cemmap
centre for microdata methods and practice

Beta-sorted portfolios

Matias D. Cattaneo
Richard K. Crump
Weining Wang

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP18/23

Economic and Social Research Council

Beta-Sorted Portfolios*

Matias D. Cattaneo[†] Richard K. Crump[‡] Weining Wang[§]

August 1, 2023

Abstract

Beta-sorted portfolios—portfolios comprised of assets with similar covariation to selected risk factors—are a popular tool in empirical finance to analyze models of (conditional) expected returns. Despite their widespread use, little is known of their statistical properties in contrast to comparable procedures such as two-pass regressions. We formally investigate the properties of beta-sorted portfolio returns by casting the procedure as a two-step nonparametric estimator with a nonparametric first step and a beta-adaptive portfolios construction. Our framework rationalizes the well-known estimation algorithm with precise economic and statistical assumptions on the general data generating process. We provide conditions that ensure consistency and asymptotic normality along with new uniform inference procedures allowing for uncertainty quantification and general hypothesis testing for financial applications. We show that the rate of convergence of the estimator is non-uniform and depends on the beta value of interest. We also show that the widely-used Fama-MacBeth variance estimator is asymptotically valid but is conservative in general, and can be very conservative in empirically-relevant settings. We propose a new variance estimator which is always consistent and provide an empirical implementation which produces valid inference. In our empirical application we introduce a novel risk factor – a measure of the business credit cycle – and show that it is strongly predictive of both the cross-section and time-series behavior of U.S. stock returns.

Keywords: Beta pricing models, portfolio sorting, nonparametric estimation, partitioning, kernel regression, smoothly-varying coefficients.

---

*The authors would like to thank Nina Boyarchenko, Fernando Duarte, and Olivier Scaillet and participants at various seminars, workshops, and conferences for helpful comments and discussions. The views expressed in this paper are those of the authors and do not necessarily represent those of the Federal Reserve Bank of New York or the Federal Reserve System. Cattaneo gratefully acknowledges financial support from the National Science Foundation (SES-1947662 and SES-2241575).

[†]Department of Operations Research and Financial Engineering, Princeton University.

[‡]Macrofinance Studies, Federal Reserve Bank of New York.

[§]Department of Economics and Related Studies, University of York,Weining Wang's research is partially supported by the ESRC (Grant Reference: ES/T01573X/1)

1 Introduction

Deconstructing expected returns into idiosyncratic factor loadings and corresponding prices of risk for interpretable factors is an evergreen pursuit in the empirical finance literature. When factors are observable, there are two workhorse approaches that continue to enjoy widespread use. The first approach, Fama-MacBeth two-pass regressions, have been extensively studied in the financial econometrics literature.[1] The second approach, which we refer to as beta-sorted portfolios, has received scant attention in the econometrics literature despite its empirical popularity.[2]

Beta-sorted portfolios are commonly characterized by the following two-step procedure, which incorporates beta-adaptive portfolios construction. In a first step, time-varying risk factor exposures are estimated through (backwards-looking) weighted time-series regressions of asset returns on the observed factors. The most popular implementation uses rolling window regressions, often with a choice of a five-year window. In a second step, the estimated factor exposures, based on data up to the previous period, are ordered and used to group assets into portfolios. These portfolios then represent assets with a similar degree of exposure to the risk factors, and the degree of return differential for differently exposed assets is used to assess the compensation for bearing this common risk. Most frequently this is achieved by differencing the portfolio returns from the two most extreme portfolios. Finally, an average over time of these return differentials is taken to infer whether the risk is priced unconditionally—whether the portfolio earns systematic (and significant) excess returns. Notwithstanding the simple and intuitive nature of the methodology, little is known of the formal properties of this estimator and its associated inference procedures.

We provide a comprehensive framework to study the economic and statistical properties of beta-sorted portfolios. We first translate the two-step estimation algorithm with beta-adaptive portfolio construction into a corresponding statistical model. We show that the model has key features which are important to consider for valid interpretation of the empirical results. For example, in

[1] See, for example, Jagannathan and Wang (1998), Chen and Kan (2004), Shanken and Zhou (2007), Kleibergen (2009), Ang, Liu, and Schwarz (2020), Gospodinov, Kan, and Robotti (2014), Adrian, Crump, and Moench (2015), Bai and Zhou (2015), Bryzgalova (2015), Gagliardini, Ossola, and Scaillet (2016), Chordia, Goyal, and Shanken (2017), Kleibergen, Lingwei, and Zhan (2019), Raponi, Robotti, and Zaffaroni (2020), Giglio and Xiu (2021) and many others. For a recent survey, see Gagliardini, Ossola, and Scaillet (2020).

[2] The empirical literature using beta-sorted portfolios is extensive. For a textbook treatment, see Bali, Engle, and Murray (2016), and for a few recent papers see, for example, Boons, Duarte, De Roon, and Szymanowska (2020), Chen, Han, and Pan (2021), Eisdorfer, Froot, Ozik, and Sadka (2021), Goldberg and Nozawa (2021), and Fan, Londono, and Xiao (2022).

this setup, no-arbitrage conditions are not imposed and instead imply testable hypotheses. Within this framework, we introduce general sampling assumptions allowing for smoothly-varying factor loadings, persistent (possibly nonstationary) factors, and conditional heteroskedasticity across time and assets. We then study the asymptotic properties of the beta-sorted portfolio estimator and associated test statistics in settings with large cross-sectional and time-series sample sizes (i.e., $n, T \to \infty$).

We provide a host of new methodological and theoretical results. First, we introduce conditions that ensure consistency and asymptotic normality of the full-sample estimator of average expected returns. Importantly, we characterize precise conditions on the bandwidth sequence of the first-stage kernel regression estimator, h, and the number of portfolios, J, relative to growth in n and T. We show that the rate of convergence of the estimator depends on the value of beta that is chosen. For beta values closer to zero the rate of convergence is faster and is slower otherwise; in fact, for values of beta away from zero we show that the rate of convergence of the estimator is only $\sqrt{T}$, despite an effective sample size of the order nT, reflecting specific properties of the setting of interest. However, we also show that certain features of average expected returns such as the discrete second derivative—which represents a butterfly spread trade—can be estimated with higher precision through faster rates of convergence for all values of beta, namely, $\sqrt{nT/J}$ for a single risk factor. This result also accommodates more powerful tests for testing the null hypothesis of no-arbitrage. Finally, we also provide novel results on uniform inference for the beta-sorted portfolio estimator for both a single period and the grand mean. This facilitates the construction of uniform confidence bands which allows for inference on a variety of hypotheses of interest such as monotonicity or inference on maximum-return trading strategies.

We also uncover some limitations of current empirical practice employing beta-sorted portfolios methodology. First, as with all nonparametric estimators, the choice of tuning parameters, h and J, are key to successful performance and are dependent on the sample sizes n and T. In contrast, empirical practice often chooses window length in the first step and total portfolios in the second step irrespective of the sample size at hand. Second, we show that the widely-used Fama and MacBeth (1973) variance estimator, is not consistent in general but only when conditional expected returns are constant over time for a fixed beta. However, we show that the Fama-MacBeth variance estimator still leads to valid, albeit possibly conservative, inferences. Unfortunately, in empirically-relevant settings it appears that the Fama-MacBeth variance estimator can be very conservative. To address this limitation, we propose a new variance estimator which is always consistent and provide an empirical implementation which produces valid inference. In our empirical application we show that our new variance estimator provides much sharper inference than the Fama-MacBeth variance estimator. We also show that differential returns for a single time period, often used as inputs for assessing the time-series properties of conditional expected returns, are contaminated by an additional term when risk factors are serially correlated.

From a theoretical perspective, beta-sorted portfolios present a number of technical challenges originating from the two-step estimation algorithm with beta-adaptive portfolio construction, since it relies on two nested nonparametric estimation steps together with a portfolio construction based on a first-step nonparametric generated regressor. More precisely, the first-stage nonparametrically estimated factor loadings enter directly into the (non-smooth) partitioning scheme further complicating the analysis.[3] To our knowledge, we are the first to prove validity of such an approach.

This paper is most related to the large literature studying asset pricing models with observable factors.[4] Given our focus on conditional asset pricing models with large panels in both the cross-section and time-series dimension, this paper is most closely related to Gagliardini, Ossola, and Scaillet (2016) (see also Gagliardini, Ossola, and Scaillet, 2020). Gagliardini, Ossola, and Scaillet (2016) introduce a general framework and econometric methodology for inference in large-dimensional conditional factors under no-arbitrage restrictions. They allow for risk exposures, which are parametric functions of observable variables and provide conditions to consistently estimate, and conduct inference on the prices of risk. Although the statistical model under study shares important similarities with the setup of Gagliardini, Ossola, and Scaillet (2016), there are substantial differences, and the models explored previously in the literature do not nest our setup. For example, the classical beta-sorted portfolio estimator implies a data-generating process that does not (nec-

[3]For analysis of partitioning-based nonparametric estimators see Cattaneo, Farrell, and Feng (2020) and references therein. Partitioning-based estimators with random basis functions have been recently studied in Cattaneo, Crump, Farrell, and Schaumburg (2020) and Cattaneo, Crump, Farrell, and Feng (2022), but in those papers the conditing variables are observed, while here the conditioning variable is generated using a preliminary time-series smoothly-varying coefficients nonparametric regression, and therefore prior results are not applicable to the settings considered herein.

[4]See, for example, Goyal (2012), Nagel (2013), Gospodinov and Robotti (2013), or Gagliardini, Ossola, and Scaillet (2020) for surveys. A related literature endeavors to jointly estimate factor loadings and *latent* risk factors. See, for example, Connor and Linton (2007), Connor, Hagmann, and Linton (2012), Fan, Liao, and Wang (2016), Kelly, Pruitt, and Su (2019), Connor, Li, and Linton (2021), and Fan, Ke, Liao, and Neuhierl (2022), among others.

essarily) exclude arbitrage opportunities and supposes risk exposures which are smoothly-varying. Furthermore, we show that valid estimation and inference can be achieved without requiring an assumption of the functional form of the conditional expectation of the risk factors. See Section 2 for more details.

Our paper is also related to the financial econometrics literature on nonparametric estimation and inference. In particular, the two steps of the beta-sorted portfolio algorithm align individually with Ang and Kristensen (2012), who study kernel regression estimators of time-varying alphas and betas, and Cattaneo, Crump, Farrell, and Schaumburg (2020) who study portfolio sorting estimators given observed individual characteristic variables. However, the linkage between the two steps, including the role of the generated (nonparametrically estimated) regressor in the second-stage nonparametric partitioning estimator has not been studied before. Finally, our paper is also related to Raponi, Robotti, and Zaffaroni (2020) who study estimation and inference of the ex-post risk premia. In analogy, we show that estimation and inference in our general setting are sensitive to the specific object of interest chosen. For example, we show that a faster convergence rate of the estimator can be obtained by centering at realized systematic returns rather than conditional expected returns. See Section 4 for more details.

In our empirical application we introduce a novel risk factor – a measure of the business credit cycle – and show that it is strongly predictive of both the cross-section and time-series behavior of U.S. stock returns. Moreover, we show the effectiveness of our new variance estimator as inference is much sharper relative to the Fama-MacBeth variance estimator.

This paper is organized as follows. In Section 2, we introduce our general data-generating process and show how it rationalizes the two-step algorithm used to construct beta-sorted portfolios. In Section 3, we study the theoretical properties of the first-step estimators of the time-varying risk factor exposures. Using these results, in Section 4 we establish the theoretical properties of the second-step nonparametric estimator. To facilitate feasible inference, Section 5 introduces pointwise and uniform inference procedures for the grand-mean estimator including characterizing the properties of the commonly-used Fama-MacBeth variance estimator. Section 6 presents an empirical application, and Section 7 concludes. Detailed assumptions and proofs of the results are relegated to a Supplemental Appendix (hereafter, SA).

Notation and conventions

It is useful to introduce the following notation. For a constant $k \in N$ and a vector $v = (v_1, \ldots, v_d)^\top \in \mathbb{R}^d$, we denote $|v|_k = (\sum_{i=1}^d |v_i|^k)^{1/k}$, $|v| = |v|_2$ and $|v|_\infty = \max_{i \le d} |v_i|$. For a matrix $A = (a_{ij})_{1 \le i \le m, 1 \le j \le n}$, we define the spectral norm $|A|_2 = \max_{|v|=1} |Av|$, the max norm $|A|_{\max} = \max_{1 \le i \le m, 1 \le j \le n} |a_{i,j}|$, $|A|_1 = \max_{1 \le j \le n} \sum_{i=1}^m a_{i,j}$, and $|A|_\infty = \max_{1 \le i \le m} \sum_{j=1}^n a_{i,j}$. For a function f, we denote $|f|_\infty = \sup_{x \in \mathcal{X}} |f(x)|$, where $\mathcal{X}$ denotes the support. We set $(a_n : n \ge 1)$ and $(b_n : n \ge 1)$ to be positive number sequences. We write $a_n = O(b_n)$ or $a_n \lesssim b_n$ (resp. $a_n \asymp b_n$) if there exists a positive constant C such that $a_n/b_n \le C$ (resp. $1/C \le a_n/b_n \le C$) for all large n, and we denote $a_n = o(b_n)$ (resp. $a_n \sim b_n$), if $a_n/b_n \to 0$ (resp. $a_n/b_n \to C$). Limits are taken as $n, T \to \infty$ unless otherwise stated explicitly. $\text{plim} X_n = X$ means that $X_n \to_{\mathbb{P}} X$. $\to_{\mathcal{L}}$ denotes convergence in law. Define $X_n = O_{\mathbb{P}}(a_n) : \lim_{n\to\infty} \mathbb{P}(|X_n| \ge \delta_\varepsilon a_n) \to 0 \quad \forall \varepsilon > 0$. Define $X_n = o_{\mathbb{P}}(a_n) : \forall \varepsilon, \delta > 0 \quad \exists N_{\varepsilon,\delta}$ such that $\mathbb{P}(|X_n| \ge \delta a_n) \le \varepsilon \quad \forall n > N_{\varepsilon,\delta}$. Let $X_n \lesssim_{\mathbb{P}} a_n$ means $X_n = O_{\mathbb{P}}(a_n)$.

2 Model setup

We introduce a general statistical model of asset returns and show how the proposed model naturally aligns with the two steps that comprise the beta-sorted portfolio algorithm. We discuss the relevant properties of the model especially with respect to the potential presence of arbitrage opportunities.

2.1 Modeling returns

Let R_{it} denote the return of asset i at time t.[5] We assume that asset returns are generated by the linear stochastic coefficient model,

$$R_{it} = \alpha_{it} + \beta_{it}^\top f_t + \varepsilon_{it}, \qquad i = 1, \cdots, n_t, \qquad t = 1, \cdots, T, \tag{2.1}$$

where $\alpha_{it} \in \mathbb{R}$ and $\beta_{it} \in \mathbb{R}^d$ $(d \ge 1)$ are random coefficients which are measurable to a filtration based on the past information, f_t is a vector of observable risk factors, and ε_{it} is an idiosyncratic

[5] Throughout we will assume that R_{it} represent excess returns. In the case when R_{it} represent raw returns then $\mu_t(0)$ may be interpreted as the zero-beta rate at time t.

error term.[6] We allow for an unbalanced panel, but assume that $n \leq n_t \leq n_u$ and $n \asymp n_u$, so that each cross-section contributes to the asymptotic properties of the estimator.

We define the filtration $\mathcal{F}_{n,T,t-1} = \sigma((\alpha_{it})_{i=1,t=1}^{n_t,t}, (\beta_{it})_{i=1,t=1}^{n_t,t}, (f_t)_{t=1}^{t-1}, (\varepsilon_{it})_{i=1,t=1}^{n_t,t-1})$. Hereafter, we suppress the n and T as in $\mathcal{F}_{n,T,t-1}$ and denote it as $\mathcal{F}_{t-1}$ for simplicity of notation. We define another cross-sectional invariant filtration $\mathcal{G}_{t-1}$. Suppose that $\beta_{it} = G_\beta(\eta_i, g_1, \cdots, g_{t-1}, f_1, \cdots, f_{t-1}, \omega_{it})$, where η_i is independent and identically distributed (i.i.d.) over i, g_t are i.i.d. factors over t, and ω_{it} are i.i.d. over t and i. Then, the cross-section invariant sigma field is $\mathcal{G}_t = \sigma(f_1, \cdots, f_t, g_1, \cdots, g_t)$. This setup may appear restrictive but is in fact general: we can always increase the dimension of the random variables entering the sigma field to accommodate more complex designs. Consequently, without loss of generality, we assume $\mathbb{E}(f_t|\mathcal{G}_{t-1}) = \mathbb{E}(f_t|\mathcal{F}_{t-1})$.

To obtain the structural form of our model, we denote $\mu_t(\beta)$ as the conditional expected return of an asset with risk exposure β. Thus,

$$\mathbb{E}(R_{it}|\mathcal{F}_{t-1}) = \mu_t(\beta_{it}), \tag{2.2}$$

so that using equation (2.1) we have,

$$\mu_t(\beta_{it}) = \alpha_{it} + \beta_{it}^\top \mathbb{E}(f_t|\mathcal{F}_{t-1}). \tag{2.3}$$

Finally, combining equations (2.1) and (2.3), we obtain the structural form

$$R_{it} = \mu_t(\beta_{it}) + \beta_{it}^\top (f_t - \mathbb{E}[f_t|\mathcal{F}_{t-1}]) + \varepsilon_{it}. \tag{2.4}$$

To distinguish conditional expected returns, $\mu_t(\beta_{it})$, from systematic realized returns, we define

$$M_t(\beta_{it}) = \mu_t(\beta_{it}) + \beta_{it}^\top (f_t - \mathbb{E}[f_t|\mathcal{F}_{t-1}]) = \alpha_{it} + \beta_{it}^\top f_t$$

to represent the latter object.

Equation (2.4) may be compared to the standard beta pricing model (e.g., Cochrane, 2005, Chapter 12) and generalizations thereof (e.g., Cochrane, 1996; Adrian, Crump, and Moench, 2015;

[6] For an alternative example of a random coefficient model tailored to a financial application, see Barras, Gagliardini, and Scaillet (2022).

Gagliardini, Ossola, and Scaillet, 2016). The most noteworthy difference between equation (2.4) is the presence of the (possibly) nonlinear, time-varying function $\mu_t(\beta_{it})$. When R_{it} represent excess returns then the no-arbitrage restriction implies that $\mu_t(\beta_{it}) = \beta_{it}^\top \lambda_t$ for some λ_t (Gagliardini, Ossola, and Scaillet, 2016). Our model nests, but does not require, the imposition of the absence of arbitrage opportunities so that

$$
\begin{aligned}
R_{it} &= \mu_t(\beta_{it}) + \beta_{it}^\top (f_t - \mathbb{E}[f_t|\mathcal{F}_{t-1}]) + \varepsilon_{it}, \\
&= \underbrace{(\mu_t(\beta_{it}) - \beta_{it}^\top \lambda_t)}_{\text{deviation from no-arbitrage}} + \beta_{it}^\top \lambda_t + \beta_{it}^\top (f_t - \mathbb{E}[f_t|\mathcal{F}_{t-1}]) + \varepsilon_{it}.
\end{aligned}
$$

The presence of this additional term representing the deviation from no-arbitrage restrictions can be motivated by appealing to structural models which feature violations of the law of one price. Such a setup as in equation (2.4) could arise, for example, in the margin-constraints model of Garleanu and Pedersen (2011) under the assumption that the security's margin is a nonlinear function of its past beta.

To see why equation (2.4) rationalizes the beta-sorted portfolio algorithm, consider the two steps in the case when $d = 1$.

Step 1: Estimation of α_{it} and β_{it}. For each individual asset, we calculate the local constant estimator for α_{it} and β_{it} as,

$$
\left(\widehat{\alpha}_{it_0}, \widehat{\beta}_{it_0}\right)^\top = \Big(\sum_{t=1}^{t_0-1} K((t-t_0)/(Th)) X_t X_t^\top\Big)^{-1} \Big(\sum_{t=1}^{t_0-1} K((t-t_0)/(Th)) X_t R_{it}\Big), \tag{2.5}
$$

where $X_t = (1, f_t)^\top$, $K(\cdot)$ is a kernel function and h a positive bandwidth determining the length of the rolling window. This construction purposely does not have "look-ahead bias"; moreover, the estimators $\widehat{\alpha}_{it_0}$ and $\widehat{\beta}_{it_0}$ do not use data from time t_0 in their construction (a "leave-one-out" estimator). This estimation of the time-varying random coefficients can be interpreted as a kernel regression of equation (2.1) for each cross-section unit. When $K(\cdot)$ takes on a constant value for the most recent prior H time periods, and zero otherwise, we obtain the familiar rolling window regression estimator with window size H. ■

Step 2: Sorting portfolios using estimated β_{it}. To see that this comprises cross-sectional nonparametric estimation observe that, for fixed t, Equation (2.2) is the conditional mean of interest.[7] We define $\mathcal{B}=[\beta_l, \beta_u]$ as the support of the possible realizations of β_{it} across i and t. For each $t = 1, \ldots, T$, let us define a beta-adaptive partition of this support as

$$\widehat{P}_{jt} = [\widehat{\beta}_{(\lfloor n_t(j-1)/J_t \rfloor)t}, \widehat{\beta}_{(\lfloor n_t j/J_t \rfloor)t}), \qquad j = 1, \ldots J_t - 1$$
$$\widehat{P}_{jt} = [\widehat{\beta}_{(\lfloor n_t(J-1)/J \rfloor)t}, \widehat{\beta}_{(n_t)t}], \qquad j = J_t,$$

where $\lfloor . \rfloor$ denotes the floor function and $\widehat{\beta}_{(\ell)t}$ denotes the ℓth order statistic of the estimated betas in the first step across i for fixed t, i.e., the order statistics of $\{\widehat{\beta}_{it} : i = 1, \ldots, n_t\}$. The number of portfolios J_t, and their random structure (i.e., break-point positions based on estimated β_{it}), vary for each time period. Then, define

$$\widehat{p}_{jt}(\beta) = \mathbf{1}\{\beta \in \widehat{P}_{jt}\},$$

with $\mathbf{1}\{\cdot\}$ the indicator function, and $\widehat{\Phi}_t = [\widehat{\Phi}_{i,j,t}]^\top_{n_t \times J_t}$ the matrix with element $\widehat{\Phi}_{i,j,t} = \widehat{p}_{jt}(\widehat{\beta}_{it})$. We also let $\widehat{p}_{jt}(\beta)$ be $\widehat{p}_{jt}$ in later sections. We can then obtain

$$\widehat{a}_t = (\widehat{\Phi}_t \widehat{\Phi}_t^\top)^{-1} (\widehat{\Phi}_t R_t),$$

which represent the average returns of assets in $\widehat{P}_{jt}$ for $j = 1, \ldots, J_t$ at time t. Define $\widehat{a}_{jt}$ as the jth element of $\hat{a}_t$.

Letting $\widehat{a}_{lt}$ and $\widehat{a}_{ut}$ be the portfolio returns of the two extreme portfolios, a common object of interest is the differential average returns in the most extreme portfolios:

$$\frac{1}{T}\sum_{t=1}^{T}(\widehat{a}_{ut} - \widehat{a}_{lt}) = \frac{1}{T}\sum_{t=1}^{T}\Big(\widehat{\mu}_t(\beta_u) - \widehat{\mu}_t(\beta_l)\Big),$$

where

$$\widehat{\mu}_t(\beta) = \sum_{j=1}^{J_t} \widehat{p}_{jt}(\beta)\widehat{a}_{jt}. \tag{2.6}$$

[7] Cattaneo, Crump, Farrell, and Schaumburg (2020) provide a detailed discussion of how sorted portfolios represent a nonparametric estimate of a conditional expectation. See also, Fama and French (2008), Cochrane (2011), and Freyberger, Neuhierl, and Weber (2020).

More generally, many other estimators of interest in finance can be defined as transformations of the stochastic processes $(\widehat{\mu}_t(\beta) : \beta \in \mathcal{B})$, for each cross-section.

Similarly, other estimators of interest can be considered by averaging across time. These estimators can be thought of as transformations of the stochastic process $(\widehat{\mu}(\beta) : \beta \in \mathcal{B})$ with

$$\widehat{\mu}(\beta) = \frac{1}{T}\sum_{t=1}^{T}\widehat{\mu}_t(\beta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta)\widehat{a}_{jt}. \tag{2.7}$$

For example, we can estimate conditional expected returns for all values of β rather than only values near β_l and β_u. Correspondingly, $\widehat{\mu}_t(\beta)$ and $\widehat{\mu}(\beta)$ may be directly interpreted as nonparametric estimators of expected returns. ■

A few comments are in order. First, the above two steps are completely in line with the empirical finance literature. Importantly, at no point in the two-step algorithm is the requirement to estimate the conditional expectation of the risk factors, $\mathbb{E}[f_t|\mathcal{F}_{t-1}]$, and so the researcher remains agnostic about the dynamics of these risk factors. We will revisit this issue in the next section. Second, the practice of moving-window regressions to accommodate time variation in β_{it} suggests a slowly-varying coefficient model as previously used in finance applications such as in Ang and Kristensen (2012) and Adrian, Crump, and Moench (2015). However, in contrast to these previous formulations, we do not condition on the realizations of the random processes α_{it} and β_{it}. Instead, we retain the randomness in these objects so that the second-stage beta-sorted portfolio estimator can have a well-defined limit as $n, T \to \infty$. Third, an alternative to the smoothly-varying coefficients approach is to specify β_{it} as a functions of individual characteristics and possibly also of economy-wide variables (see, for example, Gagliardini, Ossola, and Scaillet, 2020, and references therein). Our approach can straightforwardly accommodate such settings by modifying the kernel regressions appropriately.

Finally, the more general estimation approach described in equations (2.6) and (2.7), with more details in Section 4, does not constitute spurious generality. The conventional implementation of beta-sorted portfolios relies on a constant choice of $J_t = J \ \forall t$ and so averages J portfolios across all time periods. However, if the cross-sectional distribution of the β_{it} are changing over time then there is no guarantee that each portfolio represents assets with sufficiently similar betas. For example, it may be that assets with values of β near 1/2 fall in the sixth portfolio at times and the fifth portfolio at other times and so on. Thus, the conventional estimator will be, in general, both more biased and more variable than the estimators suggested in equations (2.6) and (2.7), all else equal. This is of special importance when we are interested in expected returns for intermediate values of betas and also in situations where tests of monotonicity or shape restrictions are of interest.

3 First step: rolling regressions

The first step involves a kernel regression of a linear stochastic coefficients model. Recall that $X_t = (1, f_t)$ and define $b_{it} = (\alpha_{it}, \beta_{it})$. Then, we can rewrite equation (2.1) as

$$R_{it_0} = X_{t_0}^\top b_{it_0} + \varepsilon_{it_0}.$$

We assume that $\mathbb{E}(\varepsilon_{it}|\mathcal{F}_{t-1}) = \mathbb{E}_{t-1}(\varepsilon_{it}) = 0$ and, because α_{it} and β_{it} are measurable with respect to $\mathcal{F}_{t-1}$, then α_{it_0} and β_{it_0} can be identified as

$$b_{it_0} = \mathbb{E}(X_{t_0} X_{t_0}^\top|\mathcal{F}_{t_0-1})^{-1}\mathbb{E}(X_{t_0} R_{it_0}|\mathcal{F}_{t_0-1}).$$

The kernel estimator from (2.5) is then $\widehat{b}_{it_0} = (\widehat{\alpha}_{it_0}, \widehat{\beta}_{it_0})^\top$. In order to accommodate the random coefficients we exploit the fact that $\sum_{t=1}^{t_0-1} K((t-t_0)/(Th))X_tX_t^\top$ and $\sum_{t=1}^{t_0-1} K((t-t_0)/(Th))X_tR_{it}$ are close, in the appropriate sense, to $\sum_{t=1}^{t_0-1} \mathbb{E}[K((t-t_0)/(Th))X_tX_t^\top|\mathcal{F}_{t-1}]$ and $\sum_{t=1}^{t_0-1} \mathbb{E}[K((t-t_0)/(Th))X_tR_{it}|\mathcal{F}_{t-1}]$, since their difference are summands of martingale difference sequences.

To formalize the intuition and establish uniform consistency and asymptotic normality of $\widehat{b}_{it_0}$ we require technical, but relatively standard, assumptions on the underlying data generating process. We report these assumptions in the Appendix (Assumptions 1–6) and discuss them briefly here. Assumption 1 ensures that the one-sided kernel $K(\cdot)$ satisfies standard properties such as being nonzero on a compact support and twice continuously differentiable. The one-sided kernel ensures that we do not have any look-ahead bias, so the procedure can be interpreted as real-time estimation, and also to define the appropriate conditional moments for the second step discussed in the next section. Assumption 2 imposes some structure on the time series properties of the factor f_t but is quite general and allows for certain forms of nonstationary behavior. We could relax some of these assumptions to allow for even more complex time-series properties at the expense of more detailed notation and proofs. Assumption 2 also imposes moment conditions on the idiosyncratic error term, ε_{it}. Assumption 3 ensures that b_{it_0} is well defined for all t_0. Assumptions 4 and 6 are regularity conditions on the rate of decay of the time-series dependence of the risk factors. Finally, Assumption 5 ensures that the alphas and betas, although random, are sufficiently smooth over time (i.e., satisfying a Lipschitz-type condition). Similar assumptions are generally imposed in varying coefficient models (see, for example, Zhang and Wu, 2015).

We first provide a uniform consistency results of our estimator $\widehat{b}_{it_0}$ over i and t. We require this result to precisely control the effect of estimating β_{it} in the first step when entering the second-step estimator. We establish this consistency on a compact interval of a trimmed support with trimming length $\lfloor Th \rfloor$. Let q denote the parameter in Assumption 2.

Theorem 3.1. *Suppose Assumptions 1–6 hold, and let* $r_T = (Th)^{-1}(T^{1/q} + \sqrt{Th \log T}) \to 0$, $h \to 0$, *and* $\log(n_u T)/Th \to 0$. *Then,*

$$\max_{1\leq i \leq n} \sup_{\lfloor Th \rfloor \leq t_0 \leq T - \lfloor Th \rfloor} \left| \widehat{b}_{it_0} - b_{it_0} \right| \lesssim_{\mathbb{P}} \delta_T,$$

where $\delta_T = (r_T + \sqrt{\log(n_u T)}/\sqrt{Th} + h)$.

Theorem 3.1 provides uniform rates of convergence for the first-stage kernel estimators of the betas. Naturally, these rates depend on n, T, and h but are also directly dependent on q which represents the number of bounded moments of the idiosyncratic error term. For very large q, essentially the uniformity is attained at rate only slower by a $\log(T)$ factor. Importantly, the theorem shows that we attain the same uniform rate for the leave-one-out estimator which ensures our theoretical results mimic empirical practice exactly.

Although estimation of $\mu(\cdot)$ is generally of interest, there are some situations where inference on β_{it} directly is instead the primary goal. To introduce the necessary results we need to present some additional useful notation. To allow for a flexible class of time series processes we model the factor, f_t, as a sum of two components, $f_t = \tau_t + x_t$, where τ_t is a smoothly-varying process and x_t is a strictly stationary process.[8] Then we can define $\tau_t = \tau'(t/T)$ for a smooth function $\tau' : [0, 1] \mapsto \mathbb{R}$.

[8] We could allow for even more general behavior in x_t; however, for simplicity we maintain the strict stationarity assumption.

Also define, $\Sigma_x = \mathbb{E}[(1, x_t)(1, x_t)^\top]$, $\tilde{\tau}(t_0/T) = (1, \tau'(t_0/T))^\top$. We let $\Sigma_A = \Sigma_x + \tilde{\tau}(t_0/T)\tilde{\tau}(t_0/T)^\top$, $\Sigma_B = \sigma^2_{\varepsilon,0}\mathbb{E}(X_{t_0}X_{t_0}^\top)\int_{-1}^{0} K^2(s)ds$. $\Sigma_b = \Sigma_A^{-1}\Sigma_B\Sigma_A^{-1} = \Sigma_A^{-1}\sigma^2_{\varepsilon,0}\int_{-1}^{0} K^2(s)ds$. With these definitions in place, we next show asymptotic normality of our estimator $\widehat{b}_{it_0}$.

Theorem 3.2 (Asymptotic Normality). *Let $h\sqrt{hT} \to 0$, $h \to 0$, $Th \to \infty$, $r_{AT} + r_T \to 0$ then, under Assumptions 1-6, we have that*

$$\sqrt{Th}\Sigma_b^{-1/2}(\widehat{b}_{it_0} - b_{it_0}) \to_{\mathcal{L}} \mathsf{N}(0, I). \tag{3.1}$$

where r_{AT} is defined in the Appendix.

We show in the appendix that the limiting asymptotic distribution is invariant to whether the leave-one-out or general kernel estimator is used. The results in Theorem 3.2 can to be extended to distribution results which are uniform over t; however, we don't pursue this here as our main focus is on the beta-sorted estimator. Finally, note that to construct a confidence interval for b_{it_0} based on $\widehat{b}_{it_0}$, we require a consistent estimator of the asymptotic variance of $\widehat{b}_{it_0}$. Using residuals from the initial step, i.e., $\widehat{\varepsilon}_{it}$, σ_t^2 can be estimated by

$$(\widehat{\sigma}^2_{t_0}, \widehat{\varsigma}^2_{t_0})^\top = \arg\min_{c_0, c_1} \sum_{t=1}^{t_0-1}\sum_{i=1}^{n_t} K\Big(\frac{t-t_0}{Th}\Big)\big(\widehat{\varepsilon}^2_{it} - c_0 - c_1(t-t_0)/T\big)^2.$$

So $\widehat{\Sigma}_b$ can be obtained by $TA(t_0)^{-1}\widehat{\sigma}(t_0/T)\int_0^1 K^2(w)dw$.

4 Second step: beta sorts

The second step of the estimation procedure is to sort assets by their value of $\widehat{\beta}_{it}$ obtained from the procedure described in the previous section. Recall that the structural form of our model is

$$\begin{aligned} R_{it} &= \mu_t(\beta_{it}) + \beta_{it}^\top(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})) + \varepsilon_{it} \\ &= M_t(\beta_{it}) + \varepsilon_{it}, \end{aligned}$$

and under our assumptions we have $\mathbb{E}(\varepsilon_{it}|\mathcal{F}_{t-1}, \beta_{it}) = \mathbb{E}(\varepsilon_{it}|\mathcal{F}_{t-1}) = 0$.

To gain intuition, suppose that the β_{it} were observed. The second equality makes clear that, for a fixed t, we can only nonparametrically estimate the unknown function $M_t(\cdot)$ rather than the direct object of interest $\mu_t(\cdot)$. However,

$$\frac{1}{T}\sum_{t=1}^{T} M_t(\beta) = \frac{1}{T}\sum_{t=1}^{T}\mu_t(\beta) + \frac{1}{T}\sum_{t=1}^{T}\beta^\top(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})). \tag{4.1}$$

The second term has summands, $\beta^\top(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))$, which are a martingale difference sequence with respect to $\mathcal{F}_t$ and so we would expect this sample average to converge to zero in probability; consequently, this would ensure that $T^{-1}\sum_{t=1}^{T} M_t(\beta)$ and $T^{-1}\sum_{t=1}^{T}\mu_t(\beta)$ are uniformly (in β) close in probability for large T. A further complication, of course, is introduced by using an estimated β_{it} in the second-stage nonparametric regression. Nevertheless, in this section, we will make these arguments rigorous and provide appropriate conditions for consistency and asymptotic normality for the beta-sorted portfolio estimator.

To motivate the assumptions we introduce shortly, note that we may rewrite our model as

$$R_{it} = \alpha_{it} + \beta_{it}^\top\mathbb{E}(f_t|\mathcal{F}_{t-1}) + \tilde{\varepsilon}_{it},$$

where $\tilde{\varepsilon}_{it} = \beta_{it}^\top(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})) + \varepsilon_{it}$ represents the sum of two different martingale difference sequences. This form makes clear that we require assumptions on α_{it} and β_{it} to be able to approximate the grand mean, $T^{-1}\sum_{t=1}^{T}\mu_t(\beta)$, with high probability.

We assume that $\beta_{it} = \beta_i(t/T, \mathcal{F}_{t-1})$ (c.f. Assumption 5 in the appendix), which is a smooth random function over time. We will further assume that β_{it} are, conditional on $\mathcal{G}_{t-1}$, i.i.d. over i. This sampling assumption was introduced in Andrews (2005) and has been utilized in the financial econometrics literature by Gagliardini, Ossola, and Scaillet (2016) and Cattaneo, Crump, Farrell, and Schaumburg (2020). Under this assumption, for a fixed time t, β_{it} follows a conditional distribution $F_{\beta,t}(\cdot) = \mathbb{P}(\beta_{it} \leq \cdot|\mathcal{G}_{t-1})$ for each time period t. Thus, we define the transformed variable $U_{it} = F_{\beta,t}(\beta_{it})$, which are i.i.d. uniform $[0,1]$ random variables over i conditioning on $\mathcal{G}_{t-1}$. Define $F_{\beta,n,t}(\cdot) = \frac{1}{n_t}\sum_{i=1}^{n_t}\mathbf{1}\{\beta_{it} \leq \cdot\}$ and $F_{\beta,n,t}^{-1}$ as the empirical counterparts of the conditional CDF and quantile functions $F_{\beta,t}$ and $F_{\beta,t}^{-1}$. The order statistics of β_{it} over i for fixed t is denoted as $\beta_{(i)t} = F_{\beta,n,t}^{-1}(i/n_t)$. In our setup, we have that α_{it} is a function of β_{it}, and thus a continuous function of U_{it}, that is,

$$\alpha_{it} = \alpha(\beta_{it}) = \alpha(F_{\beta,t}(U_{it}))$$

and the function $\alpha(F_{\beta,t}(.))$ will be smooth with respect to U_{it}; similarly, β_{it} can be regarded as a smooth functional of $F_{\beta,t}(.)$.

To gain intuition for the procedure, assume again (temporarily) that we could observe β_{it}. Define $\Phi^*_{i,j,t} = \mathbf{1}(F^{-1}_{\beta,t}((j-1)/J_t) \le \beta_{it} < F^{-1}_{\beta,t}((j)/J_t))$ which corresponds the event where β_{it} is realized between these two conditional quantiles.[9] These represent the (infeasible) basis functions which underpin the partitioning estimator. We can then define the population best linear predictor as,

$$a^*_t = (a^*_{1t}, \cdots, a^*_{J_t t})^\top = \arg\min_{a_{1t},\cdots,a_{J_t t}} \mathbb{E}\Big[\Big(R_{it} - \sum_{j=1}^{J_t} a_{jt}\Phi^*_{i,j,t}\Big)^2 \Big| \mathcal{G}_{t-1}\Big]. \tag{4.2}$$

We can then rewrite equation (2.4) as

$$R_{it} = \sum_{j=1}^{J_t} a^*_{jt}\Phi^*_{i,j,t} + \flat_{it} + \tilde{\varepsilon}_{it},$$

where $\flat_{it} = \mu_t(\beta_{it}) - \sum_{j=1}^{J_t} a^*_{jt}\Phi^*_{i,j,t}$ represents the approximation bias term.

In order to characterize the theoretical properties of the portfolio estimator it is necessary to introduce some additional notation to present the different basis functions which underpin our analysis. Define $\Phi_{i,j,t} = \mathbf{1}(U_{it} \in [U_{(\lfloor (j-1)n_t/J_t \rfloor),t}, U_{(\lfloor (j)n_t/J_t \rfloor),t})) = \mathbf{1}(F^{-1}_{\beta,n,t}((j-1)/J_t) \le \beta_{it} < F^{-1}_{\beta,n,t}((j)/J_t))$ and, for estimated β_{it}, its counterpart $\widehat{\Phi}_{i,j,t} = \mathbf{1}(F^{-1}_{\widehat{\beta},n,t}((j-1)/J_t) \le \widehat{\beta}_{it} < F^{-1}_{\widehat{\beta},n,t}(j/J_t))$. Finally, we denote the stacked elements as $\Phi_{i,t} = [\Phi_{i,j,t}]_j$, $\widehat{\Phi}_{i,t} = [\widehat{\Phi}_{i,j,t}]_j$ and $\Phi^*_{i,t} = [\Phi^*_{i,j,t}]_j$ for $J_t \times 1$ vectors and stack further as $J_t \times n_t$ matrices denoted by $\Phi^*_t = [\Phi^*_{i,j,t}]_{j,i}$ and similarly for $\widehat{\Phi}_t$ and Φ_t.

To obtain a feasible estimator we cannot rely on Φ^*_t but instead, as introduced in Section 2.1, we utilize $\widehat{\Phi}_t$. Recall that $\widehat{a}_t = \{\widehat{\Phi}_t\widehat{\Phi}_t^\top\}^{-1}\{\widehat{\Phi}_t R_t\}$ and for any $\beta \in [\beta_l, \beta_u]$ and the grand mean estimator $\widehat{\mu}(\beta) = T^{-1}\sum_{t=1}^T \widehat{p}_t(\beta)^\top \widehat{a}_t$ is given in (2.7). Let $\flat_t = [\flat_{it}]_i$. To analyze the rate of the beta sorted estimator $\widehat{a}_t$, we shall prove that under certain conditions,

$$\begin{aligned}
\widehat{a}_t &= \{\widehat{\Phi}_t\widehat{\Phi}_t^\top\}^{-1}\{(\widehat{\Phi}_t[\Phi^{*\top}_t a^*_t + \tilde{\varepsilon}_t + \flat_t])\}, \\
&= a^*_t + \{n_t^{-1}\Phi^*_t\Phi^{*\top}_t\}^{-1}(n_t^{-1}\Phi^*_t\tilde{\varepsilon}_t) + \{n_t^{-1}\Phi^*_t\Phi^{*\top}_t\}^{-1}(n_t^{-1}\Phi^*_t\flat_t) + o_{\mathbb{P}}(1).
\end{aligned}$$

[9] We assume that, at time point t the partition intervals are, $[(j-1)/J_t, j/J_t)$ for $j \in 1,\cdots,J_t - 1$ and $((J_t-1)/J_t, 1]$ for $j = J_t$.

To make these statements precise, we require additional assumptions (formally stated in the Appendix). Assumption 7 imposes regularity conditions on the conditional distribution of the β_{it} ensuring that it is sufficiently well behaved. These assumptions ensure the partitioning estimator is well defined with the probability of empty portfolios vanishing asymptotically and, furthermore, that $\tilde{q}_j$ is of the order J^{-1} where $\tilde{q}_{jt} = \int_{F_{\beta,t}^{-1}(\kappa_{j-1t})}^{F_{\beta,t}^{-1}(\kappa_{jt})} f_{\beta,t} d\beta$ with $k_{jt} = \lfloor n_t j / J_t \rfloor$. . Assumption 8 sets the properties of the uniform convergence rate of β_{it} and corresponds directly to the results in Theorem 3.1. Assumption 9 imposes restrictions on the relative rate of n_t and J_t . Assumption 10 assumes the smoothness of the function $\alpha(.)$, and ensures that the α is a well-behaved function of β. Assumption 11 imposes some additional moment and smoothness conditions on the conditional distribution of β_{it}. In the SA we establish that, under our assumptions, we may ignore the generated errors of the first-stage estimation of the β_{it} when analyzing the second stage portfolio sorting estimator.

We now have laid the necessary foundation to obtain a linearization of the grand mean estimator:

Theorem 4.1 (Leading term linearization)**.** *Suppose Assumptions 7-8 and 10-12 hold. Then, uniformly in β,*

$$\frac{1}{T}\sum_{t=1}^{T}\{\widehat{\mu}_t(\beta) - \mu_t(\beta)\} = \frac{1}{T}\sum_{t=1}^{T}\widehat{p}_t(\beta)^\top[\mathrm{diag}(\tilde{q}_{jt})^{-1} n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t] + O_{\mathbb{P}}((J^{-1} \vee h)) + o_{\mathbb{P}}(T^{-1/2}),$$

where the first term is the leading term and the second term is the bias term. Moreover,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\widehat{p}_t(\beta)^\top[\mathrm{diag}(\tilde{q}_{jt})^{-1} n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t] \\
&\quad = \frac{1}{T}\sum_{t=1}^{T}\widehat{p}_t(\beta)^\top \,\mathrm{diag}(\tilde{q}_{jt})^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\Phi_{i,t}^*\varepsilon_{it} \\
&\qquad\quad + \frac{1}{T}\sum_{t=1}^{T}\widehat{p}_t(\beta)^\top \,\mathrm{diag}(\tilde{q}_{jt})^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\Phi_{i,t}^*\beta_{it}(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})).
\end{aligned}$$

Theorem 4.1 introduces the key properties of the grand mean estimator. Importantly, the theorem shows the leading term is comprised of two elements: the first term would appear in any generic nonparametric problem whereas the second term is specific to the asset pricing setup. Importantly, the second term is of the order $O_{\mathbb{P}}(T^{-1/2})$ representing the summation of the product of the conditional beta and the deviation of the factor from its conditional mean, $(f_t - \mathbb{E}\left[f_t|\mathcal{F}_{t-1}\right])$.

Thus, despite an effective sample size of nT in concert with a nonparametric procedure with tuning parameter J, the grand mean estimator, for some values of β, achieves only a $\sqrt{T}$ rate of convergence. That said, for values of β near zero, the second term becomes degenerate and the first term dominates leading to a faster rate of convergence, namely, $O_{\mathbb{P}}(\sqrt{J/nT})$.

Remark 4.2. *An alternative approach that could be considered is to center the estimator,* $T^{-1}\sum_{t=1}^{T}\widehat{\mu}_t(\beta)$ *at* $T^{-1}\sum_{t=1}^{T}M_t(\beta)$*, rather than* $T^{-1}\sum_{t=1}^{T}\mu_t(\beta)$*. This would have the advantage of producing a uniform rate of convergence of the estimator as the second term in equation (4.1) is removed from the asymptotic distribution. This is analogous to centering the estimator at the ex-post risk premia (see, e.g., Raponi, Robotti, and Zaffaroni (2020)) and can be thought of as centering at average realized* systematic *returns. However, inference on this object appears to be of less interest, in general, and so we do not pursue this approach further.*

Next we provide a pointwise central limit theorem for $\widehat{\mu}(\beta)$ which allows us to make pointwise inference on the estimator of $T^{-1}\sum_{t=1}^{T}\mu_t(.)$. We define $E_{n_t,j} = \tilde{q}_{jt}^{-1}\mathbb{E}(\Phi^*_{i,j,t}\beta_{it}|\mathcal{G}_{t-1})$. Recall that $\mathbb{E}_{t-1}(\Phi^*_{i,j_t,t}) = \tilde{q}_{jt} = \mathbb{E}(\Phi^*_{i,j_t,t}|\mathcal{G}_{t-1})$.

Theorem 4.3 (Pointwise central limit theorem)**.** *Suppose Assumptions 1-12 hold. Then, pointwise in* β*,*

$$\frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\{\widehat{\mu}_t(\beta)-\mu_t(\beta)\} - bias(\beta)}{\mathbb{E}(\widehat{\sigma}(\beta))^{1/2}} \to_{\mathcal{L}} \mathsf{N}(0,1),$$

where

$$\widehat{\sigma}(\beta) = \sigma_f(\beta) + \sigma_\varepsilon(\beta)$$

with

$$\sigma_f(\beta) = T^{-1}\sum_{t=1}^{T}(\sum_{j=1}^{J_t}\widehat{p}_{j,t}(\beta)E^2_{n_t,j}\,Var(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})),$$

$$\sigma_\varepsilon(\beta) = T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}n_t^{-2}\sum_{i}\widehat{p}_{j,t}(\beta)\tilde{q}_{jt}^{-2}\mathbb{E}(\Phi^*_{i,j,t}\varepsilon^2_{it}|\mathcal{G}_{t-1}),$$

and

$$bias(\beta) = T^{-1}\sum_{t=1}^{T}\widehat{p}_t(\beta)^\top(n_t^{-1}\Phi_t^*\Phi_t^{*\top})^{-1}(n_t^{-1}\Phi_t^*\flat_t) + T^{-1}\sum_{t=1}^{T}\widehat{p}_t(\beta)^\top \mathrm{diag}(\tilde{q}_j)^{-1}(n_t^{-1}(\widehat{\Phi}_t\widehat{\Phi}_t^\top - \widehat{\Phi}_t\Phi_t^{*\top})a_t^*),$$

which is a term of order $(J^{-1} \vee h)$.

The theorem provides the basis of our feasible inference procedures discussed in the next section. It is important to note that the two components of the appropriate standard deviation, $\sigma_f(\beta)$ and $\sigma_\epsilon(\beta)$, are orthogonal. We will exploit this property to conduct feasible inference without assuming a specific functional form for $\mathbb{E}(f_t|\mathcal{F}_{t-1})$.

Remark 4.4 (Extension to multivariate β)**.** *The algorithm and proof are written for the case of* $d = 1$. *It would not be hard to develop for multivariate* β *corresponding to fixed* $d > 1$. *For multiple-characteristic portfolios we can adopt the Cartesian products of marginal intervals. That is, we first partition each characteristic into* J_t *intervals, using its marginal quantiles, and then form* J_t^d *portfolios by taking the Cartesian products of all such intervals. The pointwise convergence results of the beta-sorted portfolio estimator could be extended to this general case. Note also, the curse of dimensionality could be avoided by assuming additively separability.*

5 Feasible (uniform) inference for the grand mean

In order to conduct feasible inference on the grand mean we require a consistent estimator of the asymptotic variance given by Theorem 4.3. In existing empirical applications, the so-called Fama-MacBeth variance estimator is utilized. In this section we discuss the properties of both the Fama-MacBeth variance estimator and also a simple new plug-in variance estimator. We show that both variance estimators provide valid inference. The plug-in variance estimator is as simple to implement as the Fama-MacBeth variance estimator and, as we show in the next section, appears to provide much more precise inference in empirical applications.

Accurate inference is vital in order to assess whether observed realized returns of a specific trading strategy withstand statistical scrutiny. We link these practical questions with rigorous formation of uniform statistical tests. We highlight three important types of uniform inference hypotheses that corresponds to trading a specific, high-minus-low or butterfly trade portfolio respectively. We provide a valid uniform inference procedure for the grand mean estimator using the Fama-MacBeth variance estimator. This will allow us to conduct inference on more complex hypotheses of interest such as tests of monotonicity or tests of nonzero differential expected returns across the support of β_{it}.

5.1 A plug-in variance estimator

We can use the results in Theorems 4.1 and 4.3 to construct a plug-in variance estimator. To see the logic of our approach, first consider the case where we observe $\mathbb{E}(f_t|\mathcal{F}_{t-1})$. Then a natural plug-in variance estimator is,

$$\begin{aligned}\tilde{\sigma}_{\texttt{PI}}(\beta) =& T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t} n_t^{-2}\hat{q}_{jt}^{-2}\left(\sum_{i=1}^{n_t}\widehat{p}_{jt}(\beta)\widehat{p}_{jt}(\beta_{it})\beta_{it}\right)^2 (f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1}))^2 \\ &+ T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t} n_t^{-2}\hat{q}_{jt}^{-2}\sum_{i=1}^{n_t}\widehat{p}_{jt}(\beta)\widehat{p}_{jt}(\beta_{it})\hat{\varepsilon}_{it}^2, \end{aligned} \tag{5.1}$$

where $\hat{q}_{jt} = n_t^{-1}\sum_{i=1}^{n_t}\widehat{p}_{jt}(\beta_{it})\mathbf{1}\{\sum_{i=1}^{n_t}\widehat{p}_{jt}(\beta_{it}) \neq 0\}$. Of course, $\tilde{\sigma}_{\texttt{PI}}(\beta)$ is infeasible. As a feasible alternative, consider

$$\begin{aligned}\hat{\sigma}_{\texttt{PI}}(\beta) =& T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t} n_t^{-2}\hat{q}_{jt}^{-2}\left(\sum_{i=1}^{n_t}\widehat{p}_{jt}(\beta)\widehat{p}_{jt}(\beta_{it})\beta_{it}\right)^2 (f_t-h_{t-1}(\hat{\vartheta}))^2 \\ &+ T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t} n_t^{-2}\hat{q}_{jt}^{-2}\sum_{i=1}^{n_t}\widehat{p}_{jt}(\beta)\widehat{p}_{jt}(\beta_{it})\hat{\varepsilon}_{it}^2, \end{aligned} \tag{5.2}$$

where $h_{t-1}(\hat{\vartheta})$ is a feasible parametric estimate of $\mathbb{E}(f_t|\mathcal{F}_{t-1})$ using only information contained in $\mathcal{G}_{t-1}$. Importantly, even if this estimator is misspecified $\hat{\sigma}_{\texttt{PI}}(\beta)$ will still produce valid, albeit conservative, inference because of the asymptotic orthogonality of the two terms in Theorems 4.1 and 4.3.

5.2 The Fama-MacBeth estimator

Recall the definition of $\widehat{\mu}(\beta)$ and $\widehat{\mu}_t(\beta)$ as in equation (2.6) and (2.7). The Fama-Macbeth variance estimator may then be constructed as

$$\widehat{\sigma}_{\texttt{FM}}(\beta) = \frac{1}{T}\sum_{t=1}^{T}\Big(\widehat{\mu}_t(\beta) - \widehat{\mu}(\beta)\Big)^2. \tag{5.3}$$

The estimator may be motivated by the classical sample variance estimator where $\widehat{\mu}_t(\beta)$ for $t = 1, \ldots, T$ serve as the sample "observations." Following similar reasoning, we shall denote an estimator of $\mathrm{Cov}(\widehat{\mu}(\beta_1), \widehat{\mu}(\beta_2))$ as the following. For any β_1, β_2,

$$\widehat{\sigma}_{\texttt{FM}}(\widehat{\mu}(\beta_1), \widehat{\mu}(\beta_2)) = \frac{1}{T}\sum_{t=1}^{T}\Big(\widehat{\mu}_t(\beta_1) - \widehat{\mu}(\beta_1)\Big)\Big(\widehat{\mu}_t(\beta_2) - \widehat{\mu}(\beta_2)\Big). \tag{5.4}$$

Note that in the special case of $\beta_1 = \beta_2$ we obtain $\widehat{\sigma}_{\texttt{FM}}(\beta)$. To discuss the asymptotic properties of this variance estimator we need to define some specific population counterparts.

First, define

$$\sigma_f(\beta_1, \beta_2) = T^{-1}\sum_{t=1}^{T}\mathbb{E}\Big[\{\widehat{p}_t(\beta_1)^\top \operatorname{diag}(\tilde{q}_{jt})^{-1}\mathbb{E}(\Phi_{it}^*\beta_{it}|\mathcal{G}_{t-1})\} \\ \times \{\widehat{p}_t(\beta_2)^\top \operatorname{diag}(\tilde{q}_{jt})^{-1}\mathbb{E}(\Phi_{it}^*\beta_{it}|\mathcal{G}_{t-1})\}\mathbb{E}\{(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))^2\}\Big].$$

The quantity $\sigma_f(\beta_1, \beta_2)$ represents the first-order asymptotic variance utilizing the results in Theorems 4.1 and 4.3. We also need to define the additional population objects, $\sigma_\mu(\beta) = T^{-1}\sum_{t=1}^{T}\mathbb{E}[(\mu_t(\beta) - T^{-1}\sum_{t=1}^{T}\mu_t(\beta))^2]$ and $\sigma_\mu(\beta_1, \beta_2) = T^{-1}\sum_{t=1}^{T}\mathbb{E}[\mu_t(\beta_1) - T^{-1}\sum_{t=1}^{T}\mu_t(\beta_1))(\mu_t(\beta_2) - T^{-1}\sum_{t=1}^{T}\mu_t(\beta_2))]$. Then, $\sigma_\mu(\beta)$ and $\sigma_\mu(\beta_1, \beta_2)$ represent the population average time variation and co-variation in the conditional expected returns. Finally, denote $\mathbb{J}$ as a set which collects the appropriate js (identity of the relevant bin) over time. Thus, $\mathbb{J}_1$ indicates the bins that β_1 falls into for each time t for $t = 1, \ldots, T$. With these objects defined, we can now state the following properties of the Fama-MacBeth variance estimator.

Lemma 5.1. *Under the conditions of Theorem B.6 and 4.1, for* β_1*,* β_2 *corresponding to* $\mathbb{J}_1$*,* $\mathbb{J}_2$

respectively, we have

$$\sup_{\beta_1,\beta_2\in[\beta_l,\beta_u]} |\widehat{\sigma}_{\mathtt{FM}}(\widehat{\mu}(\beta_1),\widehat{\mu}(\beta_2)) - \sigma_f(\beta_1,\beta_2) - \sigma_\mu(\beta_1,\beta_2)| = o_{\mathbb{P}}(1/\sqrt{\log J}).$$

Note that the above results implies that

$$\sup_{\beta\in[\beta_l,\beta_u]} |\widehat{\sigma}_{\mathtt{FM}}(\beta) - \sigma_f(\beta) - \sigma_\mu(\beta)| = o_{\mathbb{P}}(1/\sqrt{\log J}).$$

Lemma 5.1 shows that the asymptotic limit of the Fama-MacBeth variance estimator is comprised of two terms. The first term is the population target, $\sigma_f(\beta)$ and $\sigma_f(\beta_1,\beta_2)$, respectively, which represents the limiting variance from Theorem 4.3. The second term is an extraneous term, $\sigma_\mu(\beta)$ and $\sigma_\mu(\beta_1,\beta_2)$, respectively, which are non-negative by definition. Intuitively, we can understand this result from the following decomposition of the summands of the Fama-MacBeth variance estimator:

$$\widehat{\mu}_t(\beta) - \frac{1}{T}\sum_{t=1}^{T}\widehat{\mu}_t(\beta) = (\widehat{\mu}_t(\beta) - \mu_t(\beta)) - \frac{1}{T}\sum_{t=1}^{T}(\widehat{\mu}_t(\beta) - \mu_t(\beta)) + \Big(\mu_t(\beta) - \frac{1}{T}\sum_{t=1}^{T}\mu_t(\beta)\Big)$$

It is the third term in this equation that contributes the additional term to the probability limit of $\widehat{\sigma}_{\mathtt{FM}}(\beta)$. Consequently, the Fama-MacBeth variance estimator overstates the true asymptotic variance of the estimator by exactly this extraneous term. In the special case when $\mu_t(\beta)$ is constant over time then $\sigma_\mu(\beta)$ and $\sigma_\mu(\beta_1,\beta_2)$ are equal to zero and Lemma 5.1 establishes uniform consistency of the Fama-MacBeth variance estimator. Otherwise, the variance estimator will overstate the true variance and lead to a conservative inference.

However, Lemma 5.1 has the positive implication that the Fama-MacBeth variance estimator is consistent for the asymptotic variance of the expression $T^{-1}\sum_{t=1}^{T}(\widehat{\mu}_t(\beta) - \mu(\beta))$ where $\mu(\beta) = \text{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T}\widehat{\mu}_t(\beta)$. That said, this facilitates inference only on an object, $\mu(\beta)$, that is arguably of less interest than $T^{-1}\sum_{t=1}^{T}\mu_t(\beta)$. This latter object, representing the sample average conditional expected returns, is of more direct relevance to economic inference since there can be no further information available for a given sample of T time series observations.

Remark 5.2. *It is important to note that Lemma 5.1 also implies that valid inference may be conducted on the average conditional expected returns without the need to stipulate the form of* *the conditional expectation of the risk factors. This stands in contrast to alternative estimation approaches in, for example, Adrian, Crump, and Moench (2015) and Gagliardini, Ossola, and Scaillet (2016), where a first-order Markovian structure is imposed. In practice, specifying the correct functional form including the appropriate conditioning variables for the risk factor dynamics is a challenge. This is one notable advantage of the estimation approach we study here.*

5.3 Uniform inference of the grand mean estimator

The estimator of the grand-mean function offers us the chance to test the hypothesis regarding price anomaly. In this subsection, we provide a rigorous formulation of a uniform test for the grand-mean estimator. This facilitates us to conduct various uniform tests related to the grand-mean estimator. Namely, we aim to test the following null hypothesis,

$$\mathsf{H}_0 : \mu(\beta) = 0, \forall \beta \in [\beta_l, \beta_u],$$

against the alternative,

$$\mathsf{H}_A : \mu(\beta) \neq 0, \text{for some } \beta.$$

Before we present a theorem which gives us the critical value of the uniform test, we shall discuss the estimator of the elements of variance-covariance matrix for the grand mean estimator in the strong approximation theorem. As indicated by Theorem 4.1, the long-run variance of the lead term $T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta)(f_t - \mathbb{E}(f_t))E_{n_t,j} + T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta)^\top \operatorname{diag}(\tilde{q}_{jt})^{-1}{n_t}^{-1}\sum_{i=1}^{n_t}\Phi^*_{i,j,t}\varepsilon_{it}$ is defined as follows

$$\begin{aligned}
\tilde{\sigma}(\beta_1,\beta_2) &= \sigma_f(\beta_1,\beta_2) \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{J_t}\sum_{l=1}^{J_s}\mathbb{E}\Big[\big(\mathbb{E}(f_t|\mathcal{F}_{t-1}) - \mathbb{E}(f_t)\big)\big(\mathbb{E}(f_s|\mathcal{F}_{s-1}) - \mathbb{E}(f_s)\big)E_{n_t,j}E_{n_s,l}\widehat{p}_{jt}(\beta_1)\widehat{p}_{ls}(\beta_2)\Big] \\
&\quad + T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}n_t^{-2}\sum_i \mathbb{E}\{\widehat{p}_{jt}(\beta_1)\tilde{q}_{jt}^{-2}\mathbb{E}(\Phi^*_{i,j,t}\varepsilon^2_{it}|\mathcal{G}_{t-1})\mathbf{1}\{\beta_1 = \beta_2\}\} \\
&= \sigma_f(\beta_1,\beta_2) + \sigma_\mu(\beta_1,\beta_2) + T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}n_t^{-2}\sum_i \mathbb{E}\{\widehat{p}_{jt}(\beta_1)\tilde{q}_{jt}^{-2}\mathbb{E}(\Phi^*_{i,j,t}\varepsilon^2_{it}|\mathcal{G}_{t-1})\mathbf{1}\{\beta_1 = \beta_2\}\}.
\end{aligned}$$

And $\sigma(\beta_1,\beta_2) = \sigma_f(\beta_1,\beta_2) + T^{-1}\sum_{t=1}^T\sum_{j=1}^{J_t} n_t^{-2}\sum_i \mathbb{E}\{\widehat{p}_{jt}(\beta_1)\tilde{q}_{jt}^{-2}\mathbb{E}(\Phi^*_{i,j,t}\varepsilon_{it}^2|\mathcal{G}_{t-1})\}\mathbf{1}\{\beta_1 = \beta_2\}$. In particular, the variance is $\tilde{\sigma}(\beta) = \tilde{\sigma}(\beta,\beta)$, with $\sigma(\beta) = \sigma(\beta,\beta)$ and $\sigma_\mu(\beta) = \sigma_\mu(\beta,\beta)$. We define $\mathbb{E}[\sigma(\beta)]^{-1/2}G_T(\beta)$ as a Gaussian process on a proper probability space with covariance $\mathbb{E}[\sigma(\beta_1)]^{-1/2}\mathbb{E}[\sigma(\beta_1,\beta_2)]\mathbb{E}[\sigma(\beta_2)]^{-1/2}$, and similarly for $\mathbb{E}[\tilde{\sigma}(\beta)]^{-1/2}\tilde{G}_T(\beta)$ relative to $\mathbb{E}[\tilde{\sigma}(\beta_1)]^{-1/2}\mathbb{E}[\tilde{\sigma}(\beta_1,\beta_2)]\mathbb{E}[\tilde{\sigma}(\beta_2)]^{-1/2}$. We define $\widehat{\sigma}(\beta)$ $(\widehat{\sigma}(\beta_1,\beta_2))$ as an estimator of $\sigma(\beta)$ $(\sigma(\beta_1,\beta_2))$ as well. From Lemma 5.1 the Fama-Macbeth variance is close to $\tilde{\sigma}(\beta_1,\beta_2)$ rather than to $\sigma(\beta_1,\beta_2)$. We now provide a corollary facilitating the uniform inference on the function $\mu(\beta)$ or $T^{-1}\sum_{t=1}^T\mu_t(\beta)$.

Lemma 5.3. *Under the conditions of Theorem B.6, 4.1, Assumption 15 and $\sqrt{T}(1/J\vee h)\to 0$, we have,*

$$\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\sup_{\beta\in\mathcal{B}}\Big|\frac{1}{\sqrt{T\mathbb{E}[\widehat{\sigma}(\beta)]}}\sum_{t=1}^T(\widehat{\mu}_t(\beta)-\mu_t(\beta))\Big|\le x\Big)-\mathbb{P}\Big(\sup_{\beta\in\mathcal{B}}|G_T(\beta)|\le x\Big)\Big|\to 0. \quad (5.5)$$

By the above lemma, we shall expect the asymptotic distribution of $\sup_{\beta\in\mathcal{B}}\Big|\frac{1}{\sqrt{T\mathbb{E}[\widehat{\sigma}(\beta)]}}\sum_{t=1}^T(\widehat{\mu}_t(\beta))\Big|$ under the null hypothesis to be approximated by the one of $\sup_{\beta\in\mathcal{B}}|G_T(\beta)|$. This result facilitates constructing a critical value of our proposed test statistic. Lemma 5.3 allows us to form an asymptotically valid uniform inference for the grand mean function. We can obtain uniform confidence bands and test the hypothesis H_0. To construct the uniform confidence band, it is implied from the Lemma 5.3 that if we define $L_T(\beta) = \widehat{\mu}(\beta) - \sigma(\beta)^{1/2}q_\alpha/\sqrt{T}$, and $U_T(\beta) = \widehat{\mu}(\beta) + \sigma(\beta)^{1/2}q_\alpha/\sqrt{T}$. We have, with a prefixed confidence level α,

$$\mathbb{P}\Big(\frac{1}{T}\sum_{t=1}^T\mu_t(\beta)\in[L_T(\beta),U_T(\beta)],\ \text{for all } \beta\in\mathcal{B}\Big)\to 1-\alpha.$$

Therefore, $[L_T(\cdot),U_T(\cdot)]$ is the uniform confidence band of the estimator $\widehat{\mu}(\cdot)$. To make inference on $\mu(\cdot)$, and functionals thereof, we can replace the variance estimator $\hat{\sigma}(\beta)$ by the corresponding variance estimator of $\tilde{\sigma}(\beta)$. In addition to the confidence interval, Lemma 5.3 also provides formal justification for a uniform inference procedure. In particular, the critical value to test H_0 utilizing the statistics $\sup_\beta\widehat{\mu}(\beta)/\widehat{\sigma}(\beta)^{1/2}$ can be obtained by simulating the quantile of the maximum of a Gaussian random vector. The Gaussian random vector shares the same variance-covariance structure as $G_T(\beta)$ on a set of preselected discrete points. Therefore to make inference on $\mu(\beta)$ to test H_0 we can follow the procedure:

Algorithm 1 Uniform inference for averages over $\widehat{\mu}(\beta)$.

Require: $n_t, T \geq 0$

1: Estimate $\tilde{\Sigma}$ as $\hat{\tilde{\Sigma}}$.
2: Simulate standard normal random variables $Z^{(s)}$ of $J_a \times 1$ dimension for $s = 1, \cdots, S$ times, where S is the number of bootstrap samples.
3: Multiply $\tilde{Z}^{(s)} = \hat{\tilde{\Sigma}}^{1/2} Z^{(s)}$, where $\hat{\tilde{\Sigma}} = \mathrm{diag}(\hat{\tilde{\Sigma}})^{-1/2} \hat{\tilde{\Sigma}}\, \mathrm{diag}(\hat{\tilde{\Sigma}})^{-1/2}$.
4: Obtain the $1-\alpha$ quantile of $|\tilde{Z}|_\infty$ from the above sample, and we denote it as $\widehat{q}_{1-\alpha}$.
5: Create the confidence band $[\widehat{\mu}(\beta) - \widehat{\sigma}(\beta)^{1/2}\widehat{q}_{1-\alpha}/\sqrt{T}, \widehat{\mu}(\beta) + \widehat{\sigma}(\beta)^{1/2}\widehat{q}_{1-\alpha}/\sqrt{T}]$, where $\widehat{L}_T(\beta) = \widehat{\mu}(\beta) - \widehat{\sigma}(\beta)^{1/2}\widehat{q}_{1-\alpha}/\sqrt{T}$ and $\widehat{U}_T(\beta) = \widehat{\mu}(\beta) - \widehat{\sigma}(\beta)^{1/2}\widehat{q}_{1-\alpha}/\sqrt{T}$. If 0 is within the confidence band we cannot reject H_0.

5.4 Uniform inference for the high-minus-low estimator

Besides the test regarding the simple null hypothesis H_0, we further show several additional tests that utilize the grand mean estimator. The most common inference procedure in the empirical finance literature is to compare the time-average of returns from the two extreme portfolios (i.e., the portfolios which encompass the evaluation points β_l and β_u) as discussed in Section 2. The goal is to assess whether a long-short portfolio trading strategy earns statistically significant returns, i.e., has a nonzero unconditional risk premium. However, we can use our general framework and new theoretical results to formulate a more powerful test to assess the properties of expected returns. In particular, consider the following null and alternative hypotheses,

$$\mathsf{H}_0^{(1)} : \sup_{\beta \in \mathcal{B}} \mu(\beta) - \inf_{\beta \in \mathcal{B}} \mu(\beta) = 0, \tag{5.6}$$

versus,

$$\mathsf{H}_A^{(1)} : \sup_{\beta \in \mathcal{B}} \mu(\beta) - \inf_{\beta \in \mathcal{B}} \mu(\beta) \neq 0. \tag{5.7}$$

In words, under the null hypothesis there is no profitable long-short strategy available. In the special case when $\mu(\beta)$ is monotonic, then this null hypothesis is equivalent to $\mu(\beta_u) - \mu(\beta_l) = 0$. Thus, we nest the popular high minus low portfolio inference approach but instead test for the presence of *any* profitable long-short strategy.

The high-minus-low statistics can also be re-expressed in the following form,

$$\begin{aligned}\sup_{\beta\in\mathcal{B}}\widehat{\mu}(\beta)-\inf_{\beta\in\mathcal{B}}\widehat{\mu}(\beta) &= \sup_{\beta\in\mathcal{B}}\widehat{\mu}(\beta)+\sup_{\beta\in\mathcal{B}}-\widehat{\mu}(\beta)\\ &= \sup_{\beta\in\mathcal{B}}\Big(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta)\widehat{a}_{jt}\Big)+\sup_{\beta\in\mathcal{B}}\Big(-\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta)\widehat{a}_{jt})\Big),\end{aligned}$$

where we denote β^* as the point attaining $\sup_{\beta\in\mathcal{B}}(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta)\widehat{a}_{jt})$ and β^{**} as the point attaining $\sup_{\beta\in\mathcal{B}}(-\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta)\widehat{a}_{jt}))$. Similar to the previous section, we can obtain a strong approximation results which implies a critical value test $\mathsf{H}_0^{(1)}$ and a uniform confidence band for the proposed high-minus-low estimator. To this end, we define the statistics, $\mathcal{T}_T = [T^{-1}(\sum_{t=1}^{T}\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta^{**})(\widehat{a}_{j,t}-a^*_{j,t}))-T^{-1}(\sum_{t=1}^{T}\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta^{*})(\widehat{a}_{jt}-a^*_{jt}))]/(\widehat{\sigma}(\beta^*)+\widehat{\sigma}(\beta^{**})-2\widehat{\sigma}(\beta^*,\beta^{**}))^{1/2}$ and $\tilde{T}_z = [G_T(\beta^{**})\sigma(\beta^{**})-G_T(\beta^*)\sigma(\beta^*)]/(\sigma(\beta^*)+\sigma(\beta^{**})-2\sigma(\beta^*,\beta^{**}))^{1/2}$.

Lemma 5.4. *Under the conditions of Theorem B.6, 4.1, Assumption 15 and* $\sqrt{T}(1/J\vee h)\to 0$, *we have,*

$$\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\Big|\mathcal{T}_T\Big|\le x\Big)-\mathbb{P}\Big(|\tilde{T}_z|\le x\Big)\Big|\to 0. \tag{5.8}$$

We note that the above lemma is implied by Corollary B.8.1. The above results also imply that we can approximate the quantile of $\Big|\mathcal{T}_T\Big|$ by the quantile of $|\tilde{T}_z|$ uniformly well. Therefore the test based on the statistics $\sup_{\beta\in\mathcal{B}}\widehat{\mu}(\beta)-\inf_{\beta\in\mathcal{B}}\widehat{\mu}(\beta)$ can obtain critical values by simulation from their Gaussian counterparts. The confidence interval can also be obtained similarly from the previous section. We summarise the test procedure in the following algorithm. In short, the algorithm remains quite similar to the previous section, except that the quantiles are obtained from a different vector of Gaussian vectors corresponding to the lemma above.

5.5 Uniform inference for the difference-in-difference estimator

In this section we introduce one final testing setup with the associated test statistic. The null hypothesis and the test statistic can be motivated by a "butterfly" trading strategy which is a generalization of the long-short trading strategy, which represents a discrete first derivative, to that of a discrete second derivative. As discussed earlier, the discrete second derivative also directly links the model to the presence (or absence) of arbitrage opportunities. Moreover, along with the

Algorithm 2 Algorithms for inference of the high-minus-low portfolio.

Require: $n_t, T \geq 0$

1: Estimate $\tilde{\Sigma}$.
2: Simulate standard normal random variables $Z^{(s)}$ of $J_a \times 1$ dimension for $s = 1, \cdots, S$ times.
3: Obtain $\widehat{\mu}(\beta_{\max}) = \max_\beta T^{-1} \sum_{t=1}^{T} \widehat{\mu}_t(\beta_{\max})$ and $\widehat{\mu}(\beta_{\min}) = \min_\beta T^{-1} \sum_{t=1}^{T} \widehat{\mu}_t(\beta_{\min})$. Denote $\mathbb{J}$ as the indices of j over time corresponding to a specific value of β. Obtain $\widehat{\mathbb{J}}^*$ and $\widehat{\mathbb{J}}^{**}$ correspondingly. And $\widehat{\mu}_g = \widehat{\mu}(\beta_{\max}) - \widehat{\mu}(\beta_{\min})$.
4: Multiply $\widehat{\tilde{\Sigma}}^{1/2}$, we get $\tilde{Z}^{(s)} = \widehat{\tilde{\Sigma}}^{1/2} Z^{(s)}$. Obtain $\tilde{Z}^{(s)}_{\widehat{\mathbb{J}}^*} - \tilde{Z}^{(s)}_{\widehat{\mathbb{J}}^{**}}$.
5: Obtain the $1-\alpha$ quantile of $|\tilde{Z}^{(s)}_{\widehat{\mathbb{J}}^*} - \tilde{Z}^{(s)}_{\widehat{\mathbb{J}}^{**}}|$ from the above sample, and we denote as $\widehat{q}_{1-\alpha}$. ($\tilde{Z}^{(s)}_{\widehat{\mathbb{J}}^*}$ $\tilde{Z}^{(s)}_{\widehat{\mathbb{J}}^{**}}$ are the Gaussian limit corresponding to $\beta_{\max}$ and β_{min} repectively.)
6: Create the confidence band $[\widehat{\mu}(\beta) - \widehat{\sigma}(\beta)^{1/2}\widehat{q}_{1-\alpha}/\sqrt{T}, \widehat{\mu}(\beta) + \widehat{\sigma}(\beta)^{1/2}\widehat{q}_{1-\alpha}/\sqrt{T}]$, where $\widehat{L}_T(\beta) = \widehat{\mu}(\beta) - \widehat{\sigma}(\beta)^{1/2}\widehat{q}_{1-\alpha}/\sqrt{T}$ and $\widehat{U}_T(\beta) = \widehat{\mu}(\beta) - \widehat{\sigma}(\beta)^{1/2}\widehat{q}_{1-\alpha}/\sqrt{T}$. If 0 is within the confidence band we cannot reject H_0.

practical relevance of testing for the presence of a profitable butterfly trade, we observe that the statistical properties of the inference procedure are different from those of the preceding inference procedures introduced in this section. In particular, in the previous section we observe that the estimator $\widehat{\mu}(\beta)$ has a non-uniform rate of convergence. This arises for exactly the same reason that for fixed t we can only consistently estimate $M_t(\beta)$ rather than the preferred estimand $\mu_t(\beta)$. However, by taking a discrete second derivative we can eliminate this first-order term because

$$M_t(\beta_1) - M_t(\beta_2) - (M_t(\beta_2) - M_t(\beta_3)) = \mu_t(\beta_1) - \mu_t(\beta_2) - [\mu_t(\beta_2) - \mu_t(\beta_3)],$$

whenever $\beta_1 - \beta_2 = \beta_2 - \beta_3$ with three distinct points $\beta_1, \beta_2, \beta_3 \in [\beta_l, \beta_u]$. As mentioned early, this object can be interpreted as a "butterfly" trade where one goes long one unit of each of two assets (one with β_1 and one with β_3) and short two units of an asset (with β_2). The null hypothesis can then be formulated as

$$\mathsf{H}_{0,\mathtt{diff}}: \sup_{\beta_1+\beta_3-2\beta_2=0} \left| \frac{1}{T} \sum_{t=1}^{T} [\mu_t(\beta_1) + \mu_t(\beta_3) - 2\mu_t(\beta_2)] \right| = 0,$$

versus the alternative one

$$\mathsf{H}_{A,\mathtt{diff}}: \sup_{\beta_1+\beta_3-2\beta_2=0} \left| \frac{1}{T} \sum_{t=1}^{T} [\mu_t(\beta_1) + \mu_t(\beta_3) - 2\mu_t(\beta_2)] \right| \neq 0.$$

In words, under the null hypothesis, there does not exist a profitable version of this trading approach. Thus, we adopt the following test statistic involving:

$$\sup_{\beta_1+\beta_3=2\beta_2} \frac{1}{T}\sum_{t=1}^{T}\{\widehat{\mu}_t(\beta_1)+\widehat{\mu}_t(\beta_3)-2\widehat{\mu}_t(\beta_2)\}.$$

To have valid conference bands and critical values for the test, we shall study the asymptotic distribution of $\widehat{\mu}_t(\beta_1)+\widehat{\mu}_t(\beta_3)-2\widehat{\mu}_t(\beta_2)$. Under the conditions of Theorem 4.1, we have the following leading term expansion

$$\begin{aligned}&\{\widehat{\mu}_t(\beta_1)+\widehat{\mu}_t(\beta_3)-2\widehat{\mu}_t(\beta_2)-(\mu_t(\beta_1)+\mu_t(\beta_3)-2\mu_t(\beta_2))\}\\&=\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}\sum_{i=1}^{n_t}\sum_{j=1}^{J_t}(\widehat{p}_{jt}(\beta_1)+\widehat{p}_{jt}(\beta_3)-2\widehat{p}_{jt}(\beta_2))\tilde{q}_j^{-1}\Phi^*_{i,j,t}\varepsilon_{it}+O_{\mathbb{P}}(h\vee J^{-1}).\end{aligned}$$

Before we show the theoretical results implying the critical value of a test, we first define the normalized variance both in an estimated form and in its population version. As mentioned already, unlike previous subsections, since the term involving $f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1})$ is differenced out, we have a better rate of convergence for all values of β. Namely, we have $\sqrt{T}\sqrt{n_t}/\sqrt{J}$ without the rile of the $\sqrt{T}$ rate induced by the factor term as discussed above. Define the variance of $T^{-1}\sum_{t=1}^{T}\sqrt{Tn_t/J_t}(\widehat{\mu}_t(\beta_1)+\widehat{\mu}_t(\beta_3)-2\widehat{\mu}_t(\beta_2))$ is approximately by
$\hat{\sigma}_D(\beta_1,\beta_3,\beta_2)=T^{-1}\sum_{t=1}^{T}n_t^{-1}J_t^{-1}\sum_{i=1}^{n_t}(\sum_j\mathbb{E}\{(\widehat{p}_{jt}(\beta_1)+\widehat{p}_{jt}(\beta_3)-2\widehat{p}_{jt}(\beta_2))^2\tilde{q}_j^{-2}(\Phi^*_{i,j,t}\sigma_t^2))\}$. We let $\beta_{1,2,3}$ as an abbreviation for β_1,β_2,β_3 and $\beta'_{1,2,3}$ as an abbreviation for $\beta'_1,\beta'_2,\beta'_3$. We define the limit as the following

$$\mathrm{Cov}(\beta_{1,2,3},\beta'_{1,2,3})=\frac{1}{Tn_tJ_t}\sum_{t=1}^{T}\sum_{i=1}^{n_t}\sum_{j=1}^{J_t}\mathbb{E}\{\widehat{p}_{jt}(\beta_{1,2,3})\widehat{p}_{jt}(\beta'_{1,2,3})\tilde{q}_j^{-2}(\Phi^*_{i,j,t}\sigma_t^2)\},$$

where recall that σ_t^2 is defined in Assumption 14.

Define

$$\widehat{p}_{jt}(\beta_1,\beta_2,\beta_3)=\widehat{p}_{jt}(\beta_1)+\widehat{p}_{jt}(\beta_3)-2\widehat{p}_{jt}(\beta_2).$$

Recall $\sigma_j^2 = \tilde{q}_j^{-1}\sigma_t^2/J_t$. Assume that exist a $\sigma(\beta_1,\beta_2,\beta_3) < \infty$, such that

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\mathbb{E}([\widehat{p}_{jt}(\beta_1,\beta_2,\beta_3)]^2\sigma_j^2) = \sigma_D(\beta_{1,2,3}).$$

Define $\sigma_d(\beta) = T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\mathbb{E}(\sigma_t^2\tilde{q}_j^{-1}\widehat{p}_{jt}(\beta))$.

The following theorem states that we can use the quantile of the Gaussian process $\sup_{\beta_1,\beta_2,\beta_3}|G(\beta_1,\beta_2,\beta_3)|$ to approximate the distribution of our statistics of interest under the null. Thus the corresponding algorithm is listed as well afterwards.

Theorem 5.5. *Assume the conditions of Theorem B.6, 4.1, Assumption 13, 15 and* $\sqrt{Tn_u/J}(1/J\vee h)\to 0$. *Also we define* $G(\beta_1,\beta_2,\beta_3)$ *as a Gaussian process with a finite number of jumps corresponding to the value of* β_1,β_2,β_3, *within each piece a standard normal distribution and across different points of the process has correlation*
$Cov(\beta_{1,2,3},\beta'_{1,2,3})/(\sigma_D(\beta_{1,2,3})^{1/2}\sigma_D(\beta'_{1,2,3})^{1/2})$.

$$\sup_x|\mathbb{P}(|\sup_{\beta_1,\beta_2,\beta_3}|T^{-1}\sum_{t=1}^{T}\sqrt{Tn_t/J_t}\{\widehat{\mu}_t(\beta_1)+\widehat{\mu}_t(\beta_3)-2\widehat{\mu}_t(\beta_2)$$
$$-(\mu_t(\beta_1)+\mu_t(\beta_3)-2\mu_t(\beta_2))\}/\widehat{\sigma}_D(\beta_{1,2,3})^{1/2}\geq x)-\mathbb{P}(\sup_{\beta_1,\beta_2,\beta_3}|G(\beta_1,\beta_2,\beta_3)|\geq x)|\to 0,$$

Algorithm 3 Algorithms for inference for the difference-in-difference estimator.

Require: $n_t, T \geq 0$

1: Estimate $\sigma_D(\beta_{1,2,3})$ and $\mathrm{Cov}(\beta_{1,2,3},\beta'_{1,2,3})$. Select a grid of β_1 (J_a) and $\beta_3(\neq\beta_1)$ (J_a) and then fix the relationship $2\beta_2 = \beta_1+\beta_3$.
2: Simulate standard normal random variables $Z^{(s)}$ of $J_a(J_a-1)\times 1$ dimension for $s=1,\cdots,S$ times.
3: Obtain $\widehat{\mu}(\beta_1) = T^{-1}\sum_{t=1}^{T}\widehat{p}_{jt}(\beta_1)\widehat{a}_{jt}$(similar for β_2 and β_3). And $\tilde{Z}_T(\beta_{1,2,3}) = \sup_{\beta_{1,2,3}}\{\widehat{\mu}(\beta_1)+\widehat{\mu}(\beta_3)-2\widehat{\mu}(\beta_2)\}$.
4: Multiplying $\hat{\hat{\Sigma}}_D$($\hat{\hat{\Sigma}}_D$ is a matrix with element as the correlation $\widehat{\mathrm{Cov}}(\beta_{1,2,3},\beta'_{1,2,3})/(\widehat{\sigma}_D^{1/2}(\beta_{1,2,3})\widehat{\sigma}_D^{1/2}(\beta'_{1,2,3})))$, we get $\tilde{Z}^{(s)} = \hat{\hat{\Sigma}}^{1/2}Z^{(s)}$. Obtain $|\tilde{Z}^{(s)}|_{\max}$.
5: Obtain the $1-\alpha$ quantile of $|\tilde{Z}^{(s)}|_{\max}$ from the above sample, and we denote as $\widehat{q}_{1-\alpha}$.
6: Create the confidence interval $[\tilde{Z}_T(\beta_{1,2,3})-\widehat{\sigma}_D(\beta_{1,2,3})^{1/2}\min_t\widehat{q}_{1-\alpha}\sqrt{J_t}/\sqrt{nT}$,
7: $\tilde{Z}_T(\beta_{1,2,3})+\widehat{\sigma}_D(\beta_{1,2,3})^{1/2}\max_t\widehat{q}_{1-\alpha}\sqrt{J_t}/\sqrt{nT}]$. We cannot reject $H_{0,diff}$ if 0 is contained in the confidence interval.

Remark 5.6. *The grand mean allows for inference on* unconditional *risk premia but we would also like to accommodate inference on* conditional *risk premia. For example, a risk factor may be associated with a significant risk premium in certain time periods but unconditionally earns no risk premium. Conversely, the conditional risk premium may be zero under some conditions but not unconditionally. Drawing inferences about conditional risk premia can provide additional information to understand the economic mechanisms underpinning the risk-return trade-off. In particular, we aim to test the following null hypothesis:*

$$\mathsf{H}_0^{diff} : \mu_t(\beta_1) - \mu_t(\beta_2) - [\mu_t(\beta_2) - \mu_t(\beta_3)] = 0. \tag{5.9}$$

Corollary B.11.1 in the Appendix provides justification for the inference procedure below to test H_0^{diff}*. Essentially, the test statistic is a maximum constructed over a finite number of points* $\beta_j, \beta_{j'}$*. As for the previous tests, we shall work with a fixed grid* $[\beta_j]_j$*. Then there are* $J_t \times (J_t - 1)/2$ *total number of such points. Thus, for a vector of* β_v *which are taken in every partition, we can formulate the test statistics as* $|B_{J_t}\widehat{M}_t|_\infty$*, where* $\widehat{M}_t = [\widehat{M}_t(\beta_j)]_j$*, where* B_{J_t} *is a* $J_t(J_t - 1)/2 \times J_t$ *dimensional matrix with row entries corresponding to the linear combinations of* $\left|\widehat{M}_t(\beta_1) - \widehat{M}_t(\beta_2) - \left[\widehat{M}_t(\beta_2) - \widehat{M}_t(\beta_3)\right]\right| = \widetilde{M}_t(\beta_{1,2,3})$.

Algorithm 4 Algorithm for uniform inference concerning H_0^{diff}

Require: $n_t, T \geq 0$

1: Pick $\beta_1 \in \beta_v$ and $\beta_3 (\neq \beta_1) \in \beta_v$, then β_2 follows . Calculate the residuals $\widehat{\varepsilon}_{it}$. Obtain $\widehat{\sigma}_t(\beta_{1,2,3})$.

2: Simulate standard normal random variables $B_{J_t}[\text{diag}(\widehat{\sigma}_j)]Z^{(s)}$ of $J_t \times 1$ dimension for $s = 1, \cdots, S$ times, where S is the number of bootstrap samples. Obtain the $1 - \alpha$ quantile of $|B_{J_t}[\text{diag}(\widehat{\sigma}_j)]Z^{(s)}|_\infty$ from the above sample, and we denote as $\widehat{q}_{1-\alpha,t}$.

3: Create the confidence band for $\mu_t(\beta_1) - \mu_t(\beta_2) - [\mu_t(\beta_2) - \mu_t(\beta_3)]$, i.e.

4: $[\widetilde{M}_t(\beta_{1,2,3}) - \widehat{q}_{1-\alpha,t}\sqrt{J_t}/\sqrt{n_t}, \widetilde{M}_t(\beta_{1,2,3}) + \widehat{q}_{1-\alpha,t}\sqrt{J_t}/\sqrt{n_t}]$. We reject the null $\mathsf{H}_0^{(diff)}$ if $|\sqrt{n_t}B_{J_t}\widehat{M}_t/\sqrt{J_t}|_\infty > \widehat{q}_{1-\alpha,t}$.

---

6 Empirical Application

In this section, we introduce a novel risk factor and show that it is strongly predictive of both the cross-section and time-series behavior of U.S. stock returns. We also utilize this application to illustrate the practical advantages of the novel theoretical results presented earlier in the paper.

Our risk factor is a new measure of the business credit cycle. The business credit cycle is commonly evaluated by means of ratios of credit aggregates to measures of output. Although theoretically appealing, a drawback to these approaches is that it is difficult to parse out movements in credit ratios that are arising from composition changes in the aggregates as compared to all other movements. Here we take a different approach. We rely on the Federal Reserve's Senior Loan Officer Opinion Survey[10] (SLOOS) as our proxy for the "credit" portion of the ratio and the ISM Manufacturing Index as our measure of the "output" portion. This has three distinct advantages. First, as the SLOOS and ISM are both diffusion indices, they have uniform behavior across their history even in the face of changes in the structure of the economy. Second, they are much more timely than credit aggregates and national accounts data which tend to be released with a substantial lag. Third, they are not subject to revision. Thus we have a timely factor which we can evaluate in real time with no look-ahead bias.

Our factor is simply constructed as

$$\mathrm{CCW}_t = \left(\frac{1}{2} \cdot \mathrm{SLOOS}_t + 50\right) + \mathrm{ISM}_t, \tag{6.1}$$

where SLOOS_t is the net percentage of large domestic banks tightening standards for commercial and industrial loans to all firms and ISM_t is the ISM index.[11] Although both the SLOOS and the ISM are diffusion indices, they are scaled differently and so the affine transformation of the SLOOS is implemented so that they are both on the same scale (between 0 and 100). To understand why this is (the inverse of) a credit-to-output type measure note that a fall in the SLOOS corresponds to easier lending standards (higher credit growth) and a fall in the ISM to less output. Thus, when the CCW variable is low, credit-to-output is high. A similar logic applies for when the CCW variable is high. Our factor is available starting in January 1965 when the first SLOOS was implemented.

As a preliminary check for the validity of our factor we assess its ability to predict future market returns. Specifically, a implication of our setup (see equation (2.1)) is that if the factor is

[10]The properties of the SLOOS were first studied in Schreft and Owens (1991), Lown, Morgan, and Rohatgi (2000), and Lown and Morgan (2002, 2006). See also Crump and Luck (2023).

[11]The Senior Loan Officer Opinion Survey is currently conducted on a quarterly basis. To construct a monthly series we keep the SLOOS value constant until a new value is available. For the period from 1984m1 through 1990m1, the credit standards question was not included in the SLOOS. For this period we use as a replacement the net willingness to make consumer installment loans by large domestic banks.

serially correlated then lagged values should be predictive of future equity returns. To show this, we consider the standard predictive regression setup and run predictive regressions of the form,

$$r_{t+h} = a + b \cdot z_t + v_t. \tag{6.2}$$

We utilize the standard predictors obtained from Welch and Goyal (2008) as a benchmark comparison along with our risk factor. In Table 1 we present in-sample R^2 from predictive regressions for forecast horizons of 1, 3, 6, and 12 months ahead. The first fourteen rows present the results for the benchmark predictors investigated in Welch and Goyal (2008). The next row, labelled "CGP" reports results using only the SLOOS portion of our risk factor as in Chava, Gallmeyer, and Park (2015). Finally, The last row, labelled "CCW" provides the results for our new risk factor. The results are stark. The in-sample R^2 from our new risk factor far outstrips that of the other predictors considered. To ensure our results are not a consequence of overfitting, in Table 2 we present out-of-sample R^2 results using a training sample up to the end of 1989. Again, the results are stark with our risk factor outperforming each of the other predictors by a wide margin.

We can now investigate how our risk factor performs in explaining the cross-section of equity returns. We implement our estimators as described in Sections 3 and 4. We use monthly data from the Center for Research in Security Prices (CRSP) over the sample period January 1926 to December 2019. We restrict these data to those firms listed on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), or Nasdaq and use only returns on common shares (i.e., CRSP share code 10 or 11). To deal with delisting returns we follow the procedure described in Bali, Engle, and Murray (2016). When forming market equity we use quotes when closing prices are not available and set to missing all observations with 0 shares outstanding. For our risk factor we use a measure of the business credit cycle described in equation (6.1). For simplicity, we utilize five-year rolling regressions to estimate betas and we choose the number of portfolios as $J_t = J_1 \cdot (n_t/n_{\max})^{\frac{1}{2}}$ where $J_1 = 10$. The latter choice can be motivated by appealing to Cattaneo, Crump, Farrell, and Schaumburg (2020) as the optimal choice of portfolios under the simplifying assumption that all β_{it} were known.

We first consider pointwise inference. Figure 1 presents our estimate of the grand mean, $\hat{\mu}(\beta) = T^{-1}\sum_{t=1}^{T}\hat{\mu}_t(\beta)$ in the black line. There is a clear downward slope in the relationship between β

Table 1: **In-sample Predictive Regressions:** R^2 This table reports R^2 (in percent) from predictive regressions of excess stock returns on an individual predictor variables from Welch and Goyal (2008) for horizons of 1, 3, 6, and 12 months ahead. The row labelled "CGP" reports results for the SLOOS only portion of our risk factor as studied in citeCGP2015. The row labelled "CCW" reports results for our proposed risk factor. The sample period is 1965m1–2019m12.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
(log) Dividend Price Ratio	0.09	0.30	0.72	1.41
(log) Dividend Yield	0.11	0.32	0.75	1.44
(log) Earnings Price Ratio	0.03	0.04	0.06	0.26
(log) Dividend Payout Ratio	0.02	0.20	0.57	0.69
Stock Variance	1.06	0.13	0.13	0.66
Book-to-Market Ratio	0.00	0.01	0.06	0.12
Net Equity Expansion	0.14	0.21	0.44	1.24
Treasury Bill Yield	0.40	0.81	1.13	1.58
Long Term Treasury Yield	0.13	0.18	0.17	0.00
Long Term Treasury Return	1.08	0.66	1.82	1.31
Term Spread	0.51	1.39	2.44	7.30
Default Yield Spread	0.25	0.78	2.13	3.06
Default Return Spread	0.30	0.25	0.30	0.03
(lagged) Inflation	0.01	0.48	1.88	2.21
CGP	1.28	3.06	3.65	4.30
CCW	2.87	7.81	10.55	13.21

and expected returns – although it does not appear to be linear. The grey vertical lines in Figure 1 depict pointwise confidence intervals at each selected point in the support of β. The top chart in Figure 1 uses the plug-in variance estimator we introduced in equation (5.2) whereas the bottom chart uses the Fama and MacBeth (1973) variance estimator. To implement our plug-in variance estimator we use an AR(1) specification in our risk factor. We can clearly see the difference in the precision for drawing inferences from the data. The confidence intervals based on our new plug-in variance estimator are substantially shorter than those of the FM variance estimator. This shows clear evidence that the conservativeness of the FM estimator that was proven in Section 5 has practical implications for empirical work. We can see this even more clearly in Table 3 where we present the point estimate for selected values of β along with lower and upper bounds for confidence intervals constructed with the two different variance estimators. The results are striking. Across all values of β and for both nominal coverage rates, the confidence intervals formed using our plug-in variance estimator are approximately 30% of the length of those using the FM variance estimator.

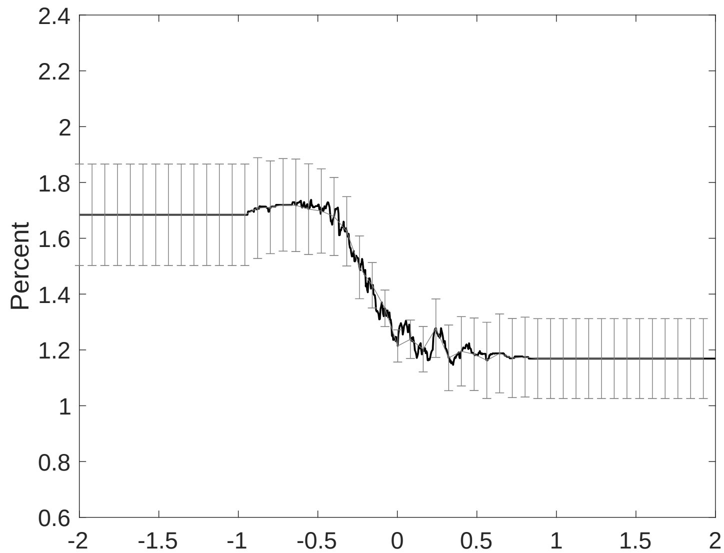
The improved precision of our new variance estimator is replicated when we shift to uniform

Table 2: **Out-of-Sample Predictive Regressions:** R^2 This table reports out-of-sample R^2 from expanding window predictive regressions of excess stock returns on an individual predictor from Welch and Goyal (2008) for horizons of 1,3,6, and 12 months ahead. The row labelled "CGP" reports results for the SLOOS only portion of our risk factor as studied in Chava, Gallmeyer, and Park (2015). The row labelled "CCW" reports results for our proposed risk factor. Positive values have been bolded. The training period is 1965m1–1989m12 and the evaluation sample is 1990m1–2019m12.

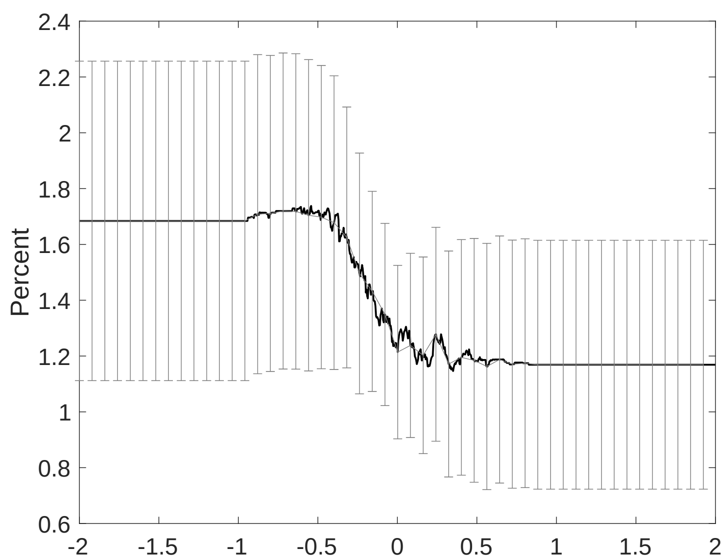
	$h = 1$	$h = 3$	$h = 6$	$h = 12$
(log) Dividend Price Ratio	-1.75	-3.78	-6.75	-12.77
(log) Dividend Yield	-1.83	-3.73	-6.83	-12.27
(log) Earnings Price Ratio	-0.96	-1.98	-3.27	-6.91
(log) Dividend Payout Ratio	-1.33	-1.23	-0.07	**0.82**
Stock Variance	-0.91	-0.50	-0.52	**0.67**
Book-to-Market Ratio	-0.58	-1.15	-2.07	-5.30
Net Equity Expansion	-2.34	-6.43	-13.70	-20.44
Treasury Bill Yield	-0.00	0.77	**1.72**	**2.36**
Long Term Treasury Yield	-0.02	-0.22	-0.65	-4.85
Long Term Treasury Return	-1.18	-0.59	-1.08	-1.66
Term Spread	-0.88	-1.15	**0.49**	**6.94**
Default Yield Spread	-2.28	-3.85	-4.12	-3.04
Default Return Spread	-1.10	**0.70**	**0.45**	**0.22**
(lagged) Inflation	-0.17	**1.44**	**3.88**	**3.50**
CGP	**1.77**	**4.51**	**5.22**	**4.75**
CCW	**3.76**	**9.62**	**11.66**	**8.74**

Figure 1: **Pointwise Inference on Expected Returns.** This figure shows the grand mean estimate, $\hat{\mu}(\beta) = T^{-1}\sum_{t=1}^{T}\hat{\mu}_t(\beta)$ (black line) with associated pointwise confidence intervals (grey vertical lines). The top chart constructs confidence intervals using the plug-in variance estimator introduced in equation (5.2) while the bottom chart uses the Fama-MacBeth variance estimator. The nominal coverage is 95%. The sample period is 1965m1–2019m12.

(a) *Plug-In Variance Estimator*



(b) *Fama-MacBeth Variance Estimator*



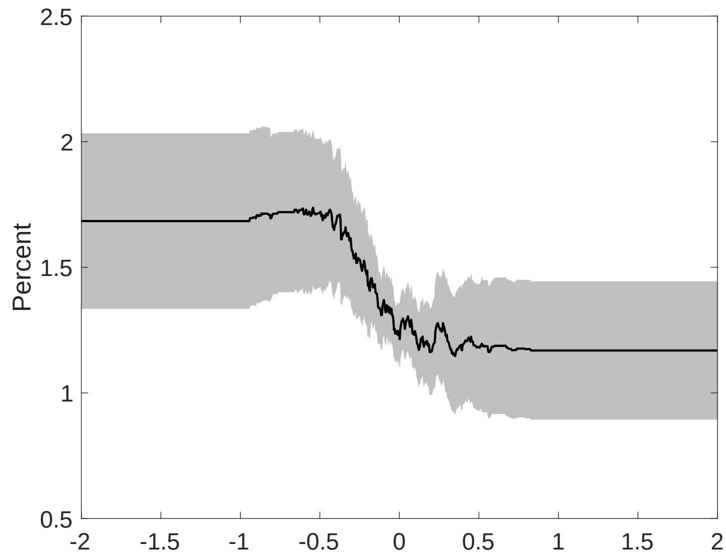
confidence bands for the grand mean rather than pointwise confidence intervals. Figure 2 is the counterpart to Figure 1 with uniform confidence bands constructed as discussed in Section 5. The bottom chart shows the uniform confidence bands formed using the FM variance estimator. The bands are so wide as to be essentially uninformative. In contrast, the top chart displays the confidence bands formed using our new variance estimator. We can reject a constant function and also clearly reject a monotonically decreasing relationship. In contrast, we fail to reject a monontonically decreasing relationship – either linear or nonlinear.

Table 3: **Pointwise Inference**: This table presents the grand mean estimate, $\hat{\mu}(\beta) = T^{-1}\sum_{t=1}^{T}\hat{\mu}_t(\beta)$ along with pointwise upper and lower bounds for nominal coverage of 95% and 99%. Confidence intervals constructed using the plug-in variance estimator introduced in equation (5.2) are denoted by "PI-LB" abd "PI-UB" whereas those using the Fama-MacBeth variance estimator are denoted by "FM-LB" and "FM-UB".The sample period is 1965m1–2019m12.

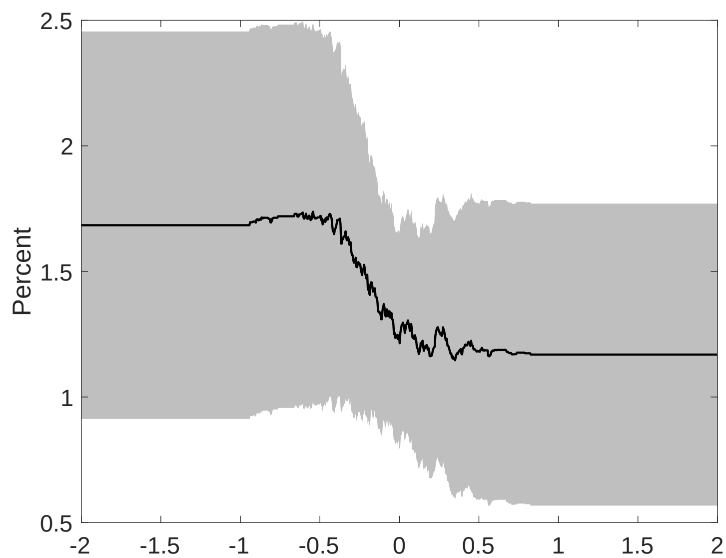
		90% Coverage				*95% Coverage*			
β	$\hat{\mu}(\beta)$	PI-LB	PI-UB	FM-LB	FM-UB	PI-LB	PI-UB	FM-LB	FM-UB
-1.00	1.68	1.50	1.87	1.00	2.26	1.47	1.90	1.00	2.37
-0.50	1.72	1.57	1.87	1.06	2.27	1.54	1.90	1.06	2.38
-0.25	1.53	1.41	1.65	1.01	1.97	1.39	1.67	1.01	2.05
0.00	1.24	1.19	1.30	0.88	1.55	1.18	1.31	0.88	1.61
0.25	1.26	1.15	1.37	0.80	1.65	1.13	1.39	0.80	1.72
0.50	1.18	1.05	1.31	0.66	1.62	1.03	1.34	0.66	1.70
1.00	1.17	1.03	1.31	0.64	1.61	1.00	1.34	0.64	1.70

Figure 2: **Uniform Inference on Expected Returns.** This figure shows the grand mean estimate, $\hat{\mu}(\beta) = T^{-1}\sum_{t=1}^{T}\hat{\mu}_t(\beta)$ (black line) with associated uniform confidence bands (shaded area). The top chart constructs uniform confidence bands using the plug-in variance estimator introduced in equation (5.2) while the bottom chart uses the Fama-MacBeth variance estimator. The nominal coverage is 95%. The sample period is 1965m1–2019m12.

(a) *Plug-In Variance Estimator*



(b) *Fama-MacBeth Variance Estimator*



7 Conclusion

Beta-sorted portfolios are a commonly used empirical tool in asset pricing. In a first step, time-varying factor exposures are estimated by weighted regressions of asset returns on an observable risk factor to ascertain how returns co-move with the variable of interest. In a second step, individual assets are grouped into portfolios by similar factor exposures and differential returns are assessed as a function of differential exposures. Yet the simple and intuitively appealing algorithm belies a more complicated statistical setting involving a two-step estimation procedure where each stage involves non-parametric estimation.

We provide a comprehensive statistical framework which rationalizes this commonly-used estimator. Armed with this foundation we study the theoretical properties of beta-sorted portfolios linking directly to the choice of estimation window in the first step and the number of portfolios in the second step which serves as the tuning parameters for each nonparametric estimator. We introduce conditions that ensure consistency and asymptotic normality for a single cross-section and for the grand mean estimator. We also introduce new uniform inference procedures which allow for more general and varied hypothesis testing than currently available in the literature. However, we also discover some limitations of current practices and provide new guidance on appropriate implementation and interpretation of empirical results.

References

Adrian, T., R. K. Crump, and E. Moench (2015): "Regression-based estimation of dynamic asset pricing models," *Journal of Financial Economics*, 118(2), 211–244.

Andrews, D. W. K. (2005): "Cross-Section Regression with Common Shocks," *Econometrica*, 73(5), 1551–1585.

Ang, A., and D. Kristensen (2012): "Testing conditional factor models," *Journal of Financial Economics*, 106(1), 132–156.

Ang, A., J. Liu, and K. Schwarz (2020): "Using Stocks or Portfolios in Tests of Factor Models," *Journal of Financial and Quantitative Analysis*, 55(3), 709–750.

Bai, J., and G. Zhou (2015): "Fama–MacBeth two-pass regressions: Improving risk premia estimates," *Finance Research Letters*, 15, 31–40.

Bali, T. G., R. F. Engle, and S. Murray (2016): *Empirical Asset Pricing: The Cross Section of Stock Returns*. John Wiley & Sons, Inc., Hoboken, NJ.

Barras, L., P. Gagliardini, and O. Scaillet (2022): "Skill, Scale, and Value Creation in the Mutual Fund Industry," *Journal of Finance*, 77(1), 601–638.

Boons, M., F. Duarte, F. De Roon, and M. Szymanowska (2020): "Time-varying inflation risk and stock returns," *Journal of Financial Economics*, 136(2), 444–470.

Bryzgalova, S. (2015): "Spurious factors in linear asset pricing models," *LSE manuscript*, 1(3).

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng (2022): "On binscatter," *arXiv preprint arXiv:1902.09608*.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and E. Schaumburg (2020): "Characteristic-sorted portfolios: Estimation and inference," *Review of Economics and Statistics*, 102(3), 531–551.

Cattaneo, M. D., M. H. Farrell, and Y. Feng (2020): "Large sample properties of partitioning-based series estimators," *Annals of Statistics*, 48(3), 1718–1741.

Chava, S., M. Gallmeyer, and H. Park (2015): "Credit conditions and stock return predictability," *Journal of Monetary Economics*, 74, 117–132.

Chen, R., and R. Kan (2004): "Finite sample analysis of two-pass cross-sectional regressions," *The Review of Financial Studies (forthcoming)*.

Chen, Y., B. Han, and J. Pan (2021): "Sentiment Trading and Hedge Fund Returns," *Journal of Finance*, 76(4), 2001–2033.

Chordia, T., A. Goyal, and J. A. Shanken (2017): "Cross-sectional asset pricing with individual stocks: betas versus characteristics," *Available at SSRN 2549578*.

Cochrane, J. (2005): *Asset Pricing*. Princeton University Press, Princeton, Revised edition.

Cochrane, J. H. (1996): "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy*, 104(3), 572–621.

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

Connor, G., M. Hagmann, and O. Linton (2012): "Efficient semiparametric estimation of the Fama–French model and extensions," *Econometrica*, 80(2), 713–754.

Connor, G., S. Li, and O. B. Linton (2021): "A Dynamic Semiparametric Characteristics-based Model for Optimal Portfolio Selection," Research Paper 21-1, Michael J. Brennan Irish Finance Working Paper Series.

Connor, G., and O. Linton (2007): "Semiparametric estimation of a characteristic-based factor model of common stock returns," *Journal of Empirical Finance*, 14(5), 694–717.

Crump, R. K., and S. Luck (2023): "The SLOOS and Economic Downturns," Working paper.

Eisdorfer, A., K. Froot, G. Ozik, and R. Sadka (2021): "Competition Links and Stock Returns," *Review of Financial Studies*.

Fama, E. F., and K. R. French (2008): "Dissecting anomalies," *The Journal of Finance*, 63(4), 1653–1678.

Fama, E. F., and J. D. MacBeth (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81(3), 607–636.

Fan, J., Z. T. Ke, Y. Liao, and A. Neuhierl (2022): "Structural Deep Learning in Conditional Asset Pricing," .

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

Fan, Z., J. M. Londono, and X. Xiao (2022): "Equity tail risk and currency risk premiums," *Journal of Financial Economics*, 143(1), 484–503.

Freyberger, J., A. Neuhierl, and M. Weber (2020): "Dissecting Characteristics Nonparametrically," *Review of Financial Studies*, 33(5), 2326–2377.

Gagliardini, P., E. Ossola, and O. Scaillet (2016): "Time-Varying Risk Premium in Large Cross-Sectional Equity Data Sets," *Econometrica*, 84(3), 985–1046.

——— (2020): "Estimation of large dimensional conditional factor models in finance," *Handbook of Econometrics*, 7, 219–282.

Garleanu, N., and L. H. Pedersen (2011): "Margin-based asset pricing and deviations from the law of one price," *The Review of Financial Studies*, 24(6), 1980–2022.

Giglio, S., and D. Xiu (2021): "Asset pricing with omitted factors," *Journal of Political Economy*, 129(7), 000–000.

Goldberg, J., and Y. Nozawa (2021): "Liquidity Supply in the Corporate Bond Market," *Journal of Finance*, 76(2), 755–796.

Gospodinov, N., R. Kan, and C. Robotti (2014): "Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors," *The Review of Financial Studies*, 27(7), 2139–2170.

Gospodinov, N., and C. Robotti (2013): "Asset Pricing Theories, Models, and Tests," in *Portfolio Theory and Management*, ed. by H. K. Baker, and G. Filbeck, chap. 3. Oxford University Press.

Goyal, A. (2012): "Empirical cross-sectional asset pricing: a survey," *Financial Markets and Portfolio Management*, 26(1), 3–38.

Jagannathan, R., and Z. Wang (1998): "An asymptotic theory for estimating beta-pricing models using cross-sectional regression," *The Journal of Finance*, 53(4), 1285–1309.

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

Kleibergen, F. (2009): "Tests of risk premia in linear factor models," *Journal of econometrics*, 149(2), 149–173.

Kleibergen, F., K. Lingwei, and Z. Zhan (2019): "Identification robust testing of risk premia in finite samples," *The Journal of Financial Econometrics*.

Lown, C., and D. P. Morgan (2006): "The Credit Cycle and the Business Cycle: New Findings Using the Loan Officer Opinion Survey," *Journal of Money, Credit and Banking*, 38(6), 1575–1597.

Lown, C. S., and D. P. Morgan (2002): "Credit Effects in the Monetary Mechanism," *FRBNY Economic Policy Review*, 8(1), 217–235.

Lown, C. S., D. P. Morgan, and S. Rohatgi (2000): "Listening to Loan Officers: The Impact of Commercial Credit Standards on Lending and Output," *FRBNY Economic Policy Review*, 6(2), 1–16.

Nagel, S. (2013): "Empirical cross-sectional asset pricing," *Annu. Rev. Financ. Econ.*, 5(1), 167–199.

Raponi, V., C. Robotti, and P. Zaffaroni (2020): "Testing Beta-Pricing Models Using Large Cross-Sections," *The Review of Financial Studies*, 33(6), 2796–2842.

Schreft, S. L., and R. E. Owens (1991): "Survey Evidence of Tighter Credit Conditions: What Does It Mean?," *Federal Reserve Bank of Richmond Economic Quarterly*, 77, 29–34.

Shanken, J., and G. Zhou (2007): "Estimating and testing beta pricing models: Alternative methods and their performance in simulations," *Journal of Financial Economics*, 84(1), 40–86.

Welch, I., and A. Goyal (2008): "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21, 1455–1508.

Zhang, T., and W. B. Wu (2015): "Time-varying nonlinear regression models: nonparametric estimation and model selection," *The Annals of Statistics*, 43(2), 741–768.

In this subsection, we show the theoretical properties of the first-step estimation of α_{it} and β_{it} as in the model (2.1). For a time series $X_{t,l}$, we define $X_{t,l} = g_l(\xi_t, \xi_{t-1}, \cdots, \xi_{-\infty})$. $\delta_{q,m}(X_{.,l}) = \max_t \|X_{t,l} - X^*_{t-m,l}\|_q$ with $X^*_{t,l}$ as a process with ξ_0 replaced by ξ^*_0 (i.i.d. copy of ξ_0). We define $\Theta_{m^0,q}(f_.) = \sum_{m\geq m^0} \delta_{q,m}(f_.)$, with m^0 as an integer. Without loss of generality, we shall assume that $n \leq n_t \leq n_u$ (and without loss of generality n_t and n_u are of the same order) throughout the section. Define $\tilde{A}(t_0) = \sum_t \mathbb{E}(X_t X_t^\top|\mathcal{F}_{t-1}) w(t,t_0)$ and $\tilde{B}_i(t_0) = \sum_t \mathbb{E}(X_t R_{it}|\mathcal{F}_{t-1}) w(t,t_0)$. Define $A(t_0) = \sum_t X_t X_t^\top w(t,t_0)$ and $B_i(t_0) = \sum_t X_t R_{it} w(t,t_0)$. Recall that $w(t,t_0) = h^{-1}K((t-t_0)/(Th))$. We suppress the dependency of t_0 by the elements as in $\tilde{A}(t_0), \tilde{B}_i(t_0), A(t_0)$. We define $\overline{B}_i(t_0) = \overline{B}_i(t_0) = \sum_t \mathbb{E}(X_t X_t^\top|\mathcal{F}_{t-1}) w(t,t_0) b_{it_0} = \tilde{A} b_{it_0}$. We define $r_T = (Th)^{-1}(T^{1/q} + (Th\log T)^{1/2})$, for an integer $q > 4$.

Assumption 1. *We assume that the bounded differentiable one-sided kernel function $K(.)$ takes support in [-1,0], and satisfies $\int_{-1}^0 K(u)du = 1$. $K(.) \in C^2[-1,0]$. $T \to \infty, h \to 0, Th \to \infty$.*

Assumption 2. *Assume that $f_t = \tau'(t/T) + x_t$, where x_t is a stationary process. The trend function $\tau'(.)$ is bounded by c_τ and is second order differentiable. $\mathbb{E}(x_t) = 0$. We assume that $\mathbb{E}(\varepsilon_{it}|\mathcal{F}_{t-1}, x_t) = 0$ and it implies that $\mathbb{E}(\varepsilon_{it}|\mathcal{F}_{t-1}) = 0$. We define $\mathbb{E}(\varepsilon_{it}^2|\mathcal{F}_{t-1}, x_t) = \sigma(\mathcal{F}_{t-1}, t/T)^2 = \sigma_t^2$, and define $\mathbb{E}x_t = 0$. $\|\varepsilon_{it}\|_{2q} \leq c$. $\mathbb{E}(\varepsilon_{it}^2|\mathcal{F}_{t-1}, x_{t-1}) = \mathbb{E}(\varepsilon_{it}^2|\mathcal{F}_{t-1})$. $\sigma^2_{\varepsilon,0} = \mathbb{E}(\varepsilon^2_{t_0}) = \mathbb{E}(\sigma^2_{t_0}) > 0$. The error term ε_{it_0} has finite qth moment with $q > 4$. $(T)^{2/q-1} n_u^{2/q} \ll h$.*

It shall be noted that $T^{2/q-1} n_u^{2/q} \ll h$ favors a high moment condition for example if $q = 8$, we just need $T^{-3/4} n^{1/4} \ll h$.

Assumption 3. *There exists a constant $c, C_{A,\max}$ such that $\min_{t_0} \lambda_{\min}(T^{-1}\mathbb{E}(\tilde{A}(t_0))) > c^{-1} > 0$, $c^{-1} < \min_{t_0} \lambda_{min}(T^{-1}\mathbb{E}A(t_0)) \leq \max_{t_0} \lambda_{max}(T^{-1}\mathbb{E}A(t_0)) < C_{A,\max}$ and $\min_{t_0} \lambda_{\min}(T^{-1}\mathbb{E}(B_i(t_0))) > c^{-1} > 0$. $\max_{t_0} \lambda_{\max}(T^{-1}\mathbb{E}(B_i(t_0))) < C_B$ for a positive constant C_B.*

Assumption 4. *We define $\sigma_x^2 = \mathbb{E}(x_t^2) < \infty$ and $\sigma_{x4} = \mathbb{E}(x_t^4) < \infty$, and we assume that they are both bounded by a positive constant $c_{x,q}$. We define the constant $c_x = 2\sigma_x^2 \int_{-1}^0 K(s)^2 ds$. We denote $\Sigma_x = \mathbb{E}[(1,x_t)^\top(1,x_t)]$, $\tilde{\tau}(t_0/T) = (1, \tau'(t_0/T))^\top$. We let $\Sigma_A = \Sigma_x + \tilde{\tau}(t_0/T)\tilde{\tau}(t_0/T)^\top$, $\Sigma_B = \sigma^2_{\varepsilon,0}\mathbb{E}(X_{t_0}X_{t_0}^\top)\int_{-1}^0 K^2(s)ds$. Assume that both Σ_A and Σ_B has eigenvalues bounded from the below and the above.*

Assumption 5. *(Lipschitz condition) We assume that the $\alpha_{it} = \alpha_i(t/T, \mathcal{F}_{t-1})$ and $\beta_{it} = \beta_i(t/T, \mathcal{F}_{t-1})$ satisfying for any $t, t' \in [Th, T-Th]$, $|\alpha_{it} - \alpha_{it'}| \leq C_\alpha(\mathcal{F}_{t-1})|t-t'|/T$ and $|\beta_{it} - \beta_{it'}| \leq C_\beta(\mathcal{F}_{t-1})|t-t'|/T$, where $C_\alpha(\mathcal{F}_{t-1}), C_\beta(\mathcal{F}_{t-1})$ are two $\mathcal{F}_{t-1}$ measurable functions. Moreover,*
$\max_t |C_\alpha(\mathcal{F}_{t-1})|, \max_t |C_\beta(\mathcal{F}_{t-1})|$ are bounded by constants C_α, C_β. Assume α_{it}, β_{it} are bounded uniformly over i, t .

Assumption 6. *In particular for any positive integer m, we assume that $\Theta_{m,2q}(f_t) = m^{-v}$ and $\max_t \|f_t\|_{2q}$ is bounded, for $2v > 1/2 - 1/q$, with $q > 4$, for a positive constant v.*

Define $r_T = (Th)^{-1}(T^{1/q} + (Th\log T)^{1/2})$. Let $q_n = J_u \vee n_u \vee T$. $h \vee r_T \vee \sqrt{\log(q_n)}/\sqrt{Th} = \delta_T$.

Assumption 7. *$F_{\beta,t}(x)$ is continuously differentiable on the compact interval $B_\delta = [\beta_l - \delta_T, \beta_u + \delta_T]$ (δ_T is a positive constant). β_{it} are i.i.d. conditioning on $\mathcal{G}_{t-1}$(sigma field of time invariant factors). ε_{it} are independent conditioning on all three of the filtration $\cup(\mathcal{F}_{t-1}, \sigma(f_t))$, $\mathcal{G}_{t-1}$ and $\mathcal{F}_{t-1}$ respectively. The condition density of β_{it} is denoted as $f_{\beta,t}(x)$. $C_{\beta,max} > \max_t \max_{x\in B_\delta} f_{\beta,t}(x) > \min_t \min_{x\in B_\delta} f_{\beta,t}(x) > c_{\beta,min} > 0$, which is also first order continuous differentiable with bounded derivatives. $\mathbb{E}(\beta_{it}\Phi^*_{it}|\mathcal{G}_{t-1}) \asymp_p J_t^{-1}$, $\min_{j_t} \mathbb{E}_{t-1}(\Phi^*_{i,j_t,t}\varepsilon^2_{it}) \asymp_p J_t^{-1}$ and $\tilde{q}_{j_t} = \int_{F^{-1}_{\beta,t}(\kappa_{j-1})}^{F^{-1}_{\beta,t}(\kappa_j)} f_{\beta,t} d\beta \asymp_p J_t^{-1}$.*

Define $a_{nT} = \max_t(\sqrt{\delta_T} c_{n_u T}/\sqrt{n_t}) \vee \sqrt{\log T}\sqrt{n_t}^{-1}$, where $c_{n_u T}$ is a positive constant.

Assumption 8. *We let $\max_t n_t \leq n_u$, $n \leq n_t$, $J \leq J_t \leq J_u$, $J_u \asymp J$ and $n_u \asymp n$. Recall $r_T = (Th)^{-1}(T^{1/q} + (Th\log T)^{1/2})$, with $q > 4$. We assume that Assumptions 1 to 6 are maintained such that $\max_i \sup_t |\widehat{\beta}_{it} - \beta_{it}| \lesssim_{\mathbb{P}} \delta_T$ and $J_u^{-1} \gg \frac{\sqrt{\log(q_n)}}{\sqrt{n}}$. $\delta_T \to 0$, $r_T \to 0$. We assume that $a_{nT} \to 0$.*

We note that the above assumption implies that $J \log q_n/n \ll 1$. Define $\delta = (\delta_T + a_{nT})^{1/2}\sqrt{\log q_n}/\sqrt{n}$.

Assumption 9. $J(\delta_T + a_{nT})^{1/2}\sqrt{\log q_n}/\sqrt{n} \ll 1$.

Assumption 10. *We assume that $\alpha(\beta)$ is continuously differentiable of the first order, and with the first derivative bounded from the above by positive constant c_α.*

We first order observations as $\ell = \ell(i,t) = \sum_{t_0=1}^{t-1} n_t + i$, $1 \le i \le n_t, 1 \le t_0 \le T$. We let $\mathcal{F}_{\ell-1}^{\beta}$ denote the sigma field of β_ℓ up to the order of $\ell - 1$.

Assumption 11. *We let $\tilde{\beta}_{\ell,j} = \beta_{\ell,j} - F_{\beta,t}^{-1}(\kappa_j)$. $\tilde{\beta}_{\ell,j}$ are different over time, however the dependence can still decay as the series of β_{it}. We assume that $\mathbf{1}(-u \le \tilde{\beta}_{\ell,j} < u) = \psi_{\ell,j}(u)$. We assume that $\max_{\ell,j}\mathbb{E}(\beta_{\ell,j}^2\mathbf{1}(-u \le \tilde{\beta}_{\ell,j} < u)|\mathcal{F}_{\ell-1}^{\beta}) \le C\beta_u^2 u$, $\max_{\ell,j}\mathbb{E}(\varepsilon_\ell^2\mathbf{1}(-u \le \tilde{\beta}_{\ell,j} < u)|\mathcal{F}_{\ell-1}^{\beta}) \le C\beta_u^2 u$, for a constant C. Moreover we assume that $\|\max_j \mathbb{E}(\psi_{\cdot,j}(u)|\mathcal{F}_{\cdot-1}^{\beta})\|_{q,\zeta} \le u^{1/q}C_{q,\zeta}$, $\|\max_j \mathbb{E}(\psi_{\cdot,j}(u)\beta_{\cdot,j}|\mathcal{F}_{\cdot-1}^{\beta})\|_{q,\zeta} \le u^{1/q}C'_{q,\zeta}$, for an integer $q > 4$.*

Define $\overline{\delta}_T = \sqrt{\log q_n}/\sqrt{nT} \vee h \vee \sqrt{\log(q_n)}/\sqrt{nTh}$.

Assumption 12. $(a_{nT}/\sqrt{T} + \overline{\delta}_T) \vee J(a_{nT} + \delta_T)^{1/2}/\sqrt{nT} \lesssim h \vee J^{-1} \ll \sqrt{T}^{-1}$.

We shall give an example on the plausible rate of n, T, J, h which is admissible to the above assumptions. For example, we can assume that $n = T$, $J = O(T^{1/3})$ and $h = O(T^{-1/3})$.

We notice that the conditions $\|\max_j \mathbb{E}(\psi_{\cdot,j}(u)\beta_{\cdot}|\mathcal{F}_{\cdot-1}^{\beta})\|_{q,\zeta} \le u^{1/q}C'_{q,\zeta}$ are easily satisfied. Let us illustrate for the stationary case of β_ℓ. For example if we assume that $f_{\beta,t}(\beta|\mathcal{F}_{t-1}^{\beta})$ is differentiable with respect to β and its i.i.d. innovation ε_0 (slightly abuse of notation), then we can derive that,

$$
\begin{aligned}
& \mathbb{E}(\psi_{\cdot,j}(u)\beta_{\ell,j}|\mathcal{F}_{\ell-1}^{\beta}) - \mathbb{E}(\psi_{\cdot,j}(u)\beta_{\ell,j}|\mathcal{F}_{\ell-1}^{\beta*}) \\
= & \int_{-u+F_{\beta,t}^{-1}(\kappa_j)}^{u+F_{\beta,t}^{-1}(\kappa_j)} \beta f_{\beta,t}(\beta|\mathcal{F}_{\ell-1}^{\beta})d\beta - \int_{-u+F_{\beta,t}^{-1}(\kappa_j)}^{u+F_{\beta,t}^{-1}(\kappa_j)} \beta f_{\beta,t}(\beta|\mathcal{F}_{\ell-1}^{\beta*})d\beta \\
\le & \; 2u|\varepsilon_0 - \varepsilon_0^*||\beta_u \partial f(\tilde{\beta}|\tilde{\mathcal{F}}_{\ell-1})/(\partial\varepsilon_0\partial\beta)|,
\end{aligned}
$$

where $\tilde{\beta}$ is a point between the intersection of $\cap_j(-u+F_{\beta,t}^{-1}(\kappa_j), u+F_{\beta,t}^{-1}(\kappa_j))$, and $\tilde{\mathcal{F}}_{\ell-1}$ is the filtration with ε_0 replaced by some value. We take the $\|.\|_q$ norm of the above object. If we can ensure that $|\varepsilon_0 - \varepsilon_0^*||\beta_u|\|\partial f(\tilde{\beta}|\tilde{\mathcal{F}}_{\ell-1})/(\partial\varepsilon_0\partial\beta)\|_q$ decrease sufficient fast according to the lag ℓ, then the conditions holds. We let $\delta' = \Delta_T/\sqrt{T}$. Define $\mathbb{N}_J = \bigcup_{t=1\cdots,T} N_{J_t}$. $\overline{\delta}_T = \sqrt{\log q_n}/\sqrt{nT} \vee h \vee \sqrt{\log(q_n)}/\sqrt{nTh}$.

Assumption 13. *Let $\widehat{\sigma}_D^{1/2}(\beta_{1,2,3})$ be a consistent estimator for $\sigma_D(\beta_{1,2,3})$ satisfying $\widehat{\sigma}_D(\beta_{1,2,3})^{1/2} - \sigma_D(\beta_{1,2,3})^{1/2} = o_{\mathbb{P}}(r_{1,2,3})$, and the rate $r_{1,2,3} \to 0$. $\sigma_d(\beta)$ is bounded from the below and the above uniformly over β. $\max_{j,t}\|\Phi_{i,j,t}^*\varepsilon_{it}\tilde{q}_j^{-1}\|_q \lesssim J^{1-1/q}$.*

Since we have $1/J_t * {J_t}^2({J_t}^{-1}(\delta_T + a_{nT}) + \sqrt{\delta_T}T^{1/2q}(n_t^{-1/2}{J_t}^{-1/2})) = o_{\mathbb{P}}(\delta_2)$ and $(1/J_t(l_{n,T}/n_t + J_t^{-1})J_t^2(l_{n,T}/n_t)J_t^{-1}J_t^2) = o_{\mathbb{P}}(\delta_1)$. Recall that $\delta_1 = \sqrt{a_{nT} + \delta_T}J_t/\sqrt{n_t} + a_{nT} + \delta_T$, $\delta_2 = \delta_T + a_{nT} + \sqrt{\delta_T}T^{1/(2q)}n_t^{-1/2}J_t^{1/2}$.

Assumption 14. *Let $\varepsilon_{it} =_d \sigma_t\eta_{it}$ conditional on $\mathcal{F}_{t-1}$, with $\sigma_t^2 = \mathbb{E}(\varepsilon_{it}^2|\mathcal{F}_{t-1})$, and η_{it} be a standard Gaussian random variable defined on a proper probability space. Exists two positive constants $c, C > 0$,*

$$c \le \min_j \sigma_j \le \max_j \sigma_j \le C.$$

And we have $\delta_{1T} + \delta_{2T} << 1/\sqrt{\log J}$. $n_t^{-1/2+1/(2q)}\sqrt{J_t} << \sqrt{J_t}^{-1}$. Moreover, $\sqrt{n_t/J_t}(h \vee J_t^{-1}) \to 0$.

Assumption 15. *Assume that with probability one, $E_{n_t,j}/\sigma_j s$ are bounded from the below and the above for all t, j. $\sup_\beta\{\widehat{\sigma}(\beta)^{1/2} - \sigma(\beta)^{1/2}\} = O_{\mathbb{P}}(r_\sigma)$ for some constant r_σ. Assume $r_\sigma T^{-1/2+1/2q}J_a^{1/2q} \to 0$ and $r_\sigma \to 0$. $c \le \inf_\beta \sigma(\beta) = \sup_\beta \sigma(\beta) \le C$. $c \le \min_\beta \tilde{\sigma}(\beta) = \sup_\beta \tilde{\sigma}(\beta) \le C$. $\|E_{n_t,j} - E_{n_t,j-1}\|_{2q} \le cJ_t^{-1}$, for a positive constant $c > 0$. $\|f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})\|_{2q} < C$, for a positive $C > 0$. Assume that the grid is $\overline{\beta}_v = [\beta_1, \beta_2, \cdots, \beta_{c/\tilde{\delta}}]$. This corresponds to $J_{a,\delta}$ distinct value of β. We shall assume that for avoiding singularity of the variance covariance matrix. We need to ensure that $\sigma(\beta_j, \beta_{j'}) \ne \sigma(\beta_j, \beta_j)$ or $\sigma(\beta_{j'}, \beta_{j'})$. Let $[\widehat{\sigma}(\beta_i, \beta_{i'})/\widehat{\sigma}^{1/2}(\beta_i)\widehat{\sigma}(\beta_{i'})^{1/2}]_{i,i'} = \Sigma_{J_{a,\delta},J_{a,\delta}}$. We shall assume that $c < \lambda_{\min}(\Sigma_{J_a,\delta}) < \lambda_{\max}(\Sigma_{J_a,\delta}) < C'$, with $C', c > 0$. Let $J_{a,\delta} \lesssim \exp(T^{\varepsilon'})$, with $\varepsilon' = 1/9$.*

Beta-Sorted Portfolios
Supplemental Appendix*

Matias D. Cattaneo[†] Richard K. Crump[‡] Weining Wang[§]

August 2, 2023

*Cattaneo gratefully acknowledges financial support from the National Science Foundation (SES-1947662 and SES-2241575).

[†]Department of Operations Research and Financial Engineering, Princeton University.

[‡]Macrofinance Studies, Federal Reserve Bank of New York.

[§]Department of Economics and Related Studies, University of York,Weining Wang's research is partially supported by the ESRC (Grant Reference: ES/T01573X/1)

Contents

1 Setup

1.1 Assumptions

In this subsection, we show the theoretical properties of the first-step estimation of α_{it} and β_{it}. For a time series $X_{t,l}$, we define $X_{t,l} = g_l(\xi_t, \xi_{t-1}, \cdots, \xi_{-\infty})$. $\delta_{q,m}(X_{.,l}) = \max_t \|X_{t,l} - X^*_{t-m,l}\|_q$ with $X^*_{t,l}$ as a process with ξ_0 replaced by ξ^*_0 (i.i.d. copy of ξ_0). We define $\Theta_{m^0,q}(f_.) = \sum_{m \geq m^0}^{\infty} \delta_{q,m}(f_.)$, with m^0 as an integer. Without loss of generality, we shall assume that $n \leq n_t \leq n_u$ (and without loss of generality n_t and n_u are of the same order) throughout the section. Define $\tilde{A}(t_0) = \sum_{t=1}^T \mathbb{E}(X_t X_t^\top | \mathcal{F}_{t-1}) w(t, t_0)$ and $\tilde{B}_i(t_0) = \sum_{t=1}^T \mathbb{E}(X_t R_{it} | \mathcal{F}_{t-1}) w(t, t_0)$. Define $A(t_0) = \sum_{t=1}^T X_t X_t^\top w(t, t_0)$ and $B_i(t_0) = \sum_{t=1}^T X_t R_{it} w(t, t_0)$. Recall that $w(t, t_0) = h^{-1} K((t - t_0)/(Th))$. We suppress the dependency of t_0 by the elements as in $\tilde{A}(t_0), \tilde{B}_i(t_0), A(t_0)$. We define $\overline{B}_i(t_0) = \overline{B}_i(t_0) \stackrel{\text{def}}{=} \sum_{t=1}^T \mathbb{E}(X_t X_t^\top | \mathcal{F}_{t-1}) w(t, t_0) b_{it_0} = \tilde{A} b_{it_0}$. We define $r_T = (Th)^{-1}(T^{1/q} + (Th \log T)^{1/2})$, for an integer $q > 4$.

Assumption 1. *We assume that the bounded differentiable one-sided kernel function $K(.)$ takes support in [-1,0], and satisfies $\int_{-1}^0 K(u)du = 1$. $K(.) \in C^2[-1, 0]$. $T \to \infty, h \to 0, Th \to \infty$.*

Assumption 2. *Assume that $f_t = \tau'(t/T) + x_t$, where x_t is a stationary process. The trend function $\tau'(.)$ is bounded by c_τ and is second order differentiable. $\mathbb{E}(x_t) = 0$. We assume that $\mathbb{E}(\varepsilon_{it} | \mathcal{F}_{t-1}, x_t) = 0$ and it implies that $\mathbb{E}(\varepsilon_{it} | \mathcal{F}_{t-1}) = 0$. We define $\mathbb{E}(\varepsilon_{it}^2 | \mathcal{F}_{t-1}, x_t) = \sigma(\mathcal{F}_{t-1}, t/T)^2 \stackrel{\text{def}}{=} \sigma_t^2$. $\|\varepsilon_{it}\|_{2q} \leq c$, and $c > 0$. $\mathbb{E}(\varepsilon_{it}^2 | \mathcal{F}_{t-1}, x_{t-1}) = \mathbb{E}(\varepsilon_{it}^2 | \mathcal{F}_{t-1})$. $\sigma^2_{\varepsilon,0} = \mathbb{E}(\varepsilon^2_{t_0}) = \mathbb{E}(\sigma^2_{t_0}) > 0$. The error term ε_{it_0} has finite qth moment with $q > 4$. $(T)^{2/q-1} n_u^{2/q} \ll h$.*

It shall be noted that $T^{2/q-1} n_u^{2/q} \ll h$ favors a high moment condition for example if $q = 8$, we just need $T^{-3/4} n^{1/4} \ll h$.

Assumption 3. *There exists a constant $c, C_{A,\max}$ such that $\min_{t_0} \lambda_{\min}(T^{-1} \mathbb{E}(\tilde{A}(t_0))) > c^{-1} > 0$, $c^{-1} < \min_{t_0} \lambda_{min}(T^{-1} \mathbb{E}[A(t_0)]) \leq \max_{t_0} \lambda_{max}(T^{-1} \mathbb{E}[A(t_0)]) < C_{A,\max}$ and $\min_{t_0} \lambda_{\min}(T^{-1} \mathbb{E}(B_i(t_0))) > c^{-1} > 0$. $\max_{t_0} \lambda_{\max}(T^{-1} \mathbb{E}(B_i(t_0))) < C_B$ for a positive constant C_B.*

Assumption 4. *We define $\sigma_x^2 = \mathbb{E}(x_t^2) < \infty$ and $\sigma_{x4} = \mathbb{E}(x_t^4) < \infty$, and we assume that they are both bounded by a positive constant $c_{x,q}$. We define the constant $c_x = 2\sigma_x^2 \int_{-1}^0 K(s)^2 ds$. We denote $\Sigma_x = \mathbb{E}[(1, x_t)^\top (1, x_t)]$, $\tilde{\tau}(t_0/T) = (1, \tau'(t_0/T))^\top$. We let $\Sigma_A = \Sigma_x + \tilde{\tau}(t_0/T) \tilde{\tau}(t_0/T)^\top$, $\Sigma_B = \sigma^2_{\varepsilon,0} \mathbb{E}(X_{t_0} X_{t_0}^\top) \int_{-1}^0 K^2(s) ds$. Assume that both Σ_A and Σ_B has eigenvalues bounded from the below and the above.*

Assumption 5. *(Lipschitz condition) We assume that the $\alpha_{it} = \alpha_i(t/T, \mathcal{F}_{t-1})$ and $\beta_{it} = \beta_i(t/T, \mathcal{F}_{t-1})$ satisfying for any $t, t' \in [Th, T - Th]$, $|\alpha_{it} - \alpha_{it'}| \leq C_\alpha(\mathcal{F}_{t-1}) |t - t'|/T$ and $|\beta_{it} - \beta_{it'}| \leq C_\beta(\mathcal{F}_{t-1}) |t - t'|/T$, where $C_\alpha(\mathcal{F}_{t-1}), C_\beta(\mathcal{F}_{t-1})$ are two $\mathcal{F}_{t-1}$ positive measurable functions. Moreover, $\max_t |C_\alpha(\mathcal{F}_{t-1})|, \max_t |C_\beta(\mathcal{F}_{t-1})|$ are bounded by constants C_α, C_β. Assume α_{it}, β_{it} are bounded uniformly over i, t .*

Remark 1.1. Note that under our Assumptions we have that

$$\mathbb{E}T^{-1}\sum_{t=1}^{T}x_t^2 w(t,t_0)=\sigma_x^2 T^{-1}\sum_{t=1}^{T}w(t,t_0)=\sigma_x^2. \tag{1.1}$$

And

$$T^{-1}\sum_{t=1}^{T}\tau'(t/T)^2 w(t,t_0)=\tau'(t_0/T)^2+O(h). \tag{1.2}$$

$$\mathrm{Var}(\sqrt{h/T}^{-1}\sum_{t=1}^{T}w(t,t_0)\varepsilon_{it}f_t)=\sigma_{\varepsilon,0}^2(\sigma_x^2+\tau'(t_0/T)^2)\int_{-1}^{0}K(s)^2ds+O(h). \tag{1.3}$$

Thus we conclude $T^{-1}\mathbb{E}(A)\overset{\text{def}}{=}\Sigma_x+\tilde{\tau}(t_0/T)\tilde{\tau}(t_0/T)^\top+O(h)$. ┘

Remark 1.2 (Admissible processes of β)**.** Suppose that $\beta_{it}=G_\beta(\eta_i,g_1,\cdots,g_{t-1},f_1,\cdots,f_{t-1},\omega_{it-1})$, where η_i is iid over i, g_t is i.i.d. factors over t, and ω_{it-1} are i.i.d. over t and i. We denote the common factor sigma field as $\mathcal{G}_{t-1}=\sigma(f_1,\cdots,f_{t-1},g_1,\cdots,g_{t-1})$. The function β_{it} shall be smooth over time, and conditional i.i.d. conditioning on the sigma field $\mathcal{G}_{t-1}$. ┘

Assumption 6. *In particular for any positive integer m, we assume that $\Theta_{m,2q}(f_t)=m^{-v}$ and $\max_t\|f_t\|_{2q}$ is bounded, for $2v>1/2-1/q$, with $q>4$, for a positive constant v.*

Remark 1.3. We shall see that Assumption 1 ensures the basic property of the kernel functions. The one-sided kernel is adopted to avoid the look- ahead bias term. Also it is important for us to adopt the one-sided kernel in this step, as in the second step, our theoretical results works with conditioning on the filtration in the past. The generated error induced by estimating beta is thus fixed with respect to the conditioned filtration. Assumption 2 poses some general structure of the factor f_t and also impose some basic Assumptions of the error term ε_{it} to ensure the validity of our estimator. We set the factor to be a trend stationary process, and we assume homoscedasticity over time for the simplicity of our analysis. It would be not hard to extend to more general settings. Moreover, we also assume finite $2q$ th moment. Assumption 3 sets the proper behavior of the matrix in the population. It is related to the identification of our estimator. Assumptions 4- 6 are regular conditions on the decay rate of the dependency for factors. Note that we can relax the Assumption on the boundedness of α_{it} and β_{it}. Assumption 6 implies that $\Theta_{m,q}(\mathbb{E}(f_\cdot|f_{\cdot-1}))$, $\Theta_{m,q}(\mathbb{E}(f_\cdot|f_{\cdot-1}))$ are bounded by constants c_θ, $\Theta_{m,q}(f_\cdot^2)=m^{-v}$, for a positive constant v. In sum, we shall assume that the alphas and betas are sufficiently smooth over time, and the temporal dependency of factors shall be weak. ┘

Define $r_T=(Th)^{-1}(T^{1/q}+(Th\log T)^{1/2})$. Let $q_n=J_u\vee n_u\vee T$. $h\vee r_T\vee\sqrt{\log(q_n)}/\sqrt{Th}\overset{\text{def}}{=}\delta_T$.

Assumption 7. *$F_{\beta,t}(x)$ is continuously differentiable on the compact interval $B_\delta=[\beta_l-\delta_T,\beta_u+\delta_T]$ and $B=[\beta_l,\beta_u]$. β_{it} are i.i.d. conditioning on $\mathcal{G}_{t-1}$(sigma field of time invariant factors). ε_{it}*

are independent conditioning on all three of the filtration $\cup(\mathcal{F}_{t-1}, \sigma(f_t))$, $\mathcal{G}_{t-1}$ *or* $\mathcal{F}_{t-1}$ *respectively. The condition density of* β_{it} *is denoted as* $f_{\beta,t}(x)$. $C_{\beta,max} > \max_t \max_{x\in B_\delta} f_{\beta,t}(x) > \min_t \min_{x\in B_\delta} f_{\beta,t}(x) > c_{\beta,min} > 0$, *which is also first order continuous differentiable with bounded derivatives.* $\mathbb{E}(\beta_{it}\Phi^*_{it}|\mathcal{G}_{t-1}) \asymp_p J_t^{-1}$, $\min_{j_t} \mathbb{E}_{t-1}(\Phi^*_{i,j_t,t}\varepsilon^2_{it}) \asymp_p J_t^{-1}$ *and* $\tilde{q}_{j_t} \overset{\text{def}}{=} \int_{F^{-1}_{\beta,t}(\kappa_{j-1})}^{F^{-1}_{\beta,t}(\kappa_j)} f_{\beta,t} d\beta \asymp_p J_t^{-1}$.

Remark 1.4 (Discussion of the conditional iid Assumptions for β_{it}.)**.** Suppose Z_{it} and Z_t are observable. Let us look the simplest model

$$R_{it} = \alpha_{it} + \beta_{it} f_t + \varepsilon_{it}, \tag{1.4}$$

where α_{it} is a function of β_{it}. We see that Z_t and $Z_{t-1}(U_{it})$ are measurable with respect to $\mathcal{F}_{t-1}$. Suppose that U_{it} are iid random variables over time (independent of any object as the above). In Gagliardini, Ossola, and Scaillet (2019), they assume,

$$\beta_{it}(U_{it}) = B(U_{it})Z_{t-1} + C(U_{it})Z_{t-1}(U_{it}), \tag{1.5}$$

where $B(U_{it}), C(U_{it})$ are parameters which are solely functions of U_{it}.

We can observe that conditioning on $\mathcal{G}_{t-1} = \sigma(Z_{t-1}(.), Z_{t-1})$, $\beta_{it}(U_{it})$ will have only source of randomness from $(U_{it})_{i,t}$ and therefore are iid over i. We can express $\beta_{it}(U_{it}) = g(U_{it}, Z_{t-1}, Z_{t-1}(.))$. To ensure that $g(U_{it}, Z_{t-1}, Z_{t-1}(.))$ is smooth over t. Furthermore the smoothness Assumption on Z_{t-1} and $Z_{t-1}(U_{it})$ over t is required for us and but NOT for required for Gagliardini, Ossola, and Scaillet (2019). However, the structure in Equation (1.5) is not necessary for us, therefore there exist models where we can cover but not covered by Gagliardini, Ossola, and Scaillet (2019). For example if we set, β_t is a vector of β_{it}. $n_t = n$,

$$\beta_t = c\eta_{t-2} + a(t/T)\eta_{t-1}, \tag{1.6}$$

where η_t are iid random vectors over t, and forms random smooth curve over t. $a(t/T)$ is a smooth trend function over time, such as $a(t/T) = (t/T)^2$. η_{t-1}, η_{t-2} are measurable to $\mathcal{F}_{t-1}$. The structure in Gagliardini, Ossola, and Scaillet (2019) is then violated. ⌟

Define $a_{nT} = \max_t(\sqrt{\delta_T} c_{n_u T}/\sqrt{n_t}) \vee \sqrt{\log T}\sqrt{n_t}^{-1}$, where $c_{n_u T}$ is a positive constant.

Assumption 8. *We let* $\max_t n_t \leq n_u$, $n \leq n_t$, $J \leq J_t \leq J_u$, $J_u \asymp J$ *and* $n_u \asymp n$. *Recall* $r_T = (Th)^{-1}(T^{1/q} + (Th\log T)^{1/2})$, *with* $q > 4$. *We assume that Assumptions 1 to 6 are maintained such that* $\max_i \sup_t |\widehat{\beta}_{it} - \beta_{it}| \lesssim_{\mathrm{P}} \delta_T$ *and* $J_u^{-1} \gg \frac{\sqrt{\log(q_n)}}{\sqrt{n}}$. $\delta_T \to 0$, $r_T \to 0$. *We assume that* $a_{nT} \to 0$. *We note that the above Assumption implies that* $J \log q_n/n \ll 1$. *Define* $\delta = (\delta_T + a_{nT})^{1/2}\sqrt{\log q_n}/\sqrt{n}$. $J(\delta_T + a_{nT})^{1/2}\sqrt{\log q_n}/\sqrt{n} \ll 1$.

Assumption 9. *We assume that* $\alpha(\beta)$ *is continuously differentiable of the first order, and with the first derivative bounded from the above by a positive constant* c_α.

Remark 1.5. Assumption 7 is regarding the property of the density of β_{it}. Assumption 8 sets the property of the uniform rate of β_{it}. We see that it corresponds to Theorem 2.6. Assumption 8 assumes the relative rate of n_t and J_t . Assumption 9 assumes the smoothness of the function $\alpha(.)$. ⌟

Remark 1.6. (Discussion of rate) We assume that q to be sufficiently large and the choice of h and n_u are satisfied so that

$$\sqrt{\log(n_u T)} \gg T^{1/q}(Th)^{-1/2},$$

then $\delta_T = h + \sqrt{\log(n_u T)}/\sqrt{Th}$. Assume that

$$\sqrt{\log(n_u T)}/\sqrt{n} \ll 1,$$

then $a_{nT} = \sqrt{h/n} + \sqrt{\log T}/\sqrt{n}$. Thus $\delta_T + a_{nT} = \sqrt{\log T}/\sqrt{n} \vee h \vee \sqrt{\log(n_u T)}/\sqrt{Th}$. Assumption 9 thus assumes that

$$(\sqrt{\log T}/\sqrt{n} \vee h \vee \sqrt{\log(n_u T)}/\sqrt{Th})^{1/2} J/\sqrt{n_t} \ll 1.$$

⌟

We first order observations as $\ell = \ell(i,t) \overset{\text{def}}{=} \sum_{t_0=1}^{t-1} n_t + i$, $1 \le i \le n_t, 1 \le t_0 \le T$. We let $\mathcal{F}_{\ell-1}^{\beta}$ denote the sigma field of β_ℓ up to the order of $\ell - 1$.

Assumption 10. *We let* $\tilde{\beta}_{\ell,j} = \beta_{\ell,j} - F_{\beta,t}^{-1}(\kappa_j)$. $\tilde{\beta}_{\ell,j}$ *are different over time, however the dependence can still decay as the series of* β_{it}. *We assume that* $\mathbf{1}(-u \le \tilde{\beta}_{\ell,j} < u) \overset{\text{def}}{=} \psi_{\ell,j}(u)$. *We assume that* $\max_{\ell,j} \mathbb{E}(\beta_{\ell,j}^2 \mathbf{1}(-u \le \tilde{\beta}_{\ell,j} < u)|\mathcal{F}_{\ell-1}^{\beta}) \le C\beta_u^2 u$, $\max_{\ell,j} \mathbb{E}(\varepsilon_\ell^2 \mathbf{1}(-u \le \tilde{\beta}_{\ell,j} < u)|\mathcal{F}_{\ell-1}^{\beta}) \le C\beta_u^2 u$, *for a constant* C. *Moreover we assume that* $\|\max_j \mathbb{E}(\psi_{.,j}(u)|\mathcal{F}_{.-1}^{\beta})\|_{q,\zeta} \le u^{1/q} C_{q,\zeta}$, $\|\max_j \mathbb{E}(\psi_{.,j}(u)\beta_{.,j}|\mathcal{F}_{.-1}^{\beta})\|_{q,\zeta} \le u^{1/q} C'_{q,\zeta}$, *for an integer* $q > 4$. *Define* $\overline{\delta}_T = \sqrt{\log q_n}/\sqrt{nT} \vee h \vee \sqrt{\log(q_n)}/\sqrt{nTh}$.

Assumption 11. $(a_{nT}/\sqrt{T} + \overline{\delta}_T) \vee J(a_{nT} + \delta_T)^{1/2}/\sqrt{nT} \lesssim h \vee J^{-1} \ll \sqrt{T}^{-1}$.

Remark 1.7. Following the rate discussion of Remark 1.6, the above conditions can be implied by

$$\sqrt{\log q_n}/(\sqrt{nT}) \ll h^{3/2} \vee h^{1/2} J^{-1},$$

$$\{(\log q_n)^{1/4}/n^{1/4} \vee h^{1/2} \vee (\sqrt{\log q_n}/\sqrt{Th})^{1/2}\} \lesssim \sqrt{nT}(J^{-1}h \vee J^{-2}) \ll \sqrt{n}.$$

⌟

We shall give an example on the plausible rate of n, T, J, h which is admissible to the above Assumptions. For example, we can assume that $n = T$, $J = O(T^{1/3})$ and $h = O(T^{-1/3})$.

For a positive constant $C'_{q,\zeta}$, we notice that the conditions $\| \max_j \mathbb{E}(\psi_{\cdot,j}(u)\beta_{\cdot}|\mathcal{F}^{\beta}_{\cdot-1})\|_{q,\zeta} \leq u^{1/q}C'_{q,\zeta}$ are easily satisfied. Let us illustrate for the stationary case of β_ℓ. Let $f_{\beta,t}(\beta|.), F_{\beta,t}(\beta|.)$ be density and distribution function corresponding to different filtrations. For example if we assume that $f_{\beta,t}(\beta|\mathcal{F}^{\beta}_{t-1})$ is differentiable with respect to β and its i.i.d. innovation ε_0 (slightly abuse of notation), then we can derive that,

$$\begin{aligned}
& \mathbb{E}(\psi_{\cdot,j}(u)\beta_{\ell,j}|\mathcal{F}^{\beta}_{\ell-1}) - \mathbb{E}(\psi_{\cdot,j}(u)\beta_{\ell,j}|\mathcal{F}^{\beta*}_{\ell-1}) \\
= \ & \int_{-u+F^{-1}_{\beta,t}(\kappa_j|\mathcal{F}^{\beta}_{\ell-1})}^{u+F^{-1}_{\beta,t}(\kappa_j|\mathcal{F}^{\beta}_{\ell-1})} \beta f_{\beta,t}(\beta|\mathcal{F}^{\beta}_{\ell-1})d\beta - \int_{-u+F^{-1}_{\beta,t}(\kappa_j|\mathcal{F}^{\beta}_{\ell-1})}^{u+F^{-1}_{\beta,t}(\kappa_j|\mathcal{F}^{\beta}_{\ell-1})} \beta f_{\beta,t}(\beta|\mathcal{F}^{\beta*}_{\ell-1})d\beta \\
\leq \ & 2u|\varepsilon_0 - \varepsilon_0^*||\beta_u \partial f(\tilde{\beta}|\tilde{\mathcal{F}}_{\ell-1})/(\partial\varepsilon_0\partial\beta)|,
\end{aligned}$$

where $\tilde{\beta}$ is a point between the intersection of $\cap_j(-u + F^{-1}_{\beta,t}(\kappa_j|\mathcal{F}^{\beta}_{\ell-1}), u + F^{-1}_{\beta,t}(\kappa_j|\mathcal{F}^{\beta}_{\ell-1}))$, and $\tilde{\mathcal{F}}_{\ell-1}$ is the filtration with ε_0 replaced by some value. We take the $\|.\|_q$ norm of the above object. If we can ensure that $|\varepsilon_0 - \varepsilon_0^*||\beta_u||\partial f(\tilde{\beta}|\tilde{\mathcal{F}}_{\ell-1})/(\partial\varepsilon_0\partial\beta)\|_q$ decrease sufficient fast according to the lag ℓ, then the conditions holds. We let $\delta' = \delta_T/\sqrt{T}$. Define $\mathbb{N}_J = \bigcup_{t=1\cdots,T} N_{J_t}$. $\overline{\delta}_T = \sqrt{\log q_n}/\sqrt{nT} \vee h \vee \sqrt{\log(q_n)}/\sqrt{nTh}$.

Remark 1.8. Following the rate discussion of Remark 1.6, the above conditions can be implied by

$$\sqrt{\log q_n}/(\sqrt{nT}) \ll h^{3/2} \vee h^{1/2}J^{-1},$$

$$\{(\log q_n)^{1/4}/n^{1/4} \vee h^{1/2} \vee (\sqrt{\log q_n}/\sqrt{Th})^{1/2}\} \lesssim \sqrt{nT}(J^{-1}h \vee J^{-2}) \ll \sqrt{n}.$$

$\lrcorner$

Recall that $\widehat{\sigma}_D^{1/2}(\beta_{1,2,3})$ is defined to be $\hat{\sigma}_D(\beta_1,\beta_3,\beta_2) = T^{-1}\sum_{t=1}^{T} n_t^{-1}J_t^{-1}\sum_{i=1}^{n_t}(\sum_j \mathbb{E}\{(\widehat{p}_{jt}(\beta_1) + \widehat{p}_{jt}(\beta_3) - 2\widehat{p}_{jt}(\beta_2))^2\tilde{q}_j^{-2}(\Phi^*_{i,j,t}\sigma_t^2))\}$.

Assumption 12. *Let $\widehat{\sigma}_D^{1/2}(\beta_{1,2,3})$ be a consistent estimator for $\sigma_D(\beta_{1,2,3})$ satisfying $|\widehat{\sigma}_D(\beta_{1,2,3})^{1/2} - \sigma_D(\beta_{1,2,3})^{1/2}| = o_{\mathrm{P}}(r_{1,2,3})$, and the rate $r_{1,2,3} \to 0$. $\sigma_D(\beta_{1,2,3})$ is bounded from the below and the above uniformly over $\beta_1, \beta_2, \beta_3$. $\max_{j,t}\|\Phi^*_{i,j,t}\varepsilon_{it}\tilde{q}_j^{-1}\|_q \lesssim J^{1-1/q}$.*

Remark 1.9. (Inconsistency of $\widehat{\mu}_t(\beta)$ for fixed t.) To facilitate the inference for fix t we shall consider the following procedure. Since by Theorem B.6,

$$|\widehat{a}_t - a_t^* - \mathrm{diag}(\tilde{q}_{jt})^{-1}n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t|_\infty = o_{\mathrm{P}}(1). \tag{1.7}$$

We see that

$$
\begin{aligned}
& [n_t^{-1}\Phi_t^{*}\tilde{\varepsilon}_t]_j = n_t^{-1}\sum_{i=1}^{n_t}\Phi_{i,j,t}^{*}\tilde{\varepsilon}_{it}, \\
= \quad & n_t^{-1}\sum_{i=1}^{n_t}\Phi_{i,j,t}^{*}\beta_{it}(f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1}))+n_t^{-1}\sum_{i=1}^{n_t}\Phi_{i,j,t}^{*}\varepsilon_{it}, \\
= \quad & n_t^{-1}\sum_{i=1}^{n_t}\{\Phi_{i,j,t}^{*}\beta_{it}-\mathbb{E}(\Phi_{i,j,t}^{*}\beta_{it}|\mathcal{G}_{t-1})\}(f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1}))+n_t^{-1}\sum_{i=1}^{n_t}\Phi_{i,j,t}^{*}\varepsilon_{it} \\
& +n_t^{-1}\sum_{i=1}^{n_t}\{\mathbb{E}(\Phi_{i,j,t}^{*}\beta_{it}|\mathcal{G}_{t-1})\}(f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1})).
\end{aligned}
$$

Denote $\tilde{f}_t \stackrel{\text{def}}{=} (f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1}))$ and $\psi_{i,j,t} \stackrel{\text{def}}{=} \{\Phi_{i,j,t}^{*}\beta_{it}-\mathbb{E}[(\Phi_{i,j,t}^{*}\beta_{it})|\mathcal{G}_{t-1}]\}$, then we have the leading term in the $\widehat{a}_{j,t}-a_{j,t}$ as $\{\mathbb{E}(\Phi_{i,j,t}^{*}\beta_{it}|\mathcal{G}_{t-1})\}(f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1}))$, which is of order $O_{\mathrm{P}}(1)$. So this term explains the inconsistency of the estimator $\widehat{\mu}_t(\beta)$ to $\mu_t(\beta)$. ⌟

Remark 1.10 (Omitted factors)**.** The omitted factor bias issue has been studied in Giglio and Xiu (2021). In case of misspecification or mispricing, there exists non-smooth and non-exogenous components in α_{it}. We shall consider the following alternative procedure. Similar to Gagliardini, Ossola, and Scaillet (2016), we can impose the following structure for the conditional expectation of the factor f_t i.e., $\mathbb{E}\left[f_t|\,\mathcal{G}_{t-1}\right]=\zeta_t+\Psi_t^{\top}z_{t-1}^{f}$, where z_{t-1}^{f} is a vector of underlying factors, Ψ_t is the loading, and μ_t is a time-varying mean. Thus we have

$$
\begin{aligned}
R_{it} &= \mu_t\left(\beta_{it}\right)+\beta_{it}^{\top}\left(f_t-\mathbb{E}\left[f_t|\,\mathcal{G}_{t-1}\right]\right)+\varepsilon_{it} \\
&= \mu_t\left(\beta_{it}\right)+\beta_{it}^{\top}\left(f_t-\zeta_t-\Psi_t^{\top}z_{t-1}^{f}\right)+\varepsilon_{it} \\
&= \underbrace{\left[\mu_t\left(\beta_{it}\right)-\beta_{it}^{\top}\zeta_t\right]}_{\text{smooth constant}}+\underbrace{\beta_{it}^{\top}f_t}_{\text{term of interest}}-\underbrace{\left[\beta_{it}^{\top}\Psi_t^{\top}z_{t-1}^{f}\right]}_{\text{control variables}}+\varepsilon_{it}.
\end{aligned}
$$

Following the above model, we can modify the estimation procedure by controlling for the factors $\Psi_t^{\top}z_{t-1}^{f}$. Namely we run kernel regression of R_{it} on $\left(1,f_t^{\top},z_{t-1}^{f\top}\right)^{\top}$. Then the second step is the same as the previous steps, we can sort portfolios based on $\widehat{\beta}_{it}$ and take averages over time.

⌟

Remark 1.11 (Leave t_0 out estimator.)**.** In the Beta-sorting step, we shall use a leave t_0 out estimator to ensure that $\widehat{p}_t(\beta)$ is purely measurable to $\mathcal{F}_{t-1}$. This is a theoretical arrangement to facilitate our derivation of the property of the Beta-sorted estimator. In this remark we show that this would not change the statistical property of the estimator in the first step. We define the leave-one-out estimator to be

$$\widehat{b}_{i(-t_0)} = [\sum_{t\neq t_0, t=1}^{T} (w(t,t_0)X_tX_t^\top)]^{-1}\{\sum_{t\neq t_0, t=1}^{T} w(t,t_0)X_tR_{it}\}.$$

Since we are using a one sided kernel, the estimator $\widehat{b}_{i(-t_0)}$ only use information up to time t_0-1. Compared to the estimator

$$\widehat{b}_{it_0} = [\sum_{t=1}^{T}(w(t,t_0)X_tX_t^\top)]^{-1}\{\sum_{t=1}^{T} w(t,t_0)X_tR_{it}\}.$$

We can derive that

$$\begin{aligned}
&\max_{i,t_0\in[\lfloor Th\rfloor, T-\lfloor Th\rfloor]} |\widehat{b}_{it_0} - \widehat{b}_{i(-t_0)}|_\infty \\
&\leq \sup_{t_0} |(Th)^{-1}X_{t_0}X_{t_0}^\top|_\infty |T[\sum_{t\neq t_0}(w(t,t_0)X_tX_t^\top)]^{-1}|_\infty |T[\sum_{t=1}^{T}(w(t,t_0)X_tX_t^\top)]^{-1}|_\infty \\
&\max_{i,t_0} |T^{-1}\sum_{t\neq t_0}(w(t,t_0)X_tR_{it})|_\infty \\
&+\sup_{t_0}\max_{i} |(Th)^{-1}X_{t_0}R_{it_0}|_\infty |[T^{-1}\sum_{t\neq t_0} w(t,t_0)X_tX_t^\top]^{-1}|_\infty \\
&\lesssim_{\mathrm{P}} (Th)^{-1}T^{1/q} \lesssim r_T.
\end{aligned}$$

We now show that $\widehat{b}_{i(-t_0)}$ is close $\widehat{b}_{i(-t_0)}$ in a uniform sense. Thus Theorem 2.6 and 2.7 still hold under the same conditions. ⌟

Since we have $1/J_t * J_t{}^2(J_t{}^{-1}(\delta_T + a_{nT}) + \sqrt{\delta_T}T^{1/2q}(n_t^{-1/2}J_t^{-1/2})) = o_{\mathrm{P}}(\delta_2)$ and $(1/J_t(l_{n,T}/n_t + J_t^{-1})J_t^2(l_{n,T}/n_t)J_t^{-1}J_t^2) = o_{\mathrm{P}}(\delta_1)$. Recall that $\delta_1 = \sqrt{a_{nT}+\delta_T}J_t/\sqrt{n_t} + a_{nT} + \delta_T$, $\delta_2 = \delta_T + a_{nT} + \sqrt{\delta_T}T^{1/(2q)}n_t^{-1/2}J_t^{1/2}$.

Recall $\tilde{q}_{j_t} = \mathbb{E}_{t-1}(\Phi^*_{i,j_t,t}) = \mathbb{E}(\Phi^*_{i,j_t,t}|\mathcal{G}_{t-1})$, and $\mathbb{E}(\Phi^*_{i,j_t,t}\varepsilon_{it}^2)|\mathcal{G}_{t-1}) = \sigma_t^2\tilde{q}_{j_t}$. We denote for any random variable X $\mathbb{E}_{t-1}(X) = \mathbb{E}(X|\mathcal{G}_{t-1})$. Define the variance estimator for a fixed time point t as $\hat{\sigma}^2_{j_t} = n_t/J_t(\sum_i \hat{\Phi}_{i,j_t,t}\hat{\varepsilon}_{it}^2)(\sum_i \hat{\Phi}_{i,j_t,t})^{-2}$. We see that $\sigma^2_{j_t} = 1/J_t(\mathbb{E}_{t-1}(\Phi^*_{i,j_t,t}\varepsilon_{it}^2))(\mathbb{E}_{t-1}(\Phi^*_{i,j_t,t}))^{-2} = J_t^{-1}\sigma_t^2\tilde{q}_{j_t}^{-1}$. The variance estimator of $\hat{L}_t(\beta)$ on the whole support of β is defined as $\hat{\sigma}_t(\beta) = \sum_{j_t}\hat{p}_{j_t}(\beta)\hat{\sigma}^2_{j_t}$ targeting at $\sigma_t(\beta) = \sum_{j_t}\hat{p}_{j_t}(\beta)\sigma^2_{j_t}$ in the population.

Assumption 13. *Let $\varepsilon_{it} =_d \sigma_t\eta_{it}$ conditional on $\mathcal{F}_{t-1}$, with $\sigma_t^2 = \mathbb{E}(\varepsilon_{it}^2|\mathcal{F}_{t-1}) = \mathbb{E}(X|\mathcal{G}_{t-1})$, and η_{it} be a standard Gaussian random variable defined on a proper probability space. Exists two positive constants $c, C > 0$,*

$$c \leq \min_{j_t}\sigma_{j_t} \leq \max_{j_t}\sigma_{j_t} \leq C.$$

And we have $\delta_1+\delta_2 \ll 1/\sqrt{\log J}$. $n_t^{-1/2+1/(2q)}\sqrt{J_t} \ll \sqrt{J_t}^{-1}$. Moreover, $\sqrt{n_t/J_t}(h\vee J_t^{-1}) \to 0$.

Remark 1.12. Following the remark 1.6, $\delta_T + a_{nT} = \sqrt{\log T}/\sqrt{n} \vee h \vee \sqrt{\log(n_u T)}/\sqrt{Th}$. The above Assumption implies that

$$J_t^{3/2}\sqrt{a_{nT} + \delta_T}/\sqrt{n_t} \ll 1,$$

and

$$a_{nT} + \delta_T \ll \sqrt{J_t}^{-1}.$$

┘

Recall that $\widehat{\sigma}(\beta)$ is defined to be

$$\widehat{\sigma}(\beta) \stackrel{\text{def}}{=} T^{-1}\sum_{t=1}^{T}(\sum_{j=1}^{J_t}\widehat{p}_{j,t}(\beta)E_{n_t,j}^2 \text{Var}(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})) + T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t} n_t^{-2}\sum_i \widehat{p}_{j,t}(\beta)\tilde{q}_{jt}^{-2}\mathbb{E}(\Phi^*_{i,j,t}\varepsilon_{it}^2|\mathcal{G}_{t-1}).$$

Assumption 14. *Assume that with probability one, $E_{n_t,j}/\sigma_j s$ are bounded from the below and the above for all t, j. $\sup_\beta |\widehat{\sigma}(\beta)^{1/2} - \sigma(\beta)^{1/2}| = O_{\mathrm{P}}(r_\sigma)$ for some constant r_σ. Assume $r_\sigma[T^{-1/2+1/2q}J_a^{1/2q}]^{-1} \to 0$ and $r_\sigma \to 0$. $c \le \inf_\beta \sigma(\beta) = \sup_\beta \sigma(\beta) \le C$. $c \le \min_\beta \tilde{\sigma}(\beta) = \sup_\beta \tilde{\sigma}(\beta) \le C$. $\|E_{n_t,j} - E_{n_t,j-1}\|_{2q} \le cJ_t^{-1}$, for a positive constant $c > 0$. $\|f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})\|_{2q} < C$, for a positive $C > 0$. Assume that the grid is $\overline{\beta}_v = [\beta_1, \beta_2, \cdots, \beta_{c/\tilde{\delta}}]$. This corresponds to $J_{a,\delta}$ distinct value of β. We shall assume that for avoiding singularity of the variance covariance matrix. We need to ensure that $\sigma(\beta_j, \beta_{j'}) \ne \sigma(\beta_j, \beta_j)$ or $\sigma(\beta_{j'}, \beta_{j'})$. Let $[\widehat{\sigma}(\beta_i, \beta_{i'})/\widehat{\sigma}^{1/2}(\beta_i)\widehat{\sigma}(\beta_{i'})^{1/2}]_{i,i'} \stackrel{\text{def}}{=} \Sigma_{J_{a,\delta},J_{a,\delta}}$. We shall assume that $c < \lambda_{\min}(\Sigma_{J_a,\delta}) < \lambda_{\max}(\Sigma_{J_a,\delta}) < C'$, with $C', c > 0$. Let $J_{a,\delta} \lesssim \exp(T^{\varepsilon'})$, with $\varepsilon' = 1/9$.*

Assumption 15. *Assume that $\sup_\beta |\widehat{\sigma}(\beta)/\mathbb{E}(\widehat{\sigma}(\beta)) - 1| = O_p((\log(nT))^{-1})$ and with probability 1, $0 < c \le \min_\beta |\widehat{\sigma}(\beta)/\mathbb{E}(\widehat{\sigma}(\beta))| \le \max_\beta |\widehat{\sigma}(\beta)/\mathbb{E}(\widehat{\sigma}(\beta)| \le C$.*

2 Results

2.1 Preliminary Technical Lemmas

We first show some useful lemmas.

Lemma 2.1 (Burkholder (1988); Rio (2009))**.** *Let $q > 1, q' = \min\{q, 2\}$, and $M_T = \sum_{t=1}^T \xi_t$, where $\xi_t \in \mathcal{L}^q$ are martingale differences. Then*

$$\|M_T\|_q^{q'} \le K_q^{q'}\sum_{t=1}^{T}\|\xi_T\|_q^{q'}, \quad \textit{where } K_q = \max\{(q-1)^{-1}, \sqrt{q-1}\}.$$

Lemma 2.2 (Freedman's inequality)**.** *Let $\xi_{a,i}$ be a martingale difference sequence, $\mathcal{F}_i$ be the filtra-*

tion, $V_a = \sum_{i=1}^n \mathbb{E}(\xi_{a,i}^2|\mathcal{F}_{i-1})$ *and* $M_a = \max_{1\le l\le n}\sum_{i=1}^l \xi_{a,i}$*, we have,*

$$\mathrm{P}(\max_{a\in\mathcal{A}}|M_a|\ge z) \le \sum_{i=1}^n \mathrm{P}(\max_{a\in\mathcal{A}}\xi_{a,i}\ge u) + 2\,\mathrm{P}(\max_{a\in\mathcal{A}}V_a\ge v) + 2|\mathcal{A}|e^{-z^2/(2zu+2v)}. \tag{2.1}$$

For a $p-$ dimensional random variable X_{jt}, we define $\|X_{j,\cdot}\|_{q,\varsigma} = \sup_{m_0}(m_0+1)^\varsigma \sum_{m\ge m_0}\delta_{q,m}(X_{j,\cdot}) < \infty$.

Lemma 2.3 (Theorem 6.2 of Zhang and Wu (2017) Tail probabilities for high dimensional partial sums)**.** *For a zero-mean p-dimensional random variable* $X_t \in \mathbb{R}^p$*, let* $S_n = \sum_{t=1}^n X_t$ *and assume that* $\||X_\cdot|_\infty\|_{q,\varsigma} < \infty$*, where* $q > 4$ *and* $\varsigma \ge 0$*, and* $\Phi_{2,\varsigma} = \max_{1\le j\le p}\|X_{j,\cdot}\|_{2,\varsigma} < \infty$*.*
i) If $\varsigma > 1/2 - 1/q$*, then for* $x \gtrsim \sqrt{n\log p}\,\Phi_{2,\varsigma} + n^{1/q}(\log p)^{3/2}\||X_\cdot|_\infty\|_{q,\varsigma}$*,*

$$\mathrm{P}(|S_n|_\infty \ge x) \le \frac{C_{q,\varsigma}n(\log p)^{q/2}\||X_\cdot|_\infty\|_{q,\varsigma}^q}{x^q} + C_{q,\varsigma}\exp\left(\frac{-C_{q,\varsigma}x^2}{n\Phi_{2,\varsigma}^2}\right).$$

ii) If $0 < \varsigma < 1/2 - 1/q$*, then for* $x \gtrsim \sqrt{n\log p}\,\Phi_{2,\varsigma} + n^{1/2-\varsigma}(\log p)^{3/2}\||X_\cdot|_\infty\|_{q,\varsigma}$*,*

$$\mathrm{P}(|S_n|_\infty \ge x) \le \frac{C_{q,\varsigma}n^{q/2-\varsigma q}(\log p)^{q/2}\||X_\cdot|_\infty\|_{q,\varsigma}^q}{x^q} + C_{q,\varsigma}\exp\left(\frac{-C_{q,\varsigma}x^2}{n\Phi_{2,\varsigma}^2}\right).$$

Lemma 2.4 (Uniform rate)**.** *By Assumptions 1- 6, we have*

$$\sup_{\lfloor Th\rfloor\le t_0\le T-\lfloor Th\rfloor} T^{-1}|A(t_0) - \mathbb{E}A(t_0)|_{\max} \lesssim_{\mathrm{P}} (Th)^{-1}(T^{1/q} + (Th\log T)^{1/2}), \tag{2.2}$$

$$\sup_{\lfloor Th\rfloor\le t_0\le T-\lfloor Th\rfloor} T^{-1}|\tilde{A}(t_0) - \mathbb{E}\tilde{A}(t_0)|_{\max} \lesssim_{\mathrm{P}} (Th)^{-1}(T^{1/q} + (Th\log T)^{1/2}), \tag{2.3}$$

$$\max_i \sup_{\lfloor Th\rfloor\le t_0\le T-\lfloor Th\rfloor} T^{-1}|\tilde{B}_i(t_0) - \mathbb{E}B_i(t_0)|_{\max} \lesssim_{\mathrm{P}} (Th)^{-1}(T^{1/q} + (Th\log T)^{1/2}), \tag{2.4}$$

$$\sup_{\lfloor Th\rfloor\le t_0\le T-\lfloor Th\rfloor} T^{-1}|A(t_0) - \tilde{A}(t_0)|_{\max} \lesssim_{\mathrm{P}} \sqrt{\log T}/(\sqrt{T}\sqrt{h}), \tag{2.5}$$

$$\max_i \sup_{\lfloor Th\rfloor\le t_0\le T-\lfloor Th\rfloor} T^{-1}|B_i(t_0) - \tilde{B}_i(t_0)|_{\max} \lesssim_{\mathrm{P}} \sqrt{\log(n_u T)}/(\sqrt{T}\sqrt{h}). \tag{2.6}$$

Lemma 2.5 (Uniform rate)**.** *Assume* u_t *and* v_{it} *which are martingale differences over* t *(*$i \in 1,\cdots,n$*,* $t\in 1,\cdots,T$*). Let* $2v > 1/2 - 1/q$*. For a positive constant* C_v*, assume* $\Theta_{m,2q}(u_\cdot) < C_v$ *, and* $\max_t\|u_t\|_{2q} < C_v$ *for* $q > 4$*.* $\Theta_{m,2q}(v_\cdot) < C_v$*. Assume* $\Theta_{m,2q}(v_{i\cdot}) < C_v$ *, and* $\max_t\|v_{it}\|_{2q} < M$ *for* $q > 4$*.* $\Theta_{m,2q}(v_\cdot) < C_v$*.*

If we assume that $T^{-1+2/q} \ll h$*, then we have,*

$$\sup_{\lfloor Th\rfloor\le t_0\le T-\lfloor Th\rfloor} T^{-1}|\sum_{t=1}^T w(t,t_0)u_t| \lesssim_{\mathrm{P}} (Th)^{-1/2}(\log T)^{1/2}, \tag{2.7}$$

moreover if we have $(T)^{2/q-1}n^{2/q} \ll h$*, then we have,*

$$\max_i \sup_{\lfloor Th \rfloor \le t_0 \le T - \lfloor Th \rfloor} T^{-1} |\sum_{t=1}^T w(t,t_0) v_{it}| \lesssim_{\mathrm{P}} (Th)^{-1/2} (\log T)^{1/2}. \tag{2.8}$$

Recall that $\hat{a}_t = \{\hat{\Phi}_t \hat{\Phi}_t^\top\}^{-1} \{\hat{\Phi}_t R_t\}$, we now derive a linearization for the estimator a_t^* at each time point t. Recall that $\tilde{q}_{jt} = \int_{F_{\beta,t}^{-1}(\kappa_{j-1t})}^{F_{\beta,t}^{-1}(\kappa_{jt})} f_{\beta,t} d\beta$. $[]_j$ denote the jth element of a vector.

2.2 Main Theorems

Theorem 2.6. *Suppose Assumptions 1–6 hold, and recall that* $r_T = (Th)^{-1}(T^{1/q} + \sqrt{Th \log T}) \to 0$, $h \to 0$*, and* $\log(n_u T)/Th \to 0$*. Then,*

$$\max_{1 \le i \le n} \sup_{\lfloor Th \rfloor \le t_0 \le T - \lfloor Th \rfloor} \left| \widehat{b}_{it_0} - b_{it_0} \right| \lesssim_{\mathrm{P}} \delta_T,$$

where recall that $\delta_T = (r_T + \sqrt{\log(n_u T)}/\sqrt{Th} + h)$*.*

Recall that $\sqrt{\log T/(Th)} \overset{\text{def}}{=} r_{AT}$.

Theorem 2.7 (Asymptotic Normality)**.** *Let* $h\sqrt{hT} \to 0$, $h \to 0$, $Th \to \infty$, $r_{AT} + r_T \to 0$ *then, under Assumptions 1-6, we have that*

$$\sqrt{Th} \Sigma_b^{-1/2} \left(\widehat{b}_{it_0} - b_{it_0} \right) \to_{\mathcal{L}} \mathsf{N}(0, I). \tag{2.9}$$

Lemma 2.8 (Rate of $\hat{\beta}_{(k_j),t}$)**.** *Conditional on* $\mathcal{G}_{t-1}$*, we have under Assumptions 7 and 8,*

$$\max_{t,j} (\hat{\beta}_{(k_{jt}),t} - F_{\hat{\beta},t}^{-1}(\kappa_{jt})) \lesssim_{\mathrm{P}} a_{nT}, \tag{2.10}$$

$$\max_{t,j} (\hat{\beta}_{(k_{jt}),t} - F_{\beta,t}^{-1}(\kappa_{jt})) \lesssim_{\mathrm{P}} a_{nT} \vee \delta_T. \tag{2.11}$$

In the following lemma, we show the uniform rate of the order statistics of $\hat{\beta}_{it}$. It shall be noted that due to the Assumption 7, β_{it}s are conditionally independent and identically distributed conditioning on $\mathcal{G}_{t-1}$. Recall that $\tilde{q}_{jt} = \int_{F_{\beta,t}^{-1}(\kappa_{j-1t})}^{F_{\beta,t}^{-1}(\kappa_{jt})} f_{\beta,t} d\beta \lesssim_{\mathrm{P}} J_t^{-1} \le J^{-1}$ due to Assumption 7. In the following, we show a few useful lemmas that facilitates further derivation. Denote $\tilde{\delta}_T = (\sqrt{\log T}/\sqrt{n}) \vee h \vee (\sqrt{\log(n_u T)}/\sqrt{nTh})$.

To derive the rate of $\widehat{\mu}_t$, we shall define a few objects for the ease of derivations. We define $k_{jt} = \lfloor n_t j / J_t \rfloor$, $\kappa_j = j/J_t$. The following Assumptions are imposed to ensure the proper rate of our estimator. We denote $n_a = \sum_{t=1}^T n_t$. Assume $n_a \asymp nT$, and n, n_u are of the same order.

Lemma 2.9. *Conditional on $\mathcal{G}_{t-1}$, given Assumptions 7 and 8, we have,*

$$\mathrm{P}(\min_j |\beta_{(k_{jt}),t} - \beta_{(k_{j-1t}),t}| > 2c_{\beta,\min}^{-1} 1/J_t) \to 0,$$

$$\mathrm{P}(\min_t \min_j |\beta_{(k_{jt}),t} - \beta_{(k_{j-1t}),t}| > 2c_{\beta,\min}^{-1} 1/J) \to 0.$$

Also we have the bias term, with a slow varying term c_{n_t},

$$\max_{t,j} |\sum_{i=1}^{n_t} (\mathbf{1}(\hat{\beta}_{it} \in \hat{P}_{jt}) - \mathbf{1}(\beta_{it} \in P_{jt}))| \lesssim_{\mathrm{P}} \max_t \sqrt{n_t}\sqrt{a_{nT} + \tilde{\delta}_T} c_{n_t} + n_t(a_{nT} + \tilde{\delta}_T)/J_t \stackrel{\text{def}}{=} l_{n,T},$$

$$\max_t |\flat_t|_\infty \lesssim_{\mathrm{P}} 1/J_t \leq 1/J.$$

Now we show the estimation accuracy of a few partial sums with the plugged in Beta. Recall that $\delta = (\delta_T + a_{nT})^{1/2}\sqrt{\log q_n}/\sqrt{n}$. The following lemma is for fixed time point t, and conditioning on $\mathcal{G}_{t-1}$.

Lemma 2.10. *Conditional on $\mathcal{G}_{t-1}$, under Assumptions 7-9, we have,*

$$\begin{aligned}
H_1 = \quad & n_t^{-1}\hat{\Phi}_t\hat{\beta}_t - n_t^{-1}\Phi_t^*\beta_t = n_t^{-1}\sum_{i=1}^{n_t}(\hat{\Phi}_{i,t}\hat{\beta}_{it} - \Phi_{i,t}^*\beta_{it}) \lesssim_{\mathrm{P}} \delta + (\tilde{\delta}_T + a_{nT})/J \stackrel{\text{def}}{=} h_1,\\
H_2 = \quad & n_t^{-1}\hat{\beta}_t^\top\hat{\beta}_t - n_t^{-1}\beta_t{}^\top\beta_t = n_t^{-1}\sum_{i=1}^{n_t}(\hat{\beta}_{it}^2 - \beta_{it}^2) \lesssim_{\mathrm{P}} \tilde{\delta}_T \stackrel{\text{def}}{=} h_2,\\
H_3 = \quad & n_t^{-1}\hat{\Phi}_t^\top\tilde{\varepsilon}_t - n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t = n_t^{-1}\sum_{i=1}^{n_t}(\hat{\Phi}_{i,t}\tilde{\varepsilon}_{it} - \Phi_{i,t}^*\tilde{\varepsilon}_{it}) \lesssim_{\mathrm{P}} \delta + (\tilde{\delta}_T + a_{nT})/J = h_1.
\end{aligned}$$

We take the $\|.\|_q$ norm of the above object. If we can ensure that $|\varepsilon_0 - \varepsilon_0^*||\beta_u|\|\partial f(\tilde{\beta}|\tilde{\mathcal{F}}_{\ell-1})/(\partial\varepsilon_0\partial\beta)\|_q$ decrease sufficient fast according to the lag ℓ, then the conditions holds. We let $\delta' = \delta/\sqrt{T}$. Define $\mathbb{N}_J = \bigcup_{t=1\cdots,T} N_{J_t}$. Recall that $\overline{\delta}_T = \sqrt{\log q_n}/\sqrt{nT} \vee h \vee \sqrt{\log(q_n)}/\sqrt{nTh}$.

Lemma 2.11. *Under Assumptions 7-10,*

$$\begin{aligned}
\tilde{H}_1 = \quad & \sup_z T^{-1}\sum_t \{n_t^{-1}\hat{p}_t(z)^\top\hat{\Phi}_t\hat{\beta}_t - n_t^{-1}\hat{p}_t(z)^\top\Phi_t^*\beta_t\}\\
& = \sup_z T^{-1}\sum_t n_t^{-1}\sum_i \hat{p}_t(z)^\top(\hat{\Phi}_{i,t}\hat{\beta}_{it} - \Phi_{i,t}^*\beta_{it}) \lesssim_{\mathrm{P}} \delta' \vee (a_{nT}/\sqrt{T} + \overline{\delta}_T)/J \stackrel{\text{def}}{=} h_1',\\
\tilde{H}_2 = \quad & T^{-1}\sum_t n_t^{-1}\{\hat{\beta}_t^\top\hat{\beta}_t - \beta_t{}^\top\beta_t\} = T^{-1}\sum_t n_t^{-1}\sum_i(\hat{\beta}_{it}^2 - \beta_{it}^2) \lesssim_{\mathrm{P}} \overline{\delta}_T,\\
\tilde{H}_3 = \quad & \sup_z T^{-1}\sum_t \{n_t^{-1}\hat{p}_t(z)^\top\hat{\Phi}_t^\top\tilde{\varepsilon}_t - n_t^{-1}\hat{p}_t(z)^\top\Phi_t^*\tilde{\varepsilon}_t\}\\
& = \sup_z T^{-1}\sum_t n_t^{-1}\sum_i \hat{p}_t(z)^\top(\hat{\Phi}_{i,t}\tilde{\varepsilon}_{it} - \Phi_{i,t}^*\tilde{\varepsilon}_{it}) \lesssim_{\mathrm{P}} \delta' \vee (a_{nT}/\sqrt{T} + \overline{\delta}_T)/(\sqrt{T}J) = h_3',
\end{aligned}$$

$$\tilde{H}_4 = \quad |\max_{\mathbb{J}\in B_{\mathbb{J}}} T^{-1}\sum_t\sum_{i=1}^{n_t}(\mathbf{1}(\hat{\beta}_{it}\in\hat{P}_{jt}) - \mathbf{1}(\beta_{it}\in P_{jt}))| \lesssim_{\mathrm{P}} h_1'.$$

Theorem 2.12. *Conditional on $\mathcal{G}_{t-1}$, under Assumptions 7-9,*

$$[\hat{a}_t - a_t^* - diag(\tilde{q}_{jt})^{-1} n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t]_j = o_p(1). \tag{2.12}$$

In the following, we present a lemma which shows the estimator $\widehat{\sigma}(\beta)$ does not affect the accuracy when plug in the the partial sums.

Lemma 2.13. *Under Assumption 15, we have*

$$\sup_\beta |\frac{\frac{1}{\sqrt{T}}\sum_{t=1}^T\{\widehat{\mu}_t(\beta)-\mu_t(\beta)\} - bias(\beta)}{\widehat{\sigma}(\beta)^{1/2}} - \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^T\{\widehat{\mu}_t(\beta)-\mu_t(\beta)\} - bias(\beta)}{[\mathbb{E}\{\widehat{\sigma}(\beta)\}]^{1/2}}| \to_p 0.$$

Theorem 2.14 (Leading term linearization)**.** *Suppose Assumptions 7-11 hold. Then, uniformly in β,*

$$\frac{1}{T}\sum_{t=1}^T\{\widehat{\mu}_t(\beta)-\mu_t(\beta)\} = \frac{1}{T}\sum_{t=1}^T \widehat{p}_t(\beta)^\top[\mathrm{diag}(\tilde{q}_{jt})^{-1}n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t] + O_{\mathrm{P}}((J^{-1}\vee h)) + o_{\mathrm{P}}(T^{-1/2}),$$

where the first term is the leading term and the second term is the bias term. Moreover,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^T \widehat{p}_t(\beta)^\top[\mathrm{diag}(\tilde{q}_{jt})^{-1}n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t] \\
&= \frac{1}{T}\sum_{t=1}^T \widehat{p}_t(\beta)^\top \mathrm{diag}(\tilde{q}_{jt})^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\Phi_{i,t}^*\varepsilon_{it} \\
&\quad + \frac{1}{T}\sum_{t=1}^T \widehat{p}_t(\beta)^\top \mathrm{diag}(\tilde{q}_{jt})^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\Phi_{i,t}^*\beta_{it}(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})).
\end{aligned}$$

Theorem 2.15 (Pointwise central limit theorem)**.** *Suppose Assumptions 1-11 hold. Then, pointwise in β,*

$$\frac{\frac{1}{\sqrt{T}}\sum_{t=1}^T\{\widehat{\mu}_t(\beta)-\mu_t(\beta)\} - bias(\beta)}{\mathbb{E}(\hat{\sigma}(\beta))^{1/2}} \to_{\mathcal{L}} \mathsf{N}(0,1),$$

where recall that

$$\widehat{\sigma}(\beta) \stackrel{\mathrm{def}}{=} T^{-1}\sum_{t=1}^T(\sum_{j=1}^{J_t}\widehat{p}_{j,t}(\beta)E_{n_t,j}^2\,Var(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})) + T^{-1}\sum_{t=1}^T\sum_{j=1}^{J_t} n_t^{-2}\sum_i \widehat{p}_{j,t}(\beta)\tilde{q}_{jt}^{-2}\mathbb{E}(\Phi_{i,j,t}^*\varepsilon_{it}^2|\mathcal{G}_{t-1}),$$

and let

$$\widehat{\sigma}(\beta) \stackrel{\mathrm{def}}{=} \sigma_f(\beta) + \sigma_\varepsilon(\beta),$$

and

$$bias(\beta) \stackrel{\text{def}}{=} (T^{-1}\sum_{t=1}^{T}\widehat{p}_t(\beta)^\top(n_t^{-1}\Phi_t^*\Phi_t^{*\top})^{-1}(n_t^{-1}\Phi_t^*\flat_t) + T^{-1}\sum_{t=1}^{T}\widehat{p}_t(\beta)^\top\text{diag}(\tilde{q}_j)^{-1}(n_t^{-1}(\widehat{\Phi}_t\widehat{\Phi}_t^\top - \widehat{\Phi}_t\Phi_t^{*\top})a_t^*))$$

which is a term of order $(J^{-1} \vee h)$.

2.3 On the accuracy of the variance estimator.

Lemma 2.16. *Under the conditions of Theorem B.6 and 2.14, for* β_1*,* β_2 *corresponding to* $\mathbb{J}_1$*,* $\mathbb{J}_2$ *respectively, we have*

$$\sup_{\beta_1,\beta_2\in[\beta_l,\beta_u]}|\widehat{\sigma}_{\texttt{FM}}(\widehat{\mu}(\beta_1),\widehat{\mu}(\beta_2)) - \sigma_f(\beta_1,\beta_2) - \sigma_\mu(\beta_1,\beta_2)| = o_{\mathrm{P}}(1/\sqrt{\log J}).$$

We define the residual as $\hat{\varepsilon}_{it} = R_{it} - \hat{\mu}_t(\beta_{it})$. The first lemma shows that the residuals are uniformly consistent to the true one and the variance estimator is valid as shown in the next lemma.

Lemma 2.17. *Under conditions of Theorem 2.6 and* $T^{1/2q} \ll Th, T^{1/q} \ll \sqrt{Th}$*, we have the residuals satisfying,*

$$\max_{t_0}|\hat{\varepsilon}_{it_0} - \varepsilon_{it_0}| \lesssim_p \delta_T T^{1/q} = o(1). \tag{2.13}$$

Recall that we define $\mathcal{G}_{t-1}$ as a filtration and β_{it} is conditionally iid conditioning on it. We define $\mathbb{E}(\varepsilon_{it}^2|\mathcal{F}_{t-1}) = \mathbb{E}(\varepsilon_{it}^2|\mathcal{G}_{t-1}) = \sigma_t^2$. Let $\tilde{q}_{j_t} = \mathbb{E}_{t-1}(\Phi_{i,j_t,t}^*) = \mathbb{E}(\Phi_{i,j_t,t}^*|\mathcal{G}_{t-1})$, and $\mathbb{E}(\Phi_{i,j_t,t}^*\varepsilon_{it}^2)|\mathcal{G}_{t-1}) = \sigma_t^2\tilde{q}_{j_t}$. We denote for any random variable X, $\mathbb{E}_{t-1}(X) = \mathbb{E}(X|\mathcal{G}_{t-1})$. Define the variance estimator for a fixed time point t as $\hat{\sigma}_{j_t}^2 = n_t/J_t(\sum_i\hat{\Phi}_{i,j_t,t}\hat{\varepsilon}_{it}^2)(\sum_i\hat{\Phi}_{i,j_t,t})^{-2}$. We see that $\sigma_{j_t}^2 = 1/J_t(\mathbb{E}_{t-1}(\Phi_{i,j_t,t}^*\varepsilon_{it}^2))(\mathbb{E}_{t-1}(\Phi_{i,j_t,t}^*))^{-2} = J_t^{-1}\sigma_t^2\tilde{q}_{j_t}^{-1}$. The variance estimator of $\hat{L}_t(\beta)$ on the whole support of β is defined as $\hat{\sigma}_t(\beta) = \sum_{j_t}\hat{p}_{j_t}(\beta)\hat{\sigma}_{j_t}^2$ targeting at $\sigma_t(\beta) = \sum_{j_t}\hat{p}_{j_t}(\beta)\sigma_{j_t}^2$ in the population.

From Assumptions 7, we have $\min_{j_t} J_t\tilde{q}_{j_t} \geq c$, and $J_t\min_{j_t}\mathbb{E}_{t-1}(\Phi_{i,j_t,t}^*\varepsilon_{it}^2) \geq c$ with probability 1. Let $\delta_1 = \sqrt{a_{nT}+\delta_T}J_t/\sqrt{n_t} + a_{nT} + \delta_T$, and $\delta_2 = \delta_T + a_{nT} + \sqrt{\delta_T}T^{1/(2q)}n_t^{-1/2}J_t^{1/2}$.

Lemma 2.18. *Under the Assumptions of Theorem 2.12,* $\sqrt{\delta_T}T^{1/(2q)}n_t^{-1/2}J_t^{1/2} \to 0$*, we have,*

$$\max_{1\leq t\leq T}\sup_{\beta\in B}|\sigma_t(\beta) - \hat{\sigma}_t(\beta)| \lesssim_{\mathrm{P}} \delta_1 + \delta_2 = o(1).$$

2.4 Test and theory on the difference estimator.

Next we discuss the procedure to test $\sup_{\beta_1+\beta_3=2\beta_2}|\mu_t(\beta_1) - \mu_t(\beta_2) - [\mu_t(\beta_2) - \mu_t(\beta_3)]| = 0$. Let $\hat{\sigma}_t(\beta_{1,2,3}) = \sum_{j_t}[\hat{p}_{j_t}(\beta_1) + \hat{p}_{j_t}(\beta_3) - 2\hat{p}_{j_t}(\beta_2)]^2\hat{\sigma}_{j_t}^2$. Also $\sigma_t(\beta_{1,2,3}) = \sum_{j_t}[\hat{p}_{j_t}(\beta_1) + \hat{p}_{j_t}(\beta_3) -$

$2\hat{p}_{j_t}(\beta_2)]^2\sigma_{j_t}^2$. We shall consider the test statistics

$$\begin{aligned}
T_{n,t}(\beta_{1,2,3}) &\stackrel{\text{def}}{=} \sup_{\beta_1,\beta_2,\beta_3:\beta_1+\beta_3=2\beta_2} |\hat{L}_t(\beta_1) - \hat{L}_t(\beta_2) - \Big[\hat{L}_t(\beta_2) - \hat{L}_t(\beta_3)\Big] \\
&\quad -\{L_t(\beta_1) - L_t(\beta_2) - [L_t(\beta_2) - L_t(\beta_3)]\}|/[\hat{\sigma}_t(\beta_{1,2,3})^{1/2}]|, \\
T_t(\beta_{1,2,3}) &\stackrel{\text{def}}{=} \sup_{\beta_1,\beta_2,\beta_3:\beta_1+\beta_3=2\beta_2} |(\hat{p}_t^\top(\beta_1) - 2\hat{p}_t^\top(\beta_2) + \hat{p}_t^\top(\beta_3))\{\mathrm{diag}[\tilde{q}_{jt}]\}^{-1}\{n_t^{-1}\Phi_t^*\varepsilon_t\} \\
&\quad /[\sigma_t(\beta_{1,2,3})^{1/2}]|, \\
Z_t(\beta_{1,2,3}) &\stackrel{\text{def}}{=} \sup_{\beta_1,\beta_2,\beta_3:\beta_1+\beta_3=2\beta_2} |\sum_{j_t}(\hat{p}_{j_t}^\top(\beta_1) - 2\hat{p}_{j_t}^\top(\beta_2) + \hat{p}_{j_t}^\top(\beta_3))Z_{j_t}\sigma_{j_t} \\
&\quad /[\sigma_t(\beta_{1,2,3})^{1/2}]|,
\end{aligned}$$

where Z_{j_t} is a standard normal random variable.

Theorem 2.19. *Under conditions of Theorem B.6 and Assumption 13, we have*

$$\begin{aligned}
\sup_{\beta\in[\beta_l,\beta_u]} |\frac{\widehat{M}_t(\beta) - M_t(\beta)}{\sqrt{J_t\widehat{\sigma}_t(\beta)/n_t}}| &= \sup_{\beta\in[\beta_l,\beta_u]} |\widehat{p}_t^\top(\beta)\{\mathrm{diag}[\tilde{q}_{jt}]\sigma_t(\beta)^{1/2}\}^{-1}\{(\sqrt{n}_t\sqrt{J}_t)^{-1}\Phi_t^*\varepsilon_t\}| \\
&\quad + o_{\mathrm{P}}(1/\sqrt{J_t} \vee \sqrt{n_t}h/\sqrt{J_t} \vee \sqrt{n_t/J_t}J_t^{-1}).
\end{aligned}$$

To approximate the quantile, we have,

$$\sup_{x\in\mathbb{R}} \Big| \mathrm{P}\Big(\sup_{\beta\in[\beta_l,\beta_u]} Z_{n_t}(\beta) \le x\Big) - \mathrm{P}\Big(\sup_{\beta\in[\beta_l,\beta_u]} Z_t(\beta) \le x\Big)\Big| \to 0.$$

Corollary 2.19.1. *Under the conditions of Theorem 2.19 and Assumption 13, we have, conditional on $\mathcal{F}_{t-1}$,*

$$\begin{aligned}
\sup_{\beta_1,\beta_2,\beta_3:\beta_1+\beta_3=2\beta_2} \sqrt{n_t/J_t}|T_{n,t}(\beta_{1,2,3}) - T_t(\beta_{1,2,3})| &\lesssim_{\mathrm{P}} 1/\sqrt{J_t} \vee \sqrt{n_t}h/\sqrt{J_t} \vee \sqrt{n_t/J_t}J_t^{-1}, \\
\sup_{\beta_1,\beta_2,\beta_3:\beta_1+\beta_3=2\beta_2} \sqrt{n_t/J_t}|T_{n,t}(\beta_{1,2,3}) - Z_t(\beta_{1,2,3})| &\lesssim_{\mathrm{P}} 1/\sqrt{J_t} \vee \sqrt{n_t}h/\sqrt{J_t} \vee \sqrt{n_t/J_t}J_t^{-1}.
\end{aligned}$$

And

$$\sup_{x\in\mathbb{R}} |\mathrm{P}(\sup_{\beta_1,\beta_2,\beta_3:\beta_1+\beta_3=2\beta_2} Z_{n_t}(\beta_{1,2,3}) \le x) - \mathrm{P}(\sup_{\beta_1,\beta_2,\beta_3:\beta_1+\beta_3=2\beta_2} Z_t(\beta_{1,2,3}) \le x)| \to 0. \tag{2.14}$$

Adopting the conditions and proofs as in Theorem 2.12. It is not hard to derive that, $|\{\hat{\Phi}_t\hat{\Phi}_t^\top\}^{-1}\{\hat{\Phi}_t(\Phi_t^{*\top}L_t^*)\} - L_t^*|_{\max} \lesssim_{\mathrm{P}} J_t l_{n,T}/n_t$, $|\{\hat{\Phi}_t\hat{\Phi}_t^\top\}^{-1}\{\hat{\Phi}_t C_t|_{\max} \lesssim_{\mathrm{P}} 1/J_t$, and $|\{\hat{\Phi}_t\hat{\Phi}_t^\top\}^{-1}[\{\hat{\Phi}_t\varepsilon_t\} - \{\Phi_t^*\varepsilon_t\}]|_{\max} \lesssim_{\mathrm{P}} J_t l_{n,T}/n_t$. $|(\{n_t^{-1}\hat{\Phi}_t\hat{\Phi}_t^\top\}^{-1} - \mathrm{diag}\{\tilde{q}_{jt}\}^{-1})\{n_t^{-1}\Phi_t^*\varepsilon_t\}|_{\max} \lesssim |n_t^{-1}\Phi_t^*\varepsilon_t|_{\max}\sqrt{\log J_t}J_t^{3/2}/\sqrt{n_t}$. The following corollary provides the theoretical support of the uniform confidence band.

We define $Z_t(\beta) = \sum_{j_t} \hat{p}_{j_t,t}(\beta) Z_{j_t}$, where Z_{j_t}s are standard normal random variables.

Corollary 2.19.2. *Under conditions of Theorem 2.12, and Assumption 13,*

$$\mathrm{P}(\sup_\beta \sqrt{n_t/J_t} T_{c,t}(\beta) - Z_t(\beta) \geq \delta_{lt} \vee \sqrt{J_t}^{-1}) \to 0,$$

$$\mathrm{P}(\sup_\beta \sqrt{n_t/J_t} (\hat{L}_t(\beta) - \hat{L}_t(\beta)^{H^1})/\hat{\sigma}_t^{1/2}(\beta) - Z_t(\beta) \geq \delta_{lt} \vee \sqrt{J_t}^{-1}) \to 0,$$

$$\mathrm{P}(\sup_\beta \sqrt{n_t/J_t} (\hat{L}_t(\beta) - \hat{L}_t(\beta)^{H^2})/\hat{\sigma}_t^{1/2}(\beta) - Z_t(\beta) \geq \delta_{lt} \vee \sqrt{J_t}^{-1}) \to 0.$$

Moreover, the above the results implies that,

$$\sup_x \mathrm{P}(\sup_\beta (\sqrt{n_t/J_t} T_{c,t}(\beta) \geq x) - \mathrm{P}(\sup_\beta Z_t(\beta \geq x) \to 0, \tag{2.15}$$

$$\sup_x \mathrm{P}(\sup_\beta \sqrt{n_t/J_t} (\hat{L}_t(\beta) - \hat{L}_t(\beta)^{H^1})/\hat{\sigma}_t(\beta)^{1/2} \geq x) - \mathrm{P}(\sup_\beta Z_t(\beta) \geq x) \to 0, \tag{2.16}$$

$$\sup_x \mathrm{P}(\sup_\beta \sqrt{n_t/J_t} (\hat{L}_t(\beta) - \hat{L}_t(\beta)^{H^2})/\hat{\sigma}_t(\beta)^{1/2} \geq x) - \mathrm{P}(\sup_\beta Z_t(\beta) \geq x) \to 0. \tag{2.17}$$

In the following we discuss how to make uniform inference on $T^{-1}\sum_t \mu_t(\beta)$ by applying a strong approximation of the leading term $T^{-1}\sum_t \sum_{j_t} \hat{p}_{j_t,t}(\beta)(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t}$ of $T^{-1}\sum_t\{\hat{\mu}_t(\beta) - \mu_t(\beta)\}$, (and recall that $E_{n_t,j} = \tilde{q}_{jt}^{-1}\mathbb{E}(\Phi^*_{i,j,t}\beta_{it}|\mathcal{G}_{t-1})$). As we can see that it is a partial sum of martingale difference sequence. The term $\hat{\mu}(\beta) - \mu(\beta)$ is dominated by $T^{-1}\sum_t \sum_{j_t} \hat{p}_{j_t,t}(\beta)(f_t - \mathbb{E}(f_t))E_{n_t,j_t}$ though. Let $\mathbb{J}_1$ and $\mathbb{J}_2$ be two sequence of indices corresponding to two distinct evaluation point β_1 and β_2. We shall provide results on approximating the maximum over a finite number of points.

Lemma 2.20. *We define $\mathbb{E}_{\mathbb{J}_1} = (E_{n_t,j_t})_t$, where $j_t \in \mathbb{J}_1$. And $\mathbb{E}_{\mathbb{J}_2} = (E_{n_t,j_t})_t$ $(T \times 1)$, where $j_t \in \mathbb{J}_2$. $\Sigma_f = diag(Var(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})))$. Therefore $C_{\mathbb{J}_1,\mathbb{J}_2} = lim_{T\to\infty} T^{-1}\mathbb{E}_{\mathbb{J}_1}^\top \Sigma_f \mathbb{E}_{\mathbb{J}_2} > 0$. Recall that $J_a = |B_\mathbb{J}|$ are the cardinality of $B_\mathbb{J}$. So let $\tilde{\Sigma}$ be a matrix of dimension $J_a \times J_a$ with element $C_{\mathbb{J}_1,\mathbb{J}_2}$. $C_{\mathbb{J},\mathbb{J}} = diag(\tilde{\Sigma})$. Assume that $\tilde{Z}_\mathbb{J}$ follows a normal distribution with $N(0, diag(\tilde{\Sigma})^{-1/2}\tilde{\Sigma} diag(\tilde{\Sigma})^{-1/2})$. Under the conditions of Theorem 2.12, 2.14 and $\sqrt{T}(1/J \vee h) \to 0$, we have*

$$sup_x |\,\mathrm{P}(\max_{\mathbb{J}\in B_\mathbb{J}} |T^{-1/2}\sum_t (C_{\mathbb{J},\mathbb{J}})^{-1/2}(\hat{a}_{j_t,t} - a^*_{j_t,t})| \geq x) - \mathrm{P}(\max_{\mathbb{J}\in B_\mathbb{J}} |\tilde{Z}_\mathbb{J}| \geq x)| \to 0. \tag{2.18}$$

2.5 Theorems on the uniform inference for time average estimators

Lemma 2.21. *Under the conditions of Theorem B.6, 2.14, Assumption 14 and* $\sqrt{T}(1/J \vee h) \to 0$, *we have,*

$$\sup_{x\in\mathbb{R}} \Big| \mathrm{P}\left(\sup_{\beta\in\mathcal{B}} \Big| \frac{1}{\sqrt{T\mathbb{E}[\widehat{\sigma}(\beta)]}} \sum_{t=1}^{T} \left(\widehat{\mu}_t(\beta) - \mu_t(\beta)\right) \Big| \le x \right) - \mathrm{P}\left(\sup_{\beta\in\mathcal{B}} \big| G_T(\beta) \big| \le x \right) \Big| \to 0. \tag{2.19}$$

Lemma 2.22. *Under the conditions of Theorem B.6, 2.14, Assumption 14 and* $\sqrt{T}(1/J \vee h) \to 0$, *we have,*

$$\sup_{x\in\mathbb{R}} \Big| \mathrm{P}\left(\big| \mathcal{T}_T \big| \le x \right) - \mathrm{P}\left(\big| \tilde{T}_z \big| \le x \right) \Big| \to 0. \tag{2.20}$$

The following theorem states that we can use the quantile of the Gaussian process $\sup_{\beta_1,\beta_2,\beta_3} |G(\beta_1,\beta_2,\beta_3)|$ to approximate the distribution of our statistics of interest under the null. Thus the corresponding algorithm is listed as well afterwards.

Theorem 2.23. *Assume the conditions of Theorem B.6, 2.14, Assumption 12 , 14 and* $\sqrt{Tn_u/J}(1/J\vee h) \to 0$. *Also we define* $G(\beta_1,\beta_2,\beta_3)$ *as a Gaussian process with a finite number of jumps corresponding to the value of* β_1,β_2,β_3, *within each piece a standard normal distribution and across different points of the process has correlation* $Cov(\beta_{1,2,3},\beta'_{1,2,3})/(\sigma_D(\beta_{1,2,3})^{1/2}\sigma_D(\beta'_{1,2,3})^{1/2})$.

$$\sup_x |\mathrm{P}(|\sup_{\beta_1,\beta_2,\beta_3} |T^{-1}\sum_{t=1}^{T}\sqrt{Tn_t/J_t}\{\widehat{\mu}_t(\beta_1)+\widehat{\mu}_t(\beta_3)-2\widehat{\mu}_t(\beta_2)$$
$$-(\mu_t(\beta_1)+\mu_t(\beta_3)-2\mu_t(\beta_2))\}/\widehat{\sigma}_D(\beta_{1,2,3})^{1/2} \ge x) - \mathrm{P}(\sup_{\beta_1,\beta_2,\beta_3}|G(\beta_1,\beta_2,\beta_3)| \ge x)| \to 0.$$

3 Proofs

3.1 Proof of Lemma 2.4 and 2.5.

Proof. As we have

$$\begin{aligned}
& T^{-1} \sup_{\lfloor Th\rfloor \le t_0 \le T-\lfloor Th\rfloor} |A_i(t_0) - \mathbb{E}A_i(t_0)|_{\max} \\
= \; & \sup_{t_0}|T^{-1}\sum_{t=1}^{T}\{x_t x_t^\top - \mathbb{E}(x_t x_t^\top)\}w(t,t_0)|_{\max} + \sup_{t_0}|T^{-1}\sum_{t=1}^{T}\{x_t \lambda'(t/T)^\top\}w(t,t_0)|_{\max} \\
& +\sup_{t_0}|T^{-1}\sum_{t=1}^{T}\{\lambda'(t/T)x_t^\top\}w(t,t_0)|_{\max} \\
= \; & I_1 + I_2 + I_3.
\end{aligned}$$

Also by Assumption 6 we assume that $\Theta_{m,2q}(x_{.,l}) = m^{-v}$, for $2v > 1/2 - 1/q$, then we have by Lemma A.3 as in Zhang and Wu (2012), we define $r_T = (Th)^{-1}(T^{1/q} + (Th \log T)^{1/2})$

$$I_1, I_2, I_3 \lesssim_{\mathrm{P}} r_T.$$

We note that this results holds for nonstationary x_t as well. Similarly (2.3) holds. (2.4) holds give then fact that $\|x_{.}\varepsilon_{i.} - x^*_{.}\varepsilon^*_{i.}\|_q \le \|\varepsilon_{it}\|_{2q}\|x_t - x^*_t\|_{2q}$ for $t \neq 0$. Then we have $\|\varepsilon_{it}\|_{2q} \le c$.

The rate in (2.5) and (2.6) can be similarly proved by the Freedman's inequality in 2.2 and similar bounding technique following lemma 2.5. □

Proof. We have by summation by part

$$\begin{aligned}\max_{t_0} |(\sum_{t=1}^{T} u_t w(t,t_0))| \quad &\le \max_{t_0} \max_{t=t_0-\lfloor Th \rfloor \le \ell \le t_0-1} |\sum_{t=1}^{\ell} u_t|| \sum_{t_0-\lfloor Th \rfloor}^{t_0-1} [b_{it}w(t,t_0) - b_{it-1}w(t-1,t_0)]| \\ &+ \max_{t_0} |\sum_{t=t_0-\lfloor Th \rfloor}^{t_0} u_t| w(t_0,t_0).\end{aligned}$$

Now since $|\sum_{t_0-\lfloor Th \rfloor}^{t_0-1}[b_{it}w(t,t_0) - b_{it-1}w(t-1,t_0)]| \le \lfloor Th \rfloor/(Th) \lesssim 1$, due to the Lipschitz property of b_{it}, and the definition of kernel. Now apply the Freedman's inequality in 2.2, to the term $\max_{t_0} |\sum_{t=t_0-\lfloor Th \rfloor}^{t_0} u_t|$ and $\max_{t_0} \max_{t_0-\lfloor Th \rfloor \le \ell \le t_0-1} |\sum_{t=1}^{\ell} u_t|$, then the results follows from the Assumptions.

Let λ be a positive constant. Since we know that for martingales, we have
$\mathrm{P}(\max_{t_0-\lfloor Th \rfloor \le \ell \le t_0-1} |\sum_{t=t_0-\lfloor Th \rfloor}^{\ell} u_t| \ge 2\lambda) \le \mathrm{P}(|\sum_{t=t_0-\lfloor Th \rfloor}^{\ell} u_t| \ge \lambda)$, c.f. for example Theorem 2.4 Hall and Heyde (2014). Thus
$\max_{t_0} \max_{t_0-\lfloor Th \rfloor \le \ell \le t_0-1} |\sum_{t=1}^{\ell} u_t| \lesssim \max_{t_0 \in 2\lfloor Th \rfloor, 3\lfloor Th \rfloor, \cdots, T-\lfloor Th \rfloor} |\sum_{t=t_0-\lfloor Th \rfloor}^{t_0} u_t|$. □

3.2 Proof of Theorem 2.6.

Proof. We shall abbreviate $\sup_{\lfloor Th \rfloor \le t_0 \le T-\lfloor Th \rfloor}$ as $\sup_{t_0}$ in the following steps. Since
$\tilde{B}_i = \sum_t \mathbb{E}(X_t R_{it}|\mathcal{F}_{t-1})w(t,t_0) = \sum_t \mathbb{E}(X_t X_t^\top|\mathcal{F}_{t-1})b_{it}w(t,t_0)$, due to the Assumption 2. And by summation by part $\max_i T^{-1}|\tilde{B}_i - \overline{B}_i|_\infty \lesssim |T^{-1}\sum_t \mathbb{E}(X_t X_t^\top|\mathcal{F}_{t-1})w(t,t_0)|_\infty$
$\max_i |\sum_{t_0-\lfloor Th \rfloor \le t \le t_0-1}(b_{it} - b_{i(t-1)})|$. Because $\max_i |\sum_{t_0-\lfloor Th \rfloor \le t \le t_0-1}(b_{it} - b_{i(t-1)})| \le hC_\beta$ by Assumption 5, we have $T^{-1}|\tilde{B}_i - \overline{B}_i|_\infty \lesssim T^{-1}|\sum_t \mathbb{E}(X_t X_t^\top|\mathcal{F}_{t-1})C_\beta w(t,t_0)|_\infty h$.
As by Assumption 1 to 5,

$$\sup_{t_0} |\mathbb{E}[T^{-1}\sum_t \mathbb{E}[X_t X_t^\top|\mathcal{F}_{t-1}]C_\beta w(t,t_0)]|_\infty \le C, \tag{3.1}$$

where C is positive constant only depend on c_λ,C_β, and $c_{x,q}$.
Moreover, $|T^{-1}\sum_t\{\mathbb{E}[X_t X_t^\top|\mathcal{F}_{t-1}] - \mathbb{E}(X_t X_t^\top)\}w(t,t_0)|_\infty \lesssim_{\mathrm{P}} (Th)^{-1/2} \vee (Th)^{-1+1/q}$, by Assump-

tion 6, and theorem 2 of Wu and Wu (2016). The $\lesssim_{\mathrm{P}}$ depends on c_θ. And similar uniform arguments as in Lemma 2.4, $\sup_{t_0} |T^{-1}\sum_t\{\mathbb{E}[X_tX_t^\top|\mathcal{F}_{t-1}] - \mathbb{E}(X_tX_t^\top)\}w(t,t_0)|_\infty \lesssim_{\mathrm{P}} r_T$. Thus we know $\max_i \sup_{t_0} T^{-1}|\tilde{B}_i - \overline{B}_i|_\infty = O_p(h)$, provided that $r_T = o_p(1)$. Also we have

$$\max_i \sup_{t_0} |\tilde{A}^{-1}(\tilde{B}_i - \overline{B}_i)|_\infty \lesssim |\tilde{A}^{-1}\tilde{A}|_\infty \max_i |T^{-1} \sum_{t_0-\lfloor Th\rfloor \le t \le t_0-1} (b_{it} - b_{i(t-1)})|_\infty \le hC_\beta. \quad (3.2)$$

Then we look the following term,

$$A^{-1}B_i - \tilde{A}^{-1}\tilde{B}_i = -A^{-1}(A-\tilde{A})\tilde{A}^{-1}B_i + A^{-1}(B_i - \tilde{B}_i). \quad (3.3)$$

We denote $|A|_{\max} = \max_{i,j}|(A)_{i,j}|$, where $()_{i,j}$ denote the element on ith row and jth column of a matrix, and $|A|_\infty = \max_i \sum_j (A)_{i,j}$.

Recall Assumption 2 and 4. We define the constant $c_x = 2\sigma_x^2\int_{-1}^0 K(s)^2 ds$. $s_{f1} = \sum_t (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))w(t,t_0)$, and $s_{f2} = \sum_t (f_t^2 - \mathbb{E}(f_t^2|\mathcal{F}_{t-1}))w(t,t_0)$.

$$A - \tilde{A} = [0, s_{f1}; s_{f1}, s_{f2}]. \quad (3.4)$$

Following Freedman's inequality as in Lemma 2.2 and similar argument as in Lemma 2.4 in the Appendix, we have,

$$\sup_{t_0} |\sum_t (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))w(t,t_0)|| \lesssim_{\mathrm{P}} c_T\sqrt{c_x}\sqrt{T\log T}/\sqrt{h}, \quad (3.5)$$

for some large enough constant c_T.

$$\begin{aligned} s_{f2} &= \sum_t 2(x_t - \mathbb{E}_{t-1}(x_{t-1}))\lambda'(t/T)w(t,t_0) \\ &\quad + \sum_t \{x_t^2 - \mathbb{E}(x_t^2|\mathcal{F}_{t-1})\}w(t,t_0), \\ &\lesssim_{\mathrm{P}} \lambda'(t_0/T)\sqrt{T/h}\sqrt{c_x}c_T + \sqrt{T/h}\sqrt{\mathbb{E}(x_t^4)\int_{-1}^0 K(s)^2 ds}\, c_T. \end{aligned}$$

Moreover $\sup_{t_0} s_{f2} \lesssim_{\mathrm{P}} \lambda'(t_0/T)\sqrt{T\log T/h}\sqrt{c_x}c_T + \sqrt{T\log T/h}\sqrt{\mathbb{E}(x_t^4)\int_{-1}^0 K(s)^2 ds}\, c_T$ as derived in Lemma 2.4. Also

$$\sup_{t_0} T^{-1}|A - \tilde{A}|_\infty \lesssim_{\mathrm{P}} c'\sqrt{\log T/(Th)} \overset{\text{def}}{=} r_{AT}, \quad (3.6)$$

where c' depends on c_T, $\sqrt{c_x}$, $\sqrt{\mathbb{E}(x_t^4)\int_{-1}^0 K(s)^2 ds}$ and $\lambda'(t_0/T)$.

Now we look at $B_i - \tilde{B}_i$. Assume $\mathbb{E}(f_t\varepsilon_{it}|\mathcal{F}_{t-1}) = 0$, (implied by $\mathbb{E}(\varepsilon_{it}|\mathcal{F}_{t-1}, x_t) = 0$ and

$\mathbb{E}(\varepsilon_{it}|\mathcal{F}_{t-1}) = 0$ in Assumption 2). We can see that similarly for $B_i - \tilde{B}_i$, we have,

$$B_i - \tilde{B}_i = \sum_t X_t \varepsilon_{it} w(t, t_0) + \sum_t (X_t X_t^\top - \mathbb{E}_{t-1} X_t X_t^\top) w(t, t_0) b_{it}. \tag{3.7}$$

$\sum_t X_t \varepsilon_{it} w(t, t_0)$ is a summand of martingale difference sequence. Recall that we denote $\sigma_\varepsilon^2(i/T) = \mathbb{E}(\varepsilon_i^2)$, $\sigma_{\varepsilon,0}^2 = \mathbb{E}(\varepsilon_{t_0}^2)$. Thus again by Lemma 2.2 and 2.4, we have

$$\max_i \sup_{t_0} |\sum_{t=1}^T \varepsilon_{it} w(t, t_0)| \lesssim_{\mathrm{P}} c_T \sqrt{\sigma_{\varepsilon,0}^2 \frac{T \log(nT)}{h} \int_{-1}^0 K^2(s) ds}. \tag{3.8}$$

And according to Assumption 2, Lemma 2.2 and 2.4, we have $\mathbb{E}(\varepsilon_{it}^2 | \mathcal{F}_{t-1}, x_t) = \sigma_\varepsilon^2(t/T, \mathcal{F}_{t-1})$, and $\mathbb{E} x_t = 0$.

$$\max_i \sup_{t_0} |\sum_{t=1}^T f_t \varepsilon_{it} w(t, t_0)| \lesssim_{\mathrm{P}} c_T \{2\sigma_{\varepsilon,0} \tau'(t_0/T) \vee \sigma_x \sigma_{\varepsilon,0}\} \sqrt{\int_{-1}^0 K^2(s) ds} \sqrt{T \log(nT)/h}. \tag{3.9}$$

Moreover if $r_{AT} \to 0$,

$$\max_i \sup_{t_0} |T^{-1} \sum_t (X_t X_t^\top - \mathbb{E}_{t-1} X_t X_t^\top) w(t, t_0) b_{it} - (A - \tilde{A}) b_{it_0}|_{\max} \le \sup_{t_0} |A - \tilde{A}|_{\max} h = O_p(r_{AT} h) \tag{3.10}$$

by Assumption 5.

Thus we have

$$\max_i \sup_{t_0} |B_i - \tilde{B}_i|_2 \lesssim_{\mathrm{P}} r_{AT},$$

due to the rate of $|(A - \tilde{A})|_2$ and the boundedness of b_{it_0} as in Assumption 5. Assume that $\lambda_{\min}(T^{-1}\mathbb{E}(\tilde{A})) > c^{-1} > 0$, for a constant c. Since A and $\mathbb{E}(A)$ are symmetric real matrices. We then have $\mathrm{P}(|TA^{-1}|_2 \le \lambda_{\min}(T^{-1}A)^{-1} \le c) \le \mathrm{P}(\lambda_{\min}(T^{-1}A) \ge c^{-1})$. Now since $\lambda_{\min}(T^{-1}A) = \min_{|v|_2=1} |T^{-1}Av|_2 \ge \min_{|v|_2=1} |\mathbb{E}(T^{-1}A)v|_2 - \max_{|v|_2=1} |T^{-1}\{A - \mathbb{E}(A)\}v|_2$. Since $\min_{|v|_2=1} |T^{-1}\mathbb{E}(A)v|_2 = \lambda_{min}(T^{-1}\mathbb{E}(A)) > c^{-1}$, we need to show that $\max_{|v|_2=1} |T^{-1}\{A - \mathbb{E}(A)\}v|_2 \ge c^{-1}/2$ with probability approach 1. This is shown in Lemma 2.4.

We now write

$$\max_i \sup_{t_0} |A^{-1}B_i - \tilde{A}^{-1}\tilde{B}_i|_2 \le 2 \max_i \sup_{t_0} T|A^{-1}|_2 T^{-1}|A - \tilde{A}|_2 T|\tilde{A}^{-1}|_2 T^{-1}|B_i|_2 T|A^{-1}|_2$$
$$+2 \max_i \sup_{t_0} T|A^{-1}|_2 T^{-1}|B_i - \tilde{B}_i|_2.$$

The rate of $T^{-1}|A - \tilde{A}|_2$ is in view of equation (3.11), note that for fixed dimension matrix, the norm as equivalent. Also we have $T|A^{-1}|_2 \le (\lambda_{min}(T^{-1}A))^{-1} \le c$ by Assumption 3. $\max_{t_0} T|\tilde{A}^{-1}|_2 \le (\min_{t_0} \lambda_{min}(T^{-1}A) - \max_{t_0} \lambda_{max}(T^{-1}(\tilde{A} - A)))^{-1} \lesssim_{\mathrm{P}} (c^{-1} - r_{AT})^{-1}$, which is bounded by a positive constant as $T \to \infty$. Regarding the rate of $\sup_{t_0} \max_i T^{-1}|B_i|_2$, we have

$$\begin{aligned} & \sup_{t_0} \max_i T^{-1}|B_i|_2 \\ \leq \quad & \sup_{t_0} \max_i T^{-1}|B_i - \tilde{B}_i|_2 + \sup_{t_0} \max_i T^{-1}|\tilde{B}_i - \mathbb{E}B_i|_2 + \sup_{t_0} \max_i \mathbb{E}(B_i) \\ \lesssim_{\mathrm{P}} \quad & r_{AT} + r_T + C_B, \end{aligned}$$

which is due to Lemma 2.4 and Assumption 3.

Thus we have $\max_i \sup_{t_0} |A^{-1}B_i - \tilde{A}^{-1}\tilde{B}_i|_2 \lesssim_{\mathrm{P}} (r_T + \sqrt{\log(nT)}/\sqrt{Th} + h)$.

□

3.3 Proof of Theorem 2.7

Next we provide a pointwise central limit theorem for $\widehat{\mu}(\beta)$ which allows us to make pointwise inference on the estimator of $T^{-1}\sum_{t=1}^{T} \mu_t(.)$. We define $E_{n_t,j} = \tilde{q}_{jt}^{-1}\mathbb{E}(\Phi^*_{i,j,t}\beta_{it}|\mathcal{G}_{t-1})$, which may degenerate for some j. Recall that $\mathbb{E}_{t-1}(\Phi^*_{i,j_t,t}) = \tilde{q}_{jt} = \mathbb{E}(\Phi^*_{i,j_t,t}|\mathcal{G}_{t-1})$.

Let $\tilde{f}_t = f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})$,

$$\eta_{t,n_t}(\beta) = \mathbb{E}(\widehat{\sigma}(\beta))^{-1/2} \sum_{j_t} \hat{p}_{j_t,t}(\beta)(\tilde{f}_t E_{n_t,j_t} + \tilde{q}_{jt}^{-1} n_t^{-1} \sum_i \Phi^*_{i,j_t,t}\varepsilon_{it}).$$

Define

$$T_{n,1}(\beta) = \sum_{j_t} \hat{p}_{j_t,t}(\beta)\mathbb{E}(\widehat{\sigma}(\beta))^{-1/2}\sqrt{T}^{-1} \sum_t (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))\tilde{q}_{jt}^{-1} n_t^{-1} \sum_i \Phi^*_{i,j_t,t}\beta_{it}$$

and

$$T_{n,2}(\beta) = \sum_{j_t} \hat{p}_{j_t,t}(\beta)\mathbb{E}(\widehat{\sigma}(\beta))^{-1/2}\sqrt{T}^{-1} \sum_t \tilde{q}_{jt}^{-1} n_t^{-1} \sum_i \Phi^*_{i,j_t,t}\varepsilon_{it}.$$

Then we have

$$\begin{aligned} T_{n,1}(\beta) + T_{n,2}(\beta) &= \mathbb{E}(\widehat{\sigma}(\beta))^{-1/2}\sqrt{T}^{-1} \sum_t \sum_{j_t} \hat{p}_{j_t,t}(\beta)(\tilde{f}_t E_{n_t,j_t} + \tilde{q}_{jt}^{-1} n_t^{-1} \sum_i \Phi^*_{i,j_t,t}\varepsilon_{it}) \\ &= \sqrt{T}^{-1} \sum_t \eta_{t,n_t}(\beta). \end{aligned}$$

Define an integer $q \geq 2 + 2\delta$. Recall that $\delta_{q,m}(X_.) = \max_t \|X_t - X^*_{t-m}\|_q$, where X^*_{t-m} is replaced with an iid copy at time point 0. Define $\Theta_{m_0,q}(X_.) = \sum_{m \geq m_0} \delta_{q,m}(X_.)$. Let $\xi > 0$ be a positive constant. Recall the definition of the dependence adjusted norm:

$\max_{j_t} \|\hat{p}_{j_t,.}(\beta)\tilde{f}_. E_{n_t,j_t}\|_{q,\xi} = \max_{j_t} \sup_{m_0 \geq 0} m_0^{\xi} \Theta_{m_0,q}(\hat{p}_{j_t,.}(\beta)\tilde{f}_. E_{n_t,j_t})$ and similarly for $\max_{i,j_t} \|\Phi^*_{i,j_t,.}\varepsilon_{i.}\|_{q,\xi}$.

Proof. Since we have already proved that $\max_i \sup_{t_0} |\tilde{A}^{-1}(\tilde{B}_i - \overline{B}_i)|_\infty \leq hC_\beta$. We just need look at the term,

$$\begin{aligned}
&(\hat{b}_{it_0} - b_{it_0})\\
&= A^{-1}B_i - \tilde{A}^{-1}\tilde{B}_i = -A^{-1}(A-\tilde{A})\tilde{A}^{-1}B_i + A^{-1}(B_i - \tilde{B}_i) + O_p(h),\\
&= -A^{-1}(A-\tilde{A})\tilde{A}^{-1}B_i + A^{-1}(B_i - \tilde{B}_i) + O_p(h),\\
&= -A^{-1}(A-\tilde{A})\tilde{A}^{-1}(B_i - \tilde{A}b_{it_0}) - A^{-1}\{(A-\tilde{A})b_{it_0} - (B_i - \tilde{B}_i)\} + O_p(h),\\
&\stackrel{\text{def}}{=} I_{11} + I_{12} + O_p(h).
\end{aligned}$$

From the proof of Theorem 2.6, we have $|I_{11}|_2 \lesssim_{\mathrm{P}} r_{AT}h$. $I_{12} = -A^{-1}\sum_t w(t,t_0)\varepsilon_{it}X_t + O_p(r_{AT}h)$. Thus we shall apply a martingale central limit theorem on the term $-(T^{-1}A)^{-1}\sqrt{h/T}\sum_t w(t,t_0)\varepsilon_{it}X_t$, which correspond to the leading term of $\sqrt{Th}(\hat{b}_{it_0} - b_{it_0})$. We shall prove that it is close to $-(T^{-1}\mathbb{E}(A))^{-1}\sqrt{h/T}\sum_t w(t,t_0)\varepsilon_{it}X_t$.

For this purpose we check,

$$T(A^{-1} - (\mathbb{E}A)^{-1}) = -(I + (T^{-1}\mathbb{E}A)^{-1}T^{-1}(A - \mathbb{E}A))^{-1}(T^{-1}\mathbb{E}A)^{-1}T^{-1}(A-\mathbb{E}A)(T^{-1}\mathbb{E}A)^{-1}.$$

By Assumption 3, $c^{-1} < \lambda_{min}(T^{-1}\mathbb{E}A) \leq \lambda_{max}(T^{-1}\mathbb{E}A) < C_{A,\max}$. Since we proved that $T^{-1}|A - \mathbb{E}A|_2 \lesssim_{\mathrm{P}} r_{AT} \vee r_T$ by Lemma 2.4. So we have

$$T^{-1}|A^{-1} - (\mathbb{E}A)^{-1}|_2 \lesssim c^2 r_{AT}(1 - c(T^{-1}|A - \mathbb{E}A|_2)^{-1}) \lesssim_{\mathrm{P}} c^2 r_{AT} \vee r_T.$$

Thus $\sqrt{T/h}| - A^{-1}\sum_t w(t,t_0)\varepsilon_{it}X_t - \mathbb{E}(A)^{-1}\sum_t w(t,t_0)\varepsilon_{it}X_t|_2 \lesssim_{\mathrm{P}} r_{AT} \vee r_T$. As $O_p((r_{AT} \vee r_T \vee h) = o_p(1)$, by the Assumption of this theorem. Then we have the elements of $\sqrt{Th}\Sigma_b^{-1/2}(\hat{b}_{it_0} - b_{it_0}) = \sqrt{h/T}\Sigma_b^{-1/2}\Sigma_A^{-1}\sum_t X_t w(t,t_0)\varepsilon_{it} + O_p(1)$ as a martingale difference with respect to $\mathcal{F}_{t-1}$ by Assumption 2.
We shall use Corollary 3.1 in Hall and Heyde (2014), with η therein as 1. The following two Assumption are needed to be verified. For a constant $c > 0$,

i) $\mathbb{E}|h/T\sum_t \mathbb{E}(\varepsilon_{it}^2 w(t,t_0)^2(e_l^\top\Sigma_b^{-1/2}(\Sigma_A)^{-1}X_t)^2|\mathcal{F}_{t-1}) - 1| \to 0$,

ii) $\sum_t h/T\mathbb{E}[(\varepsilon_{it}^2 w(t,t_0)^2(e_l^\top\Sigma_b^{-1/2}(\Sigma_A)^{-1}X_t)^2\mathbf{1}\{\sqrt{h}\varepsilon_{it}w(t,t_0)(e_l^\top\Sigma_b^{-1/2}(\Sigma_A)^{-1}X_t)/\sqrt{T} > c\})|\mathcal{F}_{t-1}] \to 0$.

For i) we have,

$$\begin{aligned}
&h/T\sum_t \mathbb{E}(\varepsilon_{it}^2 w(t,t_0)^2(e_l^\top\Sigma_b^{-1/2}X_t)^2|\mathcal{F}_{t-1})\\
&= e_l^\top\Sigma_b^{-1/2}\Sigma_b\Sigma_b^{-1/2}e_l + O(h)\\
&= 1 + O(h).
\end{aligned}$$

Therefore i) holds under Assumption 1- 6 and $h\sqrt{hT} \to 0$. ii) holds obviously when $1/Th \to 0$ under Assumption 1- 6 and $w(t,t_0)^2 \lesssim h^{-2}$. The desired results follow.

□

3.4 Proof of Lemma 2.8.

Proof. Recall from Theorem 2.6 we have the rate $\max_i \sup_t |\hat{\beta}_{it} - \beta_{it}| \lesssim_{\mathrm{P}} c_{n_u}(h + (Th)^{-1/2}(\log T)^{1/2}) = \delta_T$. Recall that we let $k_{jt} = \lfloor n_t j/J_t \rfloor$, $\kappa_{jt} = j/J_t$. According to for example Corollary 21.5, page 307, Van der Vaart (2000).

$$\hat{\beta}_{(k_{jt}),t} - F^{-1}_{\hat{\beta},t}(\kappa_{jt}) = -(F_{\hat{\beta},n,t}(F^{-1}_{\hat{\beta},t}(\kappa_{jt})) - F_{\hat{\beta},t}(F^{-1}_{\hat{\beta},t}(\kappa_{jt})))/f_{\hat{\beta},t}(F^{-1}_{\hat{\beta},t}(\kappa_{jt})) + o_p(1/\sqrt{n_t}). \quad (3.11)$$

Let the set $\mathbb{N}_{J_t} = \{F^{-1}_{\hat{\beta},t}(\kappa_{jt})\}_{jt=1,\cdots,J_t}$. We denote $G_{n,\beta,t}(x) = \sqrt{n_t}(F_{\beta,n,t}(x) - F_{\beta,t}(x))$. For sufficient large constant $c_{n_u t}$, we have,

$$\begin{aligned} & |(F_{\hat{\beta},n,t}(F^{-1}_{\hat{\beta},t}(\kappa_{jt})) - F_{\hat{\beta},t}(F^{-1}_{\hat{\beta},t}(\kappa_{jt})))/f_{\hat{\beta},t}(F^{-1}_{\hat{\beta},t}(\kappa_j))| \\ \leq \quad & \max_t c^{-1}_{\beta,\min} \sup_{|u|\leq\delta_T, x\in\mathbb{N}_{J_t}} (n_t^{-1/2}|G_{n,\beta,t}(x+u) - G_{n,\beta,t}(x)| + n_t^{-1/2}|G_{n,\beta,t}(x)|) \\ \lesssim_{\mathrm{P}} \quad & \max_t c^{-1}_{\beta,\min} \sqrt{\delta_T \log J_t} c_{n_u t}/\sqrt{n_t} + O_p(\sqrt{\log(T \vee J_t)}\sqrt{n_t}^{-1}) = a_{nT}, \end{aligned}$$

where the second last line is due to Assumption 7 and the fact that $|F^{-1}_{\hat{\beta},t}(\kappa_{jt}) - F^{-1}_{\beta,t}(\kappa_{jt})| \lesssim_{\mathrm{P}} \delta_T$; the last line is due to the modulus of continuity by Lemma 2.3 as in Stute (1982), the Assumption 7 and the uniform inequality in Massart (1990) with the union bound. Recall that we define $a_{nT} = \max_t(\sqrt{\delta_T \log J_t} c_{n_u t}/\sqrt{n_t}) \vee \sqrt{\log(T \vee J_t)}\sqrt{n_t}^{-1}$. □

3.5 Proof of Lemma 2.9

Proof. **Step 1** We define that $U_{(\tilde{k}_{jt}),t}$ is between $U_{(k_{jt}),t}$ and $U_{(k_{j-1t}),t}$, and we have for the last line $1/J_t - 2/(J_t(n_t+1)) \leq (k_{jt} - k_{j-1t})/(n_t+1) \leq 1/J_t + 2/(J_t(n_t+1))$. Since $U_{(k_{jt}),t} - U_{(k_{j-1t}),t}$ follow beta distribution with parameter $(k_{jt} - k_{j-1t}, n_t + 1 - (k_{jt} - k_{j-1t}))$, therefore $\mathbb{E}[(U_{(k_{jt}),t}) - (U_{(k_{j-1t}),t})] = (k_{jt} - k_{j-1t})/(n_t+1)$. Following B.10, in Bobkov, Gentil, and Ledoux (2001),

$$\begin{aligned} & \mathrm{P}(\min_t \min_j |\beta_{(k_{jt}),t} - \beta_{(k_{j-1t}),t}| > 2c^{-1}_{\beta,\min} 1/J_t) \\ & \leq \mathrm{P}(\max_t \max_j |\beta_{(k_{jt}),t} - \beta_{(k_{j-1t}),t}| > 2c^{-1}_{\beta,\min} 1/J_t) \\ & \leq \sum_t \sum_{j=1}^{J_t} \mathrm{P}(|F^{-1}_{\beta,t}(U_{(k_{jt}),t}) - F^{-1}_{\beta,t}(U_{(k_{j-1t}),t})| \geq 2c^{-1}_{\beta,\min} 1/J_t) \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_t \sum_{j=1}^{J_t} \mathrm{P}(f_{\beta,t}^{-1}(F_{\beta,t}^{-1}(U_{(\tilde{k}_{jt}),t}))|(U_{(k_j),t}) - (U_{(k_{j-1t}),t})| \geq 2c_{\beta,\min}^{-1} 1/J_t) \\
&\leq \sum_t \sum_{j=1}^{J_t} \mathrm{P}(c_{\beta,\min}^{-1}|(U_{(k_{jt}),t}) - (U_{(k_{j-1t}),t}) - \mathbb{E}((U_{(k_{jt}),t}) - (U_{(k_{j-1t}),t}))| \\
&\geq 2c_{\beta}^{-1} 1/J_t - c_{\beta,\min}^{-1}\mathbb{E}((U_{(k_{jt}),t}) - (U_{(k_{j-1t}),t}))) \\
&\leq \sum_t \sum_{j=1}^{J_t} \mathrm{P}(c_{\beta,\min}^{-1}|(U_{(k_{jt}),t}) - (U_{(k_{j-1t}),t}) - \mathbb{E}((U_{(k_{jt}),t}) - (U_{(k_{j-1t}),t}))| \\
&\geq 2c_{\beta}^{-1} 1/J_t - c_{\beta,\min}^{-1}(k_{jt} - k_{j-1t})/(n_t+1)) \\
&\leq 2TJ_u \max_t \exp(-C'(n_t+1)(c_{\beta,\min}^{-1} J_t^{-1}(1-(n_t+1)^{-1}))^2),
\end{aligned}
$$

the above term is tending to 0 provided the fact that by Assumption 8, we have $J^{-1} \gg \frac{\sqrt{\log TJ}}{\sqrt{n+1}}$, which implies that $J_t^{-1}(1-(n_t+1)^{-1}) \gg \frac{\sqrt{\log J_t}}{\sqrt{n_t+1}}$.

Conditioning on the event $\{\max_t \max_j |\beta_{(k_{jt}),t} - \beta_{(k_{j-1t},t)}| \leq 1/J\}$, which we can prove that the probability is tending to 1, we have the following inequality,

$$
\begin{aligned}
& \max_{t,j} |\sum_{i=1}^{n_t} (\mathbf{1}(\hat{\beta}_{it} \in \hat{P}_{jt}) - \mathbf{1}(\beta_{it} \in P_{jt}))| \\
\leq\quad & \max_{t,j} \sqrt{n_t}|G_{n,\hat{\beta},t}(\hat{\beta}_{(k_{jt}),t}) - G_{n,\beta,t}(\beta_{(k_{jt}),t})| + \sqrt{n_t} \max_{t,j} |G_{n,\hat{\beta},t}(\hat{\beta}_{(k_{j-1t}),t}) - G_{n,\beta,t}(\beta_{(k_{j-1t}),t})| \\
& + \sum_{i=1}^{n_t} (\mathrm{P}(\hat{\beta}_{it} \in \hat{\mathrm{P}}_{jt}) - \mathrm{P}(\beta_{it} \in P_{jt})), \\
\leq\quad & \max_t 2\sqrt{n_t} \sup_{|u| \leq a_{nT} + \delta_T, x \in \mathbb{N}_{J_t}} |G_{n,\beta,t}(x+u) - G_{n,\beta,t}(x)| \\
& + \sum_{i=1}^{n_t} (\mathrm{P}(\hat{\beta}_{it} \in \hat{\mathrm{P}}_{jt}) - \mathrm{P}(\beta_{it} \in P_{jt})), \\
\lesssim_{\mathrm{P}}\quad & \max_t \sqrt{n_t}\sqrt{a_{nT} + \delta_T} c_{n_u t} + n_t(a_{nT} + \tilde{\delta}_T)/J_t = l_{n,T},
\end{aligned}
$$

where the inequality is based on the rate of $\max_{j,t} |\hat{\beta}_{(k_{jt}),t} - \beta_{(k_{jt}),t}| \lesssim_{\mathrm{P}} a_{nT}$, and $\max_{i,t} |\hat{\beta}_{it} - \beta_{it}| \lesssim_{\mathrm{P}} \delta_T$. So according to the modulus of continuity by Lemma 2.3 as in Stute (1982).

Step 2 Recall that our population target is defined as,

$$
a_t^* = [\mathbb{E}(\Phi_t^* \Phi_t^{*\top} | \mathcal{G}_{t-1})]^{-1} [\mathbb{E}(\Phi_t^* \mu_t(\beta_{.t})^\top | \mathcal{G}_{t-1})].
$$

Define $\sigma(\beta_{.t}))$ as the sigma field of $[\beta_{it}]_i$. Also we define an intermediate version of the estimator as,

$$
\tilde{a}_t^* = [\Phi_t^* \Phi_t^{*\top}][\mathbb{E}(\Phi_t^* R_{.t}^\top | \mathcal{G}_{t-1}, \sigma(\beta_{.t}))].
$$

Let $\mu_t(\beta_{.t})$ be $1 \times n_t$ vector of $[\mu_t(\beta_{it})]_i^\top$. Moreover, we see the conditional version equals to,

$$\mathbb{E}(\Phi_t^* R_{.t}^\top | \mathcal{G}_{t-1}, \sigma(\beta_{.t})) = \mathbb{E}(\Phi_t^* R_{.t}^\top | \mathcal{G}_{t-1}, \sigma(\beta_{.t})) = \Phi_t^* \mu_t(\beta_{.t})^\top.$$

Thus we have the intermediate estimator defined as,

$$\tilde{a}_t^* = [\Phi_t^* \Phi_t^{*\top}]^{-1} [\Phi_t^* \mu_t(\beta_{.t})].$$

The bias term is thus expressed as,

$$\flat_t(\beta) = \mu_t(\beta) - \hat{p}_t(\beta)^\top a_t^* = \mu_t(\beta) - \hat{p}_t(\beta)^\top \tilde{a}_t^* + \hat{p}_t(\beta)^\top \tilde{a}_t^* - \hat{p}_t(\beta)^\top a_t^*,$$

since

$$|\mu_t(\beta) - p_t(\beta)^\top \tilde{a}_t^*| \leq \max_j |\mu_t(\beta_{(k_{j,t}),t}) - \mu_t(\beta_{(k_{j-1,t}),t})|,$$

$$\begin{aligned}\sup_\beta |\flat_t(\beta)| &\leq \max_j |\mu_t(\beta_{(k_{j,t}),t}) - \mu_t(\beta_{(k_{j-1,t}),t})| + |a_t^* - \tilde{a}_t^*|_\infty \\ &\lesssim_p 1/J_t.\end{aligned}$$

$$\begin{aligned}|a_t^* - \tilde{a}_t^*|_\infty &= |[\mathbb{E}(\Phi_t^* \Phi_t^{*\top} | \mathcal{G}_{t-1})]^{-1} [\mathbb{E}(\Phi_t^* \mu_t(\beta_{.t})^\top | \mathcal{G}_{t-1})] - [\Phi_t^* \Phi_t^{*\top}]^{-1} [\mathbb{E}(\Phi_t^* R_{.t}^\top | \mathcal{G}_{t-1}, \sigma(\beta_{.t}))]|_\infty \\ &= |[\mathbb{E}(\Phi_t^* \Phi_t^{*\top} | \mathcal{G}_{t-1})]^{-1} (\Phi_t^* \Phi_t^{*\top} - \mathbb{E}(\Phi_t^* \Phi_t^{*\top} | \mathcal{G}_{t-1})) [\Phi_t^* \Phi_t^{*\top}]^{-1} [\mathbb{E}(\Phi_t^* R_{.t}^\top | \mathcal{G}_{t-1}, \sigma(\beta_{.t}))] + \\ &\quad |[\mathbb{E}(\Phi_t^* \Phi_t^{*\top} | \mathcal{G}_{t-1})]^{-1}] [-\Phi_t^* \mu_t(\beta_{.t}) + \mathbb{E}(\Phi_t^* \mu_t(\beta_{.t})^\top | \mathcal{G}_{t-1})]|_\infty \\ &\lesssim_p J_t^{-1},\end{aligned}$$

where the above inequality is due to Bernstein inequalities. Moreover, we have,

$$\begin{aligned}&\max_{t,j} |\mu_t(\beta_{(k_{jt}),t}) - \mu_t(\beta_{(k_{j-1t}),t})|, \\ \leq\ &\max_{t,j} c_{\alpha,\min} |\beta_{(k_{jt}),t} - \beta_{(k_{j-1t}),t}| \lesssim_{\mathrm{P}} 1/J,\end{aligned}$$

where the last inequality follows from Assumption 8. And the last claim follows from the above derivation. Thus the statement is proved.

Step 3 Analyzing $\sum_{i=1}^{n_t} (\mathrm{P}(\hat{\beta}_{it} \in \hat{P}_{jt}) - \mathrm{P}(\beta_{it} \in P_{jt}))$.

We shall let

$$\hat{\beta}_{it}-\beta_{it}=-e_2^\top(T^{-1}E(A))^{-1}\sqrt{h/T}\sum_{t'}w(t',t)\varepsilon_{it'}X_{t'}+O_p(r_{AT}h)+e_2^\top b_i'(t/T)E(X_tX_t^\top)h=v_{it}+w_{it},$$

where the first term corresponds to the variance and the second term corresponds to the bias term. v_{it} is of order $\sqrt{Th}^{-1}$ and w_{it} is of order h.

$$\begin{aligned}&\sum_{i=1}^{n_t}(\mathrm{P}(\hat{\beta}_{it}\in\hat{P}_{jt})-\mathrm{P}(\beta_{it}\in P_{jt}))\\ =&\sum_{i=1}^{n_t}(\mathrm{P}(\beta_{it}\le F^{-1}_{n,\hat{\beta},t}(j/J_t)+(\beta_{it}-\hat{\beta}_{it}))-\mathrm{P}(\beta_{it}\le F^{-1}_{n,\beta,t}(j/J_t)))\\ -&\sum_{i=1}^{n_t}(\mathrm{P}(\beta_{it}\le F^{-1}_{n,\hat{\beta},t}((j-1)/J_t)+(\beta_{it}-\hat{\beta}_{it}))-\mathrm{P}(\beta_{it}\le F^{-1}_{\beta,t}((j-1)/J_t))).\end{aligned}$$

Since,

$$\begin{aligned}&\sum_i[\mathrm{P}(\beta_{it}\le F^{-1}_{n,\hat{\beta},t}(j/J_t)+(\beta_{it}-\hat{\beta}_{it}))-\mathrm{P}(\beta_{it}\le F^{-1}_{n,\beta,t}(j/J_t))]\\ =&\sum_i[\mathrm{P}(\beta_{it}\le F^{-1}_{\beta,t}(j/J)+F^{-1}_{n,\hat{\beta},t}(j/J_t)-F^{-1}_{\beta,t}(j/J_t)+(\beta_{it}-\hat{\beta}_{it}))-\mathrm{P}(\beta_{it}\le F^{-1}_{\beta,t}(j/J_t))].\end{aligned}$$

We let

$$F^{-1}_{n,\hat{\beta},t}(j/J_t)-F^{-1}_{\beta,t}(j/J_t)+(\beta_{it}-\hat{\beta}_{it})=c_{n,j,t}+v_{it}+w_{it}.$$

$$c_{n,j,t}=n_t^{-1}\sum_i[\mathbf{1}(\hat{\beta}_{it}\le F^{-1}_{\beta,t}(j/J_t))-P(\hat{\beta}_{it}\le F^{-1}_{\beta,t}(j/J_t))]+o_p(\sqrt{n_t}^{-1}).$$

Thus let $\tilde{c}_{n,j_t}$ be a middle point between $F^{-1}_{n,\hat{\beta},t}(j/J_t)+(\beta_{it}-\hat{\beta}_{it})$ and $F^{-1}_{\beta,t}(j/J_t)$.

$$\begin{aligned}&n_t^{-1}\sum_{i=1}^{n_t}(\mathrm{P}(\beta_{it}\le F^{-1}_{n,\hat{\beta},t}(j/J_t)+(\beta_{it}-\hat{\beta}_{it}))-\mathrm{P}(\beta_{it}\\ &\le F^{-1}_{\beta,t}(j/J_t)))=n_t^{-1}\sum_i(f_{\beta,t}(\tilde{c}_{i,j_t})(c_{n,j,t}+v_{it}+w_{it}))+o_p(h+\sqrt{n_tTh}^{-1}+\sqrt{n_t}^{-1}),\\ &=n_t^{-1}\sum_i(f_{\beta,t}(\tilde{c}_{i,j_t})(c_{n,j,t}))+n_t^{-1}\sum_i(f_{\beta,t}(\tilde{c}_{i,j_t})v_{it})\\ &+n_t^{-1}\sum_i(f_{\beta,t}(\tilde{c}_{i,j_t})w_{it})+o_p(h+\sqrt{n_tTh}^{-1}+\sqrt{n_t}^{-1}).\end{aligned}$$

So we have,

$$
\begin{aligned}
&\sum_{i=1}^{n_t}(\mathrm{P}(\hat{\beta}_{it}\in\hat{P}_{jt})-\mathrm{P}(\beta_{it}\in P_{jt}))\\
&=\sum_{i=1}^{n_t}(\mathrm{P}(\beta_{it}\le F^{-1}_{n,\hat{\beta},t}(j/J_t)+(\beta_{it}-\hat{\beta}_{it}))-\mathrm{P}(\beta_{it}\le F^{-1}_{n,\beta,t}(j/J_t)))\\
&-\sum_{i=1}^{n_t}(\mathrm{P}(\beta_{it}\le F^{-1}_{n,\hat{\beta},t}((j-1)/J_t)+(\beta_{it}-\hat{\beta}_{it}))+\mathrm{P}(\beta_{it}\le F^{-1}_{\beta,t}((j-1)/J_t)))\\
&=n_t^{-1}[\sum_i f'_{\beta,t}(c_{i,j,t})(F^{-1}_{n,\hat{\beta},t}(j/J)-F^{-1}_{n,\hat{\beta},t}((j-1)/J))(v_{it}+w_{it})]\\
&+o_p(h/J_t+\sqrt{n_tTh}^{-1}/J_t+\sqrt{n_t}^{-1}/J_t),\\
&=n_t^{-1}[\sum_i f'_{\beta,t}(c_{i,j,t})(F^{-1}_{\beta,t}(j/J)-F^{-1}_{\beta,t}((j-1)/J)-c_{n,j,t}+c_{n,j-1,t})(v_{it}+w_{it})]\\
&+o_p(h/J_t+\sqrt{n_tTh}^{-1}/J_t+\sqrt{n_t}^{-1}/J_t).
\end{aligned}
$$

Let $\tilde{j}/J_t$ be some value between $(j-1)/J$ and j/J. The leading term corresponding to the bias of order h is as follows

$$
\begin{aligned}
&\lim_{n_t\to\infty} n_t^{-1}\textstyle\sum_i f'_{\beta,t}(c_{i,j,t})(F^{-1}_{\beta,t}(j/J)-F^{-1}_{\beta,t}((j-1)/J))w_{it}\\
&=\lim_{n_t\to\infty} n_t^{-1}\textstyle\sum_i f'_{\beta,t}(c_{i,j,t})(f_{\beta,t}(F^{-1}_{\beta,t}(\tilde{j}/J_t)))w_{it}/J_t=O_p(h/J_t).
\end{aligned}
$$

□

3.6 Proof of Lemma 2.10

Proof. We will need inequalities to bound the objects H_1, H_2, H_3. We will adopt uniform concentration inequalities as the following steps.

$$
\begin{aligned}
G_1: &\stackrel{\text{def}}{=} \{f_\beta(u_1,x):\mathbf{1}(x-u_1\le\beta\le x+u_1)\beta,|u_1|\le a_{nT}+\delta_T,x\in\mathbb{N}_{J_t}\},\\
G_2: &\stackrel{\text{def}}{=} \{f_\beta(u_1):(\beta+u_1)^2-\beta^2,|u_1|\le\delta_T\},\\
G_3: &\stackrel{\text{def}}{=} \{f_{\beta,\varepsilon}(u_1,x):\mathbf{1}(\beta\le x+u_1)\varepsilon-\mathbf{1}(\beta\le x)\varepsilon,|u_1|\le a_{nT}+\delta_T,x\in\mathbb{N}_{J_t}\}.
\end{aligned}
$$

Those are functional classes that are changing with respect to n_t, T, and G_1, G_2, G_3 are all bounded functional class, and can be written trivially by the sum or product of two VC- classes, and therefore of polynomial discrimination. Thus we let $F_1(\beta), F_2(\beta), F_3(\beta,\varepsilon)$ are envelope of the functional class G_1, G_2 and G_3. $|f,g|_1=\mathbb{E}|f-g|$, $|f,g|_{1,n}=\mathbb{E}_n|f_.-g_.|$. Therefore exist a constant V such that. Let $F_1=\mathbb{E}|F_1(\beta)|$, $F_2=\mathbb{E}|F_2(\beta)|$, and $F_3=\mathbb{E}|F_3(\beta,\varepsilon)|$. Moreover, let $F_{1,n}=|F_1(\beta)|_{1,n}$, $F_{2,n}=|F_2(\beta)|_{1,n}$, and $F_{3,n}=|F_3(\beta,\varepsilon)|_{1,n}$. $\mathcal{N}(\epsilon F_1,G_1,|.|_1)$, $\mathcal{N}(\epsilon F_2,G_2,|.|_1)$ and $\mathcal{N}(\epsilon F_3,G_3,|.|_{1,n})\lesssim(1/\epsilon)^V$, and similarly $\mathcal{N}(\epsilon F_{1,n}, G_2, |.|_{1,n})$, $\mathcal{N}(\epsilon F_{2,n}, G_2, |.|_{1,n})$ and $\mathcal{N}(\epsilon F_{3,n}, G_3, |.|_{1,n}) \lesssim_{\mathrm{P}} J_t(1/\epsilon)^V$.

We use the inequality as in Lemma 3.2 in van de Geer (2000).

We look at $\sup_{x\in\mathbb{N}_{J_t},|u_1|\leq a_{nT}+\delta_T} n_t^{-1}\sum_{t=1}^{n_t}\mathbf{1}(x\leq\beta_{it}\leq x+u)\varepsilon_{it}$(or replace ε_{it} with β_{it}) and $\sup_{x\in\mathbb{N}_{J_t},|u_1|\leq a_{nT}+\delta_T} n_t^{-1}\sum_{t=1}^{n_t}\{\mathbf{1}(x\leq\beta_{it}\leq x+u)\beta_{it}-\mathbb{E}(\mathbf{1}(x\leq\beta_{it}\leq x+u)\beta_{it})\}$. Since by Markov plus Bernstein inequality,

$$
\begin{aligned}
&\mathrm{P}(n_t^{-1}\sum_i\varepsilon_{it}\geq\xi_n)\leq\mathrm{P}(\max_i|\varepsilon_{it}|\geq M_n)+\mathrm{P}(\sum_i\varepsilon_{it}\mathbf{1}(\varepsilon_{it}\leq M_n)\geq\xi_n)\\
\leq&\sum_i\mathbb{E}(\varepsilon_{it}^2/M_n^2)+C_1\exp(-n_t\xi_n^2/(4\sigma_\varepsilon^2(t/T)+M_n\xi_n)).
\end{aligned}
$$

Take $\xi_n=\sqrt{n_t}^{-1}c_{n_t}$, and $M_n=\sqrt{n_t}c_m$, with c_{n_t} and c_m are large enough constant.

We take the metric $|f,g|_{2,n}=\sqrt{n_t^{-1}\sum_i(f(x_i)-g(x_i))^2}$. We let $G_1'\overset{\text{def}}{:=}\{f_{u_1,x}(\beta):\mathbf{1}(x\leq\beta\leq x+u_1),|u_1|\leq a_{nT}+\delta_T,x\in\mathbb{N}_{J_t}\}$.

Since $\sup_{x,|u_1|\leq\delta_T+a_{nT}} n_t^{-1}\sum_{i=1}^{n_t}\mathbf{1}(x-u_1\leq\beta_{it}\leq x+u_1)\lesssim_{\mathrm{P}}(\delta_T+a_{nT})\vee(n_t^{-1/2}(\delta_T+a_{nT})^{1/2}\sqrt{\log n_tJ_t})$. Thus $\sup_{g\in G_1}\|g\|_{n,2}\leq(\delta_T+a_{nT})^{1/2}\vee(n_t^{-1/2}(\delta_T+a_{nT})^{1/2}\sqrt{\log n_tJ_t})^{1/2}\leq(\delta_T+a_{nT})^{1/2}=R$.

Then use the inequality as in Lemma 3.2 in van de Geer (2000) with $|.|_{1,n}$ replaced by $|.|_{2,n}$. Then the rate follows by setting,

$$
\begin{aligned}
\tilde{\delta}&=Rc_{nt}c_\delta/\sqrt{n_t}, &(3.12)\\
\epsilon&=Rc_{nt}c_\epsilon/\sqrt{n_t}, &(3.13)\\
K&=M_n^2, &(3.14)
\end{aligned}
$$

$\int_{\epsilon/(4K)}^R H^{1/2}(u,G_1',|.|_{n_t,2})du\leq RH^{1/2}(\epsilon/(4K),G_1',|.|_{n_t,2})=R\sqrt{\log(V(a_{nT}+\delta_T)*4*K/\epsilon)}$. To make sure that $\sqrt{n_t}(\delta-\epsilon)\geq R\sigma_\varepsilon\sqrt{V\log((a_{nT}+\delta_T)*4*K/\epsilon)}$, we should set $c_{nt}\gg\sqrt{\log n_uJ_u}$, with proper choice of constant $c_\delta-c_\epsilon$. We can achieve

$$
\sup_{x\in\mathbb{N}_{J_t},|u_1|\leq a_{nT}+\delta_T} n_t^{-1}\sum_{t=1}^{n_t}\mathbf{1}(x\leq\beta_{it}\leq x+u_1)\varepsilon_{it}\lesssim_{\mathrm{P}}\delta. \tag{3.15}
$$

Moreover, for the term $\sup_{x\in[\beta_l,\beta_u],|u_1|\leq a_{nT}+\delta_T} n_t^{-1}\sum_{t=1}^{n_t}\{\mathbf{1}(x\leq\beta_{it}\leq x+u)\beta_{it}-\mathbb{E}(\mathbf{1}(x\leq\beta_{it}\leq x+u)\beta_{it})\}$, we can repeat the above steps by setting $M=|\beta_u|$ and thus $\sqrt{n_t}(\tilde{\delta}-\epsilon)\geq R|\beta_u|\sqrt{V\log((a_{nT}+\delta_T)*4*K/\epsilon)}$, we should set $c_{nt}\gg\sqrt{\log(n_tJ_t)}$. The choice of c_δ and c_ϵ shall be adapted. So we have

$$
\sup_{x\in\mathbb{N}_{J_t},|u_1|\leq a_{nT}+\delta_T} n_t^{-1}\sum_{t=1}^{n_t}\{\mathbf{1}(x\leq\beta_{it}\leq x+u)\beta_{it}-\mathbb{E}(\mathbf{1}(x\leq\beta_{it}\leq x+u)\beta_{it})\}\lesssim_{\mathrm{P}}\delta. \tag{3.16}
$$

Since $\sup_{|u|\leq a_{nT}+\delta_T, |x-y|\leq 1/J} |n_t^{-1}\sum_i \mathbb{E}\{(\mathbf{1}(x\leq \beta_{it}\leq x+u)-\mathbf{1}(y\leq\beta_{it}\leq y+u))\beta_{it}\}| \lesssim (a_{nT}+\delta_T)/J$, the above derivation in equation (3.15) and (3.16), and the third statement in Lemma 2.9 we can conclude that,

$$
\begin{aligned}
H_1 &= n_t^{-1}\sum_i \hat{\Phi}_{i,t}(\hat{\beta}_{it}-\beta_{it}) + n_t^{-1}\sum_i(\hat{\Phi}_{i,t}-\Phi^*_{i,t})\beta_{it},\\
&\lesssim_{\mathrm{P}} \delta_T(1/J\vee l_{n,T}/\sqrt{n}) + (\delta+\tilde{\delta}_T/J + a_{nT}/J) \leq \delta + (\tilde{\delta}_T + a_{nT})/J = h_1,\\
H_2 &\lesssim_{\mathrm{P}} \tilde{\delta}_T = h_2,\\
H_3 &\leq \sup_{|u_1|\leq\delta_T} |n_t^{-1}\sum_i \mathbf{1}(x\leq\beta_{it}\leq x+u)\beta_{it}\{f_t - \mathbb{E}[f_t|\mathcal{F}_{t-1}]\}|\\
&+ \sup_{|u_1|\leq\delta_T} |n_t^{-1}\sum_{i=1}^{n_t}\mathbf{1}(x\leq\beta_{it}\leq x+u)\varepsilon_{it}|,\\
&\lesssim_{\mathrm{P}} \delta + (a_{nT}+\tilde{\delta}_T)/J + l_{n,T}/n_t \leq \delta + (\tilde{\delta}_T + a_{nT})/J = h_1.
\end{aligned}
$$

□

3.7 Proof of Lemma 2.11

Proof. We first order observations as $\ell = \ell(i,t_0) = \sum_{t=1}^{t_0-1} n_t + i$, $1\leq i\leq n_t, 1\leq t\leq T$. So that we can order observation as the following $\varepsilon_{it}\to\varepsilon_\ell$. We now define the pooled filtration $\mathcal{F}^\beta_{\ell-1} = \sigma(\beta_l : l = 1,\cdots,\ell-1)$. Recall the definition of the functional classes G_1, G_2, G_3, which are all bounded. $j_t\in 1,\cdots,J_t$ is denoted as a particular indice of a partition interval at time point t. We define the vector of indices as $\mathbb{J} = [j_1, j_2, j_3,\cdots,j_T]$, $\mathbb{J}\in B_{\mathbb{J}}$ and we assume that $|B_{\mathbb{J}}|\leq J_a$, where $J_a = O(J)$. It is clear that if the J_t and n_t are the same for each time then we can have $J_a = J$.

We brief n_ℓ as $n_tT \asymp n_a$.

Define $f_{H_1}(z) = T^{-1}\sum_t n_t^{-1}\sum_i \hat{p}_t(z)^\top[\hat{\Phi}_{i,t}\hat{\beta}_{it} - \Phi^*_{i,t}\beta_{it}]$. Denote a B_δ as a set of points in $[\beta_l,\beta_u]$, $|B_\delta| = J_a$. We see that $\sup_z f_{H_1}(z) = \sup_z |f_{H_1}(z) - f_{H_1}(\pi(z))| + \sup_{z\in B_\delta}|f_{H_1}(z)|$, where $\pi(z)$ is the closest point of z in B_δ. Since $\sup_z|f_{H_1}(z) - f_{H_1}(\pi(z))| \leq \max_{j_t,t} T^{-1}|n_t^{-1}\sum_i[\hat{\Phi}_{i,j_t,t}\hat{\beta}_{it} - \Phi^*_{i,j_t,t}\beta_{it}] \lesssim \sup_{z\in B_\delta}|f_{H_1}(z)|$. It suffice to look at the rate of $|\sup_{|u|\leq a_{nT}+\delta_T}\max_{\mathbb{J}\in B_{\mathbb{J}}}\sum_\ell n_\ell^{-1}\{\psi_{\ell,j_t}(u)\varepsilon_\ell\}|$ and $|\sup_{|u|\leq a_{nT}+\delta_T}\max_{\mathbb{J}\in B_{\mathbb{J}}}\sum_\ell n_\ell^{-1}\{\psi_{\ell,j_t}(u)\beta_\ell - \mathbb{E}(\psi_{\ell,j_t}(u)\beta_\ell)\}|$.

We will show as follows that
$|\sup_{|u|\leq a_{nT}+\delta_T}\max_{\mathbb{J}\in B_{\mathbb{J}}}\sum_\ell n_\ell^{-1}(\psi_{\ell,j_t}(u)\beta_\ell - \mathbb{E}(\psi_{\ell,j_t}(u)\beta_\ell))| \lesssim_{\mathrm{P}} c_n\sqrt{\log(n_uJ_u)(a_{nT}+\delta_T)}/\sqrt{n_a}$. We define that $\xi_\ell(x,u) = (\mathbf{1}(-u\leq\tilde{\beta}_{\ell,j_t}\leq +u)\beta_{\ell,j_t})$, then we have, $\sum_t\sum_i T^{-1}n_t^{-1}(\mathbf{1}(-u\leq\tilde{\beta}_{\ell,j_t}\leq u)\beta_\ell) = \sum_\ell n_\ell^{-1}\xi_\ell(x,u)$. $\xi_\ell(x,u) - \mathbb{E}(\xi_\ell(x,u)|\mathcal{F}^\beta_{\ell-1})$ form a martingale difference sequence with respect to the filtration $\mathcal{F}^\beta_{\ell-1}$. The random object $\mathbf{1}(-u\leq\tilde{\beta}_{\ell,j_t}\leq u)\varepsilon_\ell$ is itself a martingale

difference, as $\mathbb{E}(\varepsilon_\ell\beta_\ell|\mathcal{F}^{\beta}_{\ell-1}) = \mathbb{E}(\beta_\ell\mathbb{E}(\varepsilon_\ell|\mathcal{F}_{\ell-1})|\mathcal{F}^{\beta}_{\ell-1}) = 0$. We now show the rate,

$$\sup_{x\in[\beta_l,\beta_u],|u|\leq a_{nT}+\delta_T} |\sum_\ell n_\ell^{-1}\xi_\ell(x,u)|.$$

We denote $n_a = \sum_t n_t$. Denote the functional class $G_{n_a,1} = \{f(.) : (\tilde{\beta},\beta,u) \to f(u) = (\mathbf{1}(-u \leq \tilde{\beta} \leq u)\beta), |u| \leq a_{nT} + \delta_T\}$. We denote the bracketing number $\mathcal{N}_{[]}(\varepsilon_{n_a}, G_{n_a,1}, |.|_{1,n_a})$, where $|.|_{1,n_a}$ is the empirical norm with $n_a^{-1}\sum_\ell |f_\ell - g_\ell|$.

We shall separate to the following steps,

Step 1 We show that $\mathrm{P}^*(\mathcal{N}_{[]}(\varepsilon_{n_a}, G_{n_a,1}, |.|_{1,n_a}) \leq M_{n_a}(\varepsilon_{n_a})/4)$, with $M_{n_a}(\varepsilon_{n_a}/4)$ being less $(4(\delta_T + a_{nT})/\varepsilon_{n_a})^V$, for a fixed constant V.

Step 2

$$\sup_{x\in[\beta_l,\beta_u],|u|\leq a_{nT}+\delta_T} |\sum_\ell n_\ell^{-1}\xi_\ell(x,u)| \leq \varepsilon_n + \sup_{\xi_\ell(.)\in G_{n,1,\varepsilon_n}} |\sum_\ell n_\ell^{-1}\xi_\ell(x,u)|,$$

where $G_{n,1,\varepsilon_n}$ collects all the ε_n lower brackets.

Step 3 We do a decomposition following
$\sup_{\xi_\ell(.)\in G_{n_a,1,\varepsilon_{n_a}}} \sum_\ell n_\ell^{-1}\xi_\ell(x,u)$
$= \sup_{\xi_\ell(.)\in G_{n_a,1,\varepsilon_{n_a}}} \sum_\ell n_\ell^{-1}(\xi_\ell(x,u) - \mathbb{E}_{\ell-1}\xi_\ell(x,u)) + \sup_{\xi_\ell(.)\in G_{n_a,1,\varepsilon_{n_a}}} \sum_\ell n_\ell^{-1}\mathbb{E}_{\ell-1}(\xi_\ell(x,u) - \mathbb{E}\xi_\ell(x,u))$. The first term can be analyzed by Lemma 2.2 and the second term can be analyzed by Lemma 2.3.

We assume the dependence adjusted norm of $\|\sup_{|u|\leq a_{nT}+\delta_T}\xi_.(x,u))\|_{q,\alpha} \lesssim (a_{nT}+\delta_T)^{1/q}$. $\max J_t \leq J_u$, also for partitions with different grids for different time, the union combination of all values shall not be exceeding J_a. Namely the partition lead to
$T^{-1}\sum_t n_t^{-1}\sum_i\sum_{j_t} p_{j_t,t}(z)(\hat{\Phi}_{i,j_t,t}\hat{\beta}_{it} - \Phi^*_{i,j_t,t}\beta_{it})$ taking on finite many values J_a exhausting $z \in [\beta_l,\beta_u]$. This is ensured by the partition to be not so different across time. Without loss of generality we shall prove with $J_t \asymp J$ for all t.

Step 1

For Step 1, we have that $\mathcal{N}_{[]}(\varepsilon_{n_a}/4, G_{n_a,1}, \mathbb{E}|.|) \lesssim (4(\delta_T + a_{nT})/\varepsilon_{n_a})^V = M_n'(\varepsilon_{n_a})$, for $M_n'(\varepsilon_{n_a})$ number of brackets $[f_l^{(k)}, f_u^{(k)}]$, $k \in 1,\cdots,M_n'(\varepsilon_{n_a})$, with bounded functions defined as $b^{(k)} = f_u^{(k)} - f_l^{(k)}$, and $\mathbb{E}|b^{(i)}| \leq \varepsilon_{n_a}/4$. Denote $b_\ell^{(i)}$ as a function taking value at point i,t. We brief $M_n'(\varepsilon_{n_a})$ to M_n'. Since the event $\limsup_{f\in G_{1,n_a}}\min_{i\in[1,\cdots,M_n']}|f - f_u^{(i)}|_{1,n_a} \leq \varepsilon_{n_a}$ implies that $\mathcal{N}_{[]}(\varepsilon_{n_a}, G_{n_a,1}, |.|_{1,n_a}) \leq M_n'$. Therefore $\mathbb{E}|f - \pi(f_u^{(i)})| \leq \varepsilon_{n_a}/4$, where $\pi(f_u^{(i)})$ is the adjacent upper bracket f.

$$\begin{aligned}&\mathrm{P}(\mathcal{N}_{[]}(\varepsilon_{n_a}, G_{n_a,1}, |.|_{1,n_a}) \leq M_n'),\\ &\geq \mathrm{P}(\lim\sup_{f\in G_{1,n_a}}\min_{i\in[1,\cdots,M_n']}|f - f_u^{(i)}|_{1,n_a} \leq \varepsilon_{n_a}),\end{aligned}$$

$$\geq \mathrm{P}(\lim \sup_{f\in G_{1,n_a}} |f-\pi(f_u^{(i)})|_{1,n_a} \leq \varepsilon_{n_a}),$$
$$\geq \mathrm{P}(\lim \max_{i\in 1,\cdots,M_n'} |b^{(i)}|_{1,n_a} \leq \varepsilon_{n_a}).$$

We shall then analyze the event $\max_{i\in 1,\cdots,M_n'} |b^{(i)}|_{1,n_a} \geq \varepsilon_{n_a}$.

$$\mathrm{P}(\max_{i\in 1,\cdots,M_n'} |b^{(i)}|_{1,n_a} - \mathbb{E}|b^{(i)}| + \mathbb{E}|b^{(i)}| \geq \varepsilon_{n_a}) \tag{3.17}$$
$$\leq \mathrm{P}(\max_{i\in 1,\cdots,M_n'} |b^{(i)}|_{1,n_a} - \mathbb{E}|b^{(i)}| \geq 3\varepsilon_{n_a}/4). \tag{3.18}$$

We let $|b^{(i)}|_{1,n_a,.-1} = \sum_\ell \mathbb{E}_{\ell-1}(b_\ell^{(i)})$. We then apply Lemma 2.2 and Lemma 2.3 again to the term $\max_{i\in 1,\cdots,M_n'} |b^{(i)}|_{1,n_a} - |b^{(i)}|_{1,n_a,.-1}$ and $\max_{i\in 1,\cdots,M_n'} |b^{(i)}|_{1,n_a,.-1} - \mathbb{E}|b_\ell^{(i)}|$.

Due the specific definition of $G_{n_a,1}$, the brackets can be selected to ensure that $\mathbb{E}(\mathbb{E}_{l-1}|b^{(i)}|^2) \asymp \varepsilon_{n_a}$. We obtain that $\max_{i\in 1,\cdots,M_n'} |b^{(i)}|_{1,n_a} - |b^{(i)}|_{1,n_a,.-1} \lesssim_\mathrm{P} \sqrt{\varepsilon_{n_a}}\sqrt{\log(J_a M_n')} c_{n_a}/\sqrt{n_a}$, where $\lesssim_\mathrm{P}$ depend on β_u, provided that $(n_a^{-1} \vee n^{-\frac{q-2}{q-1}}) \ll \varepsilon_{n_a}$. Also

$$\max_{i\in 1,\cdots,M_n'} |b^{(i)}|_{1,n_a,.-1} - \mathbb{E}|b_\ell^{(i)}| \lesssim_\mathrm{P} \sqrt{\varepsilon_{n_a}}\sqrt{\log(JM_n')} c_{n_a}/\sqrt{n_a} \vee c_{n_a} \log[M_n' J_a]^{1/2} \varepsilon_{n_a}^{1/(2q)} {n_a}^{1/q-1},$$

since the above rate $\ll 3\varepsilon_{n_a}/4$. Therefore with sufficiently large c_{n_a} we have

$$\sum_{n_a=1}^{\infty} \mathrm{P}(\max_{i\in 1,\cdots,M_n'} |b^{(i)}|_{1,n_a} - \mathbb{E}|b_\ell^{(i)}| \geq 3\varepsilon_{n_a}/4) < \infty.$$

So by the Borel Cantelli Lemma, we can ensure that $\lim \max_{i\in 1,\cdots,M_n'} |b^{(i)}|_{1,n} \leq \varepsilon_{n_a}$ happens almost surely.

Step 2 This easily follows by definition of bracketing number.

Step 3 We show that

$$\sup_{\xi_l\in G_{n,1,\varepsilon_n}} |\sum_\ell n_\ell^{-1}\xi_\ell(x,u)| = \sup_{\xi_\ell\in G_{n,1,\varepsilon_n}} |\sum_\ell n_\ell^{-1}\{\xi_\ell(x,u) - \mathbb{E}_{\ell-1}\xi_\ell(x,u)\}|$$
$$+ \sup_{\xi_\ell\in G_{n,1,\varepsilon_n}} |\sum_\ell n_\ell^{-1}\mathbb{E}_{\ell-1}\xi_\ell(x,u)\} - \mathbb{E}\xi_\ell(x,u)\}|.$$

$\max_{i\in 1,\cdots,M_n'} |b^{(i)}|_{1,n_a} - |b^{(i)}|_{1,n_a,.-1} \lesssim_\mathrm{P} \sqrt{a_{nT}+\delta_T}\sqrt{\log(J_a M_n')} c_{n_a}/\sqrt{n_a}$, where $\lesssim_\mathrm{P}$ depend on β_u. Also

$$\max_{i\in 1,\cdots,M_n'} |b^{(i)}|_{1,n_a,.-1} - \mathbb{E}|b^{(i)}| \lesssim_\mathrm{P} \sqrt{a_{nT}+\delta_T}\sqrt{\log(J_a M_n')} c_{n_a}/\sqrt{n_a} \vee c_{n_a} \log[M_n' J_a]^{1/2} (a_{nT}+\delta_T)^{1/(q)} n_a^{1/q-1}.$$

We shall pick that $n_a^{-1} \ll \varepsilon_{n_a} \ll \sqrt{a_{nT}+\delta_T}\sqrt{\log(J_a M_n')}c_{n_a}/\sqrt{n_a}$, so that $\log(J_a M_n') \leq 2\log n_a$. We assume that

$$\sqrt{a_{nT}+\delta_T}\sqrt{\log(J_a M_n')}c_{n_a}/\sqrt{n_a} \gg c_{n_a}\log[M_n' J]^{1/2}(a_{nT}+\delta_T)^{1/(q)}n_a^{1/q-1},$$

which can be implied by

$$(a_{nT}+\delta_T) \gg n_a^{-1}.$$

According to remark 1.6, this condition is implied by $\sqrt{\log T}/\sqrt{n} \vee h \vee \sqrt{\log(n_u T)}/\sqrt{Th} \gg n_a^{-1}$.

Then we have that

$$\sup_{x\in[\beta_l,\beta_u],|u|\leq a_{nT}+\delta_T} |\sum_{\ell}\xi_\ell(x,u)| \lesssim_{\mathrm{P}} \sqrt{a_{nT}+\delta_T}\sqrt{\log(n_a)}c_{n_a}/\sqrt{n_a} \stackrel{\text{def}}{=} \delta'.$$

According to remark 1.6, $\delta' = (\sqrt{\log T}/\sqrt{n} \vee h \vee \sqrt{\log(n_u T)}/\sqrt{Th})^{1/2}\sqrt{\log n_a}C_{n_a}/\sqrt{n_a}$.

Recall that $\mathbb{J} = [j_1, j_2, j_3, \cdots, j_T]$, $\mathbb{J} \in B_{\mathbb{J}}$ and $|B_{\mathbb{J}}| \leq J_a$. Denote $\hat{\Phi}_{i,j_t,t}$ as $\hat{\Phi}_{\ell,j_t}$, and similar for $\Phi_{i,j_t,t}$ and $\Phi^*_{i,j_t,t}$.

$$\begin{aligned}
& \sup_z T^{-1}\sum_t n_t^{-1}\hat{p}_t(z)^\top(\hat{\Phi}_t\hat{\beta}_t - \Phi_t^*\beta_t) \\
= \; & \sup_z T^{-1}\sum_t n_t^{-1}\sum_i\sum_{j_t}\hat{p}_{j_t,t}(z)(\hat{\Phi}_{i,j_t,t}\hat{\beta}_{it} - \Phi^*_{i,j_t,t}\beta_{it}) \\
\leq \; & \max_{\mathbb{J}\in B_{\mathbb{J}}} T^{-1}\sum_t n_t^{-1}\sum_i(\hat{\Phi}_{i,j_t,t}\hat{\beta}_{it} - \Phi^*_{i,j_t,t}\beta_{it}) \\
\leq \; & |\sup_{|u_\ell|\leq a_{nT}+\delta_T}\max_{\mathbb{J}\in B_{\mathbb{J}}}\sum_\ell n_\ell^{-1}(\hat{\Phi}_{\ell,j_t}u_\ell)| + 2|\sup_{|u_\ell|\leq a_{nT}+\delta_T}\max_{\mathbb{J}\in B_{\mathbb{J}}}\sum_\ell n_\ell^{-1}\{\psi_{\ell,j_t}(u_\ell)\beta_\ell - \mathbb{E}(\psi_{\ell,jt}(u_\ell)\beta_\ell)\}| \\
& +|T^{-1}\sum_{t=1}^T\mathbb{E}\{\hat{\beta}_{it}\mathbf{1}(F^{-1}_{n_t,\hat{\beta},t}(\kappa_{j_t-1}) \leq \beta_{it} \leq F^{-1}_{n_t,\hat{\beta},t}(\kappa_{j_t}))\} \\
& -\mathbb{E}\{\beta_{it}\mathbf{1}(F^{-1}_{\beta,t}(\kappa_{jt-1}) \leq \beta_{it} \leq F^{-1}_{\beta,t}(\kappa_{jt}))\}| \\
\lesssim_{\mathrm{P}} \; & c_{n_a}\sqrt{\log(J_a)(a_{nT}+J_a^{-1}\overline{\delta}_T + c_{n_a})}\sqrt{\log(n_a J_a)(a_{nT}+\delta_T)}/\sqrt{n_a} \\
& +J_a^{-1}(\overline{\delta}_T + a_{nT}/\sqrt{T}) \\
\leq \; & \delta' + J_a^{-1}(\overline{\delta}_T + a_{nT}/\sqrt{T})
\end{aligned}$$

where the last line of the rate is derived as above, and it follows similarly from step 3 Lemma 2.8. Next we analyze the rate of $T^{-1}\sum_t n_t^{-1}\sum_i(\hat{\beta}_{it}^2 - \beta_{it}^2)$,

$$T^{-1}\sum_t n_t^{-1}\sum_i(\hat{\beta}_{it}^2 - \beta_{it}^2)$$

$$
\begin{aligned}
&\leq \quad T^{-1}\sum_t n_t^{-1}\sum_i [v_{it}+w_{it}][2\beta_{it}+v_{it}+w_{it}] \\
&\lesssim_{\mathrm{P}} \quad (\sqrt{nTh}^{-1}+h),
\end{aligned}
$$

where the last time follows from Lemma 2.2 and 2.3.

$$
\begin{aligned}
& \sup_z T^{-1}\sum_t n_t^{-1}\sum_i\sum_{j_t}\hat{p}_{j_t,t}(z)(\hat{\Phi}_{i,j_t,t}\tilde{\varepsilon}_{it}-\Phi^*_{i,j,t}\tilde{\varepsilon}_{it}) \\
\leq \quad & \max_{\mathbb{J}\in B_{\mathbb{J}}}|T^{-1}\sum_t n_t^{-1}\sum_i(\hat{\Phi}_{i,j_t,t}\tilde{\varepsilon}_{it}-\Phi^*_{i,j_t,t}\tilde{\varepsilon}_{it})| \\
\leq \quad & 2\max_{\mathbb{J}\in B_{\mathbb{J}}}|\sum_\ell n_\ell^{-1}\{(\hat{\Phi}_{\ell,j_t}-\Phi^*_{\ell,j_t})\tilde{\varepsilon}_\ell\}| \\
& +\max_{\mathbb{J}\in B_{\mathbb{J}}}|T^{-1}\sum_t(f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1}))n_t^{-1}\sum_i(\hat{\Phi}_{i,j_t,t}-\Phi^*_{i,j_t,t})\beta_{it}| \\
\lesssim_{\mathrm{P}} \quad & c_{n_a}\sqrt{\log(n_a)(a_{nT}+\delta_T)}/\sqrt{n_a}+(\sqrt{T})^{-1}h_1, \\
\lesssim \quad & c_{n_a}\sqrt{\log(n_a)(a_{nT}+\delta_T)}/\sqrt{n_a}+(\delta_T+a_{nT})^{1/2}\sqrt{\log q_n}/\sqrt{nT}+(\bar{\delta}_T+a_{nT}/\sqrt{T})/J \\
\lesssim \quad & \delta',
\end{aligned}
$$

where the last line of the rate will be derived as the above and partly follows from 2.2, and h_1 is derived in Lemma 2.10.

□

3.8 Proof of Theorem 2.12

Proof. We see that

$$
\begin{aligned}
& [n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t]_j = n_t^{-1}\sum_{i=1}^{n_t}\Phi^*_{i,j,t}\tilde{\varepsilon}_{it} \\
= \quad & n_t^{-1}\sum_{i=1}^{n_t}\Phi^*_{i,j,t}\beta_{it}(f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1}))+n_t^{-1}\sum_{i=1}^{n_t}\Phi^*_{i,j,t}\varepsilon_{it} \\
= \quad & O_p(1/J_t)+O_p(1/\sqrt{n_tJ_t}).
\end{aligned}
$$

Recall the definition $C_{n_t}=\hat{\Phi}_t\hat{\Phi}_t^\top$ and $D_{n_t}=\hat{\Phi}_t\tilde{\varepsilon}_t$. $\tilde{C}_{n_t}=\Phi_t^*\Phi_t^{*\top}$ and $\tilde{D}_{n_t}=\Phi_t^*\tilde{\varepsilon}_t$.

$$
\hat{a}_t-a_t^* \quad = \quad C_{nt}^{-1}[\hat{\Phi}_t\Phi_t^*-C_{nt}]a_t^*+C_{nt}^{-1}D_{nt}+C_{nt}^{-1}[\hat{\Phi}_t\flat_t]. \tag{3.19}
$$

By Equation (3.19), we have that

$$
\begin{aligned}
& |\hat{a}_t - a_t^* - \mathrm{diag}(\tilde{q}_{jt})^{-1}\Phi_t^*\tilde{\varepsilon}_t|_{\max} \\
= \; & |\{\hat{\Phi}_t\hat{\Phi}_t^\top\}^{-1}\{\hat{\Phi}_t\tilde{\varepsilon}_t\} - (\Phi_t^*\Phi_t^{*\top})^{-1}\Phi_t^*\tilde{\varepsilon}_t|_{\max} \\
& + |[(n_t^{-1}\Phi_t^*\Phi_t^{*\top})^{-1} - \mathrm{diag}(\tilde{q}_{jt})^{-1}]n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t|_{\max} \\
& + |\{\hat{\Phi}_t\hat{\Phi}_t^\top\}^{-1}\{\hat{\Phi}_t\flat_t\}|_{\max} + |C_{nt}^{-1}[\hat{\Phi}_t\Phi_t^* - C_{nt}]a_t^*|_{\max}.
\end{aligned}
$$

Step 1 We show the rate of $|\{\hat{\Phi}_t\hat{\Phi}_t^\top\}^{-1}\{\hat{\Phi}_t\tilde{\varepsilon}_t\} - (\Phi_t^*\Phi_t^{*\top})^{-1}\Phi_t^*\tilde{\varepsilon}_t|_{\max} = o_p(1)$.

We define $\mathbb{E}[\Phi_{t,i,j}^*\beta_{it}|\mathcal{G}_{t-1}] = \int_{F_{\beta,t}^{-1}(\kappa_{j_t-1})}^{F_{\beta,t}^{-1}(\kappa_{j_t})}\beta dF_{\beta,t}(\beta) \stackrel{\text{def}}{=} E_{\beta,j,t}(u)$. Due to the boundedness of the density function, we have $E_{\beta,j,t}(u) \lesssim_{\mathrm{P}} J_t^{-1}|\beta_u|$.

We first check the closeness of the first component which is denoted $C_{n_t}^{-1}D_{n_t} - \tilde{C}_{n_t}^{-1}\tilde{D}_{n_t} = -C_{n_t}^{-1}(C_{n_t} - \tilde{C}_{n_t})\tilde{C}_{n_t}^{-1}D_{n_t} + \tilde{C}_{n_t}^{-1}(D_{n_t} - \tilde{D}_{n_t})$. Since the matrices $(C_{n_t}, \tilde{C}_{n_t})$ involved are diagonal, $|C_{n_t}|_{\max}$ agrees with $|C_{n_t}|_\infty$ and $|C_{n_t}|_1$.

So we prove that $|C_{n_t}^{-1}D_{n_t} - \tilde{C}_{n_t}^{-1}\tilde{D}_{n_t}|_{\max} \leq |C_{n_t}^{-1}|_{\max}|(C_{n_t} - \tilde{C}_{n_t})|_{\max}|\tilde{C}_{n_t}^{-1}|_{\max}|D_{n_t}|_{\max} + |\tilde{C}_{n_t}^{-1}|_{\max}|(D_{n_t} - \tilde{D}_{n_t})|_{\max} \ll 1$. Recall the definition of h_1, h_2 in Lemma 2.10. Define $h_4 \stackrel{\text{def}}{=} 1/J_t^2 \vee c_{n_t}/(J_t\sqrt{n_tJ_t}) \vee (c_{n_t}/\sqrt{n_tJ_t})^2$, $\tilde{h}_4 = (c_{n_t}\sqrt{\log J_t}/\sqrt{n_tJ_t}) \vee J_t^{-1}$. It is not hard to see that $\tilde{h}_4 \lesssim J_t^{-2}$ and $h_4 \lesssim J_t^{-1}$. First we have from Lemma 2.10, we have

$$
\begin{aligned}
|\tilde{C}_{n_t} - C_{n_t}|_{\max} &\leq \max_{j_t}|\sum_{i=1}^{n_t}(\mathbf{1}(\hat{\beta}_{it} \in \hat{P}_{jt}) - \mathbf{1}(\beta_{it} \in P_{jt}))| \\
\lesssim_{\mathrm{P}} \quad & l_{n,T}.
\end{aligned}
$$

Recall that $l_{n,T} = \sqrt{n_t}\sqrt{a_{nT} + \delta_T}C_{nt} + (a_{nT} + \delta_T)n_t/J_t$. Moreover, from Lemma 2.10,

$$
\begin{aligned}
|D_{n_t} - \tilde{D}_{n_t}|_{\max} &\leq |\hat{\Phi}_t\tilde{\varepsilon}_t - \Phi_t^*\tilde{\varepsilon}_t|_{\max}, \\
&\lesssim_{\mathrm{P}} h_1 n_t \ll J_t^{-1}n_t.
\end{aligned}
$$

From the rate in the Remark 1.6 and Assumption 8 the above two conditions are ensured. By Bernstein inequality, we have $\max_j |[n_t^{-1}\Phi_t^*\beta_t]_j - \tilde{q}_{jt}| \lesssim_{\mathrm{P}} c_{n_t}\sqrt{\log J_t}/\sqrt{n_tJ_t}$, where c_{n_t} is a positive constant.

Step 2 Recall that we denote $\tilde{q}_{jt} = \int_{F_{\beta,t}^{-1}(\kappa_{j-1})}^{F_{\beta,t}^{-1}(\kappa_j)} f_{\beta,t}d\beta \lesssim_{\mathrm{P}} J_t^{-1}$. We have by Bernstein inequality, $\max_j |n_t^{-1}\sum_t \mathbf{1}(\beta_{it} \in P_{jt}) - \tilde{q}_{jt}| \lesssim_{\mathrm{P}} c_{n_t}\sqrt{\log J_t}/(\sqrt{n_tJ_t})$. Therefore we have by Assumption 8,

$$
|[(n_t^{-1}\Phi_t^*\Phi_t^{*\top})^{-1} - \mathrm{diag}(\tilde{q}_{jt})^{-1}]n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t|_{\max} \lesssim_{\mathrm{P}} c_{n_t}(\sqrt{\log J_t}/\sqrt{nJ_t})(J_t^2)(J_t^{-1} \vee \sqrt{J_t\log J_t}/\sqrt{n_tJ_t}) \ll 1.
$$

We then show that the bias term $\{\hat{\Phi}_t\hat{\Phi}_t^\top\}^{-1}\{\hat{\Phi}_t\flat_t\}$ is very close to the term $\{n_t^{-1}\Phi_t^*\Phi_t^{*\top}\}^{-1}(n_t^{-1}\Phi_t^*\flat_t)$. Also by the property of the partition estimator $|\flat_t|_\infty \lesssim_{\mathrm{P}} 1/J_t$(c.f. Lemma 2.9.). Thus by similar steps from the previous derivation we have that $|\{\hat{\Phi}_t\hat{\Phi}_t^\top\}^{-1}\{\hat{\Phi}_t\flat_t\}|_{\max} \lesssim_{\mathrm{P}} 1/J$.

$|C_{nt}^{-1}[\hat{\Phi}_t\Phi_t^* - C_{nt}]a_t^*|_{\max} \ll_{\mathrm{P}} 1$ by the Assumption 9.

Thus the conclusion holds.

□

3.9 Proof of Lemma 2.13.

Proof. Let

$$\mathbb{E}(\widehat{\sigma}(\beta)) \stackrel{\text{def}}{=} T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\mathbb{E}(\widehat{p}_{jt}(\beta)E_{n_t,j}^2\mathrm{Var}(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))) + T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}n_t^{-2}\sum_{i}\mathbb{E}(\widehat{p}_{jt}(\beta)\tilde{q}_{jt}^{-2}\mathbb{E}(\Phi_{i,j,t}^*\varepsilon_{it}^2|\mathcal{G}_{t-1}))$$

Since $\sup_\beta|\widehat{\sigma}(\beta)/\mathbb{E}(\widehat{\sigma}(\beta)) - 1| = \sup_\beta|[\{\widehat{\sigma}(\beta)/\mathbb{E}(\widehat{\sigma}(\beta))\}^{1/2} - 1][\{\widehat{\sigma}(\beta)/\mathbb{E}(\widehat{\sigma}(\beta))\}^{1/2} + 1]$. The proof step follows from showing first $\frac{\frac{1}{\sqrt{T}}\sum_{t=1}^T\{\widehat{\mu}_t(\beta) - \mu_t(\beta)\} - bias(\beta)}{\widehat{\sigma}(\beta)^{1/2}}$ is close to $\frac{\frac{1}{\sqrt{T}}\sum_{t=1}^T\{\widehat{\mu}_t(\beta) - \mu_t(\beta)\} - bias(\beta)}{[\mathbb{E}\{\widehat{\sigma}(\beta)\}]^{1/2}}$.

The above two conditions imply the conclusion.

□

3.10 Proof of Theorem 2.14

Proof. We first analyze the leading term. We let $\varepsilon_t^1 = (\beta_{it}(f_t - \mathbb{E}(f_t|\mathcal{F})))_i$ and recall that $\varepsilon_t = [\varepsilon_{it}]_i$. Then we have,

$$\begin{aligned} & \sqrt{T}^{-1}\sum_{t}^{T}\widehat{p}_t(\beta)^\top[\mathrm{diag}(\tilde{q}_{jt})^{-1}n_t^{-1}\Phi_t^*]\tilde{\varepsilon}_t \\ = & \sqrt{T}^{-1}\sum_{t}^{T}\widehat{p}_t(\beta)^\top[\mathrm{diag}(\tilde{q}_{jt})^{-1}n_t^{-1}\Phi_t^*]\varepsilon_t \\ & +\sqrt{T}^{-1}\sum_{t}^{T}\widehat{p}_t(\beta)^\top[\mathrm{diag}(\tilde{q}_{jt})^{-1}n_t^{-1}\Phi_t^*]\varepsilon_t^1, \end{aligned}$$

by the Assumption 7 that $\tilde{q}_{jt} \asymp_p J_t^{-1}$, and the fact that $\{\Phi_{\ell,jt}^*\varepsilon_\ell\}$ is a martingale difference sequence with respect to $\mathcal{F}_{\ell-1}^\beta$. Apply Lemma 2.2 and by Assumption 8,

$$|T^{-1}\sum_{t}^{T}\widehat{p}_t(\beta)^\top[\mathrm{diag}(\tilde{q}_{jt})^{-1}n_t^{-1}\Phi_t^*\varepsilon_t|_{\max} \lesssim_{\mathrm{P}} \frac{1}{\sqrt{T}}J(\sqrt{\log(n_aJ)}/(\sqrt{n}J)) \lesssim \frac{1}{\sqrt{J}\sqrt{T}} \ll 1/\sqrt{T}, \tag{3.20}$$

and by Bernstein inequality and the Assumption that $\mathbb{E}(\beta_{it}\Phi^*_{it}) \asymp_p J^{-1}$,

$$|T^{-1}\sum_t^T \widehat{p}_t(\beta)^\top[\text{diag}(\tilde{q}_{jt})^{-1}n_t^{-1}\Phi_t^*\varepsilon_t^1|_{\max} = |T^{-1}\sum_t(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))[n_t^{-1}\sum_i \text{diag}(\tilde{q}_{jt})^{-1}\beta_{it}\Phi^*_{it}]_{jt}|_{\max}$$
$$\lesssim_{\mathrm{P}} J\{\sqrt{T}^{-1}(1/J + \sqrt{\log(n_u J_a)}/\sqrt{nJ}) \leq \sqrt{T}^{-1} + \sqrt{TJ}^{-1} \leq \sqrt{T}^{-1}.$$

Therefore by Assumption 8, the above object has rate $\sqrt{T}^{-1}$, then $J\sqrt{T}^{-1}(J^{-2} + \sqrt{\log(n_u J_a)}/(J\sqrt{nJ}) + (\sqrt{\log(n_u J_a)}/\sqrt{nJ})^2) \ll \frac{1}{\sqrt{T}}$. The leading term $T^{-1}\sum_t(f_t - \mathbb{E}f_t|\mathcal{F}_{t-1})n_t^{-1}\sum_i \beta_{it}\Phi^*_{it} \lesssim_{\mathrm{P}} \sqrt{T}^{-1}$.

First of all, we have that,

$$\begin{aligned}\hat{\alpha}(\beta) - \alpha(\beta) &= T^{-1}\sum_t\{\widehat{p}_t(\beta)^\top\hat{a}_t - \widehat{p}_t(\beta)^\top a_t^*\}, \\ &= T^{-1}\sum_t \widehat{p}_t(\beta)^\top(\hat{a}_t - a_t^*).\end{aligned}$$

Recall that $\mathbb{J} = [j_1, j_2, j_3, \cdots, j_T]$, $\mathbb{J} \in B_{\mathbb{J}}$ and $|B_{\mathbb{J}}| \leq J$.

We denote $a^*_{jt}(\hat{a}_{jt})$ as the j_t component of a_t^* $(\hat{a}_t)$.

We will evaluate $[\max_{\mathbb{J}\in B_{\mathbb{J}}} T^{-1}\sum_t\{\hat{a}_{jt} - a^*_{jt}\}]$ in the upcoming derivation. By Assumption 10, we have $\sqrt{a_{nT}}/\sqrt{T} + \bar{\delta}_T\sqrt{\log(nJ)}/\sqrt{n_t} \ll J^{-1}$.

Now we have

$$\begin{aligned} & T^{-1}\sum_t \hat{p}_t(\beta)^\top(\hat{a}_t - a_t^*) \\ & = T^{-1}\sum_t \hat{p}_t(\beta)^\top(C_{n_t}^{-1}D_{n_t} - \tilde{C}_{n_t}^{-1}\tilde{D}_{n_t}) \\ & -T^{-1}\sum_t^T \hat{p}_t(\beta)^\top[(\text{diag}(\tilde{q}_{jt})^{-1} - (n_t^{-1}\Phi_t^*\Phi_t^{*\top})^{-1})n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t] \\ & +T^{-1}\sum_t^T \hat{p}_t(\beta)^\top[\text{diag}(\tilde{q}_{jt})^{-1}n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t] \\ \stackrel{\text{def}}{=} & I_1 + I_2 + I_3.\end{aligned}$$

We now prove that $|I_1 + I_2|_{\max} = o_p(1/\sqrt{T})$ and $|I_3|_{\max} = O_p(1/\sqrt{T})$.

We first check the rate of I_1 which is denoted $T^{-1}\sum_t \hat{p}_t(\beta)^\top(C_{n_t}^{-1}D_{n_t} - \tilde{C}_{n_t}^{-1}\tilde{D}_{n_t}) = -T^{-1}\sum_t \hat{p}_t(\beta)^\top C_{n_t}^{-1}(C_{n_t} - \tilde{C}_{n_t})\tilde{C}_{n_t}^{-1}D_{n_t} + T^{-1}\sum_t \hat{p}_t(\beta)^\top\tilde{C}_{n_t}^{-1}(D_{n_t} - \tilde{D}_{n_t})$.

So $|T^{-1}\sum_t \hat{p}_t(\beta)^\top(C_{n_t}^{-1}D_{n_t} - \tilde{C}_{n_t}^{-1}\tilde{D}_{n_t})|_{\max} \leq \max_t\{|C_{n_t}^{-1}|_{\max}|\tilde{C}_{n_t}^{-1}|_{\max}|D_{n_t}|_{\max}\}|T^{-1}\sum_t \hat{p}_t(\beta)^\top(C_{n_t} - \tilde{C}_{n_t})|_{\max} + \max_t\{|\tilde{C}_{n_t}^{-1}|_{\max}\}|T^{-1}\sum_t \hat{p}_t(\beta)^\top(D_{n_t} - \tilde{D}_{n_t})|_{\max}$. Recall that $n_u = \max_t n_t$, $n_a \asymp n_u T$.

Thus by Lemma 2.11, we have

$$|\max_{\mathbb{J}\in B_{\mathbb{J}}} T^{-1}\sum_t [\tilde{C}_{n_t}-C_{n_t}]_{jt}| \leq |\max_{\mathbb{J}\in B_{\mathbb{J}}} T^{-1}\sum_t \sum_{i=1}^{n_t}(\mathbf{1}(\hat{\beta}_{it}\in \hat{P}_{jt})-\mathbf{1}(\beta_{it}\in P_{jt}))|$$
$$\lesssim_{\mathrm{P}} \quad h_1' n_u \ll (J^{-1}n_u).$$

Similarly by Lemma 2.11, we have

$$\begin{aligned}&|\max_{\mathbb{J}\in B_{\mathbb{J}}} T^{-1}\sum_t [D_{n_t}-\tilde{D}_{n_t}]_{jt}| \leq |\max_{\mathbb{J}\in B_{\mathbb{J}}} T^{-1}\sum_t [\hat{\Phi}_t\tilde{\varepsilon}_t - \Phi_t^*\tilde{\varepsilon}_t]_{jt}| \\ &\lesssim_{\mathrm{P}} (h_3' + h_1'(h_1+\tilde{h}_4) + \sqrt{T}^{-1}\tilde{h}_4\delta_T + \tilde{h}_4/\sqrt{T})n_u \ll J^{-1}\sqrt{T}^{-1}n_u.\end{aligned}$$

Therefore

$$\begin{aligned}|I_1| &= |-T^{-1}\sum_t \hat{p}_t(\beta)^\top C_{n_t}^{-1}(C_{n_t}-\tilde{C}_{n_t})\tilde{C}_{n_t}^{-1}D_{n_t}| + |T^{-1}\sum_t \hat{p}_t(\beta)^\top \tilde{C}_{n_t}^{-1}(D_{n_t}-\tilde{D}_{n_t})| \\ &\lesssim_{\mathrm{P}} J^{-2}\max_t |n_t^{-1}(C_{n_t}-\tilde{C}_{n_t})|_{\max}|\max_{\mathbb{J}\in B_{\mathbb{J}}} T^{-1}\sum_t n_t^{-1}[D_{n_t}]_{jt}| \\ &\ll J\sqrt{T}^{-1}\tilde{h}_4 \\ &\lesssim \sqrt{T}^{-1}.\end{aligned}$$

For the second term, we have

$$\begin{aligned}|I_2| &= \max_{t,j_t}|(\mathrm{diag}(\tilde{q}_{jt})^{-1}-(n_t^{-1}\Phi_t^*\Phi_t^{*\top})^{-1})|\max_{\mathbb{J}\in B_{\mathbb{J}}}|T^{-1}\sum_t^T [n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t]_{j_t}| \\ &\ll_p J^2(\sqrt{\log q_n}/\sqrt{nJ})\sqrt{T}^{-1}\tilde{h}_4 \\ &\lesssim \sqrt{T}^{-1}.\end{aligned}$$

So we have $|I_1|,|I_2| \ll_p \sqrt{T}^{-1}$.

Moreover, by equation (3.20), $|I_3| = O_p(\sqrt{T}^{-1})$.

Now we look at I_4. We denote $\tilde{E}_{nt} = n_t^{-1}\Phi_t^*\flat_t$, and $E_{nt} = n_t^{-1}\hat{\Phi}_t(I-P_{\hat{\beta},t})\flat_t$

$$\begin{aligned}|I_4| &\leq |I_4 - T^{-1}\sum_t \hat{p}_t(\beta)^\top (n_t^{-1}\Phi_t^*\Phi_t^{*\top})^{-1}(n_t^{-1}\Phi_t^*\flat_t)| + |T^{-1}\sum_t \hat{p}_t(\beta)^\top \tilde{D}_{nt}\tilde{E}_{nt}| \\ &\leq |T^{-1}\sum_t \hat{p}_t(\beta)^\top C_{n_t}^{-1}(C_{n_t}-\tilde{C}_{n_t})\tilde{C}_{n_t}^{-1}E_{n_t}| + |T^{-1}\sum_t \hat{p}_t(\beta)^\top \tilde{C}_{n_t}^{-1}(E_{n_t}-\tilde{E}_{n_t})| \\ &+ |T^{-1}\sum_t \hat{p}_t(\beta)^\top \tilde{D}_{nt}\tilde{E}_{nt}|.\end{aligned}$$

We know from Lemma 2.9 that $\max_t |\flat_t|_{\max} \lesssim_{\mathrm{P}} 1/J$. We have that $|T^{-1}\sum_t \hat{p}_t(\beta)^\top C_{n_t}^{-1}(C_{n_t} -$

$\tilde{C}_{n_t})\tilde{C}_{n_t}^{-1}E_{n_t} \ll_p J^2(J\sqrt{T})^{-1}(J^{-1} \vee \sqrt{\log q_n}/n)1/J \lesssim J^{-1}\sqrt{T}^{-1}$. Also $|T^{-1}\sum_t \hat{p}_t(\beta)^\top \tilde{C}_{n_t}^{-1}(E_{n_t} - \tilde{E}_{n_t})| \lesssim_{\mathrm{P}} J(\delta' \vee (a_{nT} + \tilde{\delta}_T)/J)(1/J) \lesssim J^{-1} \wedge \sqrt{T}^{-1}$. Let $\mathbf{1}_{n_t}$ be a $n_t \times 1$ vector of ones. For the term $T^{-1}\sum_t \hat{p}_t(\beta)^\top (n_t^{-1}\Phi_t^*\Phi_t^{*\top})^{-1}(n_t^{-1}\Phi_t^*\flat_t)$, we have $|\max_{\mathbb{J}\in B_{\mathbb{J}}} T^{-1}\sum_t[((n_t^{-1}\Phi_t^*\Phi_t^{*\top})^{-1}(n_t^{-1}\Phi_t^*\mathbf{1}_{n_t})]_{jt}| \lesssim_{\mathrm{P}} 1$.

Thus we have $T^{-1}\sum_t \hat{p}_t(\beta)^\top (n_t^{-1}\Phi_t^*\Phi_t^{*\top})^{-1}(n_t^{-1}\Phi_t^*\flat_t) \lesssim_{\mathrm{P}} J^{-1}$. It follows that $|I_4| \lesssim_{\mathrm{P}} J^{-1}$.

□

3.11 Proof of Theorem 2.15

Proof. Recall the definition that $\hat{\sigma}_{jt}^2 = n_t/J_t(\sum_i \hat{\Phi}_{i,j_t,t}\hat{\varepsilon}_{it}^2)(\sum_i \hat{\Phi}_{i,j_t,t})^{-2}$.
We see that $\sigma_{jt}^2 = 1/J_t(\mathbb{E}_{t-1}(\Phi_{i,j_t,t}^*\varepsilon_{it}^2))(\mathbb{E}_{t-1}(\Phi_{i,j_t,t}^*))^{-2}$. Recall that $\mathbb{E}_{t-1}(\Phi_{i,j_t,t}^*) = \tilde{q}_{jt}$.
Recall that $\sigma_t(\beta) = \sum_{j_t} \hat{p}_{j_t}(\beta)\sigma_{j_t}^2$, and $\hat{\sigma}_t(\beta) = \sum_{j_t} \hat{p}_{j_t}(\beta)\hat{\sigma}_{j_t}^2$. $\sigma_\varepsilon(\beta) = T^{-1}\sum_t J_t\sigma_t(\beta)$.

We have proved that the leading term $[\{n_t^{-1}\Phi_t^*\Phi_t^{*\top}\}^{-1}(n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t)]_j$ is close to $[\tilde{q}_{jt}^{-1}n_t^{-1}\sum_i \Phi_{t,i,j}^*\tilde{\varepsilon}_t]_j$.

Define for each t, the bin covering 0 as $B_{t,0}$. So there exists a point $\beta_{t,0} \in B_{t,0}$, $F_{\beta,t}^{-1}(\lfloor j_{t,0}/n_t \rfloor) \le 0$ and $F_{\beta,t}^{-1}(\lfloor (j_{t,0}+1)/n_t \rfloor) \ge 0$. We let $\varepsilon_{it}^1 = \beta_{it}(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))$. First we prove the normality for $\mathbf{1}(\beta \notin \cup_t B_{t,0})$. Recall that we define $E_{n_t,j_t} = \tilde{q}_{jt}^{-1}\mathbb{E}(\Phi_{i,j,t}^*\beta_{it}|\mathcal{G}_{t-1})$. Then we have

$$
\begin{aligned}
& [\sqrt{T}^{-1}\sum_t \tilde{q}_{jt}^{-1}n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t]_{j_t} \\
= \; & [\sqrt{T}^{-1}\sum_t \tilde{q}_{jt}^{-1}n_t^{-1}\sum_i \Phi_{i,j_t,t}^*\varepsilon_{it}^1 + \sqrt{T}^{-1}\sum_t \tilde{q}_{jt}^{-1}n_t^{-1}\sum_i \Phi_{i,j_t,t}^*\varepsilon_{it}] \\
= \; & [\sqrt{T}^{-1}\sum_t (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))\tilde{q}_{jt}^{-1}n_t^{-1}\sum_i \Phi_{i,j_t,t}^*\beta_{it} + \sqrt{T}^{-1}\sum_t \tilde{q}_{jt}^{-1}n_t^{-1}\sum_i \Phi_{i,j_t,t}^*\varepsilon_{it}] \\
= \; & \sqrt{T}^{-1}\sum_t (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))\tilde{q}_{jt}^{-1}n_t^{-1}\sum_i (\Phi_{i,j_t,t}^*\beta_{it} - E_{n_t,j_t}) + \sqrt{T}^{-1}\sum_t \tilde{q}_{jt}^{-1}n_t^{-1}\sum_i \Phi_{i,j_t,t}^*\varepsilon_{it} \\
& + \sqrt{T}^{-1}\sum_t (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t} \\
= \; & O_p(J\sqrt{\log q_n}/\sqrt{nJ}) + O_p(\sqrt{J}/\sqrt{n}) + \sqrt{T}^{-1}\sum_t (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t} \\
= \; & \sqrt{T}^{-1}\sum_t (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t} + o_p(1).
\end{aligned}
$$

By the Assumption 2 we have $\mathbb{E}(\varepsilon_{it}|\mathcal{F}_{t-1}) = 0$ and since we have $\mathbb{E}(\tilde{f}_t|\mathcal{F}_{t-1}) = 0$. Then

$$\mathbb{E}(\eta_{t,n_t}(\beta)|\mathcal{F}_{t-1}) = 0,$$

$$\mathbb{E}(|\eta_{t,n_t}(\beta)|) < \infty.$$

Therefore by Assumption 7, $\eta_{t,n_t}(\beta)$ are MDS with respect to $\mathcal{F}_{t-1}$.

Now we shall verify the following, according to Corollary 3.1 in Hall and Heyde (2014),

$$\mathbb{E}|1/T\sum_t \mathbb{E}(\eta_{t,n_t}(\beta)^2|\mathcal{F}_{t-1}) - 1| \to 0, \tag{3.21}$$

and for $\varepsilon > 0$

$$1/T\sum_t \mathbb{E}(\eta_{t,n_t}(\beta)^2\mathbf{1}(\eta_{t,n_t}(\beta)/\sqrt{T} > \varepsilon)|\mathcal{F}_{t-1}) \to_p 0. \tag{3.22}$$

Since the two terms $T_{n,1}(\beta)$ and $T_{n,2}(\beta)$ are uncorrelated, due to Assumption 7,

$$\begin{aligned}\mathbb{E}(\eta_{t,n_t}(\beta)^2|\mathcal{F}_{t-1}) &= \mathbb{E}(\mathbb{E}(\sigma(\beta))^{-1}\sum_{j_t}\hat{p}_{j_t,t}(\beta)\tilde{f}_t^2 E^2_{n_t,j_t}|\mathcal{F}_{t-1})\\ &\quad + \mathbb{E}(\mathbb{E}(\sigma(\beta))^{-1}\sum_{j_t}\hat{p}_{j_t,t}(\beta)\tilde{q}_{jt}^{-2}n_t^{-2}\sum_i \Phi^*_{i,j_t,t}\varepsilon^2_{it}|\mathcal{F}_{t-1}),\end{aligned}$$

and

$$\mathbb{E}(\sigma(\beta))^{-1}\sum_{j_t}\hat{p}_{j_t,t}(\beta)\tilde{q}_{jt}^{-2}n_t^{-2}\sum_i \Phi^*_{i,j_t,t}\varepsilon^2_{it}|\mathcal{F}_{t-1}) = \mathbb{E}(\sigma(\beta))^{-1}\sum_{j_t}\hat{p}_{j_t,t}(\beta)\tilde{q}_{jt}^{-2}n_t^{-2}\sum_i \Phi^*_{i,j_t,t}\mathbb{E}(\varepsilon^2_{it}|\mathcal{F}_{t-1}).$$

We can proceed with the following steps. Thus the equation (3.21) holds by the definition of $\sigma(\beta)$, Assumption 7 and the following steps. Let $x_{t,n_t} = \mathbb{E}(\eta_{t,n_t}(\beta)^2|\mathcal{F}_{t-1}) - 1$. We shall operate an MDS decompoisiton with respect to $\mathcal{G}_{t-1}$. Thus we have $\mathbb{E}|1/T\sum_t \mathbb{E}(\eta_{t,n_t}(\beta)^2|\mathcal{F}_{t-1}) - 1| \leq [\mathbb{E}|1/T\sum_t \mathbb{E}(\eta_{t,n_t}(\beta)^2|\mathcal{F}_{t-1}) - 1|^{(1+\delta)}]^{1/(1+\delta)} \leq \sqrt{T}(\max_{j_t}\|\hat{p}_{j_t,.}(\beta)\tilde{f}.E_{n_t,j_t}\|_{q,\xi} \vee \max_{i,j_t}\|\Phi^*_{i,j_t,.}\varepsilon_{i.}\|_{q,\xi})T^{-1} \to 0$, by Buckerholder inequality. Now we verify equation in (3.22).

We first show that,

$$\begin{aligned}\mathbb{E}(\eta_{t,n_t}(\beta)^{2+2\delta}) &\leq c\mathbb{E}[\mathbb{E}(\sigma(\beta))^{-1/2}\sum_{j_t}\hat{p}_{j_t,t}(\beta)(\tilde{f}_t E_{n_t,j_t})]^{2+2\delta}+\\ &\quad + c\mathbb{E}[\mathbb{E}(\sigma(\beta))^{-1/2}\sum_{j_t}\hat{p}_{j_t,t}(\beta)\tilde{q}_{jt}^{-1}n_t^{-1}\sum_i \Phi^*_{i,j_t,t}\varepsilon_{it}]^{2+2\delta}(Buckerholder)\\ &\lesssim c + n_t^{-(2+2\delta)}(\sigma_f(\beta) + n_t{}^{-1-\delta})^{-1}(\sqrt{n_t})^{2+2\delta},\end{aligned}$$

where the last line is due to Lemma 2.1 and Assumption 7. Then the central limit theorem follows.

Then it holds that for all $\varepsilon > 0$, $\sum_t \mathbb{E}\left[(\eta_{t,n_t}(\beta)/\sqrt{T})^2 I\left(\eta_{t,n_t}(\beta)/\sqrt{T} > \varepsilon\right)\right] \to 0$, this hold obviously due to $\mathbb{E}\left[(\eta_{t,n_t}(\beta)/\sqrt{T})^2 I\left(\eta_{t,n_t}(\beta)/\sqrt{T} > \varepsilon\right)\right] \leq T^{-1-\delta}\mathbb{E}(|\eta_{t,n_t}(\beta)|^{2+2\delta}/\varepsilon^{2\delta})$ due to Markov inequalities. Then we have $1/T\sum_t \mathbb{E}(\eta_{t,n_t}(\beta)^2\mathbf{1}(\eta_{t,n_t}(\beta) > T\varepsilon)|\mathcal{F}_{t-1}) \to_p 0$ due to $(1/T\sum_t\{\mathbb{E}(\eta_{t,n_t}(\beta)^2\mathbf{1}(\eta_{t,n_t}(\beta) > T\varepsilon)|\mathcal{F}_{t-1}) - \mathbb{E}(\eta_{t,n_t}(\beta)^2\mathbf{1}(\eta_{t,n_t}(\beta) > T\varepsilon))\} \to_p 0)$. The central limit theorem then holds.

□

3.12 Proof of Lemma 2.16

Proof. We shall define $\tilde{\mu}_t(\beta) = \hat{p}_t(\beta)^\top a_t^*$. We define $e_t(\beta) = \hat{p}_t(\beta)^\top diag(\tilde{q}_{jt})^{-1} n_t^{-1} \Phi_t^* \tilde{\varepsilon}_t$.

Let $g_t(\beta) \stackrel{\text{def}}{=} \hat{\mu}_t(\beta) - T^{-1}\sum_t \hat{\mu}_t(\beta)$, $h_t(\beta) \stackrel{\text{def}}{=} \tilde{\mu}_t(\beta) + e_t(\beta) - T^{-1}\sum_t(\tilde{\mu}_t(\beta) + e_t(\beta))$. First we show that

$$\begin{aligned}
& \sup_\beta T^{-1}\sum_t[\hat{\mu}_t(\beta) - T^{-1}\sum_t\{\tilde{\mu}_t(\beta)\}]^2 - T^{-1}\sum_t[\tilde{\mu}_t(\beta) + e_t(\beta) - T^{-1}\sum_t(\tilde{\mu}_t(\beta) + e_t(\beta))]^2 \\
= & \sup_\beta T^{-1}\sum_t[g_t(\beta) - h_t(\beta)][g_t(\beta) + h_t(\beta)] \\
\leq & \sup_\beta \max_t |g_t(\beta) + h_t(\beta)|(T^{-1}\sum_t |[g_t(\beta) - h_t(\beta)]|) \\
\leq & \sup_\beta \max_t |g_t(\beta) + h_t(\beta)| 2(T^{-1}\sum_t |\hat{\mu}_t(\beta) - \tilde{\mu}_t(\beta) - e_t(\beta)|).
\end{aligned}$$

Thus under the conditions of Theorem 2.12 and 2.14, we have the rate

$$\sup_\beta \max_t |g_t(\beta) + h_t(\beta)| 2|T^{-1}\sum_t |\hat{\mu}_t(\beta) - \tilde{\mu}_t(\beta) - e_t(\beta)|| << _p 1.$$

Then we study the following,

$$\begin{aligned}
& T^{-1}\sum_t[\tilde{\mu}_t(\beta) + e_t(\beta) - T^{-1}\sum_t(\tilde{\mu}_t(\beta) + e_t(\beta))]^2, \\
= & T^{-1}\sum_t[\tilde{\mu}_t(\beta) - T^{-1}\sum_t(\tilde{\mu}_t(\beta))]^2 + T^{-1}\sum_t[e_t(\beta) - T^{-1}\sum_t(e_t(\beta))]^2 \\
& + 2T^{-1}\sum_t\{e_t(\beta) - T^{-1}\sum_t e_t(\beta)\}\{\tilde{\mu}_t(\beta) - T^{-1}\sum_t \tilde{\mu}_t(\beta)\}]. \\
= & T^{-1}\sum_t[\tilde{\mu}_t(\beta) - T^{-1}\sum_t(\tilde{\mu}_t(\beta))]^2 + \sup_\beta T^{-1}\sum_t[e_t(\beta) - T^{-1}\sum_t e_t(\beta)]^2 + o_p(1).
\end{aligned}$$

Since $\sup_\beta T^{-1}\sum_t[e_t(\beta) - T^{-1}\sum_t e_t(\beta)]^2 = \sup_\beta T^{-1}\sum_t e_t^2(\beta) - (T^{-1}\sum_t e_t(\beta))^2$.

Similar to the Lemma 2.3 under the qth moment conditions of $f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})$, we have

$$\sup_\beta (T^{-1}\sum_t e_t^2(\beta) - (T^{-1}\sum_t e_t(\beta))^2 - \sigma(\beta)) \lesssim_{\mathrm{P}} T^{-1}(J_a^{1/q} + (T\log J_a)^{1/2}) \to 0. \tag{3.23}$$

Similarly we have $T^{-1}\sum_t[\tilde{\mu}_t(\beta) - T^{-1}\sum_t(\tilde{\mu}_t(\beta))]^2 - \sigma_\mu(\beta) \ll_p 1$.

Thus the desire results follows.

We now show that the covariance estimator is consistent,

$$\begin{aligned}
& \sup_{\beta_1,\beta_2} T^{-1}\sum_t [g_t(\beta_1)g_t(\beta_2) - h_t(\beta_1)h_t(\beta_2)] \\
\leq\ & \sup_{\beta_1,\beta_2} [|T^{-1}\sum_t |g_t(\beta_1) - h_t(\beta_1)|| \vee |T^{-1}\sum_t |g_t(\beta_2) - h_t(\beta_2)||] \max_t\{|g_t(\beta_1) \vee h_t(\beta_2)|\}.
\end{aligned}$$

Moreover, due to $\mathbb{E}(\mu_t(\beta_2)e_t(\beta_1)) = 0$, for any t and β_1 and β_2, we have

$$\begin{aligned}
& \sup_{\beta_1,\beta_2} T^{-1}\sum_t h_t(\beta_1)h_t(\beta_2) \\
=\ & \sup_{\beta_1,\beta_2} T^{-1}\sum_t^T e_t(\beta_1)e_t(\beta_2) - (T^{-1}\sum_t^T e_t(\beta_1))(T^{-1}\sum_t^T e_t(\beta_2)) \\
+\ & \sup_{\beta_1,\beta_2} T^{-1}\sum_t^T \tilde{\mu}_t(\beta_1)\tilde{\mu}_t(\beta_2) - (T^{-1}\sum_t^T \tilde{\mu}_t(\beta_1))(T^{-1}\sum_t^T \tilde{\mu}_t(\beta_2)) \\
+\ & \sup_{\beta_1,\beta_2} T^{-1}\sum_t^T \tilde{\mu}_t(\beta_1)e_t(\beta_2) - (T^{-1}\sum_t^T \tilde{\mu}_t(\beta_1))(T^{-1}\sum_t^T e_t(\beta_2)) \\
+\ & \sup_{\beta_1,\beta_2} T^{-1}\sum_t^T \mu_t(\beta_2)e_t(\beta_1) - (T^{-1}\sum_t^T \tilde{\mu}_t(\beta_2))(T^{-1}\sum_t^T e_t(\beta_1)) \\
=\ & \sup_{\beta_1,\beta_2} T^{-1}\sum_t^T e_t(\beta_1)e_t(\beta_2) - (T^{-1}\sum_t^T e_t(\beta_1))(T^{-1}\sum_t^T e_t(\beta_2)) \\
+\ & \sup_{\beta_1,\beta_2} T^{-1}\sum_t^T \tilde{\mu}_t(\beta_1)\tilde{\mu}_t(\beta_2) - (T^{-1}\sum_t^T \tilde{\mu}_t(\beta_1))(T^{-1}\sum_t^T \tilde{\mu}_t(\beta_2)) + o_p(1).
\end{aligned}$$

Suppose that $\mathbb{J}_1$ and $\mathbb{J}_2$ corresponds to different bin indices according to β_1 and β_2 at time t. And similar to the argument above,

$$T^{-1}\sum_t^T e_t(\beta_1)e_t(\beta_2) - (T^{-1}\sum_t^T e_t(\beta_1))(T^{-1}\sum_t^T e_t(\beta_2)) - C_{\mathbb{J}_1,\mathbb{J}_2} \lesssim_{\mathrm{P}} T^{-1}(J_a^{2/q} + (T\log J_a^2)^{1/2}) \to 0.$$

$$\begin{aligned}
& \sup_{\beta_1,\beta_2} T^{-1}\sum_t^T \tilde{\mu}_t(\beta_1)\tilde{\mu}_t(\beta_2) - (T^{-1}\sum_t^T \tilde{\mu}_t(\beta_1))(T^{-1}\sum_t^T \tilde{\mu}_t(\beta_2)) - \sigma_\mu(\beta_1,\beta_2) \\
& \lesssim_{\mathrm{P}} T^{-1}(J_a^{2/q} + (T\log J_a^2)^{1/2}) + 1/J \to 0.
\end{aligned}$$

□

3.13 Proof of Lemma 2.17 and 2.18

Proof. We shall prove that the residual $\hat{\varepsilon}_{it_0}$ is close to ε_{it_0} in a uniform manner over t_0.

$$\begin{aligned}
\max_{t_0}(\hat{\varepsilon}_{it_0}-\varepsilon_{it_0}) &= \max_{t_0}[1,f_{t_0}]^\top([\sum_{i=1}^T w(t,t_0)X_tX_t^\top]^{-1}\sum_{i=1}^T w(t,t_0)X_tR_{it}-b_{i(-t_0)}),\\
&= \max_{t_0}[1,f_{t_0}]^\top([\sum_{i=1}^T w(t,t_0)X_tX_t^\top]^{-1}\sum_{i=1}^T w(t,t_0)X_tX_t^\top(b_{it}-b_{i(-t_0)})\\
&\quad + \max_{t_0}[1,f_{t_0}]^\top[\sum_{i=1}^T w(t,t_0)X_tX_t^\top]^{-1}(\sum_{i=1}^T w(t,t_0)X_t\varepsilon_{it}),\\
&\lesssim \delta_T \max_{1\le t_0\le T}|f_{t_0}|_\infty,
\end{aligned}$$

where recall that $h\vee r_T\vee\sqrt{\log(q_n)}/\sqrt{Th}=\delta_T$, and $r_T=(Th)^{-1}(T^{1/q}+(Th\log T)^{1/2})$.

$$\max_{t_0}(\hat{\varepsilon}_{it_0}-\varepsilon_{it_0})\lesssim_p \delta_T T^{1/q}. \tag{3.24}$$

□

Proof to Lemma 2.18.

Proof. To prove the consistency of $\hat{\sigma}^2_{j_t}$, we have,

$$\begin{aligned}
&\max_{j_t}(\hat{\sigma}^2_{j_t}-\sigma^2_{j_t}),\\
= \;&(\max_{j_t}\hat{\sigma}^2_{j_t}-n_t/J_t(\sum\Phi^*_{i,j_t,t}\varepsilon^2_{it})(\sum_i\Phi^*_{i,j_t,t})^{-2})\\
&+(\max_{j_t}n_t/J_t(\sum\Phi^*_{i,j_t,t}\varepsilon^2_{it})(\sum_i\Phi^*_{i,j_t,t})^{-2}-\sigma^2_{j_t}).
\end{aligned}$$

The second term follows from Bernstein inequality. For the first term, we have,

$$\begin{aligned}
&\max_{j_t}\hat{\sigma}^2_{j_t}-n_t/J_t(\sum\Phi^*_{i,j_t,t}\varepsilon^2_{it})(\sum_i\Phi^*_{i,j_t,t})^{-2},\\
&=\max_{j_t}n_t/J_t(\sum_i\hat{\Phi}_{i,j_t,t}\hat{\varepsilon}^2_{it})(\sum_i\hat{\Phi}_{i,j_t,t})^{-2}-n_t/J_t(\sum_i\hat{\Phi}_{i,j_t,t}\hat{\varepsilon}^2_{it})(\sum_i\Phi^*_{i,j_t,t})^{-2}\\
&+\max n_t/J_t(\sum_i\hat{\Phi}_{i,j_t,t}\hat{\varepsilon}^2_{it})(\sum_i\Phi^*_{i,j_t,t})^{-2}-n_t/J_t(\sum\Phi^*_{i,j_t,t}\varepsilon^2_{it})(\sum_i\Phi^*_{i,j_t,t})^{-2},\\
&=I+II.
\end{aligned}$$

Among I and II, we derive the bound for the terms respectively,

$$
\begin{aligned}
&I = \\
&\max_{j_t} 1/J_t(n_t^{-1}\sum_i \hat{\Phi}_{i,j_t,t}\hat{\varepsilon}_{it}^2)(n_t^{-1}\sum_i \hat{\Phi}_{i,j_t,t})^{-2}\{(n_t^{-1}\sum_i \hat{\Phi}_{i,j_t,t})^2 \\
&-(n_t^{-1}\sum_i {}_i\Phi^*_{i,j_t,t})^2\}(n_t^{-1}\sum_i \Phi^*_{i,j_t,t})^{-2}, \\
&= 1/J_t(n_t^{-1}\sum_i \hat{\Phi}_{i,j_t,t}\hat{\varepsilon}_{it}^2)(n_t^{-1}\sum_i \hat{\Phi}_{i,j_t,t})^{-2}\{(n_t^{-1}\sum_i \hat{\Phi}_{i,j_t,t}) \\
&-(n_t^{-1}\sum_i \Phi^*_{i,j_t,t})\}\{(n_t^{-1}\sum_i \hat{\Phi}_{i,j_t,t}) + (n_t^{-1}\sum_i \Phi^*_{i,j_t,t})\}(n_t^{-1}\sum_i \Phi^*_{i,j_t,t})^{-2}, \\
&= O_p(1/J_t(l_{n,T}/n_t + J_t^{-1})J_t^2(l_{n,T}/n_t)J_t^{-1}J_t^2) = O_p(\sqrt{a_{nT}+\delta_T}J_t/\sqrt{n_t} + a_{nT} + \delta_T) = O_p(\delta_1),
\end{aligned}
$$

which is due to the derivations in Lemma 2.9, 2.10 and 2.17.

$$
\begin{aligned}
&II \\
&= \max_{j_t} n_t/J_t(\sum_i \Phi^*_{i,j_t,t})^{-2}\sum_i \hat{\Phi}_{i,j_t,t}\hat{\varepsilon}_{it}^2 - \Phi^*_{i,j_t,t}\varepsilon_{it}^2), \\
&= n/J_t(\sum_i \Phi^*_{i,j_t,t})^{-2}\{\sum_i(\hat{\Phi}_{i,j_t,t} - \Phi^*_{i,j_t,t})\varepsilon_{it}^2 + \sum_i \Phi^*_{i,j_t,t}(\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2)\}, \\
&= n_t/J_t(\sum_i \Phi^*_{i,j_t,t})^{-2}\{\sum_i(\hat{\Phi}_{i,j_t,t} - \Phi^*_{i,j_t,t})\varepsilon_{it}^2 + \sum_i \Phi^*_{i,j_t,t}(\hat{\varepsilon}_{it} - \varepsilon_{it})(\hat{\varepsilon}_{it} + \varepsilon_{it})\}, \\
&= 1/J_t(n_t^{-1}\sum_i \Phi^*_{i,j_t,t})^{-2}\{n_t^{-1}\sum_i(\hat{\Phi}_{i,j_t,t} - \Phi^*_{i,j_t,t}\varepsilon_{it}^2) \\
&+ \max_{t_0}(|n_t^{-1}\sum_i \Phi^*_{i,j_t,t}(\hat{\varepsilon}_{it} + \varepsilon_{it}))(\hat{\varepsilon}_{it_0} - \varepsilon_{it_0})|\}, \\
&\le 1/J_t(n_t^{-1}\sum_i \Phi^*_{i,j_t,t})^{-2}\{n_t^{-1}\sum_i(\hat{\Phi}_{i,j_t,t} - \Phi^*_{i,j_t,t})\varepsilon_{it}^2 + \max_{t_0} n_t^{-1}\sum_i \Phi^*_{i,j_t,t}(\hat{\varepsilon}_{it} - \varepsilon_{it})(\hat{\varepsilon}_{it_0} + \varepsilon_{it_0})\}, \\
&\lesssim_p 1/J_t * J_t^2(J_t^{-1}(\delta_T + a_{nT}) + \sqrt{\delta_T}T^{1/(2q)}n_t^{-1/2}J_t^{-1/2}), \\
&= O_p(\delta_T + a_{nT} + \sqrt{\delta_T}T^{1/(2q)}n_t^{-1/2}J_t^{1/2}), \\
&= O_p(\delta_2),
\end{aligned}
$$

which is due to the derivations in Lemma 2.9, 2.10 and 2.17.

Thus we have the uniform rate for the variance estimator as follows,

$$
\max_{j_t}(\hat{\sigma}_{j_t}^2 - \sigma_{j_t}^2)
$$

$$
\begin{aligned}
&= I + II + \max_{j_t} n_t/J_t(\sum \Phi^*_{i,j_t,t}\varepsilon^2_{it})(\sum_i \Phi^*_{i,j_t,t})^{-2} - \sigma^2_{j_t} \\
&= \max_{j_t} J_t^{-1}(n_t^{-1}\sum_i \Phi^*_{i,j_t,t}\varepsilon^2_{it} - \mathbb{E}_{t-2}(\Phi^*_{i,j_t,t}\varepsilon^2_{it}))(n_t^{-1}\sum_i \Phi^*_{i,j_t,t})^{-2} \\
&+ \max_{j_t} J_t^{-1}(n_t^{-1}\sum_i \Phi^*_{i,j_t,t})^{-2}(n_t^{-1}\sum_i \Phi^*_{i,j_t,t} - \tilde{q}_{j_t})(n_t^{-1}\sum_i \Phi^*_{i,j_t,t} + \tilde{q}_{j_t})(\tilde{q}_{j_t})^{-2}\mathbb{E}_{t-1}(\Phi^*_{i,j_t,t}\varepsilon^2_{it}) \\
&\lesssim_p c_{n_t}\sqrt{\log J_t/n_tJ_t}J_t^{-1}J_t^2 + J_t^{-1}J_t^2 c_{n_t}\sqrt{\log J_t/n_tJ_t}J_t^{-1}J_t^2J_t^{-1} \\
&\lesssim c_{n_t}\sqrt{J_t\log J_t/n_t}.
\end{aligned}
$$

Since we have the following, then the above bounds can be applied to prove the uniform rate of $\sigma_t(\beta) - \hat{\sigma}_t(\beta)$.

$$
\begin{aligned}
\sup_\beta \sigma_t(\beta) - \hat{\sigma}_t(\beta) &= \sum_{j_t} \hat{p}_{j_t}(\beta)(\sigma^2_{j_t} - \hat{\sigma}^2_{j_t}) \\
&= \max_{j_t} |\sigma^2_{j_t} - \hat{\sigma}^2_{j_t}|.
\end{aligned}
$$

Thus the results follows. □

3.14 Proof of Theorem 2.19

Recall that $\hat{\sigma}^2_{j_t} = n_t/J_t(\sum_i \hat{\Phi}_{i,j_t,t}\hat{\varepsilon}^2_{it})(\sum_i \hat{\Phi}_{i,j_t,t})^{-2}$.
We see that $\sigma^2_{j_t} = 1/J_t(\mathbb{E}_{t-1}(\Phi^*_{i,j_t,t}\varepsilon^2_{it}))(\mathbb{E}_{t-1}(\Phi^*_{i,j_t,t}))^{-2}$. Recall that $\mathbb{E}_{t-1}(\Phi^*_{i,j_t,t}) = \tilde{q}_{jt}$.
Recall that $\sigma_t(\beta) = \sum_{j_t} \hat{p}_{j_t}(\beta)\sigma^2_{j_t}$, and $\hat{\sigma}_t(\beta) = \sum_{j_t} \hat{p}_{j_t}(\beta)\hat{\sigma}^2_{j_t}$.

For the test statistics $\sqrt{n_t}/\sqrt{J_t}\{\hat{L}_t(\beta) - L_t(\beta)\}/\hat{\sigma}_t(\beta)^{1/2}$.

1 We show that the leading term $Z_{n_t}(\beta) \overset{\text{def}}{=} \sqrt{n_t}/\sqrt{J_t}\{(\hat{L}_t(\beta) - L_t(\beta))\}/\hat{\sigma}_t(\beta)^{1/2}$ is $\sum_{j_t} \hat{p}_{j_t}(\beta)\sqrt{1/(n_tJ_t)}\tilde{q}^{-1}_{j_t}\sum_i \Phi^*_{i,j_t,t}\varepsilon_{it}/\sigma_t(\beta)^{1/2}$. Namely, we have

$$
\begin{aligned}
&\sup_\beta \sqrt{n_t}/\sqrt{J_t}(\hat{L}_t(\beta) - L_t(\beta))/\hat{\sigma}_t(\beta)^{1/2} = \sup_\beta \sqrt{n_t}/\sqrt{J_t}\hat{p}_t^\top(\beta)\{\text{diag}[\tilde{q}_{jt}]\sigma_t(\beta)^{1/2}\}^{-1}\{n_t^{-1}\Phi^*_t\varepsilon_t\} \\
&+ O_p(\sqrt{n_t}/(\sqrt{J_t}J_t) \vee \sqrt{n_t}h/\sqrt{J_t} \vee 1/\sqrt{J_t}).
\end{aligned}
$$

2 We need to assume that ε_{it} are conditional iid on $\mathcal{F}_{t-1}$. We show a coupling step for the leading term in the previous step by conditioning on $\mathcal{F}_{t-1}$. Namely $\sum_{j_t} \hat{p}_{j_t}(\beta)\sqrt{1/(n_tJ_t)}\tilde{q}^{-1}_{j_t}\sum_i \Phi^*_{i,j_t,t}(\varepsilon_{it} - \sigma_t\eta_{it})/\sigma_t^{1/2}(\beta) = o_p(n_t^{-1/2+1/(2q)}\sqrt{J_t})$ is of small order by applying a Komlós-Major-Tusnády (KMT) strong approximation argument. We let $\sigma^2_{t-1} = \mathbb{E}(\varepsilon^2_{it}|\mathcal{F}_{t-1})$, and η_{it} is iid standard random variable conditional on $\mathcal{F}_{t-1}$. We can see that $\varepsilon_{it} =_d \sigma_t\eta_{it}$ conditional on $\mathcal{F}_{t-1}$.

3 We shall prove that $\sum_{j_t} \hat{p}_{j_t}(\beta)\sqrt{1/(n_tJ_t)}\tilde{q}^{-1}_{j_t}\sum_i \Phi^*_{i,j_t,t}\sigma_t\eta_{it}/\sigma_t^{1/2}(\beta)$ which is conditional Gaus-

sian distributed and is close enough to a Gaussian random variable $Z_t(\beta)$. We define $Z_t(\beta) = \sum_{j_t} \hat{p}_{j_t}(\beta) Z_{j_t}$, where Z_{j_t}s are standard Gaussian random variable. The argument is due to Gaussian maximal inequalities.

Step 1 Recall that $\delta_1 = \sqrt{a_{nT} + \delta_T} J_t/\sqrt{n_t} + a_{nT} + \delta_T$, $\delta_2 = \delta_T + a_{nT} + \sqrt{\delta_T} T^{1/(2q)} n_t^{-1/2} J_t^{1/2}$. We first show that the standardization with estimated variance is negligible.

$$\begin{aligned}
&\sup_\beta \sqrt{n_t}/\sqrt{J_t}\{\hat{L}_t(\beta) - L_t(\beta)\}/\hat{\sigma}_t^{1/2}(\beta) - \sqrt{n_t}/\sqrt{J_t}\{\hat{L}_t(\beta) - L_t(\beta)\}/\sigma_t^{1/2}(\beta) \\
&\lesssim \sup_\beta c(\delta_1 + \delta_2)|\sqrt{n_t}/\sqrt{J_t}\{\hat{L}_t(\beta) - L_t(\beta)\}/\sigma_t(\beta)^{1/2}|.
\end{aligned}$$

By the Bernstein inequality.

Let $X_1, \ldots, X_n$ be independent zero-mean random variables. Suppose that $|X_i| \leq M$ (M is a positive constant) almost surely, for all i. Then, for all positive t.

$$\mathrm{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{\frac{1}{2}t^2}{\sum_{i=1}^n \mathbb{E}\left[X_i^2\right] + \frac{1}{3}Mt}\right) \tag{3.25}$$

If we assume that the $q-$th moment of ε_{it}^2 is finite, then we shall study the inequalities on the event $\mathcal{A} = \{\max_i \varepsilon_{it}^2 \Phi_{i,j_t,t}^* \leq M\}$. $P(\mathcal{A}^c) \lesssim n_t/(M^q J_t)$, by Assumption 2.

Let $M = c_{n_t} n_t^{1/q}/J_t^{1/q}$, we have $P(\mathcal{A}^c) \leq 1/(c_{n_t}^q)$. Now apply Bernstein inequality, conditional on $\mathcal{F}_{t-1}$, we have

$$\max_{j_t} n_t^{-1} \sum_i (\Phi_{i,j_t,t}^* - \tilde{q}_{j_t}) \lesssim_p c_{n_t}\sqrt{\log J_t/n_t J_t}.$$

$$\max_{j_t} n_t^{-1} \sum_i (\Phi_{i,j_t,t}^* \varepsilon_{it}^2 - \mathbb{E}_{t-1}(\Phi_{i,j_t,t}^* \varepsilon_{it}^2)) \lesssim_p c_{n_t}(\sqrt{\log J_t/n_t J_t} + n_t^{-1+1/q} J_t^{-1/q} \log J_t).$$

It is not hard to see that for $q > 4$, if $J_t(\log J_t)^2/n_t \lesssim 1$, by Assumption 8, then we have $c_{n_t}(\sqrt{\log J_t/n_t J_t} + n_t^{-1+1/q} J_t^{-1/q} \log J_t) \leq c_{n_t}\sqrt{\log J_t/n_t J_t}$.

By Lemma 2.18, we have $\hat{\sigma}_{j_t}^2 - \sigma_{j_t}^2 \lesssim_{\mathrm{P}} \sqrt{J_t \log J_t/n_t}$

$$(1/\hat{\sigma}_t(\beta)^{1/2} - 1/\sigma_t(\beta)^{1/2}) = ((\sigma_t(\beta)^{1/2} - \hat{\sigma}_t(\beta)^{1/2})/\hat{\sigma}_t(\beta)^{1/2}\sigma_t(\beta)^{1/2}).$$

We take the following steps,

$$\begin{aligned}
&\sqrt{n_t}/\sqrt{J_t}\{\hat{L}_t(\beta) - L_t(\beta)\}/\hat{\sigma}_t(\beta)^{1/2} - \sqrt{n_t}/\sqrt{J_t}\{\hat{L}_t(\beta) - L_t(\beta)\}/\sigma_t(\beta)^{1/2}, \\
= \quad &\sqrt{n_t}/\sqrt{J_t}\{\hat{L}_t(\beta) - L_t(\beta)\}/\sigma_t(\beta)^{1/2}(\sigma_t(\beta)^{1/2}/\hat{\sigma}_t(\beta)^{1/2} - 1),
\end{aligned}$$

$$\leq \quad |\sqrt{n_t}/\sqrt{J_t}\{\hat{L}_t(\beta)-L_t(\beta)\}/\sigma_t(\beta)^{1/2}||(\sigma_t(\beta)^{1/2}/\hat{\sigma}_t^{1/2}(\beta)-1)|.$$

Thus

$$\sup_\beta \sqrt{n_t}/\sqrt{J_t}\{\hat{L}_t(\beta)-L_t(\beta)\}/\hat{\sigma}_t(\beta)^{1/2}-\sqrt{n_t}/\sqrt{J_t}\{\hat{L}_t(\beta)-L_t(\beta)\}/\sigma_t^{1/2}(\beta)$$
$$\leq \sup_\beta |\sqrt{n_t}/\sqrt{J_t}\{\hat{L}_t(\beta)-L_t(\beta)\}/\sigma_t^{1/2}(\beta)| \sup_\beta |(\sigma_t^{1/2}(\beta)/\hat{\sigma}_t^{1/2}(\beta)-1)|.$$

By Assumption 13, we have,

$$\begin{aligned}
&P(\sup_\beta |(\sigma_t(\beta)^{1/2}/\hat{\sigma}_t^{1/2}(\beta)-1)| \geq x),\\
=&P(\max_{j_t} |(\sigma_{j_t}/\hat{\sigma}_{j_t}-1)| \geq x),\\
\leq&P(\max_{j_t} |(\sigma_{j_t}/\hat{\sigma}_{j_t}-1)(\sigma_{j_t}/\hat{\sigma}_{j_t}+1)| \geq x),\\
\leq&P(\max_{j_t} |(\sigma_{j_t}^2/\hat{\sigma}_{j_t}^2-1)| \geq x).
\end{aligned}$$

Since by Lemma 2.18

$$\max_{j_t} |\hat{\sigma}_{j_t}^2-\sigma_{j_t}^2| = O_p((\delta_1+\delta_2)).$$

The event for the positive constant $c>0$ (defined in Assumption 13) happens with probability approaching 1,

$$\max_{j_t} |\hat{\sigma}_{j_t}^2-\sigma_{j_t}^2| \leq c/2.$$

We have, by Assumption 13, with probability approaching 1, $\forall\ j_t$, we have, exist a positive constant c,

$$\hat{\sigma}_{j_t}^2 \geq \sigma_{j_t}^2 - c/2 \geq c/2.$$

So we have,

$$\begin{aligned}
&\mathrm{P}(\max_{j_t} |(\sigma_{j_t}^2/\hat{\sigma}_{j_t}^2-1)| \geq x),\\
\leq&\mathrm{P}(\max_{j_t} |(\sigma_{j_t}^2-\hat{\sigma}_{j_t}^2)| \geq xc/2).
\end{aligned}$$

Thus we have

$$\sup_\beta |(\sigma_t(\beta)/\hat{\sigma}_t(\beta)-1)| \lesssim_p c(\delta_1+\delta_2).$$

It is not hard to derive that under the conditions of Theorem 2.12. Similar to the derivation,

$$(\hat{L}_t(\beta)-L_t(\beta)) = \hat{p}_t^\top(\beta)\{\mathrm{diag}[\tilde{q}_{jt}]\}^{-1}\{n_t^{-1}\Phi_t^*\varepsilon_t\} + O_p(1/J_t \vee h). \tag{3.26}$$

Then

$$\sup_\beta \sqrt{n_t}/\sqrt{J_t}(\hat{L}_t(\beta) - L_t(\beta))/\sigma_t(\beta) = \sup_\beta \sqrt{n_t}/\sqrt{J_t}\hat{p}_t^\top(\beta)\{\mathrm{diag}[\tilde{q}_{jt}]\sigma_t(\beta)^{1/2}\}^{-1}\{n_t^{-1}\Phi_t^*\varepsilon_t\}$$
$$+O_p(\sqrt{n_t}/(\sqrt{J_t}J_t) \vee \sqrt{n_t}h/\sqrt{J_t} \vee 1/\sqrt{J_t}).$$

So the first step is proved.

Step 2 We now show the steps of strong approximation. We couple the sequence $\sum_{j_t}\hat{p}_{j_t}(\beta)\sqrt{n_t/J_t}\tilde{q}_{j_t}^{-1}\sum_i \Phi^*_{i,j_t,t}\varepsilon_{it}/\sigma_t(\beta)^{1/2}$ by $\sum_{j_t}\hat{p}_{j_t}(\beta)\sqrt{n_t/J_t}\tilde{q}_{j_t}^{-1}\sum_i \Phi^*_{i,j_t,t}(\sigma_t\eta_{it})/\sigma_t(\beta)^{1/2}$, where $\varepsilon_{it} =_d \sigma_i\eta_{it}$ conditional on $\mathcal{F}_{t-1}$.

Now we cite a KMT type theorem to show that such a coupling exists.

Theorem 3.1. *(Theorem 2.1, Berkes, Liu, and Wu (2014)) Assume that $X_i \in \mathcal{L}^p$ (p is an integer.) with mean $0, p > 2$, and there exists $\alpha > p$ such that*

$$\Theta_{\alpha,p} := \sum_{j=-\infty}^{\infty} |j|^{1/2-1/\alpha}\delta_{j,p}^{p/\alpha} < \infty$$

($\delta_{j,p}$ is $\delta_{j,p}(X_i)$). Further assume that there exists a positive integer sequence $(m_k)_{k=1}^\infty$ such that

$$M_{\alpha,p} := \sum_{k=1}^{\infty} 3^{k-k\alpha/p} m_k^{\alpha/2-1} < \infty,$$
$$\sum_{k=1}^{\infty}\frac{3^{kp/2}\Theta^p_{m_k,p}}{3^k} < \infty,$$

and

$$\Theta_{m_k,p} + \min_{l\geq 0}\left(\Theta_{l,p} + l3^{k(2/p-1)}\right) = o\left(\frac{3^{k(1/p-1/2)}}{(\log k)^{1/2}}\right).$$

Then there exists a probability space $(\Omega_c, \mathcal{A}_c, \mathrm{P}_c)$ on which we can define random variables X_i^c with the partial sum process $S_n^c = \sum_{i=1}^n X_i^c$, and a standard Brownian motion $\mathbb{B}_c(\cdot)$, such that $(X_i^c)_{i\in\mathbb{Z}} \overset{\mathcal{D}}{=} (X_i)_{i\in\mathbb{Z}}$ and (2.7),

$$S_n^c - \sigma\mathbb{B}_c(n) = o_{a.s.}\left(n^{1/p}\right) \qquad \text{in } (\Omega_c, \mathcal{A}_c, \mathrm{P}_c).$$

We can see that for iid data the dependence Assumption naturally satisfies by assuming that ε_{it} has finite qth moment, with q >4, and by Assumption 2, we have, conditional on $\mathcal{F}_{t-1}$,

$$\max_{1\leq l\leq n_t-1}|\sum_i^l(\varepsilon_{it} - \sigma_t\eta_{it})| = O_p(n_t^{1/(2q)}).$$

We define

$$K(\beta, \beta_{it}) = \sum_{j_t} \hat{p}_{j_t}(\beta)\sqrt{1/(n_t J_t)}\tilde{q}_{j_t}^{-1}\Phi^*_{i,j_t,t}/\sigma_t(\beta)^{1/2}.$$

Then conditional on $\mathcal{F}_{t-1}$,by summation by part,

$$\begin{aligned}
&\sup_\beta |\sum_i K(\beta, \beta_{it})(\varepsilon_{it} - \sigma_{it}\eta_{it})| \\
&\leq \sup_\beta |\sum_i^{n_t-1}[K(\beta, \beta_{(i)t}) - K(\beta, \beta_{(i-1)t})]| \max_{1\leq l\leq n_t-1} |\sum_i^l \varepsilon_{it} - \sigma_{it}\eta_{it}| + K(\beta, \beta_{(n_t)t})|\sum_i^{n_t} \varepsilon_{it} - \sigma_t\eta_{it}| \\
&\lesssim \max_{1\leq l\leq n_t-1} |\sum_i^l(\varepsilon_{it} - \sigma_t\eta_{it})|\sqrt{J_t/n_t} \\
&\lesssim_p n_t^{-1/2+1/(2q)}\sqrt{J_t}.
\end{aligned}$$

Step 3,

We now show that the conditional normal process $\sum_i K(\beta, \beta_{it})(\sigma_t\eta_{it}) =_d \sqrt{\sum_i K(\beta, \beta_{it})^2}Z \stackrel{\text{def}}{=} Z_{n_t}$ (Z is a standard normal variable) is close enough to the Gaussian variable $Z_t(\beta)$, which we show by Gaussian maximal inequalities. Recall that we define $Z_t(\beta) = \sum_{j_t} \hat{p}_{j_t}(\beta)Z_{j_t}$ (Z_{j_t}s are iid standard normal random variable) as a conditional Gaussian process, conditional on $\mathcal{G}_{t-1}$, $Z_t(\beta)$ is conditional Gaussian process. Recall $\sigma^2_{j_t} = \tilde{q}_{j_t}^{-1}\sigma_t^2/J_t$, and $\sigma_t(\beta)^{1/2} = \sum_{j_t}\hat{p}_{j_t}(\beta)\sigma_{j_t}$. β_{j_t} lies in the interior point of $\hat{P}_{j_t}$. Recall that β_{j_t} are some interior point for each partition j_t. Since we know that by Gaussian Maximal inequality, conditional on $\mathcal{F}_{t-1}$,

$$\begin{aligned}
&\mathrm{P}(\sup_\beta |\sum_i\{K(\beta, \beta_{it})\sigma_t\eta_{it}\} - Z_t(\beta)| \geq x|\mathcal{F}_{t-1}), \\
&\lesssim \sqrt{\log J_t}\max_{j_t}|\sigma_t\sqrt{\sum_i K(\beta_{j_t}, \beta_{it})^2} - 1|, \\
&\leq \sqrt{\log J_t}\max_{j_t}|\sum_i K(\beta_{j_t}, \beta_{it})^2\sigma_t^2 - 1|, \\
&\leq \sqrt{\log J_t}\max_{j_t}(n_t)^{-1}|\sum_i(\tilde{q}_{j_t}^{-1}\Phi^*_{i,j_t,t} - 1)|, \\
&\lesssim_{\mathrm{P}} \sqrt{\log J_t}\sqrt{J_t}\sqrt{\log J_t}/\sqrt{n_t}.
\end{aligned}$$

Step 4, We shall look at

$$\sup_x \mathrm{P}(\sup_\beta Z_{n_t}(\beta) \leq x) - \mathrm{P}(\sup_\beta Z_t(\beta) \leq x). \tag{3.27}$$

Define a positive constant $\xi > 0$.

$$
\begin{aligned}
& \sup_x |\,\mathrm{P}(\sup_\beta Z_{n_t}(\beta) \leq x) - \mathrm{P}(\sup_\beta Z_t(\beta) \leq x)| \\
\leq \quad & |\,\mathrm{P}(\sup_\beta |Z_{n_t}(\beta) - Z_t(\beta)| \geq \xi)| + |\sup_x \mathrm{P}(x \leq \sup_\beta Z_t(\beta) \leq x + \xi)|.
\end{aligned}
$$

Due to the anticoncentration property of $Z(\beta)$ conditional on $\mathcal{F}_{t-1}$ as in Lemma 3.2, and the dominant convergence theorem, we have, for $\xi \lesssim \sqrt{J_t}^{-1}$

$$
\begin{aligned}
& |\sup_x \mathrm{P}(x \leq \sup_\beta Z_t(\beta) \leq x + \xi)|, \\
& |\mathbb{E}(\sup_x \mathrm{P}(x \leq \sup_\beta Z_t(\beta) \leq x + \xi | \mathcal{F}_{t-1}))|, \\
& |\sup_x \mathbb{E}\,\mathrm{P}(x \leq \sup_\beta Z_t(\beta) \leq x + \xi | \mathcal{F}_{t-1})|, \\
& \lesssim_{\mathrm{P}} \sqrt{\log J_t}/\sqrt{J_t},
\end{aligned}
$$

Notice that $\sup_\beta Z_t(\beta) = \max_{j_t} Z_{j_t}$. The above statement is derived based on Lemma 3.2. In particular, we have $X_j = Z_{j_t}$, $\sigma_j = \sigma_{j_t}$. From Assumption 13, we have $\min_{j_t} \sigma_{j_t} > 0$. Also $\epsilon = \xi > 0$ as the correspondence the Lemma. $\mathbb{E}[\max_{j_t} Z_{j_t}/\sigma_{j_t}] \lesssim \sqrt{\log J_t}$ by Gaussian maximal inequality.

Define $\mathcal{L}\left(\max_{1 \leq j \leq p} X_j, \epsilon\right) = \sup_t \mathrm{P}(|\max_j X_j - t| \leq \epsilon)$.

Theorem 3.2. *(Anti-concentration: Chernozhukov, Chetverikov, Kato, et al. (2013)). Let* $(X_1, \ldots, X_p)^T$ *be a centered Gaussian random vector in* $\mathbb{R}^p$ *with* $\sigma_j^2 := \mathbb{E}\left[X_j^2\right] > 0$ *for all* $1 \leq j \leq p$*. Moreover, let* $\underline{\sigma} := \min_{1 \leq j \leq p} \sigma_j, \bar{\sigma} := \max_{1 \leq j \leq p} \sigma_j$*, and* $a_p := \mathbb{E}\left[\max_{1 \leq j \leq p}\left(X_j/\sigma_j\right)\right]$ *(i) If the variances are all equal, namely* $\underline{\sigma} = \bar{\sigma} = \sigma$*, then for every* $\epsilon > 0$

$$
\mathcal{L}\left(\max_{1 \leq j \leq p} X_j, \epsilon\right) \leq 4\epsilon\left(a_p + 1\right)/\sigma
$$

(ii) If the variances are not equal, namely $\underline{\sigma} < \bar{\sigma}$*, then for every* $\epsilon > 0$*,*

$$
\mathcal{L}\left(\max_{1 \leq j \leq p} X_j, \epsilon\right) \leq C\epsilon\left\{a_p + \sqrt{1 \vee \log(\underline{\sigma}/\epsilon)}\right\}
$$

where $C > 0$ *depends only on* $\underline{\sigma}$ *and* $\bar{\sigma}$*.*

Since $\sup_\xi |\,\mathrm{P}(\sup_\beta |Z_{n_t}(\beta) - Z_t(\beta)| \geq \xi) - \mathrm{P}(\sup_\beta |Z_{n_t}(\beta) - Z_t(\beta)| \geq \xi | \mathcal{F}_{t-1})| \to_p 0$, which follows similar arguments as in the proof of Theorem SA-4.1, Cattaneo, Crump, Farrell, and Feng (2022).

3.15 Proof of Corollary 2.19.1 and 2.19.2

The proof is similar to the previous paper and therefore omitted.

3.16 Proof of Lemma 2.20

Proof. Since by Theorem 2.14, we have

$$T^{-1/2}\sum_t(\hat{\mu}_t(\beta)-\mu_t(\beta)) = T^{-1/2}\sum_t\sum_{j_t}\hat{p}_{j_t,t}(\beta)(f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t}+o_p(1). \tag{3.28}$$

Now we can use Theorem 3.3 to conduct our uniform inference for beta sorting estimators. The strong approximation results can be applied on the term $T^{-1/2}\sum_t\sum_{j_t}\hat{p}_{j_t,t}(\beta)(f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t}$. From any $\mathbb{J}_1,\mathbb{J}_2\in B_{\mathbb{J}}$.

Assume that $\tilde{Z}_{t,j}$ follows a normal distribution with $N(0,\mathrm{diag}(\tilde{\Sigma})^{-1/2}\tilde{\Sigma}\mathrm{diag}(\tilde{\Sigma})^{-1/2})$, $c\leq\lambda_{\min}(\tilde{\Sigma})\leq\lambda_{\max}(\tilde{\Sigma})\leq C$,

$$\max_{\mathbb{J}\in B_{\mathbb{J}}}|T^{-1/2}\sum_t C_{\mathbb{J},\mathbb{J}}^{-1/2}(f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t}-\tilde{Z}_{\mathbb{J}}| = O_p(T^{-\varepsilon'}), \tag{3.29}$$

where ε' is a constant between 1/6 to 0. $\bar{\Sigma}=\mathrm{diag}(\tilde{\Sigma})^{-1/2}\tilde{\Sigma}\mathrm{diag}(\tilde{\Sigma})^{-1/2}$.

□

3.17 Proof of Corollary 2.21

Recall that $\tilde{f}_t=f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1})$,

$$\eta_{t,n_t}(\beta)=\mathbb{E}(\hat{\sigma}(\beta))^{-1/2}\sum_{j_t}\hat{p}_{j_t,t}(\beta)(\tilde{f}_tE_{n_t,j_t}+\tilde{q}_{jt}^{-1}n_t^{-1}\sum_i\Phi_{i,j_t,t}^*\varepsilon_{it}).$$

Proof. For a constant $\delta>0$ (defined within the proof only).

Let $\tilde{\mathbb{U}}_T(\beta)=T^{-1/2}\sum_{t=1}^T(\hat{\sigma}(\beta))^{-1/2}(\hat{\mu}_t(\beta)-\hat{\mu}_t(\beta))$.

Denote

$$\mathbb{U}_T(\beta)=T^{-1/2}\sum_t(\hat{\sigma}_{FM}(\beta,\beta))^{-1/2}(\hat{\mu}_t(\beta)-\hat{\mu}(\beta)). \tag{3.30}$$

We shall derive

$$\begin{aligned}&\mathrm{P}(\sup_\beta\tilde{\mathbb{U}}_T(\beta)\geq x)-\mathrm{P}(\sup_\beta G_T(\beta)\geq x)\\&\leq\mathrm{P}(\sup_\beta|\tilde{\mathbb{U}}_T(\beta)-G_T(\beta)|\geq\delta)+\sup_x\mathrm{P}(x\leq\sup_\beta G_T(\beta)\leq x+\delta).\end{aligned}$$

Define $\mathbb{U}_T^*(\beta)=T^{-1/2}\sum_t\sigma(\beta)^{-1/2}(\hat{\mu}_t(\beta)-\mu_t(\beta))$. Next we shall study $\sup_\beta\tilde{\mathbb{U}}_T(\beta)$ with the replaced true $\sigma(\beta)$. Since $\mathrm{P}(\sup_\beta|\tilde{\mathbb{U}}_T(\beta)-G_T(\beta)|\geq\delta)\leq\mathrm{P}(\sup_\beta|\tilde{\mathbb{U}}_T(\beta)-\mathbb{U}_T^*(\beta)|\geq\delta/2)+\mathrm{P}(\sup_\beta|\mathbb{U}_T^*(\beta)-G_T(\beta)|\geq\delta/2)$. From Assumption 14, we have that $\sup_\beta|\tilde{\mathbb{U}}_T(\beta)-\mathbb{U}_T^*(\beta)|\leq$

$(\sigma(\beta)^{-1/2} - \widehat{\sigma}(\beta)^{-1/2})(T^{-1/2}\sum_t(\hat{\mu}_t(\beta) - \mu_t(\beta))) \lesssim O_p(r_\sigma)\sup_\beta|\mathbb{U}_T^*(\beta)| \lesssim r_\sigma T^{-1/2+1/2q}J_a^{1/2q} \lesssim T^{-\varepsilon'}$.

For the term $\mathrm{P}(\sup_\beta|\tilde{\mathbb{U}}_T(\beta) - G_T(\beta)| \geq \delta)$ we analyze via a linearization following Theorem 2.14 and strong approximation to the leading term. For the term $\sup_x \mathrm{P}(x \leq \sup_\beta G_T(\beta) \leq x + \delta)$ we study via anticoncentration inequalities for Gaussian processes (c.f. Lemma 3.2.) Because of Assumption 14 and there are at most J_a jumps in the limiting process (recall J_a is the number combined intervals over time), we have for $\delta \lesssim \sqrt{J}^{-1}$, $\sup_x \mathrm{P}(x \leq \sup_\beta G_T(\beta) \leq x+\delta) \lesssim \sqrt{\log J_a}/\sqrt{J}$.

(Coupling with a Normal Vector)let the limit variance covariance matrix be $\Sigma(\beta_{\mathbb{J}}) = \lim_{T\to\infty}\mathrm{Cov}(\tilde{\mathbb{U}}_T \circ \pi_{\tilde{\delta}}(\beta_{\mathbb{J}})$, denote $\mathbb{J}$ as the number of elements within $B_{\tilde{\delta}}$, there exists $G_T \circ \pi_\delta(\beta_{\mathbb{J}}) \sim N(0, \Sigma(\beta_{\mathbb{J}}))$ such that

$$r_2 = |\tilde{\mathbb{U}}_T \circ \pi_{\tilde{\delta}}(\beta_{\mathbb{J}}) - Z_{\mathbb{J}}|_\infty = O_p(T^{-\varepsilon'} \vee \sqrt{T}(J^{-1} \vee h)) = o\left(T^{-\varepsilon'}\right);$$

The follow steps prove the statement the coupling by verifying conditions of Theorem 3.3(i)-iv)). We mention the correspondence between subjects in the Theorem and our case. Recall that for the grid $\overline{\beta}_v = [\beta_1, \beta_2, \cdots, \beta_{c/\tilde{\delta}}]$. This corresponds to $J_{a,\delta}$ distinct value of β. From Assumption 14 for avoiding singularity of the variance covariance matrix. We have $\sigma(\beta_{j_t}, \beta_{j_t'}) \neq \sigma(\beta_{j_t}, \beta_{j_t})$ or $\sigma(\beta_{j_t'}, \beta_{j_t'})$. In the theorem $n = T$, $X_t = \sqrt{T}^{-1}\sum_{t=1}^T\sum_{j_t}\hat{p}_{t,j_t}(\overline{\beta}_v)E_{n_t,j_t}[f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})]/\widehat{\sigma}(\beta)^{1/2}$. Also $p = J_{a,\tilde{\delta}}$. Thus the Σ_z in the theorem therein corresponds to $[\hat{\sigma}(\beta_i, \beta_{i'})/\hat{\sigma}^{1/2}\hat{\sigma}(\beta_{i'})^{1/2}]_{i,i'}$, which is a $\Sigma_{J_{a,\delta},J_{a,\delta}}$ dimension matrix. We shall assume that $c < \lambda_{\min}(\Sigma_{J_a,\delta}) < \lambda_{\max}(\Sigma_{J_a,\delta}) < C'$, with $C', c > 0$. This proves iii). Moreover, to prove ii), we shall assume that $m = O(T^{1/6})$, $M = O(T^{1/3})$, $J_{a,\delta} \lesssim \exp(T^{\varepsilon'})$, with $\varepsilon' = 1/9$. To derive the strong approximation for $\sup_{\beta\in B_{\tilde{\delta}}} T^{-1/2}\sum_t\sum_{j_t}\hat{p}_{j_t,t}(\beta)(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t}$ uniformly over β. We note that it is equivalent to look at strong approximation for $\max_{\mathbb{J}} T^{-1/2}\sum_t \hat{p}_{j_t,t}(\beta)(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t}$. $\hat{p}_{j_t,t}(\beta)(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t}$ is involved, and it is still martingale difference sequence with respect to $\mathcal{F}_{t-1}$, as $\hat{p}_{j_t,t}(\beta)$ is measurable with respect to $\mathcal{F}_{t-1}$. It remains to check the dependence adjusted norm for $\hat{p}_{j_t,t}(\beta)(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t}$. We see that since $\hat{p}_{j_t,t}(\beta)$ is uniformly bounded by 1. For martingale differences, let $e_{t,j_t} = (\hat{p}_{j_t,t}(\beta)(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t}$, and $\delta_{q,m}(e_{.,j_t,l}) = 0$ for $l > 0$ (recall that $e_{.,j_t,l}$ is defined by $e_{.,j_t}$ replaced with an iid copy of the lth lag.) Thus due to the property of martingale differences the dependence adjusted norm is just the norm itself $\Theta_{m^0,q}(e_., j_t) = \|\hat{p}_{j_t,t}(\beta)(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t}\|_q \leq \|(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))\|_{2q}\|E_{n_t,j_t}\|_{2q}$. Then by Assumption 14, ii) is verified regarding the bounded dependence adjusted norm.

$\mathbb{E}(|\hat{p}_{j_t,t}(\beta)(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t} - \hat{p}^*_{j_t,t}(\beta)(f_t^* - \mathbb{E}(f_t|\mathcal{F}^*_{t-1}))E_{n_t,j_t}|^q)$, where we recall that $.^*$ indicates the independent copy replacement. We see that $\mathbb{E}(|\hat{p}_{j_t,t}(\beta)(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t} - \hat{p}^*_{j_t,t}(\beta)(f_t^* - \mathbb{E}(f_t|\mathcal{F}^*_{t-1}))E_{n_t,j_t}|^q) \leq \mathbb{E}[\max|(\hat{p}_{j_t,t}(\beta), \hat{p}^*_{j_t,t}(\beta))||\{(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})) - (f_t^* - \mathbb{E}(f_t|\mathcal{F}^*_{t-1}))\}E_{n_t,j_t}|]^q \leq \mathbb{E}[|\{(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1})) - (f_t^* - \mathbb{E}(f_t|\mathcal{F}^*_{t-1}))\}E_{n_t,j_t}|]^q$. Thus the dependence adjusted norm of $C_{\mathbb{J},\mathbb{J}}^{-1/2}(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t}$ stays the same as $\hat{p}^\top_{j_t,t}(\beta)C_{\mathbb{J},\mathbb{J}}^{-1/2}(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j_t}$.

Add addtional steps to verify the other terms. □

3.18 Proof of Theorem 2.23

Proof. Define the standardized process as the following,

$$\tilde{U}_T(\beta_1, \beta_2, \beta_3) \stackrel{\text{def}}{=}$$
$$|T^{-1} \sum_t \sqrt{Tn_t/J_t}\{\hat{\mu}_t(\beta_1) + \hat{\mu}_t(\beta_3) - 2\hat{\mu}_t(\beta_2) - (\mu_t(\beta_1) + \mu_t(\beta_3) - 2\mu_t(\beta_2))\}/\hat{\sigma}_D(\beta_{1,2,3})^{1/2}|.$$

$$U_T(\beta_1, \beta_2, \beta_3) \stackrel{\text{def}}{=}$$
$$|T^{-1} \sum_t \sqrt{Tn_t/J_t}\{\hat{\mu}_t(\beta_1) + \hat{\mu}_t(\beta_3) - 2\hat{\mu}_t(\beta_2) - (\mu_t(\beta_1) + \mu_t(\beta_3) - 2\mu_t(\beta_2))\}/\sigma_D(\beta_{1,2,3})^{1/2}|.$$

Define the centered and standardized process as the leading term as follows,

$$Z_T(\beta_1, \beta_2, \beta_3) = \sqrt{T}T^{-1} \sum_t \sqrt{n_t/J_t} n_t^{-1} \sum_i \sum_{j_t} (\hat{p}_{j_t,t}(\beta_1) + \hat{p}_{j_t,t}(\beta_3) - 2\hat{p}_{j_t,t}(\beta_2))$$
$$\tilde{q}_{j_t}^{-1} \Phi^*_{i,j_t,t} \varepsilon_{it} / \sigma_D(\beta_{1,2,3})^{1/2}.$$

We shall analyze the leading term of the above statistics object,

$$\begin{aligned}
& \sup_{\beta_1,\beta_2,\beta_3} |\tilde{U}_T(\beta_1, \beta_2, \beta_3)| \\
= & \sup_{\beta_1,\beta_2,\beta_3} |U_T(\beta_1, \beta_2, \beta_3)| + O_p(r_{1,2,3}), \\
= & \sup_{\beta_1,\beta_2,\beta_3} |Z_T(\beta_1, \beta_2, \beta_3)| + O_p(r_{1,2,3} \vee \sqrt{Tn_u/J}J^{-1} \vee \sqrt{Tn_u/J}h),
\end{aligned}$$

where the first equality is due to Assumption 12, and the second equality is due to Theorem 2.14. We shall then prove the following steps:

- (Coupling with a Normal Vector)let the limit variance covariance matrix be $\Sigma(\beta_{\mathbb{J}_1}, \beta_{\mathbb{J}_2}, \beta_{\mathbb{J}_3}) = \lim_{T\to\infty} \operatorname{Cov}(U_T \circ \pi_{\tilde{\delta}}(\beta_{\mathbb{J}_1}, \beta_{\mathbb{J}_2}, \beta_{\mathbb{J}_3})$, denote $\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3$ as the number of elements within $B_{\tilde{\delta}}$, there exists $G_T \circ \pi_\delta(\beta_{\mathbb{J}}) \sim N(0, \Sigma(\beta_{\mathbb{J}_1,\mathbb{J}_2,\mathbb{J}_3}))]$ such that

$$r_2' = |\tilde{\mathbb{U}}_T \circ \pi_{\tilde{\delta}}(\beta_{\mathbb{J}_1,\mathbb{J}_2,\mathbb{J}_3}) - Z_{\mathbb{J}_1,\mathbb{J}_2,\mathbb{J}_3}|_\infty = O_p(T^{-\varepsilon'} \vee \sqrt{T}(J^{-1} \vee h)) = o\left((nT)^{-\varepsilon'}\right);$$

Let $Z_T(\beta)$ be a mean zero Gaussian process, with variance $\sigma_d(\beta) = \lim_{T\to\infty} T^{-1} \sum_t \sigma_t^2 \tilde{q}_{j_t}^{-1} \hat{p}_{j_t,t}(\beta)$. Since we have

$$\begin{aligned}
& |\tilde{\mathbb{U}}_T \circ \pi_{\tilde{\delta}}(\beta_{\mathbb{J}_1,\mathbb{J}_2,\mathbb{J}_3}) - Z_{\mathbb{J}_1,\mathbb{J}_2,\mathbb{J}_3}|_\infty \\
\leq & C \sup_{\beta\in\beta_v} |\sum_t \sqrt{T}^{-1/2} \sqrt{n_t J_t}^{-1} \sum_{j_t} \Phi^*_{i,j_t,t} \tilde{q}_{j_t}^{-1} \varepsilon_{it} \hat{p}_{j_t,t}(\beta) - Z_T(\beta)| \max_\beta (\sigma^{1/2}(\beta)) / \min_\beta \sigma_D^{1/2}(\beta_{1,2,3}).
\end{aligned}$$

Apply Theorem 3.3 again, with n correspond to $\sum_t n_t \asymp n_u T$. X_t corresponds to $\Phi^*_{i,j_t,t}\varepsilon_{it}\hat{p}_{j_t,t}(\beta_v)\tilde{q}_{j_t}$. Σ_z corresponds to $\mathrm{diag}(\sigma_d(\beta_v))$. Since by Assumption 12, $\max_{\beta\in\beta_v}\sigma_d(\beta)$ is bounded from the above and $\min_{\beta\in\beta_v}\sigma_d(\beta)$ is bounded from the below. In addition, we have by Assumption 12, the qth moment of $\max_{j_t,t}\|\Phi^*_{i,j_t,t}\tilde{q}_{j_t}^{-1}\varepsilon_{it}\|_q \lesssim J^{1-1/q}$. Thus we can prove that

$$\sup_{\beta\in\beta_v}|\sum_t \sqrt{T}^{-1/2}\sqrt{n_t J_t}^{-1}\sum_{j_t}\Phi^*_{i,j_t,t}\tilde{q}_{j_t}^{-1}\varepsilon_{it}\hat{p}_{j_t,t}(\beta) - Z_T(\beta)| = O_p((n_u T)^{-\varepsilon'}), \tag{3.31}$$

for a constant $0 < \varepsilon' < 1/6$.

□

3.19 Strong approximation for weakly dependent processes

Let us first derive results for nonstationary high dimensional time series. We denote X_t ($t \in 1,\cdots,n$) as a $p-$ dimension zero mean time series, we let $X_{t,j} = H_{t,j}(\varepsilon_t,\varepsilon_{t-1},\cdots)$. We let $X_{t,j} = H_{t,j}(\varepsilon_t,\varepsilon_{t-1},\cdots,\varepsilon^*_{t-k},\cdots)$, where ε^*_{t-k} is an independent i.i.d. copy of ε_{t-k} . We can denote the dependence adjusted norm as

$$\delta_{j,k,q} \stackrel{\text{def}}{=} \sup_t \|X_{t,j} - X^{k*}_{t,j}\|. \tag{3.32}$$

We denote $\Gamma_{\alpha,q} = \sup_{l\geq 0} l^{-\alpha}(\sum_j(\sum_{k\geq l}\delta_{j,k,q})^q)^{1/q}$. And $\Theta_{q,\alpha} = \||X_\cdot|_\infty\|_{q,\alpha}(\log p)^{3/2}\wedge\Gamma_{\alpha,q}$. $\Psi_{2,\alpha} \stackrel{\text{def}}{=} \max_j \sup_{l\geq 0} l^{-\alpha}(\sum_{k\geq l}\delta_{j,k,q})$.

Theorem 3.3. *Suppose that X_t is a $p-$ dimensional mean zero nonstationary time series then on a rich probability space, there exists a Gaussian random variable Z such that i) $\Sigma_z = \lim_{n\to\infty} n^{-1}\sum_t\sum_{l\geq 0}\mathbb{E}(X_t X_{t-l}^\top)$, and $Z \sim \mathrm{N}(0,\Sigma_z)$. Assume that X_t has elementwise bounded qth moment ($q > 4$). The element of X_t is $X_{t,j}$. ii) $\log(p)^3 \leq L$ and $\log p \ll (M/m)^{1/3}$. The dependence adjusted norm for X_t, i.e. $\Theta_{q,\alpha}$ and $\Psi_{2,\alpha}$ are bounded. We define the dependence adjusted norm for Z_t as $\Phi_{\psi_1,\alpha'} \stackrel{\text{def}}{=} \max_q q^{-\alpha'}\max_j \Theta^z_{\alpha',q,j}$. Let $\beta = 2/(1+2\alpha')$. Let $\alpha' = 1$, $\beta = 2/3$. iii) We assume that for the long run variance Σ_z, $\lambda_{\max}(\Sigma_z) \leq C$ and $\lambda_{\min}(\Sigma_z) \geq c > 0$, then we have,*

$$\mathrm{P}(|\sqrt{n}^{-1/2}\sum_t X_t - Z|_\infty \geq \delta(p,L,m,q,\alpha)) \to 0. \tag{3.33}$$

Note that we shall let the small blocks $m \ll M$ and $L = \lfloor n/(m+M)\rfloor$. iv) $\delta(p,L,m,q,\alpha)\to 0$. We let $n \gg mL(\log p)^2$, and $\delta(p,L,m,q,\alpha) \lesssim L^{-1/6}(\log p)^{3/2}\vee L^{-1/2+1/q}(\log p)^{1-1/q}p^{1/q}\vee m^{-\alpha}\sqrt{\log p}\vee m^{1/2-1/q-\alpha}n^{1/q-1/2} \ll 1$ for $\alpha > 1/2 - 1/q$. And $L^{-1/2}\sqrt{\log p^3}\log(p\vee L) \ll \delta(p,L,m,q,\alpha)$. $(\log p)^{1/\beta}m^{-\alpha}\vee\sqrt{mL}(\log p)^{1/\beta} \ll \sqrt{n}$. $\{(\log p)^{1/\beta}m^{-\alpha}\vee\sqrt{mL}(\log p)^{1/\beta}\}/\sqrt{n} \lesssim \delta(p,L,m,q,\alpha) \ll 1$.

Remark 3.1. (discussion of rates) We see that q, m, M, L, p interplays with each other. Compared to the iid case, we have observation loss, with respect to blocks. The term $L^{-1/6}(\log p)^{3/2}\vee L^{-1/2+1/q}(\log p)^{1-1/q}p^{1/q}$ corresponds to the rate in the iid case with L replaced by n in Theorem

2.1 Chernozhukov, Chetverikov, and Kato (2016). We show an example, when we set q to be large enough and p to be small enough such that $L^{-1/2+1/q}(\log p)^{1-1/q}p^{1/q} \vee m^{1/2-1/q-\alpha}n^{1/q-1/2}$ are of small order. We analyze $L^{-1/6}(\log p)^{3/2} \vee m^{-\alpha}\sqrt{\log p}$. For example let $L = n^{2/3}$, then $n^{1/(9\alpha)} \ll m \ll M \lesssim n^{1/3}$. Then $L^{-1/6}(\log p)^{3/2} \vee m^{-\alpha}\sqrt{\log p} \lesssim n^{-1/9}(\log p)^{3/2}$. It would also be noted that if the dependence is rather weak α can be very large and therefore the rate of m can be small and the block size can diverge less slowly. Thus we can see that for q to be large enough, the rate $\delta(p, L, m, q, \alpha)$ can be of the following order $L^{-1/2}\sqrt{\log p^3}\log(p \vee L) \ll \delta(p, L, m, q, \alpha) \asymp L^{-1/6}(\log p)^{3/2} \vee (\log p)^{3/2}\sqrt{m/M}$. ┘

Proof. **Step 1** divide observations into block, approximate by m dependent blocks.
We denote $X_{t,m} = \mathbb{E}(X_t|\varepsilon_{t-m}, \cdots, \varepsilon_t)$. We shall divide the observations into big blocks of size M and small blocks of size m.

$$L_b = [(b-1)(M+m)+1, bM+(b-1)m] \tag{3.34}$$

$$S_b = [bM+(b-1)m+1, b(M+m)]. \tag{3.35}$$

This in total leads to $L \stackrel{\text{def}}{=} \lfloor T/(m+M) \rfloor$. We then throw away the observations in small blocks to construct approximation of the partial sums by partial sum independent blocks. Namely we denote

$$Y_{b,m} \stackrel{\text{def}}{=} \sum_{t \in L_b} X_t$$

$$T_b \stackrel{\text{def}}{=} \sum_{b=1}^{L_b} Y_{b,m}.$$

The m- dependent counterpart is denoted as

$$\tilde{Y}_{b,m} \stackrel{\text{def}}{=} \sum_{t \in L_b} X_{t,m}$$

$$\tilde{T}_b \stackrel{\text{def}}{=} \sum_{b=1}^{L_b} \tilde{Y}_{b,m}.$$

According to similar steps as in Lemma 7.1 Zhang and Wu (2017), we have that if $\Theta_{q,\alpha} < \infty$. We let $\alpha > 1/2 - 1/q$. Then we have

$$\mathrm{P}(|\sum_t X_t - \sum_{b=1}^{L_b} \tilde{Y}_{b,m}|_\infty \geq y) \leq f(y, m), \tag{3.36}$$

where $f(y,m)$ is denoted as

$$y^{-q}nm^{q/2-1-\alpha q}\Theta_{q,\alpha}^{q}+p\exp(-C_{q,\alpha}y^{2}m^{2\alpha}/n\Psi_{2,\alpha}^{2}). \tag{3.37}$$

By choose sufficient large m we can arrange $\sum_b \tilde{Y}_{b,m}$ to be sufficiently close to $\sum_t X_t$. If $\Theta_{q,\alpha}$ and $\Psi_{2,\alpha}$ are bounded, we need to assume $n^{1/q}m^{1/2-1/q-\alpha}n^{-1/2}\ll 1$ and $n^{1/2}m^{-\alpha}\sqrt{\log p}n^{-1/2}\ll 1$. y shall be set $n^{1/q}m^{1/2-1/q-\alpha}n^{-1/2}\vee n^{1/2}m^{-\alpha}\sqrt{\log p}n^{-1/2}\ll yn^{-1/2}\ll 1$.

Step 2 Approximate independent blocks by Gaussian random variables.

As $\sum_b \tilde{Y}_{b,m}$ are partial sum of independent blocks by construction, we shall apply Theorem 3.1 as in Chernozhukov, Chetverikov, and Kato (2016). We denote $Y_{b,m,j}$ as the jth element of $Y_{b,m}$. $g(\delta)\stackrel{\text{def}}{=}\frac{\log^2 p}{\delta^3\sqrt{L}}\{L_n+M_{L,1}(\delta)+M_{L,2}(\delta)\}$. We denote Z_b as a p dimension mean zero Gaussian random variable with variance covariance structure as $\mathbb{E}(Y_bY_b^\top)$, then

$$\begin{aligned}&\mathrm{P}(\max_j|\sum_b\sqrt{L}^{-1/2}\tilde{Y}_{b,m,j}|\in A)-\mathrm{P}(\max_j|\sum_b\sqrt{L}^{-1/2}Z_{b,j}|\in A^{\delta})\\&\le g(\delta),\end{aligned} \tag{3.38}$$

where the term in $g(\delta)$ involves $L_n=\max_{b,j}\mathbb{E}(|Y_{b,j}|^3)$,

$$M_{L,1}(\delta)\stackrel{\text{def}}{=}L^{-1}\sum_b\mathbb{E}[\max_j|Y_{b,j}|^3\mathbf{1}\{\max_j|Y_{b,j}|\ge\delta\sqrt{L}/\log p\}],$$

and

$$M_{L,2}(\delta)\stackrel{\text{def}}{=}L^{-1}\sum_b\mathbb{E}[\max_j|Z_{b,j}|^3\mathbf{1}\{\max_j|Z_{b,j}|\ge\delta\sqrt{L}/\log p\}].$$

As the statement in (3.38) would imply

$$\mathrm{P}(\max_j|\sqrt{L}^{-1/2}\sum_b(\tilde{Y}_{b,j}-Z_{b,j})|\ge\delta)\le g(\delta).$$

Now we can see that with properly chosen m,δ,y,

$$\mathrm{P}(\sqrt{L}^{-1/2}|\sum_t X_t-\sum_b Z_b|_\infty\ge\delta+y)\le f(y,m)+g(\delta), \tag{3.39}$$

we shall pick m M, δ such that $\delta/\sqrt{M+m}\ll 1$ and $g(\delta)\to 0$.

To analyze the rate of $g(\delta)$ we make use of the following lemma,

Lemma 3.4 (Burkholder (1988), Rio (2009)). *Let $q>1$, $q'=\min(q,2)$. Let $M_n=\sum_{t=1}^n\xi_t$; where*

$\xi_t \in \mathcal{L}^q$ *(i.e.,* $\|\xi_t\|_q < \infty$*) are martingale differences. Then*

$$\|M_n\|_q^{q'} \leq K_q^{q'} \sum_{t=1}^{n} \|\xi_t\|_q^{q'} \quad \textit{where} \quad K_q = \max((q-1)^{-1}, \sqrt{q-1}).$$

Denote $\theta_{q,j,\alpha} \stackrel{\text{def}}{=} \sup_{l \geq 0} l^{-\alpha} \sum_{k \geq l} \delta_{j,k,q}$, by Lemma 3.4,

$$L_n \quad = \quad \max_{b,j} \mathbb{E}(|Y_{b,j}|^3) \lesssim M^{3/2} \theta_{q,j,\alpha}^3.$$

$v_{L,q} \stackrel{\text{def}}{=} \max_b (\mathbb{E}(\max_j |Y_{b,j}|^q))^{1/q} \lesssim (M^{1/2} \theta_{q,j,\alpha})$.

$$\begin{aligned} & M_{L,1}(\delta) \stackrel{\text{def}}{=} (L)^{-1} \sum_b \mathbb{E} \max_j |Y_{b,j}|^3 \mathbf{1}\{\max_j |Y_{b,j}| \geq \delta\sqrt{L}/\log p\} \\ \lesssim \; & \frac{(\log p)^{q-3}}{(\delta\sqrt{L})^{q-3}} v_{L,q}^q p \lesssim \frac{(\log p)^{q-3}}{(\delta\sqrt{L})^{q-3}} (M^{q/2} \theta_{q,j,\alpha}^q) p, \end{aligned}$$

following the steps as the proof of Theorem 2.1 of Chernozhukov, Chetverikov, and Kato (2016), however we have a better bound for the $M_{nT,2}(\delta)$. For a Gaussian random variable we have, $Z_{b,j}$ follows a $N(0, \Sigma_{b,jj})$, let $\xi = \max_{1 \leq j \leq p} |Z_{b,j}|/\Sigma_{b,jj}^{1/2}$,

$$\begin{aligned} & \mathbb{E}|\xi|^3 \mathbf{1}(\xi \geq t) \\ = \; & \mathrm{P}(\xi > t) t^3 + 3 \int_t^\infty \mathrm{P}(\xi > x) x^2 dx \\ \leq \; & p/(t\sqrt{2\pi}) \exp(-t^2/2) + 6p \int_t^\infty \frac{1}{\sqrt{2\pi} s} \exp(-s^2/2) s^2 ds \\ \lesssim \; & p t^2 \exp(-t^2/2) + (\sqrt{2\pi})^{-1} 6p \int_t^\infty -d \exp(-s^2/2) \\ \lesssim \; & t^2 \exp(-t^2/2) p/\sqrt{2\pi} + 6p \exp(-t^2/2) \sqrt{2\pi}^{-1}, \end{aligned}$$

by the inequality of a Gaussian distribution

$$\mathrm{P}(|Z| > t) \leq \frac{2}{t\sqrt{2\pi}} \exp(-t^2/2). \tag{3.40}$$

It thus follows that, if we assume that $\theta_{q,j,\alpha}$ is bounded,

$$M_{L,2}(\delta) \lesssim M^{3/2} \{\delta\sqrt{L}/(\sqrt{M} \log p)\}^2 [\exp(-\{\delta\sqrt{L}/(\sqrt{M} 2 \log p^{3/2})\}^2)]. \tag{3.41}$$

Thus to ensure that $g(\delta) \to 0$ if the involved dependence adjusted norm is bounded, we need to have $T_{yn2} \stackrel{\text{def}}{=} M^{1/2} L^{-1/6} (\log p)^{2/3} \vee p^{1/(q)} M^{1/2} (\log p)^{1-1/q} / \sqrt{L}^{(q-2)/q}$ and $\delta \lesssim T_{yn2}$. $L^{-1/2} \sqrt{\log p^3} \sqrt{M} \log(p \vee L) \ll \delta$, and $\delta/\sqrt{M+m} \to 0$.

Step 3 Approximate Gaussian partial sums

We define $T_{yn1} \overset{\text{def}}{=} n^{1/q}m^{1/2-1/q-\alpha}n^{-1/2} \vee n^{1/2}m^{-\alpha}\sqrt{\log p}n^{-1/2}$ as obtained in Step 1. Let $T_{yn1} \to 0$. As both Z and $\sum_b Z_b$ are Gaussian random variables, we can derive the rate following an concentration inequality using a sub Gaussian norm. We define $\Phi_{\psi_1,\alpha'} \overset{\text{def}}{=} \max_q q^{-\alpha'} \max_j \Theta^z_{\alpha',q,j}$. Let $\beta = 2/(1+2\alpha')$. Let $\alpha' = 1$, $\beta = 2/3$.

We define $Z_n = n^{-1/2}\sum_i^n Z_i$ such that $\mathbb{E}(Z_t Z_{t-l}^\top) = \mathbb{E}(X_t X_{t-l}^\top)$. In the next step we bound

$$\mathrm{P}(|(\sqrt{n}Z_n - \sum_b Z_b)|_\infty \geq y') \leq f'(y', \delta), \tag{3.42}$$

where $f'(y',\delta)$ is taking form of $p\exp(-C_\beta(y'm^\alpha/(\sqrt{n})\Phi_{\psi_1,\alpha})^\beta) + p\exp(-C_\beta(y'/(\sqrt{mL}\Phi_{\psi_1,\alpha}))^\beta)$ by Zhang and Wu (2017)(Lemma 7.1 (ii)). Therefore, we need to ensure $(\log p)^{1/\beta}m^{-\alpha}(\sqrt{n}) \vee \sqrt{mL}(\log p)^{1/\beta} \ll y'$ to make the right hand side function $f'(y',\delta)$ tends to zero. The rate is fine tuned by combining the above three steps, $y'/\sqrt{n} \ll 1$.

□

4 References

Berkes, I., W. Liu, and W. B. Wu (2014): "Komlós–Major–Tusnády approximation under dependence," *The Annals of Probability*, 42(2), 794–817.

Bobkov, S. G., I. Gentil, and M. Ledoux (2001): "Hypercontractivity of Hamilton–Jacobi equations," *Journal de Mathématiques Pures et Appliquées*, 80(7), 669–696.

Burkholder, D. L. (1988): "Sharp inequalities for martingales and stochastic integrals," *Astérisque*, 157-158, 75–94.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng (2022): "On binscatter," *arXiv preprint arXiv:1902.09608*.

Chernozhukov, V., D. Chetverikov, and K. Kato (2016): "Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings," *Stochastic Processes and their Applications*, 126(12), 3632–3651.

Chernozhukov, V., D. Chetverikov, K. Kato, et al. (2013): "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors," *The Annals of Statistics*, 41(6), 2786–2819.

Gagliardini, P., E. Ossola, and O. Scaillet (2016): "Time-Varying Risk Premium in Large Cross-Sectional Equity Data Sets," *Econometrica*, 84(3), 985–1046.

——— (2019): "Estimation of large dimensional conditional factor models in finance," *Swiss Finance Institute Research Paper*, (19-46).

Giglio, S., and D. Xiu (2021): "Asset pricing with omitted factors," *Journal of Political Economy*, 129(7), 000–000.

Hall, P., and C. C. Heyde (2014): *Martingale limit theory and its application*. Academic press.

Massart, P. (1990): "The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality," *The annals of Probability*, pp. 1269–1283.

Rio, E. (2009): "Moment inequalities for sums of dependent random variables under projective conditions," *J Theor Probab*, 22(1), 146–163.

Stute, W. (1982): "The oscillation behavior of empirical processes," *The annals of Probability*, pp. 86–107.

van de Geer, S. (2000): *Empirical Processes in M-estimation*, vol. 6. Cambridge university press.

Van der Vaart, A. W. (2000): *Asymptotic statistics*, vol. 3. Cambridge university press.

Wu, W.-B., and Y. N. Wu (2016): "Performance bounds for parameter estimates of high-dimensional linear models with correlated errors," *Electronic Journal of Statistics*, 10(1), 352–379.

Zhang, D., and W. B. Wu (2017): "Gaussian approximation for high dimensional time series," *The Annals of Statistics*, 45(5), 1895–1919.

Zhang, T., and W. B. Wu (2012): "Inference of time-varying regression models," *The Annals of Statistics*, 40(3), 1376–1402.