,1}+I_{h,2}+I_{h,3}.
\end{aligned}
$$

Consider $I_{h,1}$. By Assumptions B and C, together with result (i) of Proposition A.2, one can readily show that $I_{h,1}=O_p(\frac{1}{\sqrt{N^\nu T}})$. For the second term, by $\frac{1}{c}\sum_{l=1}^{c}\boldsymbol{f}_{hl}=O_p(\frac{1}{\sqrt{T}})$, it is easy to show that $I_{h,2}=O_p(\frac{1}{\sqrt{N^\nu T}})$. For the third term, note that

$$
\begin{aligned}
\widehat{\boldsymbol{D}}^{-1}\frac{1}{N^\nu Tc}\sum_{l=1}^{c}\sum_{i=1}^{N}\sum_{s=1}^{T}\widehat{\boldsymbol{f}}_s u_{is}u_{i,hl} &= \widehat{\boldsymbol{D}}^{-1}\frac{1}{N^\nu Tc}\sum_{l=1}^{c}\sum_{i=1}^{N}\sum_{s=1}^{T}\widehat{\boldsymbol{f}}_s \mathbb{E}(u_{is}u_{i,hl})\\
&\quad + \widehat{\boldsymbol{D}}^{-1}\frac{1}{N^\nu Tc}\sum_{l=1}^{c}\sum_{i=1}^{N}\sum_{s=1}^{T}(\widehat{\boldsymbol{f}}_s-\boldsymbol{R}'\boldsymbol{f}_s)\zeta_{i,hl,s}\\
&\quad + \widehat{\boldsymbol{D}}^{-1}\boldsymbol{R}'\frac{1}{N^\nu Tc}\sum_{l=1}^{c}\sum_{i=1}^{N}\sum_{s=1}^{T}\boldsymbol{f}_s\zeta_{i,hl,s},
\end{aligned}
$$

where $\zeta_{i,ts}=u_{it}u_{is}-\mathbb{E}(u_{it}u_{is})$. The first term on right hand side is $O_p(\frac{N^{1-\nu}}{T})$. The second term is bounded in norm by

$$
\|\widehat{\boldsymbol{D}}^{-1}\|\Big[\frac{1}{T}\sum_{s=1}^{T}\|\widehat{\boldsymbol{f}}_s-\boldsymbol{R}'\boldsymbol{f}_s\|^2\Big]^{1/2}\Big[\frac{1}{T}\sum_{s=1}^{T}\Big\|\frac{1}{N^\nu c}\sum_{l=1}^{c}\sum_{i=1}^{N}\zeta_{i,hl,s}\Big\|^2\Big]^{1/2},
$$

which is $O_p(N^{1/2-3\nu/2}T^{-1/2})+O_p(N^{3/2-2\nu}T^{-3/2})$ due to Proposition A.2. The third term is $O_p(N^{1/2-\nu}T^{-1})$. The last two terms are both of smaller order relative to the first one. Given this, we have Proposition A.3. □

Proposition A.4. *Under the assumptions of Proposition A.1, we have*

$$
\|\widehat{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}}-\boldsymbol{R}'\Sigma_{\boldsymbol{f}|\boldsymbol{y}}\boldsymbol{R}\| = O_p(\frac{N^{1-\nu}}{T})+O_p(\frac{1}{\sqrt{T}})+O_p(\frac{1}{N^\nu}).
$$

where $\boldsymbol{R}$ is defined in Proposition A.2. Let $\boldsymbol{R}^*=(\Phi'\Phi)^{-1/2}\boldsymbol{R}_\Omega$ with $\boldsymbol{R}_\Omega$ the eigenvector matrix of

$$
(\Phi'\Phi)^{1/2}\Big[\frac{1}{H}\sum_{h=1}^{H}\mathbb{E}\Big(\boldsymbol{a}(y_{t+1})|y_{t+1}\in I_h\Big)\mathbb{E}\Big(\boldsymbol{a}(y_{t+1})|y_{t+1}\in I_h\Big)'\Big](\Phi'\Phi)^{1/2},
$$

we have that

$$\|\widehat{\Phi} - \boldsymbol{R}'\Phi\boldsymbol{R}^*\| = O_p(\frac{N^{1-\nu}}{T}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{N^\nu}).$$

PROOF OF PROPOSITION A.4. By definition,

$$\Sigma_{\boldsymbol{f}|\boldsymbol{y}} = \frac{1}{H}\sum_{h=1}^{H} \mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h)\mathbb{E}(\boldsymbol{f}_t'|y_{t+1} \in I_h)$$

and

$$\widehat{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}} = \frac{1}{H}\sum_{h=1}^{H}\Big[\frac{1}{c}\sum_{l=1}^{c}\widehat{\boldsymbol{f}}_{h,l}\Big]\Big[\frac{1}{c}\sum_{l=1}^{c}\widehat{\boldsymbol{f}}_{h,l}\Big]'.$$

We therefore have

$$\begin{aligned}
\widehat{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}} - \boldsymbol{R}'\Sigma_{\boldsymbol{f}|\boldsymbol{y}}\boldsymbol{R} &= \frac{1}{H}\sum_{h=1}^{H}\Big[\frac{1}{c}\sum_{l=1}^{c}\widehat{\boldsymbol{f}}_{h,l}\Big]\Big[\frac{1}{c}\sum_{l=1}^{c}\widehat{\boldsymbol{f}}_{h,l}\Big]' \\
&\quad - \frac{1}{H}\sum_{h=1}^{H}\boldsymbol{R}'\mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h)\mathbb{E}(\boldsymbol{f}_t'|y_{t+1} \in I_h)\boldsymbol{R} \\
&= \frac{1}{H}\sum_{h=1}^{H}\Big[\frac{1}{c}\sum_{l=1}^{c}\widehat{\boldsymbol{f}}_{h,l} - \boldsymbol{R}'\mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h)\Big]\Big[\frac{1}{c}\sum_{l=1}^{c}\widehat{\boldsymbol{f}}_{h,l} - \boldsymbol{R}'\mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h)\Big]' \\
&\quad + \boldsymbol{R}'\frac{1}{H}\sum_{h=1}^{H}\mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h)\Big[\frac{1}{c}\sum_{l=1}^{c}\widehat{\boldsymbol{f}}_{h,l} - \boldsymbol{R}'\mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h)\Big]' \\
&\quad + \frac{1}{H}\sum_{h=1}^{H}\Big[\frac{1}{c}\sum_{l=1}^{c}\widehat{\boldsymbol{f}}_{h,l} - \boldsymbol{R}'\mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h)\Big]\mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h)'\boldsymbol{R}.
\end{aligned}$$

By the fact that

$$\widehat{\boldsymbol{f}}_t - \boldsymbol{R}'\mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h) = (\widehat{\boldsymbol{f}}_t - \boldsymbol{R}'\boldsymbol{f}_t) + \boldsymbol{R}'\Big[\boldsymbol{f}_t - \mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in I_h)\Big],$$

we see that

$$\begin{aligned}
\|\widehat{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}} - \boldsymbol{R}'\Sigma_{\boldsymbol{f}|\boldsymbol{y}}\boldsymbol{R}\| &\leq 2\frac{1}{H}\sum_{h=1}^{H}\Big\|\frac{1}{c}\sum_{l=1}^{c}(\widehat{\boldsymbol{f}}_{h,l} - \boldsymbol{R}'\boldsymbol{f}_{h,l})\Big\|^2 \\
&\quad + 2\|\boldsymbol{R}\|^2\frac{1}{H}\sum_{h=1}^{H}\Big\|\frac{1}{c}\sum_{l=1}^{c}[\boldsymbol{f}_{h,l} - \mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in h)]\Big\|^2 \\
&\quad + 2\|\boldsymbol{R}\|\frac{1}{H}\sum_{h=1}^{H}\|E(\boldsymbol{f}_t|y_{t+1} \in I_h)\|\Big\|\frac{1}{c}\sum_{l=1}^{c}[\widehat{\boldsymbol{f}}_{h,l} - \boldsymbol{R}'\boldsymbol{f}_{h,l}]\Big\| \\
&\quad + 2\|\boldsymbol{R}\|^2\frac{1}{H}\sum_{h=1}^{H}\|E(\boldsymbol{f}_t|y_{t+1} \in I_h)\|\Big\|\frac{1}{c}\sum_{l=1}^{c}[\boldsymbol{f}_{h,l} - \mathbb{E}(\boldsymbol{f}_t|y_{t+1} \in h)]\Big\|.
\end{aligned}$$

Since H is finite, the first term is dominated by the third one, and the second term is dominated by the fourth one. The third term is $O_p(\frac{1}{\sqrt{N^\nu T}}) + O_p(\frac{1}{N^\nu}) + O_p(\frac{N^{1-\nu}}{T})$ according to Proposition A.3. The fourth term is $O_p(\frac{1}{\sqrt{T}})$ by the central limit theorem. Given this result,

$$\|\widehat{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}} - \boldsymbol{R}'\Sigma_{\boldsymbol{f}|\boldsymbol{y}}\boldsymbol{R}\| = O_p(\frac{N^{1-\nu}}{T}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{N^\nu}).$$

Note that

$$\begin{aligned}\Sigma_{\boldsymbol{f}|\boldsymbol{y}} &= \Phi \underbrace{\left[\frac{1}{H}\sum_{h=1}^{H}\mathbb{E}\Big(\boldsymbol{a}(y_{t+1})|y_{t+1}\in I_h\Big)\mathbb{E}\Big(\boldsymbol{a}(y_{t+1})|y_{t+1}\in I_h\Big)'\right]}_{\Omega_H}\Phi' \\ &= \Phi\Omega_H\Phi' = \Phi(\Phi'\Phi)^{-1/2}\Big[(\Phi'\Phi)^{1/2}\Omega_H(\Phi'\Phi)^{1/2}\Big](\Phi'\Phi)^{-1/2}\Phi'\end{aligned}$$

Let $\boldsymbol{R}_\Omega$ be the eigenvector matrix of $(\Phi'\Phi)^{1/2}\Omega_H(\Phi'\Phi)^{1/2}$ with D_Ω the associated eigenvalue diagonal matrix, that is $(\Phi'\Phi)^{1/2}\Omega_H(\Phi'\Phi)^{1/2} = \boldsymbol{R}_\Omega D_\Omega \boldsymbol{R}_\Omega'$. With this, we have

$$\Sigma_{\boldsymbol{f}|\boldsymbol{y}} = \Phi(\Phi'\Phi)^{-1/2}\boldsymbol{R}_\Omega D_\Omega \boldsymbol{R}_\Omega'(\Phi'\Phi)^{-1/2}\Phi' = \Phi\boldsymbol{R}^* D_\Omega \boldsymbol{R}^{*\prime}\Phi',$$

where $\boldsymbol{R}^* = (\Phi'\Phi)^{-1/2}\boldsymbol{R}_\Omega$. Given this, we therefore have

$$\|\widehat{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}} - \boldsymbol{R}'\Phi\boldsymbol{R}^* D_\Omega \boldsymbol{R}^{*\prime}\Phi'\boldsymbol{R}\| = O_p(\frac{N^{1-\nu}}{T}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{N^\nu}). \tag{A.3}$$

Note that

$$\begin{aligned}(\boldsymbol{R}^{*\prime}\Phi'\boldsymbol{R})(\boldsymbol{R}'\Phi\boldsymbol{R}^*) &= \boldsymbol{R}^{*\prime}\Phi'(\boldsymbol{R}\boldsymbol{R}' - I_r)\Phi\boldsymbol{R}^* + \boldsymbol{R}^{*\prime}\Phi'\Phi\boldsymbol{R}^* \\ &= \boldsymbol{R}^{*\prime}\Phi'(\boldsymbol{R}\boldsymbol{R}' - I_r)\Phi\boldsymbol{R}^* + \boldsymbol{R}_\Omega'(\Phi'\Phi)^{-1/2}\Phi'\Phi(\Phi'\Phi)^{-1/2}\boldsymbol{R}_\Omega \\ &= I_r + \boldsymbol{R}^{*\prime}\Phi'(\boldsymbol{R}\boldsymbol{R}' - I_r)\Phi\boldsymbol{R}^* = I_r + O_p(\frac{1}{N^\nu}) + O_p(\frac{N^{1-\nu}}{T}),\end{aligned}$$

where the last result is due to the second result of Proposition A.2. So the matrix $\boldsymbol{R}'\Phi\boldsymbol{R}^*$ can be asymptotically viewed as the eigenvector matrix of $\boldsymbol{R}'\Sigma_{\boldsymbol{f}|\boldsymbol{y}}\boldsymbol{R}$ since the error term, $O_p(\frac{1}{N^\nu}) + O_p(\frac{N^{1-\nu}}{T})$, is dominated by the estimation error of $\widehat{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}}$ given in (A.3). Given this, we have

$$\|\widehat{\Phi} - \boldsymbol{R}'\Phi\boldsymbol{R}^*\| = O_p(\frac{N^{1-\nu}}{T}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{N^\nu}).$$

This completes the proof. □

Proof of Theorem 1. Theorem 1 is the direct consequence of result (a) of Proposition A.2, and the last result of Proposition A.4. □

Proposition A.5. *Let $\widetilde{\boldsymbol{D}}$ be the diagonal matrix whose diagonal elements are the largest r eigenvalues of $\frac{1}{N^\nu T}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}$ with $\widetilde{\boldsymbol{X}} = (\widehat{\gamma}_i x_{it})_{N\times T}$, where $\widehat{\gamma}_i$ is the R-square of the regression of x_{it} on $\widehat{\Phi}'\widehat{\boldsymbol{f}}_t$. Under Assumptions A-D, as $N, T \to \infty$ and $N^{1-\nu}/T \to 0$,*

$$\tau_i(\widetilde{\boldsymbol{D}}) - \tau_i\Big(\frac{1}{N^\nu}\sum_{i=1}^N \gamma_i^2 \boldsymbol{b}_i\boldsymbol{b}_i'\Big) = o_p(1),$$

where $\gamma_i = \boldsymbol{b}_i' P_\Phi \boldsymbol{b}_i$ and $\tau_i(\cdot)$ denotes the i-th largest eigenvalue of its input. Furthermore, let $\widetilde{\boldsymbol{R}} = \frac{1}{N^\nu T}\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{b}_i\boldsymbol{b}_i'\boldsymbol{F}'\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{D}}^{-1}$ where $\widetilde{\boldsymbol{F}}$ is defined in Proposition A.6 below, we therefore have $\widetilde{\boldsymbol{R}} = O_p(1)$. Suppose that $\widetilde{\boldsymbol{R}}'\widetilde{\boldsymbol{R}} = I_r + o_p(1)$ as asserted in result (c) of Proposition A.6, we also have $\widetilde{\boldsymbol{R}}^{-1} = O_p(1)$.

Proof of Proposition A.5. Let $\overline{\boldsymbol{X}}$ be the matrix with its (i,t)th entry equal to $\gamma_i x_{it}$. We first show that

$$\tau_i(\widetilde{\boldsymbol{D}}) = \tau_i\Big(\frac{1}{N^\nu T}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}\Big) = \tau_i\Big(\frac{1}{N^\nu T}\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\Big) + o_p(1). \tag{A.4}$$

We next further show that

$$\tau_i\Big(\frac{1}{N^\nu T}\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\Big) = \tau_i\Big(\frac{1}{N^\nu}\sum_{i=1}^N \gamma_i^2 \boldsymbol{b}_i\boldsymbol{b}_i'\Big) + o_p(1). \tag{A.5}$$

To prove (A.4), by Weyl Theorem,

$$\Big|\tau_i\Big(\frac{1}{N^\nu T}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}\Big) - \tau_i\Big(\frac{1}{N^\nu T}\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\Big)\Big| \le \Big\|\frac{1}{N^\nu T}(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}} - \overline{\boldsymbol{X}}'\overline{\boldsymbol{X}})\Big\|_2.$$

Using the arguments in the proof of Proposition A.6, we can show that the right hand side is $o_p(1)$. Actually, the derivation here is easier since we only need to show the $o_p(1)$ result instead of giving the explicit convergence rates as in Proposition A.6. So we omit the details. Equation (A.5) can be proved in the same way as Proposition A.1. Given this, we have the first result of this proposition. The remaining two results can be proved by the same method in Corollary A.1. This completes the proof. □

Proposition A.6. *Let $\widetilde{\boldsymbol{F}}$ be the PC estimator for the scaled matrix $\widetilde{\boldsymbol{X}}$. Under the assumptions of Proposition A.1, as $N \to \infty, T \to \infty$ and $N^{1-\nu}/T \to 0$, we have*

$$(a)\quad \frac{1}{\sqrt{T}}\|\widetilde{\boldsymbol{F}} - \boldsymbol{F}\widetilde{\boldsymbol{R}}\| = O_p(\frac{1}{\sqrt{N^\nu}}) + O_p(\frac{1}{T}) + O_p(\frac{N^{1-\nu}}{T\sqrt{T}}),$$

$$(b)\quad \frac{1}{T}\|\boldsymbol{F}'(\widetilde{\boldsymbol{F}} - \boldsymbol{F}\widetilde{\boldsymbol{R}})\| = O_p(\frac{1}{N^\nu}) + O_p(\frac{1}{T}) + O_p(\frac{N^{1-\nu}}{T\sqrt{T}}),$$

$$(c)\quad \widetilde{\boldsymbol{R}}'\widetilde{\boldsymbol{R}} - I_r = O_p(\frac{1}{N^\nu}) + O_p(\frac{1}{T}) + O_p(\frac{N^{1-\nu}}{T\sqrt{T}}),$$

$$(d)\quad \frac{1}{c}\sum_{l=1}^{c}(\widetilde{\boldsymbol{f}}_{h,l}-\widetilde{\boldsymbol{R}}'\boldsymbol{f}_{h,l}) = O_p(\frac{1}{\sqrt{N^{\nu}T}}) + O_p(\frac{1}{T}) + O_p(\frac{N^{1-\nu}}{T\sqrt{T}}).$$

PROOF OF PROPOSITION A.6. (a) The proof is based on the scaled version, i.e. $\boldsymbol{X}_i'\boldsymbol{X}_i/T = 1$. For each i, we run the regression

$$x_{it} = \theta_i'\widehat{\Phi}'\widehat{\boldsymbol{f}}_t + e_{it}, \qquad \text{or} \qquad \boldsymbol{X}_i = \widehat{\boldsymbol{F}}\widehat{\Phi}\theta_i + e_i.$$

Let $\widehat{\gamma}_i$ be the R-square of the above regression. According to the definition of R-square,

$$\begin{aligned}
\widehat{\gamma}_i &= \frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i}\boldsymbol{X}_i'\widehat{\boldsymbol{F}}\widehat{\Phi}(\widehat{\Phi}'\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\widehat{\Phi})^{-1}\widehat{\Phi}'\widehat{\boldsymbol{F}}'\boldsymbol{X}_i = \frac{1}{T}\frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i}\boldsymbol{X}_i'\widehat{\boldsymbol{F}}P_{\widehat{\Phi}}\widehat{\boldsymbol{F}}'\boldsymbol{X}_i \\
&= \frac{1}{T}\frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i}\boldsymbol{X}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i + \frac{1}{T}\frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i}\boldsymbol{X}_i'\boldsymbol{F}\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i \\
&\quad + \frac{1}{T}\frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i}\boldsymbol{X}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}\boldsymbol{R}'\boldsymbol{F}'\boldsymbol{X}_i + \frac{1}{T}\frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i}\boldsymbol{X}_i'\boldsymbol{F}\Big(\boldsymbol{R}P_{\widehat{\Phi}}\boldsymbol{R}' - P_{\Phi}\Big)\boldsymbol{F}'\boldsymbol{X}_i \\
&\quad + \frac{1}{T}\frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i}\boldsymbol{X}_i'\boldsymbol{F}P_{\Phi}\boldsymbol{F}'\boldsymbol{X}_i
\end{aligned}$$

where $P_X = X(X'X)^{-1}X'$. Substituting $\boldsymbol{X}_i = \boldsymbol{F}\boldsymbol{b}_i + \boldsymbol{U}_i$ into the last term of the above expression, we have

$$\begin{aligned}
\widehat{\gamma}_i &= \frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i/T}\boldsymbol{b}_i'P_{\Phi}\boldsymbol{b}_i + \frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i}\boldsymbol{b}_i'P_{\Phi}\boldsymbol{F}'\boldsymbol{U}_i + \frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i}\boldsymbol{U}_i'\boldsymbol{F}'P_{\Phi}\boldsymbol{b}_i + \frac{1}{T}\frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i}\boldsymbol{U}_i'\boldsymbol{F}P_{\Phi}\boldsymbol{F}'\boldsymbol{U}_i \\
&\quad + \frac{1}{T}\frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i}\boldsymbol{X}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i + \frac{1}{T}\frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i}\boldsymbol{X}_i'\boldsymbol{F}\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i \\
&\quad + \frac{1}{T}\frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i}\boldsymbol{X}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}\boldsymbol{R}'\boldsymbol{F}'\boldsymbol{X}_i + \frac{1}{T}\frac{1}{\boldsymbol{X}_i'\boldsymbol{X}_i}\boldsymbol{X}_i'\boldsymbol{F}\Big(\boldsymbol{R}P_{\widehat{\Phi}}\boldsymbol{R}' - P_{\Phi}\Big)\boldsymbol{F}'\boldsymbol{X}_i
\end{aligned}$$

Since we normalize each predictor, $\boldsymbol{X}_i'\boldsymbol{X}_i = T$ for each i. Then the above expression can be simplified as

$$\begin{aligned}
\widehat{\gamma}_i &= \boldsymbol{b}_i'P_{\Phi}\boldsymbol{b}_i + 2\frac{1}{T}\boldsymbol{b}_i'P_{\Phi}\boldsymbol{F}'\boldsymbol{U}_i + \frac{1}{T^2}\boldsymbol{U}_i'\boldsymbol{F}P_{\Phi}\boldsymbol{F}'\boldsymbol{U}_i + \frac{1}{T^2}\boldsymbol{X}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i \\
&\quad + 2\frac{1}{T^2}\boldsymbol{X}_i'\boldsymbol{F}\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i + \frac{1}{T^2}\boldsymbol{X}_i'\boldsymbol{F}\Big(\boldsymbol{R}P_{\widehat{\Phi}}\boldsymbol{R}' - P_{\Phi}\Big)\boldsymbol{F}'\boldsymbol{X}_i \\
&= \gamma_i + \varepsilon_i,
\end{aligned}$$

where $\gamma_i = \boldsymbol{b}_i'P_{\Phi}\boldsymbol{b}_i$ and ε_i is the remaining expression, which consists of five terms. If $i \notin \mathcal{I}_o$, we have $\gamma_i = 0$ because of either $P_{\Phi}\boldsymbol{b}_i = 0$ for i in $\mathcal{I}_b$ or $\boldsymbol{b}_i = 0$ for i in $\mathcal{I}\backslash(\mathcal{I}_o \cup \mathcal{I}_b)$.

By definition, we have

$$\left[\frac{1}{N^\nu T}\boldsymbol{F}\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{b}_i\boldsymbol{b}_i'\boldsymbol{F}' + \frac{1}{N^\nu T}\boldsymbol{F}\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{b}_i\boldsymbol{U}_i' + \frac{1}{N^\nu T}\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{U}_i\boldsymbol{b}_i'\boldsymbol{F}' + \frac{1}{N^\nu T}\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{U}_i\boldsymbol{U}_i'\right]\widetilde{\boldsymbol{F}} = \widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{D}}.$$

Let $\widetilde{\boldsymbol{R}} = \frac{1}{N^\nu T}\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{b}_i\boldsymbol{b}_i'\boldsymbol{F}'\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{D}}^{-1}$. With this, we have

$$\begin{aligned}\frac{\|\widetilde{\boldsymbol{F}} - \boldsymbol{F}\widetilde{\boldsymbol{R}}\|}{\sqrt{T}} &= \frac{1}{N^\nu T^{3/2}}\left\|\boldsymbol{F}\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{b}_i\boldsymbol{U}_i'\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{D}}^{-1}\right\| + \frac{1}{N^\nu T^{3/2}}\left\|\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{U}_i\boldsymbol{b}_i'\boldsymbol{F}'\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{D}}^{-1}\right\| \\ &\quad + \frac{1}{N^\nu T^{3/2}}\left\|\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{U}_i\boldsymbol{U}_i'\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{D}}^{-1}\right\| = I_1 + I_2 + I_3. \end{aligned} \tag{A.6}$$

Propositions A.5 and A.6 have shown that $\|\widetilde{\boldsymbol{R}}\| = O_p(1)$ and $\|\widetilde{\boldsymbol{D}}^{-1}\| = O_p(1)$, which will be used in the following derivation. Now we investigate the above three terms one by one. Consider I_1.

$$\begin{aligned}&\frac{1}{N^\nu T^{3/2}}\left\|\boldsymbol{F}\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{b}_i\boldsymbol{U}_i'\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{D}}^{-1}\right\| \\ &\quad \le \frac{1}{N^\nu T^{3/2}}\left\|\boldsymbol{F}\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{b}_i\boldsymbol{U}_i'(\widetilde{\boldsymbol{F}} - \boldsymbol{F}\widetilde{\boldsymbol{R}})\widetilde{\boldsymbol{D}}^{-1}\right\| + \frac{1}{N^\nu T^{3/2}}\left\|\boldsymbol{F}\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\right\| \\ &\quad \le \frac{\|\boldsymbol{F}\|}{\sqrt{T}}\frac{\|\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{b}_i\boldsymbol{U}_i'\|}{N^\nu\sqrt{T}}\frac{\|\widetilde{\boldsymbol{F}} - \boldsymbol{F}\widetilde{\boldsymbol{R}}\|}{\sqrt{T}}\|\widetilde{\boldsymbol{D}}^{-1}\| + \frac{\|\boldsymbol{F}\|}{\sqrt{T}}\frac{\|\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F}\|}{N^\nu T}\|\boldsymbol{R}\|\|\widetilde{\boldsymbol{D}}^{-1}\|.\end{aligned}$$

For the first expression, by $\widehat{\gamma}_i = \gamma_i + \varepsilon_i$, we have

$$\frac{1}{N^\nu\sqrt{T}}\left\|\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{b}_i\boldsymbol{U}_i'\right\| = \frac{1}{N^\nu\sqrt{T}}\left\|\sum_{i\in\mathcal{I}_o}\gamma_i^2 \boldsymbol{b}_i\boldsymbol{U}_i'\right\| + \frac{1}{N^\nu\sqrt{T}}\left\|\sum_{i\in\mathcal{I}_o\cup\mathcal{I}_b}(2\gamma_i\varepsilon_i + \varepsilon_i^2)\boldsymbol{b}_i\boldsymbol{U}_i\right\|.$$

The first term is $O_p(\frac{1}{\sqrt{N^\nu}})$ since one can readily verify that $\mathbb{E}(\|\sum_{i\in\mathcal{I}_o}\gamma_i^2\boldsymbol{b}_i\boldsymbol{U}_i'\|^2) = O(N^\nu T)$. The second term is $o_p(1)$ by the definition of ε_i (see (A.8) below) and Proposition A.2. So the first expression is of smaller order relative to $\frac{1}{\sqrt{T}}\|\widetilde{\boldsymbol{F}} - \boldsymbol{F}\widetilde{\boldsymbol{R}}\|$, the left hand side of (A.6). So it is negligible. Consider the second expression, note that

$$\begin{aligned}\frac{1}{N^\nu T}\sum_{i=1}^N \widehat{\gamma}_i^2 \boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F} &= \frac{1}{N^\nu T}\sum_{i=1}^N(\gamma_i^2 + 2\gamma_i\varepsilon_i + \varepsilon_i^2)\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F} \\ &= \frac{1}{N^\nu T}\sum_{i\in\mathcal{I}_o}\gamma_i^2\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F} + 2\frac{1}{N^\nu T}\sum_{i\in\mathcal{I}_o}\gamma_i\varepsilon_i\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F} + \frac{1}{N^\nu T}\sum_{i\in\mathcal{I}_o\cup\mathcal{I}_b}\varepsilon_i^2\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F}\end{aligned} \tag{A.7}$$

The first term on right hand side is $O_p(\frac{1}{\sqrt{N^\nu T}})$. For the second term, we note that

$$\varepsilon_i = 2\frac{1}{T}\boldsymbol{b}_i' P_\Phi \boldsymbol{F}'\boldsymbol{U}_i + \frac{1}{T^2}\boldsymbol{U}_i'\boldsymbol{F}P_\Phi\boldsymbol{F}'\boldsymbol{U}_i + \frac{1}{T^2}\boldsymbol{X}_i'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i + 2\frac{1}{T^2}\boldsymbol{X}_i'\boldsymbol{F}\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i + \frac{1}{T^2}\boldsymbol{X}_i'\boldsymbol{F}\Big(\boldsymbol{R}P_{\widehat{\Phi}}\boldsymbol{R}' - P_\Phi\Big)\boldsymbol{F}'\boldsymbol{X}_i \tag{A.8}$$

With the above result, we have

$$\begin{aligned}\frac{1}{N^\nu T}\sum_{i\in\mathcal{I}_o}\varepsilon_i\boldsymbol{b}_i\gamma_i\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F} = {} & 2\frac{1}{N^\nu T^2}\sum_{i\in\mathcal{I}_o}\boldsymbol{b}_i'P_\Phi\boldsymbol{F}'\boldsymbol{U}_i\gamma_i\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F} \\ & + \frac{1}{N^\nu T^3}\sum_{i\in\mathcal{I}_o}\boldsymbol{U}_i'\boldsymbol{F}P_\Phi\boldsymbol{F}'\boldsymbol{U}_i\gamma_i\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F} \\ & + \frac{1}{N^\nu T^3}\sum_{i\in\mathcal{I}_o}\boldsymbol{X}_i'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i\gamma_i\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F} \\ & + 2\frac{1}{N^\nu T^3}\sum_{i\in\mathcal{I}_o}\boldsymbol{X}_i'\boldsymbol{F}\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i\gamma_i\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F} \\ & + \frac{1}{N^\nu T^3}\sum_{i\in\mathcal{I}_o}\boldsymbol{X}_i'\boldsymbol{F}\Big(\boldsymbol{R}P_{\widehat{\Phi}}\boldsymbol{R}' - P_\Phi\Big)\boldsymbol{F}'\boldsymbol{X}_i\gamma_i\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F}.\end{aligned}$$

The first term is bounded in norm by

$$\frac{1}{T}\|P_\Phi\|\frac{1}{N^\nu}\sum_{i\in\mathcal{I}_o}\frac{\|\boldsymbol{F}'\boldsymbol{U}_i\|^2}{T}\|\boldsymbol{b}_i\|^2\gamma_i = O_p(\frac{1}{T}).$$

The second term is is bounded in norm by

$$\frac{1}{T\sqrt{T}}\|P_\Phi\|\frac{1}{N^\nu}\sum_{i\in\mathcal{I}_o}\frac{\|\boldsymbol{F}'\boldsymbol{U}_i\|^3}{T\sqrt{T}}\|\boldsymbol{b}_i\|\gamma_i = O_p(\frac{1}{T\sqrt{T}}).$$

The third term can be decomposed into

$$\begin{aligned}&\frac{1}{N^\nu T^3}\sum_{i\in\mathcal{I}_o}\boldsymbol{b}_i'\boldsymbol{F}'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\gamma_i\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F} \\ &\quad + \frac{1}{N^\nu T^3}\sum_{i\in\mathcal{I}_o}\boldsymbol{b}_i'\boldsymbol{F}'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\gamma_i\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F} \\ &\quad + \frac{1}{N^\nu T^3}\sum_{i\in\mathcal{I}_o}\boldsymbol{U}_i'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\gamma_i\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F} \\ &\quad + \frac{1}{N^\nu T^3}\sum_{i\in\mathcal{I}_o}\boldsymbol{U}_i'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\gamma_i\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F}.\end{aligned}$$

By Proposition A.2, the first expression is $o_p(\frac{1}{\sqrt{N^\nu T}})$, the remaining three expressions are all $o_p(\frac{1}{\sqrt{N^\nu T}}) + o_p(\frac{1}{T})$. So the third term is $o_p(\frac{1}{\sqrt{N^\nu T}}) + o_p(\frac{1}{T})$. By the same arguments, we can show that the fourth and fifth terms are both $o_p(\frac{1}{\sqrt{N^\nu T}}) + o_p(\frac{1}{T})$. Given this, we have

$$\frac{1}{N^\nu T}\sum_{i\in\mathcal{I}_o}\varepsilon_i \boldsymbol{b}_i\gamma_i\boldsymbol{b}_i\boldsymbol{U}_i'\boldsymbol{F} = o_p(\frac{1}{\sqrt{N^\nu T}}) + O_p(\frac{1}{T}).$$

The third term on the right hand side of (A.7) can be shown to be $o_p(\frac{1}{\sqrt{N^\nu T}}) + o_p(\frac{1}{T})$ similarly as the second one and the details are omitted. With these results,

$$I_1 = O_p(\frac{1}{\sqrt{N^\nu T}}) + O_p(\frac{1}{T}).$$

Consider I_2.

$$\begin{aligned}\frac{1}{\sqrt{T}}\|I_2\| &= \frac{1}{N^\nu T^{3/2}}\Big\|\sum_{i=1}^N\widehat{\gamma}_i^2\boldsymbol{U}_i\boldsymbol{b}_i'\boldsymbol{F}'\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{D}}^{-1}\Big\| = \frac{1}{N^\nu T^{3/2}}\Big\|\sum_{i=1}^N\widehat{\gamma}_i^2\boldsymbol{U}_i\boldsymbol{b}_i'\boldsymbol{F}'\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{D}}^{-1}\Big\|\\ &\le \frac{\|\sum_{i=1}^N\widehat{\gamma}_i^2\boldsymbol{U}_i\boldsymbol{b}_i'\|}{N^\nu\sqrt{T}}\frac{\|\boldsymbol{F}'\widetilde{\boldsymbol{F}}\|}{T}\|\widetilde{\boldsymbol{D}}^{-1}\|.\end{aligned}$$

It suffices to investigate $\sum_{i=1}^N\widehat{\gamma}_i^2\boldsymbol{U}_i\boldsymbol{b}_i'/(N^\nu\sqrt{T})$, which is equal to

$$\frac{1}{N^\nu\sqrt{T}}\sum_{i\in\mathcal{I}_o}\gamma_i^2\boldsymbol{U}_i\boldsymbol{b}_i' + 2\frac{1}{N^\nu\sqrt{T}}\sum_{i\in\mathcal{I}_o}\gamma_i\varepsilon_i\boldsymbol{U}_i\boldsymbol{b}_i' + \frac{1}{N^\nu\sqrt{T}}\sum_{i\in\mathcal{I}_0\cup\mathcal{I}_b}\varepsilon_i^2\boldsymbol{U}_i\boldsymbol{b}_i'.$$

The first term is $O_p(\frac{1}{\sqrt{N^\nu}})$. For the second term, by the definition of ε_i,

$$\begin{aligned}\frac{1}{N^\nu\sqrt{T}}\sum_{i\in\mathcal{I}_o}\varepsilon_i\gamma_i\boldsymbol{U}_i\boldsymbol{b}_i' =& 2\frac{1}{N^\nu\sqrt{T}T}\sum_{i\in\mathcal{I}_o}\boldsymbol{b}_i'P_\Phi\boldsymbol{F}'\boldsymbol{U}_i\gamma_i\boldsymbol{U}_i\boldsymbol{b}_i'\\ &+\frac{1}{N^\nu\sqrt{T}T^2}\sum_{i\in\mathcal{I}_o}\boldsymbol{U}_i'\boldsymbol{F}P_\Phi\boldsymbol{F}'\boldsymbol{U}_i\gamma_i\boldsymbol{U}_i\boldsymbol{b}_i'\\ &+\frac{1}{N^\nu\sqrt{T}T^2}\sum_{i\in\mathcal{I}_o}\boldsymbol{X}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i\gamma_i\boldsymbol{U}_i\boldsymbol{b}_i'\\ &+2\frac{1}{N^\nu\sqrt{T}T^2}\sum_{i\in\mathcal{I}_o}\boldsymbol{X}_i'\boldsymbol{F}\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i\gamma_i\boldsymbol{U}_i\boldsymbol{b}_i'\\ &+\frac{1}{N^\nu\sqrt{T}T^2}\sum_{i\in\mathcal{I}_o}\boldsymbol{X}_i'\boldsymbol{F}\Big(\boldsymbol{R}P_{\widehat{\Phi}}\boldsymbol{R}'-P_\Phi\Big)\boldsymbol{F}'\boldsymbol{X}_i\gamma_i\boldsymbol{U}_i\boldsymbol{b}_i'\\ =& 2II_1 + II_2 + II_3 + 2II_4 + II_5, \text{say.}\end{aligned}$$

We show the convergence rates of $II_1, \ldots, II_5$ one by one. For II_1, its (s,l)-th element

($s = 1, \ldots, T$ and $l = 1, \ldots, r$) can be decomposed into

$$\frac{1}{N^{\nu}\sqrt{T}T}\sum_{i\in\mathcal{I}_o}\boldsymbol{b}_i' P_{\Phi}\boldsymbol{F}'[\boldsymbol{U}_i u_{is} - \mathbb{E}(\boldsymbol{U}_i u_{is})]\gamma_i b_{il} + \frac{1}{N^{\nu}\sqrt{T}T}\sum_{i\in\mathcal{I}_o}\boldsymbol{b}_i' P_{\Phi}\boldsymbol{F}'\mathbb{E}(\boldsymbol{U}_i u_{is})\gamma_i b_{il},$$

which is $O_p(\frac{1}{\sqrt{N^{\nu}T}}) + O_p(\frac{1}{T\sqrt{T}})$. So we have

$$II_1 = O_p(\frac{1}{\sqrt{N^{\nu}T}}) + O_p(\frac{1}{T}).$$

For II_2, it is bounded in norm by

$$\|II_2\| \leq \frac{1}{T}\|P_{\Phi}\|\frac{1}{N^{\nu}}\sum_{i\in\mathcal{I}_o}\frac{\|\boldsymbol{U}_i'\boldsymbol{F}\|^2}{T}\frac{\|\boldsymbol{U}_i\|}{\sqrt{T}}\|\boldsymbol{b}_i\|\gamma_i = O_p(\frac{1}{T}).$$

For II_3, it can be written as

$$\begin{aligned}
&\frac{1}{N^{\nu}\sqrt{T}T^2}\sum_{i\in\mathcal{I}_o}\boldsymbol{b}_i'\boldsymbol{F}'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\gamma_i\boldsymbol{U}_i\boldsymbol{b}_i' \\
&\quad + \frac{1}{N^{\nu}\sqrt{T}T^2}\sum_{i\in\mathcal{I}_o}\boldsymbol{U}_i'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\gamma_i\boldsymbol{U}_i\boldsymbol{b}_i' \\
&\quad + \frac{1}{N^{\nu}\sqrt{T}T^2}\sum_{i\in\mathcal{I}_o}\boldsymbol{b}_i'\boldsymbol{F}'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\gamma_i\boldsymbol{U}_i\boldsymbol{b}_i' \\
&\quad + \frac{1}{N^{\nu}\sqrt{T}T^2}\sum_{i\in\mathcal{I}_o}\boldsymbol{U}_i'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\gamma_i\boldsymbol{U}_i\boldsymbol{b}_i' \\
&= III_1 + III_2 + III_3 + III_4, \qquad \text{say.}
\end{aligned}$$

For III_1, its (s, l)-th element is equal to

$$\frac{1}{N^{\nu}\sqrt{T}T^2}\sum_{i\in\mathcal{I}_o}\boldsymbol{b}_i'\boldsymbol{F}'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\gamma_i u_{is}b_{il},$$

which is bounded in norm by

$$\frac{1}{\sqrt{N^{\nu}T}}\frac{\|\boldsymbol{F}'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\|^2}{T^2}\Big\|\frac{1}{\sqrt{N^{\nu}}}\sum_{i\in\mathcal{I}_o}\boldsymbol{b}_i\gamma_i u_{is}b_{il}\boldsymbol{b}_i'\Big\| = O_p(\frac{1}{\sqrt{N^{\nu}T}})\frac{\|\boldsymbol{F}'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\|^2}{T^2}.$$

Given this, we have

$$\|III_1\| = O_p(\frac{1}{\sqrt{N^{\nu}}})\Big[\frac{\|\boldsymbol{F}'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})\|^2}{T^2}\Big].$$

For III_2, it is bounded in norm by

$$\begin{aligned}
&\|P_{\widehat{\Phi}}\| \frac{\|(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\|}{T} \frac{1}{N^\nu} \sum_{i\in\mathcal{I}_o} \frac{\|\boldsymbol{U}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\|}{T} \frac{\|\boldsymbol{U}_i\|}{\sqrt{T}} \|\boldsymbol{b}_i\|^2\gamma_i \\
&= O_p(1)\Big[\frac{1}{N^\nu}\sum_{i\in\mathcal{I}_o} \frac{\|(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\|^2}{T^2}\Big]^{1/2} \frac{\|(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\|}{T} \\
&= o_p(1)\Big[\frac{1}{N^\nu}\sum_{i\in\mathcal{I}_o} \frac{\|(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\|^2}{T^2}\Big]^{1/2}.
\end{aligned}$$

Terms III_3 and III_4 are both

$$o_p(1)\Big[\frac{1}{N^\nu}\sum_{i\in\mathcal{I}_o} \frac{\|(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\|^2}{T^2}\Big]^{1/2},$$

which can be proved by the same arguments as in III_2. With these results, we have that

$$II_3 = o_p(\frac{1}{\sqrt{N^\nu}})\frac{\|(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\|}{T} + o_p(1)\Big[\frac{1}{N^\nu}\sum_{i\in\mathcal{I}_o} \frac{\|(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\|^2}{T^2}\Big]^{1/2}.$$

Consider II_4. It can be decomposed into

$$\begin{aligned}
&\frac{1}{N^\nu T\sqrt{T}}\sum_{i\in\mathcal{I}_o} \boldsymbol{b}_i'\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\gamma_i\boldsymbol{U}_i\boldsymbol{b}_i' \\
&\quad + \frac{1}{N^\nu T^2\sqrt{T}}\sum_{i\in\mathcal{I}_o} \boldsymbol{U}_i'\boldsymbol{F}\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\gamma_i\boldsymbol{U}_i\boldsymbol{b}_i' \\
&\quad + \frac{1}{N^\nu T\sqrt{T}}\sum_{i\in\mathcal{I}_o} \boldsymbol{b}_i'\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\gamma_i\boldsymbol{U}_i\boldsymbol{b}_i' \\
&\quad + \frac{1}{N^\nu T^2\sqrt{T}}\sum_{i\in\mathcal{I}_o} \boldsymbol{U}_i'\boldsymbol{F}\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\gamma_i\boldsymbol{U}_i\boldsymbol{b}_i' \\
&= III_5 + III_6 + III_7 + III_8, \qquad \text{say.}
\end{aligned}$$

For III_5, its (s,l)-th element is equal to

$$\frac{1}{N^\nu T\sqrt{T}}\sum_{i\in\mathcal{I}_o} \boldsymbol{b}_i'\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\gamma_i u_{is} b_{il}$$

which, by $|\mathrm{tr}(A'B)| = \|A\|\|B\|$, is bounded in norm by

$$\frac{1}{\sqrt{N^\nu T}}\|\boldsymbol{R}\|\|P_{\widehat{\Phi}}\|\frac{\|(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\|}{T}\Big\|\frac{1}{\sqrt{N^\nu}}\sum_{i\in\mathcal{I}_o}\boldsymbol{b}_i\gamma_i u_{is} b_{il}\boldsymbol{b}_i'\Big\| = O_p(\frac{1}{\sqrt{N^\nu T}})\|\frac{\|(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\|}{T}.$$

With this, we therefore have

$$\|III_5\| = O_p(\frac{1}{\sqrt{N^\nu}}) \frac{\|(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\|}{T}.$$

For III_6, its (s, l)-th element is equal to

$$\frac{1}{N^\nu T^2 \sqrt{T}} \sum_{i \in \mathcal{I}_o} \boldsymbol{U}_i' \boldsymbol{F}\boldsymbol{R} P_{\widehat{\Phi}} (\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})' \boldsymbol{F} \boldsymbol{b}_i \gamma_i u_{is} b_{il}$$

which is bounded in norm by

$$\begin{aligned}
&\frac{1}{N^\nu T \sqrt{T}} \|\boldsymbol{R}\| \|P_{\widehat{\Phi}}\| \frac{\|(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\|}{T} \Big\| \sum_{i \in \mathcal{I}_o} \boldsymbol{b}_i \gamma_i u_{is} b_{il} \boldsymbol{U}_i' \boldsymbol{F} \Big\| \\
&= \Big[O_p(\frac{1}{\sqrt{N^\nu T}}) + O_p(\frac{1}{T\sqrt{T}}) \Big] \frac{\|(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\|}{T}.
\end{aligned}$$

With this result, we therefore have

$$\|III_6\| = \Big[O_p(\frac{1}{\sqrt{N^\nu T}}) + O_p(\frac{1}{T}) \Big] \frac{\|(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\|}{T}.$$

For III_7, it is bounded in norm by

$$\|\boldsymbol{R}\| \|P_{\widehat{\Phi}}\| \frac{1}{N^\nu} \sum_{i \in \mathcal{I}_o} \frac{\|(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\|}{T} \frac{\|\boldsymbol{U}_i\|}{\sqrt{T}} \|\boldsymbol{b}_i\|^2 \gamma_i = O_p(1) \Big[\frac{1}{N^\nu} \sum_{i \in \mathcal{I}_o} \frac{\|(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\|^2}{T^2} \Big]^{1/2}.$$

For III_8, it is bounded in norm by

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \|\boldsymbol{R}\| \|P_{\widehat{\Phi}}\| \frac{1}{N^\nu} \sum_{i \in \mathcal{I}_o} \frac{\|\boldsymbol{U}_i' \boldsymbol{F}\|}{\sqrt{T}} \frac{\|(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\|}{T} \frac{\|\boldsymbol{U}_i\|}{\sqrt{T}} \|\boldsymbol{b}_i\| \gamma_i \\
&= O_p(\frac{1}{\sqrt{T}}) \Big[\frac{1}{N^\nu} \sum_{i \in \mathcal{I}_o} \frac{\|(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\|^2}{T^2} \Big]^{1/2}.
\end{aligned}$$

With the results on III_5, III_6, III_7 and III_8, we have that

$$II_4 = \Big[O_p(\frac{1}{\sqrt{N^\nu}}) + O_p(\frac{1}{T}) \Big] \frac{\|(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\|}{T} + O_p(1) \Big[\frac{1}{N^\nu} \sum_{i \in \mathcal{I}_o} \frac{\|(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\|^2}{T^2} \Big]^{1/2}.$$

Consider the fifth term. One can readily show that

$$II_5 = \Big[O_p(\frac{1}{\sqrt{N^\nu}}) + O_p(\frac{1}{T}) \Big] \|\boldsymbol{R} P_{\widehat{\Phi}} \boldsymbol{R}' - P_\Phi\|.$$

Given this, we conclude that

$$\frac{1}{N^\nu\sqrt{T}}\sum_{i\in\mathcal{I}_o}\gamma_i\varepsilon_i\boldsymbol{U}_i\boldsymbol{b}_i' = o_p(\frac{1}{\sqrt{N^\nu}}) + O_p(\frac{1}{T}) + O_p(1)\Big[\frac{1}{N^\nu}\sum_{i\in\mathcal{I}_o}\frac{\|(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\|^2}{T^2}\Big]^{1/2}.$$

The term $\frac{1}{N^\nu\sqrt{T}}\sum_{i\in\mathcal{I}_0\cup\mathcal{I}_b}\varepsilon_i^2\boldsymbol{U}_i\boldsymbol{b}_i'$ is asymptotically negligible. Given this, we therefore have

$$I_2 = O_p(\frac{1}{\sqrt{N^\nu}}) + O_p(\frac{1}{T}) + O_p(\frac{N^{1-\nu}}{T\sqrt{T}}).$$

Consider I_3. Note that

$$\begin{aligned}\frac{1}{\sqrt{T}}\|I_3\| &= \frac{1}{N^\nu T^{3/2}}\Big\|\sum_{i=1}^N\widehat{\gamma}_i^2\boldsymbol{U}_i\boldsymbol{U}_i'\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{D}}^{-1}\Big\|\\ &= \frac{1}{N^\nu T^{3/2}}\Big\|\sum_{i=1}^N\widehat{\gamma}_i^2\boldsymbol{U}_i\boldsymbol{U}_i'(\widetilde{\boldsymbol{F}}-\boldsymbol{F}\widetilde{\boldsymbol{R}})\widetilde{\boldsymbol{D}}^{-1}\Big\| + \frac{1}{N^\nu T^{3/2}}\Big\|\sum_{i=1}^N\widehat{\gamma}_i^2\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\Big\|.\end{aligned}$$

We use I_a and I_b to denote the above two expressions. Further consider I_a, which is bounded by

$$I_a \le \frac{\|\sum_{i=1}^N\widehat{\gamma}_i^2\boldsymbol{U}_i\boldsymbol{U}_i'\|_2}{N^\nu T}\frac{\|\widetilde{\boldsymbol{F}}-\boldsymbol{F}\widetilde{\boldsymbol{R}}\|}{\sqrt{T}}\|\widetilde{\boldsymbol{D}}^{-1}\| \le \Big(\max_{1\le i\le N}\widehat{\gamma}_i^2\Big)\frac{\|\sum_{i=1}^N\boldsymbol{U}_i\boldsymbol{U}_i'\|_2}{N^\nu T}\frac{\|\widetilde{\boldsymbol{F}}-\boldsymbol{F}\widetilde{\boldsymbol{R}}\|}{\sqrt{T}}\|\widetilde{\boldsymbol{D}}^{-1}\|.$$

By $\widehat{\gamma}_i = \gamma_i + \varepsilon_i$, together with the maximal inequality and $E(\|u_{it}\|^8) \le \infty$, we can readily verify $\max_{1\le i\le N}\widehat{\gamma}_i^2 = O_p(1)$. This result, together with

$$\frac{1}{N^\nu T}\|\sum_{i=1}^N\boldsymbol{U}_i\boldsymbol{U}_i'\|_2 = O_p(\frac{N^{1-\nu}}{T}) + O_p(\frac{1}{N^\nu}),$$

indicates that under $\frac{N^{1-\nu}}{T}\to 0$, term I_a is of smaller order relative to $\frac{\|\widetilde{\boldsymbol{F}}-\boldsymbol{F}\widetilde{\boldsymbol{R}}\|}{\sqrt{T}}$, and is therefore negligible.

Next consider I_b, which, by $\widehat{\gamma}_i = \gamma_i + \varepsilon_i$, is bounded by

$$\begin{aligned}I_b \le{}& \frac{1}{N^\nu T^{3/2}}\Big\|\sum_{i=1}^N\gamma_i^2\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\Big\| + 2\frac{1}{N^\nu T^{3/2}}\Big\|\sum_{i=1}^N\gamma_i\varepsilon_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\Big\|\\ &+ \frac{1}{N^\nu T^{3/2}}\Big\|\sum_{i=1}^N\varepsilon_i^2\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\Big\| = II_6 + 2II_7 + II_8 \quad \text{say.}\end{aligned}$$

For II_6, we have

$$II_6 \le \frac{1}{\sqrt{N^\nu T}}\frac{\|\sum_{i=1}^N\gamma_i^2[\boldsymbol{U}_i\boldsymbol{U}_i' - \mathbb{E}(\boldsymbol{U}_i\boldsymbol{U}_i')]\boldsymbol{F}\|}{\sqrt{N^\nu}T}\|\widetilde{\boldsymbol{R}}\|\|\widetilde{\boldsymbol{D}}^{-1}\|$$

$$+\frac{1}{T}\frac{\|\sum_{i=1}^{N}\gamma_i^2\mathbb{E}(\boldsymbol{U}_i\boldsymbol{U}_i')\|_2}{N^\nu}\frac{\|\boldsymbol{F}\|}{\sqrt{T}}\|\widetilde{\boldsymbol{R}}\|\|\widetilde{\boldsymbol{D}}^{-1}\|,$$

which is $O_p(\frac{1}{\sqrt{N^\nu T}})+O_p(\frac{1}{T})$. For II_7, note that

$$\begin{aligned}
&\frac{1}{N^\nu T^{3/2}}\sum_{i=1}^{N}\varepsilon_i\gamma_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\
&\quad=2\frac{1}{N^\nu T^{5/2}}\sum_{i=1}^{N}\boldsymbol{b}_i'P_\Phi\boldsymbol{F}'\boldsymbol{U}_i\gamma_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\
&\qquad+\frac{1}{N^\nu T^{7/2}}\sum_{i=1}^{N}\boldsymbol{U}_i'\boldsymbol{F}P_\Phi\boldsymbol{F}'\boldsymbol{U}_i\gamma_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\
&\qquad+\frac{1}{N^\nu T^{7/2}}\sum_{i=1}^{N}\boldsymbol{X}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i\gamma_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\
&\qquad+2\frac{1}{N^\nu T^{7/2}}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{F}\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i\gamma_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\
&\qquad+\frac{1}{N^\nu T^{7/2}}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{F}\Big(\boldsymbol{R}P_{\widehat{\Phi}}\boldsymbol{R}'-P_\Phi\Big)\boldsymbol{F}'\boldsymbol{X}_i\gamma_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}
\end{aligned}$$

We use $III_9,\ldots,III_{13}$ to denote the above five terms. For III_9, it is bounded in norm by

$$III_9=\frac{1}{T}\|\widetilde{\boldsymbol{R}}\|\|\widetilde{\boldsymbol{D}}^{-1}\|\|P_\Phi\|\frac{1}{N^\nu}\sum_{i\in\mathcal{I}_o\cup\mathcal{I}_b}\frac{\|\boldsymbol{U}_i'\boldsymbol{F}\|^2}{T}\frac{\|\boldsymbol{U}_i\|}{\sqrt{T}}\|\boldsymbol{b}_i\|\gamma_i=O_p(\frac{1}{T}).$$

For III_{10}, it is bounded in norm by

$$\|III_{10}\|=\frac{N^{1-\nu}}{T\sqrt{T}}\|P_\Phi\|\cdot\|\widetilde{\boldsymbol{R}}\|\cdot\|\widetilde{\boldsymbol{D}}^{-1}\|\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{\|\boldsymbol{U}_i'\boldsymbol{F}\|^3}{T\sqrt{T}}\frac{\|\boldsymbol{U}_i\|}{\sqrt{T}}\gamma_i=O_p(\frac{N^{1-\nu}}{T\sqrt{T}}).$$

For III_{11}, it can be decomposed into

$$\begin{aligned}
&\frac{1}{N^\nu T^{7/2}}\sum_{i=1}^{N}\boldsymbol{b}_i'\boldsymbol{F}'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\gamma_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\
&\quad+\frac{1}{N^\nu T^{7/2}}\sum_{i=1}^{N}\boldsymbol{U}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\gamma_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\
&\quad+\frac{1}{N^\nu T^{7/2}}\sum_{i=1}^{N}\boldsymbol{b}_i'\boldsymbol{F}'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\gamma_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\
&\quad+\frac{1}{N^\nu T^{7/2}}\sum_{i=1}^{N}\boldsymbol{U}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\gamma_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}
\end{aligned}$$

For the first term, ignore $\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}$, its (s,l)-th element is equal to

$$\frac{1}{N^{\nu}T^{7/2}}\sum_{i=1}^{N}\boldsymbol{b}_i'\boldsymbol{F}'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\gamma_i u_{is}\boldsymbol{U}_i'\boldsymbol{F}_l\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}$$

which is bounded in norm by

$$\begin{aligned}&\frac{\|\boldsymbol{F}'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\|^2}{T^2}\|P_{\widehat{\Phi}}\|\frac{1}{N^{\nu}T^{3/2}}\Big\|\sum_{i=1}^{N}\boldsymbol{b}_i\gamma_i u_{is}\boldsymbol{U}_i'\boldsymbol{F}_l\boldsymbol{b}_i'\Big\|\\&=\Big[O_p(\frac{1}{T\sqrt{T}})+O_p(\frac{1}{\sqrt{N^{\nu}}T})\Big]\frac{\|\boldsymbol{F}'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\|^2}{T^2}.\end{aligned}$$

Given this, we therefore have that the first term is

$$\Big[O_p(\frac{1}{T})+O_p(\frac{1}{\sqrt{N^{\nu}T}})\Big]\frac{\|\boldsymbol{F}'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\|^2}{T^2}=o_p(\frac{1}{T})+o_p(\frac{1}{\sqrt{N^{\nu}T}}).$$

For the remaining three terms, it is easy to show that they are

$$O_p(\frac{1}{\sqrt{T}})\Big[\frac{1}{N^{\nu}}\sum_{i\in\mathcal{I}_o\cup\mathcal{I}_b}\frac{\|\boldsymbol{U}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\|^2}{T^2}\Big]^{1/2}=o_p(\frac{1}{T})+o_p(\frac{1}{\sqrt{N^{\nu}T}}).$$

Given these results, we have

$$III_{11}=o_p(\frac{1}{T})+o_p(\frac{1}{\sqrt{N^{\nu}T}}).$$

For III_{12}, it can be decomposed into

$$\begin{aligned}&\frac{1}{N^{\nu}T^{5/2}}\sum_{i=1}^{N}\boldsymbol{b}_i'\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\gamma_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\&+\frac{1}{N^{\nu}T^{5/2}}\sum_{i=1}^{N}\boldsymbol{b}_i'\boldsymbol{F}\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\gamma_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\&+\frac{1}{N^{\nu}T^{7/2}}\sum_{i=1}^{N}\boldsymbol{U}_i'\boldsymbol{F}\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\gamma_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\&+\frac{1}{N^{\nu}T^{7/2}}\sum_{i=1}^{N}\boldsymbol{U}_i'\boldsymbol{F}\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\gamma_i\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\end{aligned}$$

By the same arguments in III_{11}, the first term is

$$\Big[O_p(\frac{1}{T})+O_p(\frac{1}{\sqrt{N^{\nu}T}})\Big]\frac{\|\boldsymbol{F}'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\|}{T}=o_p(\frac{1}{T})+o_p(\frac{1}{\sqrt{N^{\nu}T}}).$$

The remaining three terms all involves $\|\boldsymbol{U}'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\widetilde{\boldsymbol{R}})\|/T$. By result (c) of Proposition

A.2, we can readily show that they are

$$O_p(\frac{1}{\sqrt{T}})\Big[\frac{1}{N^\nu}\sum_{i\in\mathcal{I}_o\cup\mathcal{I}_b}\frac{\|\boldsymbol{U}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})\|^2}{T^2}\Big]^{1/2}=o_p(\frac{1}{T})+o_p(\frac{1}{\sqrt{N^\nu T}}).$$

With these results, we have

$$III_{12}=o_p(\frac{1}{T})+o_p(\frac{1}{\sqrt{N^\nu T}}).$$

Final consider the last term III_{13}. By the same arguments, we can readily show that

$$III_{13}=\Big[O_p(\frac{1}{\sqrt{N^\nu T}})+O_p(\frac{1}{T})\Big]\|\boldsymbol{R}P_{\widehat{\Phi}}\boldsymbol{R}'-P_\Phi\|=o_p(\frac{1}{T})+o_p(\frac{1}{\sqrt{N^\nu T}}).$$

With the results on $III_9,\ldots,III_{13}$, we have

$$II_7=o_p(\frac{1}{\sqrt{N^\nu T}})+O_p(\frac{1}{T})+O_p(\frac{N^{1-\nu}}{T\sqrt{T}}).$$

Finally, consider II_8. By the definition, ε_i^2 consists of 15 terms, where 5 terms are squared ones and 10 terms are cross-product ones with a coefficient of 2. For simplicity, we consider the five squared terms. The five terms are

$$\begin{aligned}
IV_1&=\frac{1}{N^\nu T^{3/2}}\sum_{i=1}^N\Big[\frac{1}{T}\boldsymbol{b}_i'P_\Phi\boldsymbol{F}'\boldsymbol{U}_i\Big]^2\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\
IV_2&=\frac{1}{N^\nu T^{3/2}}\sum_{i=1}^N\Big[\frac{1}{T^2}\boldsymbol{U}_i'\boldsymbol{F}P_\Phi\boldsymbol{F}'\boldsymbol{U}_i\Big]^2\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\
IV_3&=\frac{1}{N^\nu T^{3/2}}\sum_{i=1}^N\Big[\frac{1}{T^2}\boldsymbol{X}_i'(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i\Big]^2\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\
IV_4&=\frac{1}{N^\nu T^{3/2}}\sum_{i=1}^N\Big[\frac{1}{T^2}\boldsymbol{X}_i'\boldsymbol{F}\boldsymbol{R}P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{R})'\boldsymbol{X}_i\Big]^2\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}\\
IV_5&=\frac{1}{N^\nu T^{3/2}}\sum_{i=1}^N\Big[\frac{1}{T^2}\boldsymbol{X}_i'\boldsymbol{F}\Big(\boldsymbol{R}P_{\widehat{\Phi}}\boldsymbol{R}'-P_\Phi\Big)\boldsymbol{F}'\boldsymbol{X}_i\Big]^2\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}
\end{aligned}$$

For IV_1, it is bounded in norm by

$$IV_1=\|P_\Phi\|^2\|\widetilde{\boldsymbol{R}}\|\|\widetilde{\boldsymbol{D}}^{-1}\|\frac{1}{N^\nu T^{3/2}}\sum_{i\in\mathcal{I}_o\cup\mathcal{I}_b}\|\boldsymbol{b}_i\|^2\frac{\|\boldsymbol{F}'\boldsymbol{U}_i\|^3}{T\sqrt{T}}\frac{\|\boldsymbol{U}_i\|}{\sqrt{T}}=O_p(\frac{1}{T^{3/2}}).$$

For IV_2, it is bounded in norm by

$$IV_2 = \|P_\Phi\|^2 \|\widetilde{\boldsymbol{R}}\| \|\widetilde{\boldsymbol{D}}^{-1}\| \frac{1}{N^\nu T^{5/2}} \sum_{i=1}^N \frac{\|\boldsymbol{F}'\boldsymbol{U}_i\|^5}{T^2\sqrt{T}} \frac{\|\boldsymbol{U}_i\|}{\sqrt{T}} = O_p(\frac{N^{1-\nu}}{T}\frac{1}{T^{3/2}}) = o_p(\frac{1}{T^{3/2}}).$$

For IV_3, it is bounded in norm by

$$\begin{aligned}
&4\frac{1}{N^\nu T^{3/2}} \sum_{i=1}^N \Big[\frac{1}{T^2}\boldsymbol{b}_i'\boldsymbol{F}'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\Big]^2 \boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1} \\
&\quad + 4\frac{1}{N^\nu T^{3/2}} \sum_{i=1}^N \Big[\frac{1}{T^2}\boldsymbol{b}_i'\boldsymbol{F}'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\Big]^2 \boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1} \\
&\quad + 4\frac{1}{N^\nu T^{3/2}} \sum_{i=1}^N \Big[\frac{1}{T^2}\boldsymbol{U}_i'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{F}\boldsymbol{b}_i\Big]^2 \boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1} \\
&\quad + 4\frac{1}{N^\nu T^{3/2}} \sum_{i=1}^N \Big[\frac{1}{T^2}\boldsymbol{U}_i'(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})P_{\widehat{\Phi}}(\widehat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{R})'\boldsymbol{U}_i\Big]^2 \boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{F}\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{D}}^{-1}
\end{aligned}$$

The first three terms are all $o_p(\frac{1}{\sqrt{N^\nu T}}) + o_p(\frac{1}{T})$, which can be proved by the similar arguments in III_{11}. The last term is $o_p(\frac{1}{N^\nu}) + o_p(\frac{1}{T})$. Given this, we have

$$IV_3 = o_p(\frac{1}{N^\nu}) + o_p(\frac{1}{T}).$$

For IV_4 and IV_5, we can show that they are both $o_p(\frac{1}{\sqrt{N^\nu T}}) + o_p(\frac{1}{T})$ by the same arguments in III_{11}. With the above results, we have

$$II_8 = o_p(\frac{1}{N^\nu}) + o_p(\frac{1}{T}).$$

Given this, we have

$$I_3 = o_p(\frac{1}{N^\nu}) + O_p(\frac{1}{\sqrt{N^\nu T}}) + O_p(\frac{1}{T}) + O_p(\frac{N^{1-\nu}}{T\sqrt{T}}).$$

Given the results on I_1, I_2 and I_3, we have the desired result. We have (a).

Consider (b). A close examination on the proof of result (a) indicates that

$$\frac{1}{\sqrt{T}}\Big\|\widetilde{\boldsymbol{F}} - \boldsymbol{F}\widetilde{\boldsymbol{R}} - \frac{1}{N^\nu T^{3/2}} \sum_{i\in\mathcal{I}_o} \gamma_i^2 \boldsymbol{U}_i\boldsymbol{b}_i'\boldsymbol{F}'\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{D}}^{-1}\Big\| = O_p(\frac{1}{N^\nu}) + O_p(\frac{1}{T}) + O_p(\frac{N^{1-\nu}}{T\sqrt{T}}).$$

With this result, we have

$$\frac{1}{T}\|\boldsymbol{F}'(\widetilde{\boldsymbol{F}} - \boldsymbol{F}\widetilde{\boldsymbol{R}})\| \leq \frac{\|\boldsymbol{F}\|}{\sqrt{T}}\frac{1}{\sqrt{T}}\Big\|\widetilde{\boldsymbol{F}} - \boldsymbol{F}\widetilde{\boldsymbol{R}} - \frac{1}{N^\nu T^{3/2}} \sum_{i\in\mathcal{I}_o} \gamma_i^2 \boldsymbol{U}_i\boldsymbol{b}_i'\boldsymbol{F}'\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{D}}^{-1}\Big\|$$

$$+\frac{1}{T}\|\frac{1}{N^{\nu}T^{3/2}}\sum_{i\in\mathcal{I}_o}\boldsymbol{F}'\gamma_i^2\boldsymbol{U}_i\boldsymbol{b}_i'\boldsymbol{F}'\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{D}}^{-1}\|$$

The first term on the right hand side is $O_p(\frac{1}{N^{\nu}})+O_p(\frac{1}{T})+O_p(\frac{N^{1-\nu}}{T\sqrt{T}})$ and the second term is $O_p(\frac{1}{\sqrt{N^{\nu}T}})$. Given this result, we have

$$\frac{1}{T}\|\boldsymbol{F}'(\widetilde{\boldsymbol{F}}-\boldsymbol{F}\widetilde{\boldsymbol{R}})\|=O_p(\frac{1}{N^{\nu}})+O_p(\frac{1}{T})+O_p(\frac{N^{1-\nu}}{T\sqrt{T}}).$$

Result (c) is a direct result of result (b) given the arguments in the proof of result (e) of Proposition A.2.

The proof of result (d) is very similar as that of Proposition A.3. Note that result (d) is only different with Proposition A.3 in the last term on the right hand side. A close examination on where the term $O_p(\frac{N^{1-\nu}}{T})$ comes from (the first term of $I_{h,3}$) indicates that this term can indeed be sharpened into $O_p(\frac{N^{1-\nu}}{T\sqrt{T}})$. So we have (d).

This completes the whole proof. □

Proposition A.7. *Under the assumptions of Proposition A.1, we have*

$$\|\widetilde{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}}-\widetilde{\boldsymbol{R}}'\Sigma_{\boldsymbol{f}|\boldsymbol{y}}\widetilde{\boldsymbol{R}}\|=O_p(\frac{1}{N^{\nu}})+O_p(\frac{1}{\sqrt{T}}).$$

where $\widetilde{\boldsymbol{R}}$ is defined in Proposition A.6. Let $\boldsymbol{R}^*$ be defined in Proposition A.4, we have that

$$\|\widetilde{\Phi}-\widetilde{\boldsymbol{R}}'\Phi\boldsymbol{R}^*\|=O_p(\frac{1}{N^{\nu}})+O_p(\frac{1}{\sqrt{T}}).$$

Proof of Proposition A.7. By definition,

$$\Sigma_{\boldsymbol{f}|\boldsymbol{y}}=\frac{1}{H}\sum_{h=1}^{H}\mathbb{E}(\boldsymbol{f}_t|y_{t+1}\in I_h)\mathbb{E}(\boldsymbol{f}_t'|y_{t+1}\in I_h)$$

and

$$\widetilde{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}}=\frac{1}{H}\sum_{h=1}^{H}\Big[\frac{1}{c}\sum_{l=1}^{c}\widetilde{\boldsymbol{f}}_{h,l}\Big]\Big[\frac{1}{c}\sum_{l=1}^{c}\widetilde{\boldsymbol{f}}_{h,l}\Big]'.$$

We therefore have

$$\begin{aligned}\widetilde{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}}-\widetilde{\boldsymbol{R}}'\Sigma_{\boldsymbol{f}|\boldsymbol{y}}\widetilde{\boldsymbol{R}}=&\frac{1}{H}\sum_{h=1}^{H}\Big[\frac{1}{c}\sum_{l=1}^{c}\widetilde{\boldsymbol{f}}_{h,l}\Big]\Big[\frac{1}{c}\sum_{l=1}^{c}\widetilde{\boldsymbol{f}}_{h,l}\Big]'\\&-\frac{1}{H}\sum_{h=1}^{H}\widetilde{\boldsymbol{R}}'\mathbb{E}(\boldsymbol{f}_t|y_{t+1}\in I_h)\mathbb{E}(\boldsymbol{f}_t'|y_{t+1}\in I_h)\widetilde{\boldsymbol{R}}\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{H}\sum_{h=1}^{H}\left[\frac{1}{c}\sum_{l=1}^{c}\widetilde{\boldsymbol{f}}_{h,l} - \widetilde{\boldsymbol{R}}'\mathbb{E}(\boldsymbol{f}_t|y_{t+1}\in I_h)\right]\left[\frac{1}{c}\sum_{l=1}^{c}\widetilde{\boldsymbol{f}}_{h,l} - \widetilde{\boldsymbol{R}}'\mathbb{E}(\boldsymbol{f}_t|y_{t+1}\in I_h)\right]' \\
&\quad + \widetilde{\boldsymbol{R}}'\frac{1}{H}\sum_{h=1}^{H}\mathbb{E}(\boldsymbol{f}_t|y_{t+1}\in I_h)\left[\frac{1}{c}\sum_{l=1}^{c}\widetilde{\boldsymbol{f}}_{h,l} - \widetilde{\boldsymbol{R}}'\mathbb{E}(\boldsymbol{f}_t|y_{t+1}\in I_h)\right]' \\
&\quad + \frac{1}{H}\sum_{h=1}^{H}\left[\frac{1}{c}\sum_{l=1}^{c}\widetilde{\boldsymbol{f}}_{h,l} - \widetilde{\boldsymbol{R}}'\mathbb{E}(\boldsymbol{f}_t|y_{t+1}\in I_h)\right]\mathbb{E}(\boldsymbol{f}_t|y_{t+1}\in I_h)'\widetilde{\boldsymbol{R}}.
\end{aligned}
$$

By the fact that

$$
\widetilde{\boldsymbol{f}}_t - \widetilde{\boldsymbol{R}}'\mathbb{E}(\boldsymbol{f}_t|y_{t+1}\in I_h) = (\widetilde{\boldsymbol{f}}_t - \widetilde{\boldsymbol{R}}'\boldsymbol{f}_t) + \widetilde{\boldsymbol{R}}'\left[\boldsymbol{f}_t - \mathbb{E}(\boldsymbol{f}_t|y_{t+1}\in I_h)\right],
$$

we see that

$$
\begin{aligned}
\|\widetilde{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}} - \widetilde{\boldsymbol{R}}'\Sigma_{\boldsymbol{f}|\boldsymbol{y}}\widetilde{\boldsymbol{R}}\| \le\ & 2\frac{1}{H}\sum_{h=1}^{H}\left\|\frac{1}{c}\sum_{l=1}^{c}(\widetilde{\boldsymbol{f}}_{h,l} - \widetilde{\boldsymbol{R}}'\boldsymbol{f}_{h,l})\right\|^2 \\
&+ 2\|\widetilde{\boldsymbol{R}}\|^2\frac{1}{H}\sum_{h=1}^{H}\left\|\frac{1}{c}\sum_{l=1}^{c}[\boldsymbol{f}_{h,l} - \mathbb{E}(\boldsymbol{f}_t|y_{t+1}\in h)]\right\|^2 \\
&+ 2\|\widetilde{\boldsymbol{R}}\|\frac{1}{H}\sum_{h=1}^{H}\|E(\boldsymbol{f}_t|y_{t+1}\in I_h)\|\left\|\frac{1}{c}\sum_{l=1}^{c}[\widetilde{\boldsymbol{f}}_{h,l} - \widetilde{\boldsymbol{R}}'\boldsymbol{f}_{h,l}]\right\| \\
&+ 2\|\widetilde{\boldsymbol{R}}\|^2\frac{1}{H}\sum_{h=1}^{H}\|E(\boldsymbol{f}_t|y_{t+1}\in I_h)\|\left\|\frac{1}{c}\sum_{l=1}^{c}[\boldsymbol{f}_{h,l} - \mathbb{E}(\boldsymbol{f}_t|y_{t+1}\in h)]\right\|.
\end{aligned}
$$

The first term on the right hand side is dominated by the third one, and therefore is negligible. The second term is dominated by the fourth term. The third term is $O_p(\frac{1}{\sqrt{N^\nu T}}) + O_p(\frac{1}{T}) + O_p(\frac{N^{1-\nu}}{T\sqrt{T}})$ due to result (d) of Proposition A.6. The fourth term is $O_p(\frac{1}{\sqrt{T}})$. Given this, under the condition $N^{1-\nu}/T \to 0$, we therefore have

$$
\|\widetilde{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}} - \widetilde{\boldsymbol{R}}'\Sigma_{\boldsymbol{f}|\boldsymbol{y}}\widetilde{\boldsymbol{R}}\| = O_p(\frac{1}{N^\nu}) + O_p(\frac{1}{\sqrt{T}}).
$$

As shown in the proof of Proposition A.4, $\Sigma_{\boldsymbol{f}|\boldsymbol{y}} = \Phi\boldsymbol{R}^* D_\Omega \boldsymbol{R}^{*\prime}\Phi'$, where $\boldsymbol{R}^*$ and D_Ω are defined in Proposition A.4. Given this, we therefore have

$$
\|\widetilde{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}} - \widetilde{\boldsymbol{R}}'\Phi\boldsymbol{R}^* D_\Omega \boldsymbol{R}^{*\prime}\Phi'\widetilde{\boldsymbol{R}}\| = O_p(\frac{1}{N^\nu}) + O_p(\frac{1}{\sqrt{T}}). \tag{A.9}
$$

Note that

$$
\begin{aligned}
(\boldsymbol{R}^{*\prime}\Phi'\widetilde{\boldsymbol{R}})(\widetilde{\boldsymbol{R}}'\Phi\boldsymbol{R}^*) &= \boldsymbol{R}^{*\prime}\Phi'(\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{R}}' - I_r)\Phi\boldsymbol{R}^* + \boldsymbol{R}^{*\prime}\Phi'\Phi\boldsymbol{R}^* \\
&= \boldsymbol{R}^{*\prime}\Phi'(\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{R}}' - I_r)\Phi\boldsymbol{R}^* + \boldsymbol{R}_\Omega'(\Phi'\Phi)^{-1/2}\Phi'\Phi(\Phi'\Phi)^{-1/2}\boldsymbol{R}_\Omega
\end{aligned}
$$

$$= I_r + \boldsymbol{R}^{*\prime}\Phi'(\widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{R}}' - I_r)\Phi\boldsymbol{R}^* = I_r + O_p(\frac{1}{N^\nu}) + O_p(\frac{1}{T}) + O_p(\frac{N^{1-\nu}}{T\sqrt{T}}),$$

where the last result is due to result (c) of Proposition A.6. So the matrix $\widetilde{\boldsymbol{R}}'\Phi\boldsymbol{R}^*$ can be asymptotically viewed as the eigenvector matrix of $\widetilde{\boldsymbol{R}}'\Sigma_{\boldsymbol{f}|\boldsymbol{y}}\widetilde{\boldsymbol{R}}$ since the error term, $O_p(\frac{1}{N^\nu}) + O_p(\frac{1}{T}) + O_p(\frac{N^{1-\nu}}{T}\sqrt{T})$, is dominated by the estimation error of $\widetilde{\Sigma}_{\boldsymbol{f}|\boldsymbol{y}}$ given in (A.9). Given this, we have

$$\|\widetilde{\Phi} - \widetilde{\boldsymbol{R}}'\Phi\boldsymbol{R}^*\| = O_p(\frac{1}{N^\nu}) + O_p(\frac{1}{\sqrt{T}}).$$

This completes the proof. □

Proof of Theorem 2. Theorem 2 is a direct result of Propositions A.6 and A.7. □

B. Real-time Vintage Data

This section provides a detailed description of the real-time macro variables we use in this paper. We first list the panel of 60 individual macroeconomic variables obtained from the Archival Federal Reserve Economic database (ALFRED). For each variable, we report the ALFRED mnemonics, a full variable description, and the transformation code (trcode) used to ensure stationarity of the underlying data series. The transformation codes generally follow the corresponding ones in the FRED-MD database as discussed in McCracken and Ng (2016), which are based on the same data source. The particular forms of the transformations are specified below. To fix notation, let $x_{i,t}^{\text{raw}}$ and $x_{i,t}^{\text{tr}}$ denote the raw and transformed version of the ith variable observed at time t, respectively, and let $\Delta = (1 - L)$, with a lag operator $Lx_{i,t}^{\text{raw}} = x_{i,t-1}^{\text{raw}}$. We then apply one of six possible transformations:

1. lvl: $x_{i,t}^{\text{tr}} = x_{i,t}^{\text{raw}}$
2. Δlvl: $x_{i,t}^{\text{tr}} = x_{i,t}^{\text{raw}} - x_{i,t-1}^{\text{raw}}$
3. Δ^2lvl: $x_{i,t}^{\text{tr}} = \Delta^2 x_{i,t}^{\text{raw}}$
4. ln: $x_{i,t}^{\text{tr}} = \ln\left(x_{i,t}^{\text{raw}}\right)$
5. Δln: $x_{i,t}^{\text{tr}} = \ln\left(x_{i,t}^{\text{raw}}\right) - \ln\left(x_{i,t-1}^{\text{raw}}\right)$
6. Δ^2ln: $x_{i,t}^{\text{tr}} = \Delta^2 \ln\left(x_{i,t}^{\text{raw}}\right)$

No.	Mnemonic	Variable description	trcode
1	INDPRO	Industrial Production Index	5
2	AWHMAN	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing	1
3	AWHNONAG	Average Weekly Hours Of Production And Nonsupervisory Employees: Total private	2
4	AWOTMAN	Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing	2
5	DSPI	Disposable Personal Income	5
6	DSPIC96	Real Disposable Personal Income	5
7	PI	Personal Income	5
8	CE16OV	Civilian Employment	5
9	CLF16OV	Civilian Labor Force	5
10	PAYEMS	All Employees: Total nonfarm	5
11	MANEMP	All Employees: Manufacturing	5
12	DMANEMP	All Employees: Durable Goods	5
13	NDMANEMP	All Employees: Nondurable goods	5
14	USCONS	All Employees: Construction	5
15	USFIRE	All Employees: Financial Activities	5
16	USGOOD	All Employees: Goods-Producing Industries	5
17	USGOVT	All Employees: Government	5
18	USMINE	All Employees: Mining and logging	5
19	USPRIV	All Employees: Total Private Industries	5
20	USSERV	All Employees: Other Services	5
21	USTPU	All Employees: Trade, Transportation & Utilities	5
22	USTRADE	All Employees: Retail Trade	5
23	USWTRADE	All Employees: Wholesale Trade	5
24	SRVPRD	All Employees: Service-Providing Industries	5
25	UEMP5TO14	Number of Civilians Unemployed for 5 to 14 Weeks	5
26	UEMP15OV	Number of Civilians Unemployed for 15 Weeks & Over	5
27	UEMP15T26	Number of Civilians Unemployed for 15 to 26 Weeks	5
28	UEMP27OV	Number of Civilians Unemployed for 27 Weeks and Over	5
29	UEMPLT5	Number of Civilians Unemployed - Less Than 5 Weeks	5
30	UEMPMEAN	Average (Mean) Duration of Unemployment	2
31	UEMPMED	Median Duration of Unemployment	2
32	UNEMPLOY	Unemployed	5
33	UNRATE	Civilian Unemployment Rate	2
34	PCE	Personal Consumption Expenditures	5
35	PCEDG	Personal Consumption Expenditures: Durable Goods	5
36	PCEND	Personal Consumption Expenditures: Nondurable Goods	5
37	PCES	Personal Consumption Expenditures: Services	5
38	HOUST	Housing Starts: Total: New Privately Owned Housing Units Started	4
39	HOUST1F	Privately Owned Housing Starts: 1-Unit Structures	4
40	HOUST2F	Housing Starts: 2-4 Units	4
41	CURRSL	Currency Component of M1	5
42	DEMDEPSL	Demand Deposits at Commercial Banks	6
43	M1SL	M1 Money Stock	6
44	OCDSL	Other Checkable Deposits	6
45	SAVINGSL	Savings Deposits - Total	6
46	STDCBSL	Small Time Deposits at Commercial Banks	6
47	STDSL	Small Time Deposits - Total	6
48	STDTI	Small Time Deposits at Thrift Institutions	6
49	SVGCBSL	Savings Deposits at Commercial Banks	6
50	SVGTI	Savings Deposits at Thrift Institutions	6
51	SVSTCBSL	Savings and Small Time Deposits at Commercial Banks	6
52	TCDSL	Total Checkable Deposits	6
53	CPIAUCSL	Consumer Price Index for All Urban Consumers: All Items	6
54	PFCGEF	Producer Price Index: Finished Consumer Goods Excluding Foods	6
55	PPICPE	Producer Price Index: Finished Goods: Capital Equipment	6
56	PPICRM	Producer Price Index: Crude Materials for Further Processing	6
57	PPIFCF	Producer Price Index: Finished Consumer Foods	6
58	PPIFGS	Producer Price Index: Finished Goods	6
59	PPIIFF	Producer Price Index: Intermediate Foods & Feeds	6
60	PPIITM	Producer Price Index: Intermediate Materials: Supplies & Components	6