Beta-Sorted Portfolios*

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

August 23, 2022

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 rationalize the well-known estimation algorithm with precise economic and statistical assumptions on the general data generating process and characterize its key features. We study beta-sorted portfolios for both a single cross-section as well as for aggregation over time (e.g., the grand mean), offering conditions that ensure consistency and asymptotic normality along with new uniform inference procedures allowing for uncertainty quantification and testing of various relevant hypotheses in financial applications. We also highlight some limitations of current empirical practices and discuss what inferences can and cannot be drawn from returns to beta-sorted portfolios for either a single cross-section or across the whole sample. Finally, we illustrate the functionality of our new procedures in an empirical application.

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

*The authors would like to thank Fernando Duarte and Olivier Scaillet along with seminar participants at the 2022 ML approaches in Finance and Management conference, 9th asset pricing workshop in York and Vienna–Copenhagen Conference on Financial Econometrics in 2022 for helpful discussions and comments. 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).

[†] 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.

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

[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).

example, in this setting, no-arbitrage conditions are not guaranteed to hold and instead imply testable hypotheses. Within this framework, we impose 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 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, 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 Fama-MacBeth 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. 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 (necessarily) exclude arbitrage opportunities and supposes risk exposures which are smoothly-varying. 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-

[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.

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.

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. In Section 6 we study pointwise and uniform inference procedures for a single cross-section and describe what properties of conditional expected returns are estimable. Section 7 presents an empirical application, and Section 8 concludes. Detailed assumptions and proofs of the results are relegated to the Appendix.

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 $X_n = o_{\mathbb{P}}(a_n) : \forall \varepsilon, \delta > 0 \quad \exists N_{\varepsilon,\delta} \quad$ 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 $\beta_{it} \in \mathbb{R}^d$ ($d \geq 1$) 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 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,

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

[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.

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

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

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

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

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

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

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

[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).

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

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

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 estimator 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 conditional 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 there estimation of the conditional expectation of the risk factors, $\mathbb{E}[f_t|\mathcal{F}_{t-1}]$, and so the researcher ostensibly 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 not controversial, 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* $\log(n_uT)/Th \to 0$. *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 $\delta_T = (r_T + \sqrt{\log(n_uT)}/\sqrt{Th} + h)$.

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 $\tau' : [0,1] \mapsto \mathbb{R}$. 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 equation (B.14) 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^2_t can be estimated by

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

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

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 $\mu_t(\cdot)$; see Remark A.9 in the Appendix for a formal discussion. 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 $\Phi^*_{i,j,t} = \mathbf{1}(F_{\beta,t}^{-1}((j-1)/J_t) \leq \beta_{it} < F_{\beta,t}^{-1}((j)/J_t))$ 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,

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

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 different basis functions which underpin our analysis.

[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 $j = J_t$.

Define $\Phi_{i,j,t} = \mathbf{1}(U_{it} \in [U_{(\lfloor (j-1)n_t/J_t \rfloor),t}, U_{(\lfloor (j)n_t/J_t \rfloor),t})) = \mathbf{1}(F^{-1}_{\beta,n,t}((j-1)/J_t) \leq \beta_{it} < F^{-1}_{\beta,n,t}((j)/J_t))$ and, for estimated β_{it}, its counterpart $\widehat{\Phi}_{i,j,t} = \mathbf{1}(F^{-1}_{\widehat{\beta},n,t}((j-1)/J_t) \leq \widehat{\beta}_{it} < F^{-1}_{\widehat{\beta},n,t}(j/J_t))$. 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,

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

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 $\tilde{q}_{jt} = \int_{F^{-1}_{\beta,t}(\kappa_{j-1t})}^{F^{-1}_{\beta,t}(\kappa_{jt})} f_{\beta,t} d\beta$ 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 Appendix, we present a number of preliminary lemmas – Lemmas A.4 to A.7 – which 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,*

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

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

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

Theorem 4.1 introduces the key properties of the grand mean estimator. Importantly, the theorem shows that the leading term of the linearization is of the order $O_{\mathbb{P}}(T^{-1/2})$ and involves a term 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 achieves only a $\sqrt{T}$ rate of convergence. In this sense, the time-series dimension is responsible for the dominant source of variation in the estimator. This comes about because we remain agnostic about the functional form of the conditional expectation $\mathbb{E}\left[f_t|\mathcal{F}_{t-1}\right]$.

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 accelerating the 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 estimated $T^{-1}\sum_{t=1}^{T}\mu_t(.)$. We define $E_{n_t,j}=\tilde{q}_{jt}^{-1}\mathbb{E}(\Phi_{i,j,t}^{*}\beta_{it}|\mathcal{G}_{t-1})$.

Theorem 4.3 (Pointwise central limit theorem)**.** *Suppose Assumptions 6, 7-8 and 10-12 hold. Then, pointwise in* β*,*

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

where $\hat{\sigma}(\beta)=T^{-1}\sum_{t=1}^{T}(\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta)E_{n_t,j}^2\,Var(f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1}))$*, 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 $(J^{-1} \vee h)$.

In the next section, we shall apply a uniform Gaussian approximation on the time average of the leading component, providing a uniform inference counterpart to equation (4.3).

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.*

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. We begin by demonstrating the theoretical properties of this variance estimator. We show that the estimator is not consistent, in general, but that it is guaranteed to produce valid, albeit possibly conservative, inference. 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 The Fama-MacBeth estimator

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

$$\widehat{\sigma}_{\mathtt{FM}}(\beta) = \frac{1}{T}\sum_{t=1}^{T}\Big(\widehat{\mu}_t(\beta) - \widehat{\mu}(\beta)\Big)^2. \tag{5.1}$$

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,

$$\widehat{\sigma}_{\mathtt{FM}}(\widehat{\mu}(\beta_1), \widehat{\mu}(\beta_2)) = \frac{1}{T} \sum_{t=1}^{T} \Big(\widehat{\mu}_t(\beta_1) - \widehat{\mu}(\beta_1)\Big)\Big(\widehat{\mu}_t(\beta_2) - \widehat{\mu}(\beta_2)\Big). \tag{5.2}$$

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

First, define

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

The quantity $\sigma(\beta_1, \beta_2)$ 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*

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

Note that the above results implies that

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

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, $\sigma(\beta)$ and $\sigma(\beta_1, \beta_2)$, 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 $\mu(\beta) = \operatorname{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T}\widehat{\mu}_t(\beta)$. 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.*

5.2 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 is $T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta)(f_t - \mathbb{E}(f_t))E_{n_t,j}$ is defined as follows

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

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 $\sigma(\beta)^{-1/2}G_T(\beta)$ as a Gaussian process on a proper probability space with covariance $\sigma(\beta_1)^{-1/2}\sigma(\beta_1,\beta_2)\sigma(\beta_2)^{-1/2}$, and similarly for $\tilde{\sigma}(\beta)^{-1/2}\tilde{G}_T(\beta)$ relative to $\tilde{\sigma}(\beta_1)^{-1/2}\tilde{\sigma}(\beta_1,\beta_2)\tilde{\sigma}(\beta_2)^{-1/2}$. 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,*

$$\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\sup_{\beta\in\mathcal{B}}\Big|\frac{1}{\sqrt{T\widehat{\sigma}(\beta)}}\sum_{t=1}^{T}\left(\widehat{\mu}_t(\beta)-\mu_t(\beta)\right)\Big|\le x\Big)-\mathbb{P}\Big(\sup_{\beta\in\mathcal{B}}|G_T(\beta)|\le x\Big)\Big|\to 0. \tag{5.3}$$

By the above lemma, we shall expect the asymptotic distribution of $\sup_{\beta\in\mathcal{B}}\left|\frac{1}{\sqrt{T\widehat{\sigma}(\beta)}}\sum_{t=1}^{T}\left(\widehat{\mu}_t(\beta)\right)\right|$ 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 $\sup_\beta \hat{\mu}(\beta)/\hat{\sigma}(\beta)^{1/2}$ 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:

Algorithms

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 $\hat{\tilde{\Sigma}} = \operatorname{diag}(\hat{\tilde{\Sigma}})^{-1/2} \hat{\tilde{\Sigma}} \operatorname{diag}(\hat{\tilde{\Sigma}})^{-1/2}$.

4 Obtain the α quantile of $|\tilde{Z}|_\infty$ from the above sample, and we denote it as $\widehat{q}_\alpha$.

5 Create the confidence band $[\widehat{\mu}(\beta) - \widehat{\sigma}(\beta)^{1/2}\widehat{q}_\alpha/\sqrt{T}, \widehat{\mu}(\beta) + \widehat{\sigma}(\beta)^{1/2}\widehat{q}_\alpha/\sqrt{T}]$, where $\widehat{L}_T(\beta) = \widehat{\mu}(\beta) - \widehat{\sigma}(\beta)^{1/2}\widehat{q}_\alpha/\sqrt{T}$ and $\widehat{U}_T(\beta) = \widehat{\mu}(\beta) - \widehat{\sigma}(\beta)^{1/2}\widehat{q}_\alpha/\sqrt{T}$. If 0 is within the confidence band we cannot reject H_0.

5.3 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,

$$\mathsf{H}_0^{(1)} : \sup_{\beta \in \mathcal{B}} \mu(\beta) - \inf_{\beta \in \mathcal{B}} \mu(\beta) = 0, \tag{5.4}$$

versus,

$$\mathsf{H}_A^{(1)} : \sup_{\beta \in \mathcal{B}} \mu(\beta) - \inf_{\beta \in \mathcal{B}} \mu(\beta) \neq 0. \tag{5.5}$$

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, $\mathcal{T}_T = [T^{-1}(\sum_{t=1}^{T}\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta^{**})(\widehat{a}_{j,t}-a^*_{j,t}))-T^{-1}(\sum_{t=1}^{T}\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta^{*})(\widehat{a}_{j,t}-a^*_{j,t}))]/(\widehat{\sigma}(\beta^*)+\widehat{\sigma}(\beta^{**})-2\widehat{\sigma}(\beta^*,\beta^{**}))^{1/2}$ 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,*

$$\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\Big|\mathcal{T}_T\Big|\le x\Big)-\mathbb{P}\Big(|\tilde{T}_z|\le x\Big)\Big|\to 0. \tag{5.6}$$

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.

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

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 $\widehat{\mu}_g = \widehat{\mu}(\beta_{\max}) - \widehat{\mu}(\beta_{\min})$.

3 Multiply $\widehat{\tilde{\Sigma}}^{1/2}$, we get $\tilde{Z}^{(s)} = \widehat{\tilde{\Sigma}}^{1/2} Z^{(s)}$. Obtain $\tilde{Z}^{(s)}_{\widehat{\mathbb{J}}^*} - \tilde{Z}^{(s)}_{\widehat{\mathbb{J}}^{**}}$.

4 Obtain the α quantile of $|\tilde{Z}^{(s)}_{\widehat{\mathbb{J}}^*} - \tilde{Z}^{(s)}_{\widehat{\mathbb{J}}^{**}}|$ from the above sample, and we denote as $\widehat{q}_\alpha$. ($\tilde{Z}^{(s)}_{\widehat{\mathbb{J}}^*}$ $\tilde{Z}^{(s)}_{\widehat{\mathbb{J}}^{**}}$ are the Gaussian limit corresponding to $\beta_{\max}$ and β_{min} repectively.)

5 Create the confidence interval $[\widehat{\mu}_g - \widehat{q}_\alpha/\sqrt{T}, \widehat{\mu}_g + \widehat{q}_\alpha/\sqrt{T}]$, where $\widehat{L}_T(\beta) = \widehat{\mu}_g - \widehat{q}_\alpha/\sqrt{T}$ and $\widehat{U}_T(\beta) = \widehat{\mu}_g + \widehat{q}_\alpha/\sqrt{T}$. We cannot reject $H_0^{(1)}$ if 0 is contained in the confidence interval.

5.4 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 practical relevance of testing for the presence of a profitable butterfly trade, we observe that the statistical properties of the inference procedure are sharply 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 rate of convergence of only $\sqrt{T}$ owing to the presence of a linear (in beta) term in the asymptotic expansion. 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

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

versus the alternative one

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

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. Namely, we have $\sqrt{T}\sqrt{n_t}/\sqrt{J}$ versus a $\sqrt{T}$ rate induced by the factor term. 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 15, 13 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*
$Cov(\beta_{1,2,3},\beta'_{1,2,3})/(\sigma_D(\beta_{1,2,3})^{1/2}\sigma_D(\beta'_{1,2,3})^{1/2})$.

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

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

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 $\hat{\tilde{\Sigma}}_D$($\hat{\tilde{\Sigma}}_D$ 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 $\tilde{Z}^{(s)} = \hat{\tilde{\Sigma}}^{1/2} Z^{(s)}$. Obtain $|\tilde{Z}^{(s)}|_{\max}$.

5 Obtain the α quantile of $|\tilde{Z}^{(s)}|_{\max}$ from the above sample, and we denote as $\widehat{q}_\alpha$.

6 Create the confidence interval $[\tilde{Z}_T(\beta_{1,2,3}) - \widehat{\sigma}_D(\beta_{1,2,3})^{1/2} \min_t \widehat{q}_\alpha \sqrt{J_t}/\sqrt{nT}$, $\tilde{Z}_T(\beta_{1,2,3}) + \widehat{\sigma}_D(\beta_{1,2,3})^{1/2} \max_t \widehat{q}_\alpha \sqrt{J_t}/\sqrt{nT}]$. We cannot reject $H_{0,diff}$ if 0 is contained in the confidence interval.

6 Feasible (uniform) inference for fixed t

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. Thus this section is devoted to studying procedures for conducting inference for a fixed cross-section at time t. A common thread throughout the discussion in this section is that, as mentioned earlier, we cannot estimate $\mu_t(\beta)$

consistently in our setting. Instead, we can consistently estimate

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

Nonetheless, we can still conduct inference on some features of $\mu_t(\beta)$ even if we cannot fully characterize this primary object of interest. For example the linearity of $M_t(\beta)$ in β can also imply the linearity of $\mu_t(\beta)$ in β. We will introduce a uniform inference procedure which allows for drawing inference on $M_t(\beta)$ and testing some related hypotheses.

Let us now discuss our hypotheses of interest. First, we have

$$\mathsf{H}_0^c : M_t(\beta) = c.$$

Under this null, conditional expected returns are constant irrespective of the value of β. If we fail to reject this null, we could conclude that the data do not provide evidence that the risk is conditionally priced – even if it might be unconditionally priced.[10] Next, we have

$$\mathsf{H}_0^{(1)} : \exists \ell_{1t} \ s.t. \ M_t(\beta) = \ell_{1t}'\beta.$$

The hypothesis $\mathsf{H}_0^{(1)}$ corresponds to testing the no -arbitrage assumption since when this assumption holds we have that $M_t(\beta)$ is linear in β. Finally, we could also consider the null hypothesis:

$$\mathsf{H}_0^{(2)} : \exists (\ell_{0t}, \ell_{1t}) \ \ s.t. \ M_t(\beta) = \ell_{0t} + \ell_{1t}'\beta.$$

This hypothesis is compelling because if we can reject $\mathsf{H}_0^{(2)}$ then it implies that $M_t(\beta)$ is not constant, nor is it affine in β. Instead conditional expected returns are a *nonlinear* function of β. Note that $\mathsf{H}_0^{(1)}$ implies $\mathsf{H}_0^{(2)}$ but the converse is not true in general. Moreover, if $\mathsf{H}_0^{(1)}$ is false but $\mathsf{H}_0^{(2)}$ is true that would imply that $M_t(\beta)$ is either constant or affine in β. More generally, we have the following relationships:

[10]Unless otherwise noted, all of the alternative hypotheses within this section may be obtained by replacing the equality in the null hypothesis by $\neq$.

$$
\begin{array}{c|c|c}
 & \left[\mathsf{H}_0^{(1)} \text{ true}\right] & \left[\mathsf{H}_0^{(1)} \text{ false}\right] \\
\hline
\left[\mathsf{H}_0^{(2)} \text{ true}\right] & \left\{\begin{array}{c} M_t(\beta) = 0 \ \forall \beta \\ \text{or} \\ M_t(\beta) \text{ is linear in } \beta \end{array}\right\} & \left\{\begin{array}{c} M_t(\beta) \text{ is constant in } \beta \\ \text{or} \\ M_t(\beta) \text{ is affine in } \beta \end{array}\right\} \\
\hline
\left[\mathsf{H}_0^{(2)} \text{ false}\right] & \{\emptyset\} & \{M_t(\beta) \text{ is nonlinear in } \beta\}
\end{array}
$$

This relationship naturally suggests a sequential testing procedure: testing $\mathsf{H}_0^{(1)}$ first and, if you reject $\mathsf{H}_0^{(1)}$, then you test $\mathsf{H}_0^{(2)}$ next. We should be able to determine an appropriate size α_1 to test $\mathsf{H}_0^{(1)}$ and α_2 to test $\mathsf{H}_0^{(2)}$ such as by a simple Bonferroni correction. Thus, although we cannot consistently estimate $\mu_t(\beta)$ we can still plausibly learn about some properties of $\mu_t(\beta)$.

We can utilize the same estimator as for $\mu_t(\beta)$ but for clarity, in this section, we will denote it as $\widehat{M}_t(\beta) = \sum_{j=1}^{J_t} \widehat{p}_{jt}(\beta)\widehat{a}_{jt}$. Under the conditions of Theorem B.6, we can derive that

$$
(\widehat{M}_t(\beta) - M_t(\beta)) = \widehat{p}_t^\top(\beta)\{\operatorname{diag}[\tilde{q}_{jt}]\}^{-1}\{n_t^{-1}\Phi_t^* \varepsilon_t\} + O_{\mathbb{P}}(J_t^{-1} \vee h).
$$

6.1 Test and theory regarding $M_t(\beta) = 0$

Since we have listed the hypotheses of interest for fixed t, we need to show the theoretical results to obtain asymptotic critical values. We shall start with the simplest hypothesis, which is associated with constructing a uniform confidence band for $\widehat{M}_t(\beta)$. To this end, we consider the following t-statistics $Z_{n_t}(\beta) = (\sqrt{n_t}/\sqrt{J_t})\{\widehat{M}_t(\beta) - M_t(\beta)\}/\widehat{\sigma}_t^{1/2}(\beta)$. Namely, $Z_t(\beta) = \sum_{j=1}^{J_t} \widehat{p}_j(\beta) Z_j$, where Z_js are standard normal distributed random variable. Let $\widehat{\sigma}_j^2 = n_t/J_t(\sum_{i=1}^{n_t} \widehat{\Phi}_{i,j,t}\widehat{\varepsilon}_{it}^2)(\sum_{i=1}^{n_t} \widehat{\Phi}_{i,j,t})^{-2}$. Denote $\mathbb{E}_{t-1}(.) = \mathbb{E}(.|\mathcal{F}_{t-1})$. We see that $\sigma_j^2 = 1/J_t(\mathbb{E}_{t-1}(\Phi_{i,j,t}^* \varepsilon_{it}^2))(\mathbb{E}_{t-1}(\Phi_{i,j,t}^*))^{-2}$. Let $\mathbb{E}_{t-1}(\Phi_{i,j,t}^*) = \tilde{q}_{jt}$, $\sigma_t(\beta) = \sum_j \widehat{p}_j(\beta)\sigma_j^2$, and $\widehat{\sigma}_t(\beta) = \sum_j \widehat{p}_j(\beta)\widehat{\sigma}_j^2$.

The following Theorem provides theoretical support on the uniform inference procedure for the estimator $\widehat{M}_t(\beta)$. Unless otherwise noted, $\sup_\beta$ is meant by $\sup_{\beta \in [\beta_l, \beta_u]}$.

Theorem 6.1. *Under conditions of Theorem B.6 and Assumption 14, we have*

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

To approximate the quantile, we have,

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

We now list the implementation steps for the uniform inference. Denote $\beta_v=[\beta_j]_j$ as a sequence of values selecting from each of the $\widehat{P}_{jt}$s (preferably one from each interval). To test

$$\mathsf{H}_0^{(0)}:M_t\left(\beta\right)=0.\quad\forall\beta\in[\beta_l,\beta_u],$$

we can utilize Theorem 6.1 which motivates the following algorithm.

Algorithm for uniform inference of $M_t(\beta)$

Step 1 Pick $[\beta_j]_j$. Calculate the residuals $\widehat{\varepsilon}_{it}$. Obtain $\widehat{\sigma}_t(\beta_v)$ (denoted as the vector with $\widehat{\sigma}_t(\beta)$ evaluated at the elementwise value of β_v, and $\widehat{M}_t(\beta_v)$ is defined accordingly).

Step 2 Simulate standard normal random vector $Z^{(s)}$ of the dimension $J_t\times 1$ for $s=1,\cdots,S$ times, where S is the number of bootstrap samples. Obtain the α quantile of $|Z^{(s)}|_\infty$ from the above sample, and we denote as $\widehat{q}_{\alpha,t}$.

Step 3 Create the confidence band $[\widehat{M}_t(\beta_v)-\widehat{\sigma}_t(\beta_v)^{1/2}\widehat{q}_{\alpha,t}\sqrt{J_t}/\sqrt{n_t},\widehat{M}_t(\beta_v)+\widehat{\sigma}_t(\beta_v)^{1/2}\widehat{q}_{\alpha,t}\sqrt{J_t}/\sqrt{n_t}]$. We cannot reject $H_0^{(0)}$ if 0 is contained in the confidence interval.

6.2 Test and theory on the difference estimator.

We shall analyse the difference estimator's theoretical properties for fixed t. In particular, we aim to test the following null hypothesis:

$$\mathsf{H}_0^{diff}: \mu_t(\beta_1) - \mu_t(\beta_2) - [\mu_t(\beta_2) - \mu_t(\beta_3)] = 0. \tag{6.1}$$

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 $|\widehat{M}_t\,(\beta_1) - \widehat{M}_t\,(\beta_2) - \left[\widehat{M}_t\,(\beta_2) - \widehat{M}_t\,(\beta_3)\right]| = \widetilde{M}_t(\beta_{1,2,3})$.

Algorithm for uniform inference concerning H_0^{diff}

Step 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 $\widehat{\sigma}_t(\beta_{1,2,3})$.

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

Step 3 Create the confidence band for $\mu_t(\beta_1) - \mu_t(\beta_2) - [\mu_t(\beta_2) - \mu_t(\beta_3)]$, i.e. $[\widetilde{M}_t(\beta_{1,2,3}) - \widehat{q}_{\alpha,t}\sqrt{J_t}/\sqrt{n_t}, \widetilde{M}_t(\beta_{1,2,3}) + \widehat{q}_{\alpha,t}\sqrt{J_t}/\sqrt{n_t}]$. We reject the null $\mathsf{H}_0^{(diff)}$ if $|\sqrt{n_t}B_{J_t}\widehat{M}_t/\sqrt{J_t}|_\infty > \widehat{q}_{\alpha,t}$.

6.3 Test algorithm and theory for $\mathsf{H}_0^c, \mathsf{H}_0^{(1)}$ and $\mathsf{H}_0^{(2)}$

As mentioned previously, testing whether the functional form of $M_t(\beta)$ falls to some specific parametric class is essential as we can associate the conclusion with the validity of the non-arbitrage assumption. This subsection discusses the theoretical results and an algorithm related to those hypotheses. In particular, we consider the cases where we test for the constancy and linearity of the function $M_t(\beta)$, namely the hypothesis H_0^c, $\mathsf{H}_0^{(1)}$ and $\mathsf{H}_0^{(2)}$. To test the constancy of $\widehat{M}_t(\beta)$, we adopt the following test statistics for H_0^c,

$$T_{c,t}(\beta) = |\widehat{p}_t(\beta)^\top \widehat{M}_t(\beta) - (\beta_u - \beta_l)^{-1} \int_{\beta_l}^{\beta_u} \widehat{p}_t(\beta)^\top \widehat{M}_t(\beta) d\beta | / [\widehat{\sigma}_t^{1/2}(\beta)]. \tag{6.2}$$

To be more specific, we can use the test statistics, $T_j^c = \widehat{M}_t(\beta_j) - J_t^{-1} \sum_{j=1}^{J_t} \widehat{M}_t(\beta_j)$. We let the matrix $A_{J_t} = I_{J_t} - 1/J_t \mathbf{1}_{J_t}$, where $I_{J_t}(J_t \times J_t)$ is an identity matrix, and $\mathbf{1}_{J_t}(J_t \times J_t)$ is a matrix of 1. The vector of T_j equals to $A_{J_t} Z_{t,T}$.

In the end we shall discuss the linear statistics. To test the null hypothesis of

$$\mathsf{H}_0^{(2)} : \exists \left(\ell_{0t}, \ell_{1t}\right) \ s.t. \ M_t\left(\beta\right) = \ell_{0t} + \ell_{1t}' \beta,$$

or

$$\mathsf{H}_0^{(1)} : \exists \ell_{1t} \ s.t. \ M_t\left(\beta\right) = \ell_{1t}' \beta.$$

Under $\mathsf{H}_0^{(1)}$ and $\mathsf{H}_0^{(2)}$, the model reduced to

$$R_{it} = \ell_{0t} + \ell_{1t}\beta_{it} + \varepsilon_{it}, \tag{6.3}$$

$$R_{it} = \ell_{1t}\beta_{it} + \varepsilon_{it}. \tag{6.4}$$

Recall that β_t, R_t are vectors of β_{it} and R_{it} respectively. And $\widehat{\beta}_t$ is denoted as a vector of $\widehat{\beta}_{it}$. Denote $\mathcal{B}_t$ as $(\mathbf{1}_{n_t}, \widehat{\beta}_t)$ $(n_t \times 2)$. The estimators are respectively $\widehat{M}_t(\beta)^{H^2} = \widehat{\ell}_{0t} + \widehat{\ell}_{1t}\beta = (1, \beta)(\mathcal{B}_t^\top \mathcal{B}_t)^{-1} \mathcal{B}_t R_t$ and $\widehat{M}_t(\beta)^{H^1} = \beta \widehat{\ell}_{1t} = \beta(\widehat{\beta}_t^\top \widehat{\beta}_t)^{-1} \widehat{\beta}_t^\top R_t$. Recall our nonparametric partition estimator $\widehat{M}_t(\beta) = \widehat{\mu}_t(\beta)$ which is defined in (2.6). Under either the two null hypothesis $\mathsf{H}_0^{(1)}$ or $\mathsf{H}_0^{(2)}$, we consider the following test statistics for the uniform inference respectively,

$$T_t^1 = \max_\beta(\sqrt{n_t/J_t})|\widehat{M}_t(\beta) - \widehat{M}_t(\beta)^{H^1} - M_t(\beta) + M_t(\beta)^{H^1}|/\widehat{\sigma}_t^{1/2}(\beta), \tag{6.5}$$

$$T_t^2 = \max_\beta(\sqrt{n_t/J_t})|\widehat{M}_t(\beta) - \widehat{M}_t(\beta)^{H^2} - M_t(\beta) + M_t(\beta)^{H^2}|/\widehat{\sigma}_t^{1/2}(\beta). \tag{6.6}$$

The theoretical results giving a valid critical value and confidence intervals are provided by Corollary B.11.2 in the Appendix B.13. Correspondingly, we shall list the following steps for the inference.

Algorithm for the uniform inference corresponding to H_0^c, ($\mathsf{H}_0^{(1)}$,$\mathsf{H}_0^{(2)}$)

Step 1 Pick $\beta_v = [\beta_j]_j$. Calculate the residuals $\widehat{\varepsilon}_{it}$. Obtain $\widehat{\sigma}_t(\beta_v)$.

Step 2 Simulate standard normal random variables $Z^{(s)}$ of $J_t \times 1$ dimension for $s = 1, \cdots, S$ times, where S is the number of bootstrap samples. Obtain the α quantile of $|Z^{(s)}|_\infty$ from the above sample, and denote it as $\widehat{q}_{\alpha,t}$.

Step 3 Create the confidence band $[T_j^c - \widehat{\sigma}_t(\beta_v)^{1/2}\widehat{q}_{\alpha,t}\sqrt{J_t}/\sqrt{n_t}, T_j^c + \widehat{\sigma}_t(\beta_v)^{1/2}\widehat{q}_{\alpha,t}\sqrt{J_t}/\sqrt{n_t}]$. (replace T_j^c by $\widehat{M}_t(\beta) - \widehat{M}_t(\beta)^{H^1}$ for $\mathsf{H}_0^{(1)}$ or replace T_j^c by $\widehat{M}_t(\beta) - \widehat{M}_t(\beta)^{H^2}$ for $\mathsf{H}_0^{(2)}$). We cannot reject the null if 0 is contained in the confidence interval.

7 Empirical Application

TO BE COMPLETED

8 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 nonparametric 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).

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*.

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.

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.

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.

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.

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.

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.

Hall, P., and C. C. Heyde (2014): *Martingale limit theory and its application*. Academic press.

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*.

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.

Rio, E. (2009): "Moment inequalities for sums of dependent random variables under projective conditions," *Journal of Theoretical Probability*, 22(1), 146–163.

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.

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

A Assumptions and Further Discussion

A.1 Assumptions and intermediate results

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. Define $w(t, t_0) = K((t - t_0)/Th)$, where $K(.)$ is a one sided kernel function taking 0 at positive values. With slight abuse of notation, b_{it_0} in the main text is denoted as $b_{i(-t_0)}$, and $b_{it_0} = \{\sum_{t=1}^T K((t - t_0)/(Th))X_t X_t^\top\}^{-1} \sum_{t=1}^T K((t - t_0)/(Th))X_t R_{it}$. Without loss of generality, we shall assume that $n_t \leq n_u$ 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 $\overline{B}_i(t_0) = \overline{B}_i(t_0) = \sum_{t=1}^T \mathbb{E}(X_t X_t^\top | \mathcal{F}_{t-1}) w(t, t_0) b_{it_0} = \tilde{A} b_{it_0}$. We define $r_T = (Th)^{-1}(T^{1/q} + (Th \log T)^{1/2})$, for an integer q.

Assumption 1. *We assume that the bounded 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 $\mathbb{E}(\varepsilon_{it}|\mathcal{F}_{t-1}) = 0$. We define $\mathbb{E}(\varepsilon^2_{it}|\mathcal{F}_{t-1}, x_t) = \sigma(\mathcal{F}_{t-1}, t/T)^2 = \sigma^2_t$, and define $\mathbb{E}x_t = 0$. $\|\varepsilon_{it}\|_{2q} \leq c$. $\mathbb{E}(\varepsilon^2_{it}|\mathcal{F}_{t-1}, x_{t-1}) = \mathbb{E}(\varepsilon^2_{it}|\mathcal{F}_{t-1})$. $\sigma^2_{\varepsilon,0} = \mathbb{E}(\varepsilon^2_{t_0}) = \mathbb{E}(\sigma^2_{t_0})$. The error term ε_{it_0} has finite qth moment with $q > 4$. $(T)^{2/q-1} n_u^{2/q} \ll h$.*

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

Assumption 4. *We define $\sigma^2_x = \mathbb{E}(x^2_t)$ and $\sigma_{x4} = \mathbb{E}(x^4_t)$, and we assume that they are both bounded by a positive constant $c_{x,q}$. We define the constant $c_x = 2\sigma^2_x \int_{-1}^0 K(s)^2 ds$. 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 .*

Remark A.1. *Note that under our assumptions we have that*

$$\mathbb{E}T^{-1} \sum_{t=1}^T x^2_t w(t, t_0) = \sigma^2_x T^{-1} \sum_{t=1}^T w(t, t_0) = \sigma^2_x. \tag{A.1}$$

And

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

$$Var(\sqrt{h/T}^{-1} \sum_{t=1}^T w(t, t_0)\varepsilon_{it} f_t) = \sigma^2_{\varepsilon,0}(\sigma^2_x + \tau'(t_0/T)^2) \int_{-1}^0 K(s)^2 ds + O(h). \tag{A.3}$$

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

Remark A.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 i.i.d. 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 $\Theta_{m,2q}(f_\cdot) = m^{-v}$ and $\max_t \|f_\cdot\|_{2q}$ is bounded, for $2v > 1/2 - 1/q$, with $q > 4$, for a positive constant v.*

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. $\mathbb{E}(\beta_{it}\Phi^*_{it}|\mathcal{G}_{t-1}) \asymp_p J_t^{-1}$, $\min_j \mathbb{E}_{t-1}(\Phi^*_{i,j,t}\varepsilon^2_{it}) \asymp_p J_t^{-1}$ and $\tilde{q}_j = \int_{F^{-1}_{\beta,t}(\kappa_{j-1})}^{F^{-1}_{\beta,t}(\kappa_j)} f_{\beta,t} d\beta \asymp_p J_t^{-1}$.*

Let $q_n = J_u \vee n_u \vee T$. Define $a_{nT} = \max_t(\sqrt{\delta_T} c_{n_uT}/\sqrt{n_t}) \vee \sqrt{\log T}\sqrt{n_t}^{-1}$, where c_{n_uT} is a large enough positive constant. $h \vee r_T \vee \sqrt{\log(q_n)}/\sqrt{Th} = \delta_T$.

Assumption 8. *We let $\max_t n_t \le n_u$, $n \le n_t$, $J \le J_t \le J_u$, $J_u \asymp J$ 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 $\Delta_T = (\delta_T + a_{nT})^{1/2}\sqrt{\log q_n}/\sqrt{n}$.

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.*

Remark A.3. *(Discussion of rate) We assume that q to be sufficiently large and the choice of h and n_u are satisfied so that*

$$\sqrt{\log(n_uT)} >> T^{1/q-1/2}h^{-1/2},$$

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

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

then $a_{nT} = \sqrt{h/n} + \sqrt{\log T}/\sqrt{n}$. Thus $\delta_T + a_{nT} = \sqrt{\log T}/\sqrt{n} \vee h \vee \sqrt{\log(n_uT)}/\sqrt{Th}$. Assumption 10 thus assumes that

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

In the following lemma, we show the uniform rate of the order statistics of $\widehat{\beta}_{it}$. It shall be noted that we condition on $\mathcal{G}_{t-1}$ and due to the Assumption 7, β_{it} will be conditionally independent and identically distributed.

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 A.4 (Rate of $\widehat{\beta}_{(k_j),t}$)**.** *Conditional on $\mathcal{G}_{t-1}$, we have under Assumptions 7 and 8,*

$$\max_{t,j}(\widehat{\beta}_{(k_{jt}),t} - F^{-1}_{\widehat{\beta},t}(\kappa_{jt})) \lesssim_{\mathbb{P}} a_{nT},$$

$$\max_{t,j}(\widehat{\beta}_{(k_{jt}),t} - F^{-1}_{\beta,t}(\kappa_{jt})) \lesssim_{\mathbb{P}} a_{nT} \vee \delta_T.$$

Recall that $\tilde{q}_{jt} = \int_{F^{-1}_{\beta,t}(\kappa_{j-1t})}^{F^{-1}_{\beta,t}(\kappa_{jt})} f_{\beta,t} d\beta \lesssim_{\mathbb{P}} J_t^{-1} \le J^{-1}$ due to Assumption 7. In the following, we show a few useful lemmas that facilitates further derivation. Denote $\tilde{\delta}_T = \sqrt{\log T}/\sqrt{n} \vee h \vee \sqrt{\log(n_uT)}/\sqrt{nTh}$.

Lemma A.5. *Conditional on $\mathcal{G}_{t-1}$, given Assumptions 7 to 9, we have,*

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

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

Also we have the bias term,

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

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

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

Lemma A.6. *Conditional on $\mathcal{G}_{t-1}$, under Assumptions 7-10, we have,*

$$H_1 = \quad n_t^{-1}\widehat{\Phi}_t\widehat{\beta}_t - n_t^{-1}\Phi_t^*\beta_t = \frac{1}{n_t}\sum_{i=1}^{n_t}(\widehat{\Phi}_{i,t}\widehat{\beta}_{it} - \Phi_{i,t}^*\beta_{it}) \lesssim_{\mathbb{P}} \delta + (\tilde{\delta}_T + a_{nT})/J = h_1, \tag{A.4}$$

$$H_2 = \quad n_t^{-1}\widehat{\beta}_t^\top\widehat{\beta}_t - n_t^{-1}\beta_t{}^\top\beta_t = \frac{1}{n_t}\sum_{i=1}^{n_t}(\widehat{\beta}_{it}^2 - \beta_{it}^2) \lesssim_{\mathbb{P}} \tilde{\delta}_T = h_2, \tag{A.5}$$

$$H_3 = \quad n_t^{-1}\widehat{\Phi}_t^\top\tilde{\varepsilon}_t - n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t = \frac{1}{n_t}\sum_{i=1}^{n_t}(\widehat{\Phi}_{i,t}\tilde{\varepsilon}_{it} - \Phi_{i,t}^*\tilde{\varepsilon}_{it}) \lesssim_{\mathbb{P}} \delta + (\tilde{\delta}_T + a_{nT})/J = h_1. \tag{A.6}$$

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 $q > 4$.*

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,

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

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 $\mathbb{N}_J = \bigcup_{t=1\cdots,T} N_{J_t}$. $\overline{\delta}_T = \sqrt{\log q_n}/\sqrt{nT} \vee h \vee \sqrt{\log(q_n)}/\sqrt{nTh}$.

Lemma A.7. *Under Assumptions 7-8 and 10-11,*

$$
\begin{aligned}
\tilde{H}_1 = \quad & \sup_z T^{-1}\sum_{t=1}^{T}\{n_t^{-1}\widehat{p}_t(z)^\top \widehat{\Phi}_t\widehat{\beta}_t - n_t^{-1}\widehat{p}_t(z)^\top \Phi_t^*\beta_t\} \\
& = \sup_z T^{-1}\sum_{t=1}^{T}\frac{1}{n_t}\sum_{i=1}^{n_t}\widehat{p}_t(z)^\top(\widehat{\Phi}_{i,t}\widehat{\beta}_{it} - \Phi_{i,t}^*\beta_{it}) \lesssim_{\mathbb{P}} \delta' \vee (a_{nT}/\sqrt{T} + \bar{\delta}_T)/J = h_1', \\
\tilde{H}_2 = \quad & T^{-1}\sum_{t=1}^{T} n_t^{-1}\{\widehat{\beta}_t^\top\widehat{\beta}_t - \beta_t{}^\top\beta_t\} = T^{-1}\sum_{t=1}^{T}\frac{1}{n_t}\sum_{i=1}^{n_t}(\widehat{\beta}_{it}^2 - \beta_{it}^2) \lesssim_{\mathbb{P}} \bar{\delta}_T, \\
\tilde{H}_3 = \quad & \sup_z T^{-1}\sum_{t=1}^{T}\{n_t^{-1}\widehat{p}_t(z)^\top \widehat{\Phi}_t^\top\tilde{\varepsilon}_t - n_t^{-1}\widehat{p}_t(z)^\top \Phi_t^*\tilde{\varepsilon}_t\} \\
& = \sup_z T^{-1}\sum_{t=1}^{T}\frac{1}{n_t}\sum_{i=1}^{n_t}\widehat{p}_t(z)^\top(\widehat{\Phi}_{i,t}\tilde{\varepsilon}_{it} - \Phi_{i,t}^*\tilde{\varepsilon}_{it}) \lesssim_{\mathbb{P}} \delta' \vee (a_{nT}/\sqrt{T} + \bar{\delta}_T)/(\sqrt{T}J) = h_3', \\
\tilde{H}_4 = \quad & |\max_{\mathbb{J}\in B_{\mathbb{J}}} T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n_t}(\mathbf{1}(\widehat{\beta}_{it}\in\widehat{P}_{jt}) - \mathbf{1}(\beta_{it}\in P_{jt}))| \lesssim_{\mathbb{P}} h_1'.
\end{aligned}
$$

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

Remark A.8. *Following the rate discussion of remark A.3, the above conditions can be implied by*

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

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

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. $\max_{j,t}\|\Phi_{i,j,t}^*\varepsilon_{it}\tilde{q}_j^{-1}\|_q \lesssim J^{1-1/q}$.*

Remark A.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,*

$$|\widehat{a}_t - a_t^* - \operatorname{diag}(\tilde{q}_{jt})^{-1}n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t|_\infty = o_{\mathbb{P}}(1). \tag{A.7}$$

We see that

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

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

Remark A.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., $\mathbb{E}[f_t|\mathcal{G}_{t-1}] = \zeta_t + \Psi_t^\top z_{t-1}^f$, 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*

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

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 A.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*

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

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

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

We can derive that

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

We now show that $\widehat{b}_{i(-t_0)}$ is close $\widehat{b}_{i(-t_0)}$ in a uniform sense. Thus Theorem 3.1 and 3.2 still hold under the same conditions.

Since we have $1/J_t * J_t^2(J_t^{-1}(\delta_T + a_{nT}) + \sqrt{\delta_T} T^{1/2q}(n_t^{-1/2} J_t^{-1/2})) = o_{\mathbb{P}}(\delta_2)$ 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 $c, C > 0$,*

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

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, $\sqrt{n_t/J_t}(h \vee J_t^{-1}) \to 0$.

Remark A.12. *Following the remark A.3,* $\delta_T + a_{nT} = \sqrt{\log T}/\sqrt{n} \vee h \vee \sqrt{\log(n_u T)}/\sqrt{Th}$*. The above assumption implies that*

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

and

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

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* $\varepsilon' = 1/9$*.*

B Proofs of Main Results

We first present some useful lemmas.

Lemma B.1 (Freedman's inequality)**.** *Let* $\xi_{a,i}$ *be a martingale difference sequence,* $\mathcal{F}_i$ *be the filtration and* $V_a = \sum_{i=1}^n \mathbb{E}(\xi_{a,i}^2|\mathcal{F}_{i-1})$ *and* $M_a = \max_{1\le l\le n}\sum_{i=1}^l \xi_{a,i}$*, we have,*

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

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 B.2 (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* $q > 2$ *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* $x \gtrsim \sqrt{n\log p}\Phi_{2,\varsigma} + n^{1/q}(\log p)^{3/2}\||X_\cdot|_\infty\|_{q,\varsigma}$*,*

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

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

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

Lemma B.3 (Uniform rate)**.** *By Assumptions 1- 6, we have*

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

$$\sup_{\lfloor Th\rfloor \le t_0 \le T-\lfloor Th\rfloor} T^{-1}|\tilde{A}(t_0) - \mathbb{E}\tilde{A}(t_0)|_{\max} \lesssim_{\mathbb{P}} (Th)^{-1}(T^{1/q} + (Th\log T)^{1/2}), \tag{B.3}$$

$$\max_i \sup_{\lfloor Th\rfloor \le t_0 \le T-\lfloor Th\rfloor} T^{-1}|\tilde{B}_i(t_0) - \mathbb{E}B_i(t_0)|_{\max} \lesssim_{\mathbb{P}} (Th)^{-1}(T^{1/q} + (Th\log T)^{1/2}), \tag{B.4}$$

$$\sup_{\lfloor Th\rfloor \le t_0 \le T-\lfloor Th\rfloor} T^{-1}|A(t_0) - \tilde{A}(t_0)|_{\max} \lesssim_{\mathbb{P}} \sqrt{\log T}/(\sqrt{T}\sqrt{h}), \tag{B.5}$$

$$\max_i \sup_{\lfloor Th\rfloor \le t_0 \le T-\lfloor Th\rfloor} T^{-1}|B_i(t_0) - \tilde{B}_i(t_0)|_{\max} \lesssim_{\mathbb{P}} \sqrt{\log(n_u T)}/(\sqrt{T}\sqrt{h}). \tag{B.6}$$

Lemma B.4 (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 $\Theta_{m,2q}(u_.) < C_v$, and $\max_t \|u_t\|_{2q} < C_v$ for $q > 4$. $\Theta_{m,2q}(v_.) < C_v$. Assume $\Theta_{m,2q}(v_{i.}) < C_v$, and $\max_t \|v_{it}\|_{2q} < M$ for $q > 4$. $\Theta_{m,2q}(v_.) < C_v$.*

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

$$\sup_{\lfloor Th \rfloor \leq t_0 \leq T - \lfloor Th \rfloor} T^{-1} | \sum_{t=1}^{T} w(t, t_0) u_t | \lesssim_{\mathbb{P}} (Th)^{-1/2} (\log T)^{1/2}, \tag{B.7}$$

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

$$\max_i \sup_{\lfloor Th \rfloor \leq t_0 \leq T - \lfloor Th \rfloor} T^{-1} | \sum_{t=1}^{T} w(t, t_0) v_{it} | \lesssim_{\mathbb{P}} (Th)^{-1/2} (\log T)^{1/2} \tag{B.8}$$

Lemma B.5 (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*

$$\|M_n\|_q^{q'} \leq K_q^{q'} \sum_{t=1}^{n} \|\xi_t\|_q^{q'} \quad \text{where} \quad K_q = \max((q-1)^{-1}, \sqrt{q-1}).$$

B.1 Proof of Theorem 3.1

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

$$\sup_{t_0} |\mathbb{E}[T^{-1} \sum_{t=1}^{T} \mathbb{E}[X_t X_t^\top | \mathcal{F}_{t-1}] C_\beta w(t, t_0)]|_\infty \leq C, \tag{B.9}$$

where C is positive constant only depend on c_τ,C_β, and $c_{x,q}$.
Moreover, $|T^{-1} \sum_{t=1}^{T} \{\mathbb{E}[X_t X_t^\top | \mathcal{F}_{t-1}] - \mathbb{E}(X_t X_t^\top)\} w(t, t_0)|_\infty \lesssim_{\mathbb{P}} (Th)^{-1/2} \vee (Th)^{-1+1/q}$, by Assumption 6, and theorem 2 of Wu and Wu (2016). The $\lesssim_{\mathbb{P}}$ depends on c_θ. And similar uniform arguments as in Lemma B.3, $\sup_{t_0} |T^{-1} \sum_{t=1}^{T} \{\mathbb{E}[X_t X_t^\top | \mathcal{F}_{t-1}] - \mathbb{E}(X_t X_t^\top)\} w(t, t_0)|_\infty \lesssim_{\mathbb{P}} r_T$. Thus we know $\max_i \sup_{t_0} T^{-1} |\tilde{B}_i - \overline{B}_i|_\infty = O_{\mathbb{P}}(h)$, provided that $r_T = o_{\mathbb{P}}(1)$. Also we have

$$\max_i \sup_{t_0} |\tilde{A}^{-1} (\tilde{B}_i - \overline{B}_i)|_\infty \lesssim |\tilde{A}^{-1} \tilde{A}|_\infty \max_i |T^{-1} \sum_{t_0 - \lfloor Th \rfloor \leq t \leq t_0 - 1} (b_{it} - b_{i(t-1)})|_\infty \leq h C_\beta. \tag{B.10}$$

Then we look the following term,

$$A^{-1} B_i - \tilde{A}^{-1} \tilde{B}_i = -A^{-1} (A - \tilde{A}) \tilde{A}^{-1} B_i + A^{-1} (B_i - \tilde{B}_i). \tag{B.11}$$

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 $c_x = 2\sigma_x^2 \int_{-1}^{0} K(s)^2 ds$. $s_{f1} = \sum_{t=1}^{T} (f_t - \mathbb{E}(f_t | \mathcal{F}_{t-1})) w(t, t_0)$, and $s_{f2} = \sum_{t=1}^{T} (f_t^2 - \mathbb{E}(f_t^2 | \mathcal{F}_{t-1})) w(t, t_0)$.

$$A - \tilde{A} = [0, s_{f1}; s_{f1}, s_{f2}]. \tag{B.12}$$

Following Freedman's inequality as in Lemma B.1 and similar argument as in Lemma B.3 in the Appendix, we have,

$$\sup_{t_0} |\sum_{t=1}^{T} (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))w(t, t_0)|| \lesssim_{\mathbb{P}} c_T \sqrt{c_x} \sqrt{T \log T} / \sqrt{h}, \tag{B.13}$$

for some large enough constant c_T.

$\sup_{t_0} s_{f2} \lesssim_{\mathbb{P}} \tau'(t_0/T)\sqrt{T \log T/h}\sqrt{c_x} c_T + \sqrt{T \log T/h}\sqrt{\mathbb{E}(x_t^4) \int_{-1}^{0} K(s)^2 ds} c_T$ as derived in Lemma B.3. Also

$$\sup_{t_0} T^{-1} |A - \tilde{A}|_\infty \lesssim_{\mathbb{P}} c' \sqrt{\log T/(Th)} = r_{AT}, \tag{B.14}$$

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

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 Asssumption 2). We can see that similarly for $B_i - \tilde{B}_i$, we have,

$$B_i - \tilde{B}_i = \sum_{t=1}^{T} X_t \varepsilon_{it} w(t, t_0) + \sum_{t=1}^{T} (X_t X_t^\top - \mathbb{E}_{t-1} X_t X_t^\top) w(t, t_0) b_{it}. \tag{B.15}$$

$\sum_{t=1}^{T} X_t \varepsilon_{it} w(t, t_0)$ 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 B.1 and B.3, we have

$$\max_i \sup_{t_0} |\sum_{t=1}^{T} \varepsilon_{it} w(t, t_0)| \lesssim_{\mathbb{P}} c_T \sqrt{\sigma_{\varepsilon,0}^2 \frac{T \log(nT)}{h} \int_{-1}^{0} K^2(s) ds}. \tag{B.16}$$

And according to Assumption 2, Lemma B.1 and B.3, we have $\mathbb{E}(\varepsilon_{it}^2|\mathcal{F}_{t-1}, x_t) = \sigma_\varepsilon^2(t/T, \mathcal{F}_{t-1})$, and $\mathbb{E} x_t = 0$.

$$\max_i \sup_{t_0} |\sum_{t=1}^{T} f_t \varepsilon_{it} w(t, t_0)| \lesssim_{\mathbb{P}} c_T \{2\sigma_{\varepsilon,0} \tau'(t_0/T) \vee \sigma_x \sigma_{\varepsilon,0}\} \sqrt{\int_{-1}^{0} K^2(s) ds} \sqrt{T \log(nT)/h}. \tag{B.17}$$

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

$$\max_i \sup_{t_0} |T^{-1} \sum_{t=1}^{T} (X_t X_t^\top - \mathbb{E}_{t-1} X_t X_t^\top) w(t, t_0) b_{it} - (A - \tilde{A}) b_{it_0}|_{\max} \le \sup_{t_0} |A - \tilde{A}|_{\max} h = O_{\mathbb{P}}(r_{AT} h) \tag{B.18}$$

by Assumption 5.

Thus we have

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

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 $\mathbb{P}(|TA^{-1}|_2 \le \lambda_{\min}(T^{-1}A)^{-1} \le c) \le \mathbb{P}(\lambda_{\min}(T^{-1}A) \ge c^{-1})$. 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 B.3.

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

□

B.2 Proof of Theorem 3.2

Proof. We now look at the asymptotic properties,

$$\mathbb{E} T^{-1} \sum_{t=1}^{T} x_t^2 w(t, t_0) = \sigma_x^2 T^{-1} \sum_{t=1}^{T} w(t, t_0) = \sigma_x^2. \tag{B.19}$$

And

$$T^{-1}\sum_{t=1}^{T}\tau'(t/T)^2 w(t,t_0) = \tau'(t_0/T)^2 + O(h). \tag{B.20}$$

$$\mathrm{Var}(\sqrt{h/T}^{-1}\sum_{t=1}^{T} w(t,t_0)\varepsilon_{it} f_t) = \sigma_{\varepsilon,0}^2(\sigma_x^2 + \tau'(t_0/T)^2)\int_{-1}^{0} K(s)^2 ds + O(h). \tag{B.21}$$

Thus we denote $T^{-1}\mathbb{E}(A) = \Sigma_x + \tilde{\tau}(t_0/T)\tilde{\tau}(t_0/T)^\top + O(h)$. Since we have already proved that $\max_i \sup_{t_0} |\tilde{A}^{-1}(\tilde{B}_i - \overline{B}_i)|_\infty \le hC_\beta$. We just need look at the term,

$$\begin{aligned}
&(\widehat{b}_{it_0} - b_{it_0})\\
&= A^{-1}B_i - \tilde{A}^{-1}\tilde{B}_i = -A^{-1}(A-\tilde{A})\tilde{A}^{-1}B_i + A^{-1}(B_i - \tilde{B}_i) + O_{\mathbb{P}}(h),\\
&= -A^{-1}(A-\tilde{A})\tilde{A}^{-1}B_i + A^{-1}(B_i - \tilde{B}_i) + O_{\mathbb{P}}(h),\\
&= -A^{-1}(A-\tilde{A})\tilde{A}^{-1}(B_i - \tilde{A}b_{it_0}) - A^{-1}\{(A-\tilde{A})b_{it_0} - (B_i - \tilde{B}_i)\} + O_{\mathbb{P}}(h),\\
&= I_{11} + I_{12} + O_{\mathbb{P}}(h).
\end{aligned}$$

From the proof of Theorem 3.1, we have $|I_{11}|_2 \lesssim_{\mathbb{P}} r_{AT}h$. $I_{12} = -A^{-1}\sum_{t=1}^{T} w(t,t_0)\varepsilon_{it}X_t + O_{\mathbb{P}}(r_{AT}h)$. Thus we shall apply a martingale central limit theorem on the term $-(T^{-1}A)^{-1}\sqrt{h/T}\sum_{t=1}^{T} w(t,t_0)\varepsilon_{it}X_t$, which correspond to the leading term of $\sqrt{Th}(\widehat{b}_{it_0} - b_{it_0})$. We shall prove that it is close to $-(T^{-1}\mathbb{E}(A))^{-1}\sqrt{h/T}\sum_{t=1}^{T} w(t,t_0)\varepsilon_{it}X_t$.

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

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

Thus $\sqrt{T/h}|-A^{-1}\sum_{t=1}^{T} w(t,t_0)\varepsilon_{it}X_t - \mathbb{E}(A)^{-1}\sum_{t=1}^{T} w(t,t_0)\varepsilon_{it}X_t|_2 \lesssim_{\mathbb{P}} r_{AT} \vee r_T$. As $O_{\mathbb{P}}((r_{AT}\vee r_T \vee h) = o_{\mathbb{P}}(1)$, by the assumption of this theorem. Then we have the elements of $\sqrt{Th}\Sigma_b^{-1/2}(\widehat{b}_{it_0} - b_{it_0}) = \sqrt{h/T}\Sigma_b^{-1/2}\Sigma_A^{-1}\sum_{t=1}^{T} X_t w(t,t_0)\varepsilon_{it} + O_{\mathbb{P}}(1)$ as a martingale difference with respect to $\mathcal{F}_{t-1}$ by Assumption 2.
The rest of proof follows from Theorem 3.5 in Hall and Heyde (2014).

□

B.3 Theorem B.6 and the Proof

Theorem B.6. *Conditional on $\mathcal{G}_{t-1}$, under Assumptions 7–10,*

$$|\widehat{a}_t - a_t^* - \mathrm{diag}(\tilde{q}_{jt})^{-1} n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t|_\infty = o_{\mathbb{P}}(1). \tag{B.22}$$

Theorem B.6 is the main building block for our study of the grand mean, $T^{-1}\sum_{t=1}^{T}\widehat{\mu}_t(\beta)$, since $\widehat{\mu}_t(\beta) = \widehat{p}_t(\beta)^\top\widehat{a}_t$. In our proofs we will work with differences of the form $T^{-1}\sum_{t=1}^{T}\{\widehat{p}_t(\beta)^\top\widehat{a}_t - \widehat{p}_t(\beta)^\top a_t^*\}$, as discussed further in the following remark.

Remark B.7. *Let $p_{jt}(\beta) = \mathbf{1}(F_{\beta,t}^{-1}((j-1)/J_t) \le \beta < F_{\beta,t}^{-1}((j)/J_t))$. It may appear more natural to utilize $p_t(\beta)$*

rather than $\widehat{p}_t(\beta)$ in the linearization presented in Theorem B.6. However, note that

$$
\begin{aligned}
&|\mathbf{1}(\beta \in \widehat{P}_{jt}) - \mathbf{1}(\beta \in P^*_{jt})| \\
&\quad = \mathbf{1}(\beta \in [\min\{F^{-1}_{\beta,t}((j-1)/J_t, \widehat{\beta}_{(\lfloor n_t(j-1)/J_t \rfloor)t}\}, \max\{F^{-1}_{\beta,t}((j-1)/J_t), \widehat{\beta}_{(\lfloor n_t(j-1)/J_t \rfloor)t}\})) \\
&\qquad + \mathbf{1}(\beta \in [\min\{F^{-1}_{\beta,t}(j/J_t), \widehat{\beta}_{(\lfloor n_t j/J_t \rfloor)t}\}, \max\{F^{-1}_{\beta,t}(j/J_t), \widehat{\beta}_{(\lfloor n_t(j)/J_t \rfloor)t}\})).
\end{aligned}
$$

Thus, around shrinking regions of $F^{-1}_{\beta,t}(j/J_t)$, uniform convergence fail for $\widehat{p}_t(\beta)$ to $p_t(\beta)$. To study the asymptotic properties over the whole support, we always work with the random evaluation point $\widehat{p}_t(\beta)$ instead.

Proof. We see that

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

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

$$
\widehat{a}_t - a^*_t \quad = \quad C^{-1}_{nt}[\widehat{\Phi}_t\Phi^*_t - C_{nt}]a^*_t + C^{-1}_{nt}D_{nt} + C^{-1}_{nt}[\widehat{\Phi}_t\flat_t]. \tag{B.23}
$$

By equation (B.23), we have that

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

Step 1 We show the rate of $|\{\widehat{\Phi}_t\widehat{\Phi}_t^\top\}^{-1}\{\widehat{\Phi}_t\tilde{\varepsilon}_t\} - (\Phi^*_t\Phi^{*\top}_t)^{-1}\Phi^*_t\tilde{\varepsilon}_t|_{\max} = o_{\mathbb{P}}(1)$.

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

We first check the closeness of the first component which is denoted $C^{-1}_{n_t}D_{n_t} - \tilde{C}^{-1}_{n_t}\tilde{D}_{n_t} = -C^{-1}_{n_t}(C_{n_t} - \tilde{C}_{n_t})\tilde{C}^{-1}_{n_t}D_{n_t} + \tilde{C}^{-1}_{n_t}(D_{n_t} - \tilde{D}_{n_t})$. 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 $|C^{-1}_{n_t}D_{n_t} - \tilde{C}^{-1}_{n_t}\tilde{D}_{n_t}|_{\max} \le |C^{-1}_{n_t}|_{\max}|(C_{n_t} - \tilde{C}_{n_t})|_{\max}|\tilde{C}^{-1}_{n_t}|_{\max}|D_{n_t}|_{\max} + |\tilde{C}^{-1}_{n_t}|_{\max}|(D_{n_t} - \tilde{D}_{n_t})|_{\max} \ll 1$. Recall the definition of h_1, h_2 in Lemma A.6. Define $h_4 = 1/J_t^2 \vee c_{n_t}/(J_t\sqrt{n_t J_t}) \vee (c_{n_t}/\sqrt{n_t J_t})^2$, $\tilde{h}_4 = (c_{n_t}\sqrt{\log J_t}/\sqrt{n_t J_t}) \vee J_t^{-1}$. 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 A.6, we have

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

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 A.6,

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

From the rate in the Remark A.3 and Assumption 9 the above two conditions are ensured. By Bernstein inequality, we have $\max_j |[n_t^{-1}\Phi^*_t\beta_t]_j - \tilde{q}_{jt}| \lesssim_{\mathbb{P}} c_{n_t}\sqrt{\log J_t}/\sqrt{n_t J_t}$, where c_{n_t} is a positive constant.

Step 2 Recall that we denote $\tilde{q}_{jt} = \int_{F_{\beta,t}^{-1}(\kappa_{j-1})}^{F_{\beta,t}^{-1}(\kappa_j)} f_{\beta,t} d\beta \lesssim_{\mathbb{P}} J_t^{-1}$. We have by Bernstein inequality, $\max_j |n_t^{-1} \sum_{t=1}^T \mathbf{1}(\beta_{it} \in P_{jt}) - \tilde{q}_{jt}| \lesssim_{\mathbb{P}} c_{n_t} \sqrt{\log J_t}/(\sqrt{n_t J_t})$. Therefore we have by assumption 9,

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

We then show that the bias term $\{\widehat{\Phi}_t \widehat{\Phi}_t^\top\}^{-1}\{\widehat{\Phi}_t \flat_t\}$ 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 $|\flat_t|_\infty \lesssim_{\mathbb{P}} 1/J_t$(c.f. Lemma A.5.). Thus by similar steps from the previous derivation we have that $|\{\widehat{\Phi}_t \widehat{\Phi}_t^\top\}^{-1}\{\widehat{\Phi}_t \flat_t\}|_{\max} \lesssim_{\mathbb{P}} 1/J$.

$|C_{nt}^{-1}[\widehat{\Phi}_t \Phi_t^* - C_{nt}] a_t^*|_{\max} <<_p 1$ by the assumption 10.

Thus the conclusion holds.

□

B.4 Proof of Theorem 4.1

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,

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

by the assumption 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 B.1 and by Assumption 8,

$$|T^{-1}\sum_t^T \widehat{p}_t(\beta)^\top [\mathrm{diag}(\tilde{q}_{jt})^{-1} n_t^{-1}\Phi_t^* \varepsilon_t|_{\max} \lesssim_{\mathbb{P}} \frac{1}{\sqrt{T}} J(\sqrt{\log(n_a J)}/(\sqrt{nJ})) \lesssim \frac{1}{\sqrt{J}\sqrt{T}} \ll 1/\sqrt{T}, \quad \text{(B.24)}$$

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

$$\begin{aligned} &|T^{-1}\sum_t^T \widehat{p}_t(\beta)^\top [\mathrm{diag}(\tilde{q}_{jt})^{-1} n_t^{-1}\Phi_t^* \varepsilon_t^1|_{\max} = |T^{-1}\sum_{t=1}^T (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))[n_t^{-1}\sum_{i=1}^{n_t} \mathrm{diag}(\tilde{q}_{jt})^{-1}\beta_{it}\Phi_{it}^*]_{jt}|_{\max} \\ &\lesssim_{\mathbb{P}} J\{\frac{1}{\sqrt{T}}(1/J + \sqrt{\log(n_u J_a)}/\sqrt{nJ}) \le \frac{1}{\sqrt{T}} + \sqrt{TJ}^{-1} \le \frac{1}{\sqrt{T}}. \end{aligned}$$

Therefore by Assumption 9, the above object has rate $\frac{1}{\sqrt{T}}$, then $J\frac{1}{\sqrt{T}}(J^{-2} + \sqrt{\log(n_u J_a)}/(J\sqrt{nJ}) + (\sqrt{\log(n_u J_a)}/\sqrt{nJ})^2) \ll \frac{1}{\sqrt{T}}$. The leading term $T^{-1}\sum_{t=1}^T (f_t - \mathbb{E}f_t|\mathcal{F}_{t-1}) n_t^{-1}\sum_{i=1}^{n_t}\beta_{it}\Phi_{it}^* \lesssim_{\mathbb{P}} \frac{1}{\sqrt{T}}$.

First of all, we have that,

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

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

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

We will evaluate $[\max_{\mathbb{J}\in B_{\mathbb{J}}} T^{-1}\sum_{t=1}^T\{\widehat{a}_{jt} - a^*_{jt}\}]$ in the upcoming derivation. By Assumption 11, we have $\sqrt{a_{nT}}/\sqrt{T} + \overline{\delta}_T\sqrt{\log(nJ)}/\sqrt{n_t} \ll J^{-1}$.

Now we have

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

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

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

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

Thus by Lemma A.7, we have

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

Similarly by Lemma A.7, we have

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

Therefore

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

For the second term, we have

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

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

Moreover, by equation (B.24), $|I_3| = O_{\mathbb{P}}(\frac{1}{\sqrt{T}})$.

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

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

We know from Lemma A.5 that $\max_t |\flat_t|_{\max} \lesssim_{\mathbb{P}} 1/J$. We have that $|T^{-1}\sum_{t=1}^T \widehat{p}_t(\beta)^\top C_{n_t}^{-1}(C_{n_t} - \tilde{C}_{n_t})\tilde{C}_{n_t}^{-1}E_{n_t} \ll_p J^2(J\sqrt{T})^{-1}(J^{-1}\vee\sqrt{\log q_n}/n)1/J \lesssim J^{-1}\frac{1}{\sqrt{T}}$. Also $|T^{-1}\sum_{t=1}^T \widehat{p}_t(\beta)^\top \tilde{C}_{n_t}^{-1}(E_{n_t} - \tilde{E}_{n_t})| \lesssim_{\mathbb{P}} J(\delta'\vee(a_{nT}+\tilde{\delta}_T)/J)(1/J) \lesssim J^{-1}\wedge\frac{1}{\sqrt{T}}$. Let $\mathbf{1}_{n_t}$ be a $n_t\times 1$ vector of ones. For the term $T^{-1}\sum_{t=1}^T \widehat{p}_t(\beta)^\top (n_t^{-1}\Phi_t^*\Phi_t^{*\top})^{-1}(n_t^{-1}\Phi_t^*\flat_t)$, we have $|\max_{\mathbb{J}\in B_{\mathbb{J}}} T^{-1}\sum_{t=1}^T [((n_t^{-1}\Phi_t^*\Phi_t^{*\top})^{-1}(n_t^{-1}\Phi_t^*\mathbf{1}_{n_t})]_{jt}| \lesssim_{\mathbb{P}} 1$.
Thus we have $T^{-1}\sum_{t=1}^T \widehat{p}_t(\beta)^\top (n_t^{-1}\Phi_t^*\Phi_t^{*\top})^{-1}(n_t^{-1}\Phi_t^*\flat_t) \lesssim_{\mathbb{P}} J^{-1}$. It follows that $|I_4| \lesssim_{\mathbb{P}} J^{-1}$.

$\square$

B.5 Proof of Theorem 4.3

Proof. Therefore the leading term $\{n_t^{-1}\widehat{p}_t(\beta)^\top \Phi_t^*\Phi_t^{*\top}\}^{-1}(n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t)$ is close to $[\tilde{q}_{jt}^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\Phi_{t,i,j}^*\tilde{\varepsilon}_t]_j$. We let $\varepsilon_{it}^1 = \beta_{it}(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))$. Recall that we define $E_{n_t,j} = \tilde{q}_{jt}^{-1}\mathbb{E}(\Phi_{i,j,t}^*\beta_{it}|\mathcal{G}_{t-1})$. Then we have

$$
\begin{aligned}
& [\frac{1}{\sqrt{T}}\sum_{t=1}^T \tilde{q}_{jt}^{-1} n_t^{-1}\Phi_t^*\tilde{\varepsilon}_t]_j \\
= & [\frac{1}{\sqrt{T}}\sum_{t=1}^T \tilde{q}_{jt}^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\Phi_{i,j,t}^*\varepsilon_{it}^1 + \frac{1}{\sqrt{T}}\sum_{t=1}^T \tilde{q}_{jt}^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\Phi_{i,j,t}^*\varepsilon_{it}] \\
= & [\frac{1}{\sqrt{T}}\sum_{t=1}^T (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))\tilde{q}_{jt}^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\Phi_{i,j,t}^*\beta_{it} + \frac{1}{\sqrt{T}}\sum_{t=1}^T \tilde{q}_{jt}^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\Phi_{i,j,t}^*\varepsilon_{it}] \\
= & \frac{1}{\sqrt{T}}\sum_{t=1}^T (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))\tilde{q}_{jt}^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\Phi_{i,j,t}^*\beta_{it} - E_{n_t,j}) + \frac{1}{\sqrt{T}}\sum_{t=1}^T \tilde{q}_{jt}^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\Phi_{i,j,t}^*\varepsilon_{it} \\
& + \frac{1}{\sqrt{T}}\sum_{t=1}^T (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j} \\
= & O_{\mathbb{P}}(J\sqrt{\log q_n}/\sqrt{nJ}) + O_{\mathbb{P}}(\sqrt{J}/\sqrt{n}) + \frac{1}{\sqrt{T}}\sum_{t=1}^T (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j} \\
= & \frac{1}{\sqrt{T}}\sum_{t=1}^T (f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j} + o_{\mathbb{P}}(1).
\end{aligned}
$$

Thus the final leading term is $\frac{1}{\sqrt{T}}\sum_{t=1}^T(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta)E_{n_t,j}$. We shall use Theorem 3.5 in Hall and Heyde (2014), with η therein as 1. The following two assumptions are needed to be verified.

i) $\mathbb{E}|1/T\sum_{t=1}^T \mathbb{E}(\Sigma_{\alpha(\beta)}^{-1}(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))^2\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta)E_{n_t,j}^2|\mathcal{F}_{t-1}) - 1| \to 0$,

ii) $1/T\max_t \mathbb{E}(\Sigma_{\alpha(\beta)}^{-1}(f_t - \mathbb{E}(f_t|\mathcal{F}_{t-1}))^2\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta)E_{n_t,j}^2|\mathcal{F}_{t-1}) \to_{\mathbb{P}} 0$.

The above two conditions is trivial to hold by Assumption 6, and the desired results follows. ☐

B.6 Proof of Theorem 6.1

Recall that $\widehat{\sigma}_j^2 = n_t/J_t(\sum_{i=1}^{n_t}\widehat{\Phi}_{i,j,t}\widehat{\varepsilon}_{it}^2)(\sum_{i=1}^{n_t}\widehat{\Phi}_{i,j,t})^{-2}$.
We see that $\sigma_j^2 = 1/J_t(\mathbb{E}_{t-1}(\Phi_{i,j,t}^*\varepsilon_{it}^2))(\mathbb{E}_{t-1}(\Phi_{i,j,t}^*))^{-2}$. Recall that $\mathbb{E}_{t-1}(\Phi_{i,j,t}^*) = \tilde{q}_{jt}$.
Recall that $\sigma_t(\beta) = \sum_j \widehat{p}_j(\beta)\sigma_j^2$, and $\widehat{\sigma}_t(\beta) = \sum_j \widehat{p}_j(\beta)\widehat{\sigma}_j^2$.

For the test statistics $\sqrt{n_t}/\sqrt{J_t}\{\widehat{M}_t(\beta) - M_t(\beta)\}/\widehat{\sigma}_t(\beta)^{1/2}$.

1 We show that the leading term $Z_{n_t}(\beta) = \sqrt{n_t}/\sqrt{J_t}\{(\widehat{M}_t(\beta) - M_t(\beta))\}/\widehat{\sigma}_t(\beta)^{1/2}$ is $\sum_j \widehat{p}_j(\beta)\sqrt{1/(n_tJ_t)}\tilde{q}_j^{-1}\sum_{i=1}^{n_t}\Phi_{i,j,t}^*\varepsilon_{it}/\sigma_t(\beta)^{1/2}$. Namely, we have

$$
\begin{aligned}
& \sup_\beta \sqrt{n_t}/\sqrt{J_t}(\widehat{M}_t(\beta) - M_t(\beta))/\widehat{\sigma}_t(\beta)^{1/2} = \sup_\beta \sqrt{n_t}/\sqrt{J_t}\widehat{p}_t^\top(\beta)\{\mathrm{diag}[\tilde{q}_{jt}]\sigma_t(\beta)^{1/2}\}^{-1}\{n_t^{-1}\Phi_t^*\varepsilon_t\} \\
& + O_{\mathbb{P}}(\sqrt{n_t}/(\sqrt{J_t}J_t) \vee \sqrt{n_t}h/\sqrt{J_t} \vee 1/\sqrt{J_t}).
\end{aligned}
$$

2 We need to assume that ε_{it} are conditional i.i.d. 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 $\sum_j \widehat{p}_j(\beta)\sqrt{1/(n_tJ_t)}\tilde{q}_j^{-1}\sum_{i=1}^{n_t}\Phi_{i,j,t}^*(\varepsilon_{it} - \sigma_t\eta_{it})/\sigma_t^{1/2}(\beta) = o_{\mathbb{P}}(n_t^{-1/2+1/(2q)}\sqrt{J_t})$ is of small order by applying a Komlós Major Tusnády (KMT) strong approximation argument. We let $\sigma_{t-1}^2 = \mathbb{E}(\varepsilon_{it}^2|\mathcal{F}_{t-1})$, and η_{it} is i.i.d. 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 $\sum_j \widehat{p}_j(\beta)\sqrt{1/(n_t J_t)}\tilde{q}_j^{-1}\sum_{i=1}^{n_t}\Phi^*_{i,j,t}\sigma_t\eta_{it}/\sigma_t^{1/2}(\beta)$ which is conditional Gaussian distributed and is close enough to a Gaussian random variable $Z_t(\beta)$. We define $Z_t(\beta)=\sum_j \widehat{p}_j(\beta)Z_j$, where Z_js are standard Gaussian random variable. The argument is due to Gaussian maximal inequalities.

B.7 Strong approximation for finite points

We shall provide results on approximating the maximum over a finite number of points. 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 discuss how to make uniform inference on $T^{-1}\sum_{t=1}^T \mu_t(\beta)$ by applying a strong approximation of the leading term $T^{-1}\sum_{t=1}^T\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta)(f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1}))E_{n_t,j}$ of $T^{-1}\sum_{t=1}^T\{\widehat{\mu}_t(\beta)-\mu_t(\beta)\}$; 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 $\widehat{\mu}(\beta)-\mu(\beta)$ is dominated by $T^{-1}\sum_{t=1}^T\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta)(f_t-\mathbb{E}(f_t))E_{n_t,j}$ though.

Lemma B.8. *We define $\mathbb{E}_{\mathbb{J}_1}=(E_{n_t,j})_t$, where $j\in\mathbb{J}_1$. And $\mathbb{E}_{\mathbb{J}_2}=(E_{n_t,j})_t$ $(T\times 1)$, where $j\in\mathbb{J}_2$. $\Sigma_f=\operatorname{diag}(Var(f_t-\mathbb{E}(f_t|\mathcal{F}_{t-1})))$. 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 as $C_{j,j'}$. $C_{\mathbb{J},\mathbb{J}}=\operatorname{diag}(\tilde{\Sigma})$. Assume that $\tilde{Z}_{\mathbb{J}}$ follows a normal distribution with $N(0,\operatorname{diag}(\tilde{\Sigma})^{-1/2}\tilde{\Sigma}\operatorname{diag}(\tilde{\Sigma})^{-1/2})$. Under the conditions of Theorem B.6, 4.1 and $\sqrt{T}(1/J\vee h)\to 0$, we have*

$$\sup_x|\mathbb{P}(\max_{\mathbb{J}\in B_{\mathbb{J}}}|T^{-1/2}\sum_{t=1}^T(C_{\mathbb{J},\mathbb{J}})^{-1/2}(\widehat{a}_{j,t}-a^*_{j,t})|\geq x)-\mathbb{P}(\max_{\mathbb{J}\in B_{\mathbb{J}}}|\tilde{Z}_{\mathbb{J}}|\geq x)|\to 0. \tag{B.25}$$

B.8 Strong approximation for high-minus-low statistics

To test the hypothesis (5.4),
recall that we define, $\mathcal{T}_T=[T^{-1}(\sum_{t=1}^T\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta^{**})(\widehat{a}_{j,t}-a^*_{j,t}))-T^{-1}(\sum_{t=1}^T\sum_{j=1}^{J_t}\widehat{p}_{jt}(\beta^*)(\widehat{a}_{j,t}-a^*_{j,t}))]/(\widehat{\sigma}(\beta^*)+\widehat{\sigma}(\beta^{**})-2\widehat{\sigma}(\beta^*,\beta^{**}))^{1/2}$. We let $\widehat{\tilde{\sigma}}^{1/2}_{\mathtt{FM}}(\beta^*,\beta^{**})=(\widehat{\sigma}_{\mathtt{FM}}(\beta^*)+\widehat{\sigma}_{\mathtt{FM}}(\beta^{**})-2\widehat{\sigma}_{\mathtt{FM}}(\beta^*,\beta^{**}))^{1/2}$. Let $\widehat{\tilde{\sigma}}(\beta^*,\beta^{**})^{1/2}=(\widehat{\sigma}(\beta^*)+\widehat{\sigma}(\beta^{**})-2\widehat{\sigma}(\beta^*,\beta^{**}))^{1/2}$ and $\tilde{\tilde{\sigma}}^{1/2}(\beta^*,\beta^{**})\mathcal{T}_T=(\tilde{\sigma}(\beta^*)+\tilde{\sigma}(\beta^{**})-2\tilde{\sigma}(\beta^*,\beta^{**}))^{1/2}$. $T_z=[\tilde{G}_T(\beta^{**})\tilde{\sigma}(\beta^{**})-\tilde{G}_T(\beta^*)\tilde{\sigma}(\beta^*)]/(\tilde{\sigma}(\beta^*)+\tilde{\sigma}(\beta^{**})-2\tilde{\sigma}(\beta^*,\beta^{**}))^{1/2}$ and $\tilde{T}_z=[G_T(\beta^{**})\sigma(\beta^{**})-G_T(\beta^*)\sigma(\beta^*)]/(\sigma(\beta^*)+\sigma(\beta^{**})-2\sigma(\beta^*,\beta^{**}))^{1/2}$.

Corollary B.8.1. *Under the conditions of Theorem B.6, 4.1, Assumption 15 and $\sqrt{T}(1/J\vee h)\to 0$, we have, implied from the equation as in (B.25), since*

$$\begin{aligned}
&\mathbb{P}(|\mathcal{T}_T-\mathcal{T}_z|\geq x/2)\\
\leq\ &\mathbb{P}(\sup_\beta|T^{1/2}\widehat{\tilde{\sigma}}(\beta^*,\beta^{**})^{-1/2}(T^{-1}\sum_{t=1}^T\{\widehat{\mu}_t(\beta)-\mu_t(\beta)\})\\
&-G_T(\beta)\sigma(\beta)^{1/2}/(\sigma(\beta^*)+\sigma(\beta^{**})-2\sigma(\beta^*,\beta^{**}))^{1/2}|\geq x/2)\to 0.
\end{aligned}$$

Proof. The proof is similar to 5.3 and therefore is omitted. □

Remark B.9. *It shall be noted that we can achieve the following strong approximation results with a different center,*

$$\begin{aligned}
&\mathbb{P}(|\mathcal{T}_T-T_z|\geq x)\\
&\leq\mathbb{P}(\sup_{\beta\in B}|T^{1/2}\widehat{\sigma}_{\mathtt{FM}}(\beta^*,\beta^{**})^{-1/2}(\widehat{\mu}(\beta)-\mu(\beta))-\tilde{G}_T(\beta)\tilde{\sigma}(\beta)^{1/2}/\tilde{\tilde{\sigma}}^{1/2}(\beta^*,\beta^{**})|\geq x/2)\to 0.
\end{aligned}$$

According to Lemma B.13, then we have the the above object tend to zero, with $x\lesssim_{\mathbb{P}}T^{-\varepsilon'}\vee\sqrt{T}(1/J\vee h)$ (with ε' is a constant $0<\varepsilon'<1/6$).

B.9 Proof of Corollary B.11.1 and B.11.2

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

B.10 Proof of Theorem 5.5

Proof. Define the standardized process as the following,

$$\tilde{U}_T(\beta_1,\beta_2,\beta_3) =$$
$$|T^{-1}\sum_{t=1}^{T}\sqrt{Tn_t/J_t}\{\widehat{\mu}_t(\beta_1)+\widehat{\mu}_t(\beta_3)-2\widehat{\mu}_t(\beta_2)-(\mu_t(\beta_1)+\mu_t(\beta_3)-2\mu_t(\beta_2))\}/\widehat{\sigma}_D(\beta_{1,2,3})^{1/2}|.$$

$$U_T(\beta_1,\beta_2,\beta_3) =$$
$$|T^{-1}\sum_{t=1}^{T}\sqrt{Tn_t/J_t}\{\widehat{\mu}_t(\beta_1)+\widehat{\mu}_t(\beta_3)-2\widehat{\mu}_t(\beta_2)-(\mu_t(\beta_1)+\mu_t(\beta_3)-2\mu_t(\beta_2))\}/\sigma_D(\beta_{1,2,3})^{1/2}|.$$

Define the centered and standardized process as the leading term as follows,

$$Z_T(\beta_1,\beta_2,\beta_3) = \sqrt{T}T^{-1}\sum_{t=1}^{T}\sqrt{n_t/J_t}n_t^{-1}\sum_{i=1}^{n_t}\sum_j(\widehat{p}_{j,t}(\beta_1)+\widehat{p}_{j,t}(\beta_3)-2\widehat{p}_{j,t}(\beta_2))$$
$$\tilde{q}_j^{-1}\Phi^*_{i,j,t}\varepsilon_{it}/\sigma_D(\beta_{1,2,3})^{1/2}.$$

We shall analyze the leading term of the above statistics object,

$$\begin{aligned}
&\sup_{\beta_1,\beta_2,\beta_3}|\tilde{U}_T(\beta_1,\beta_2,\beta_3)|\\
=\ &\sup_{\beta_1,\beta_2,\beta_3}|U_T(\beta_1,\beta_2,\beta_3)|+O_{\mathbb{P}}(r_{1,2,3}),\\
=\ &\sup_{\beta_1,\beta_2,\beta_3}|Z_T(\beta_1,\beta_2,\beta_3)|+O_{\mathbb{P}}(r_{1,2,3}\vee\sqrt{Tn_u/J}J^{-1}\vee\sqrt{Tn_u/J}h),
\end{aligned}$$

where the first equality is due to Assumption 13, and the second equality is due to Theorem 4.1. We shall then prove the following steps:

(1) (Finite-Dimensional Approximation)

$$r_1' = \sup_{\beta_{1,2,3}\in[\beta_l,\beta_u]}|U_T(\beta_1,\beta_2,\beta_3)-U_T\circ\pi_{\tilde{\delta}}(\beta_1,\beta_2,\beta_3)| = O_{\mathbb{P}}((T)^{-1/2+1/q}) = o\left((nT)^{-\varepsilon'}\right).$$

Since we have,

$$\begin{aligned}
&\sup_{\beta_{1,2,3}\in[\beta_l,\beta_u]}|U_T(\beta_1,\beta_2,\beta_3)-U_T\circ\pi_{\tilde{\delta}}(\beta_1,\beta_2,\beta_3)|\\
&\sup_{\beta_{1,2,3}}T^{-1}\sum_{t=1}^{T}n_t^{-1}\sum_{i=1}^{n_t}\sqrt{T}\sqrt{n_t/J_t}\sum_j[(\widehat{p}_{j,t}(\beta_1)+\widehat{p}_{j,t}(\beta_3)-2\widehat{p}_{j,t}(\beta_2))\tilde{q}_j^{-1}\Phi^*_{i,j,t}\varepsilon_{it}\\
&-(\widehat{p}_{j_t,t}(\beta_1^-)+\widehat{p}_{j_t,t}(\beta_3^-)-2\widehat{p}_{j,t}(\beta_2^-))\tilde{q}_j^{-1}\Phi^*_{i,j,t}\varepsilon_{it}]\\
\le\ &\sup_{\beta\in[\beta_l,\beta_u]}C|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{1/J_tn_t}\sum_{j=1}^{J_t}\sum_{i=1}^{n_t}(\widehat{p}_{j,t}(\beta)-\widehat{p}_{j,t}(\beta^-))\Phi^*_{i,j,t}\varepsilon_{it}\tilde{q}_j^{-1}/\sigma_D^{1/2}(\beta_{1,2,3})|\\
&+O_{\mathbb{P}}(\sqrt{n_u}\sqrt{T}/\sqrt{J}(J^{-1}\vee h))\\
\le\ &|\max_{\mathbb{J}\in B_{\mathbb{J}}}\max_t\frac{1}{\sqrt{T}}|\sqrt{1/J_tn_t}\sum_{i=1}^{n_t}(\Phi^*_{i,j,t}\varepsilon_{it}\tilde{q}_j^{-1}-\Phi^*_{i,j-1,t}\varepsilon_{it}\tilde{q}_{j-1}^{-1})/\sigma_D^{1/2}(\beta_{1,2,3})|\\
&+O_{\mathbb{P}}(\sqrt{n_u}\sqrt{T}/\sqrt{J}(J^{-1}\vee h)),\\
\lesssim_{\mathbb{P}}\ &J_u^{1/2}T^{1/q}T^{-1/2}c_{n_u}
\end{aligned}$$

where c_{n_u} is a large enough constant, and the leading term linearization is according to Theorem 4.1, and the summand is a martingale difference sequence with respect to $\mathcal{F}_\ell$. Recall that $\mathcal{F}_\ell = \sigma(\varepsilon_\ell, \mathcal{F}_{t-1})$, $(i,t) \to \ell = \{\ell = \sum_{t'=1}^{t-1}\sum_{i'=1}^{n_{t'}} + i\}$. The last inequality is due the condition that $\max_{t,j}(\mathbb{E}|\tilde{q}_j^{-1}\Phi^*_{i,j,t}\varepsilon_{it}|^q) \lesssim J^{q-1}$, and the Lemma B.5.

(2) (Coupling with a Normal Vector)let the limit variance covariance matrix be $\Sigma(\beta_{\mathbb{J}_1}, \beta_{\mathbb{J}_2}, \beta_{\mathbb{J}_3}) = \lim_{T\to\infty}\mathbb{C}\mathrm{ov}(U_T \circ \pi_{\tilde{\delta}}(\beta_{\mathbb{J}_1}, \beta_{\mathbb{J}_2}, \beta_{\mathbb{J}_3})$, denote $\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3$ as the number of elements within $B_{\tilde{\delta}}$, there exists $G_T \circ \pi_\delta(\beta_{\mathbb{J}}) \sim N\left(0, \Sigma(\beta_{\mathbb{J}_1,\mathbb{J}_2,\mathbb{J}_3})\right)]$ such that

$$r_2' = |\tilde{\mathbb{U}}_T \circ \pi_{\tilde{\delta}}(\beta_{\mathbb{J}_1,\mathbb{J}_2,\mathbb{J}_3}) - Z_{\mathbb{J}_1,\mathbb{J}_2,\mathbb{J}_3}|_\infty = O_{\mathbb{P}}(T^{-\varepsilon'} \vee \sqrt{T}(J^{-1} \vee h)) = o\left((nT)^{-\varepsilon'}\right);$$

Let $Z_T(\beta)$ be a mean zero Gaussian process, with variance $\sigma_d(\beta) = \lim_{T\to\infty} T^{-1}\sum_{t=1}^T \sigma_t^2 \tilde{q}_j^{-1}\widehat{p}_{jt}(\beta)$. Since we have

$$\begin{aligned} & |\tilde{\mathbb{U}}_T \circ \pi_{\tilde{\delta}}(\beta_{\mathbb{J}_1,\mathbb{J}_2,\mathbb{J}_3}) - Z_{\mathbb{J}_1,\mathbb{J}_2,\mathbb{J}_3}|_\infty \\ \leq \quad & C\sup_{\beta\in\beta_v} |\sum_{t=1}^T \sqrt{T}^{-1/2}\sqrt{n_t J_t}^{-1}\sum_{j=1}^{J_t}\Phi^*_{i,j,t}\tilde{q}_j^{-1}\varepsilon_{it}\widehat{p}_{jt}(\beta) - Z_T(\beta)| \max_\beta(\sigma^{1/2}(\beta))/\min_\beta \sigma_D^{1/2}(\beta_{1,2,3}). \end{aligned}$$

Apply Theorem B.13 again, with n correspond to $\sum_{t=1}^T n_t \asymp n_u T$. X_t corresponds to $\Phi^*_{i,j,t}\varepsilon_{it}\widehat{p}_{jt}(\beta_v)\tilde{q}_j$. Σ_z corresponds to $\mathrm{diag}(\sigma_d(\beta_v))$. Since by Assumption 13, $\max_{\beta\in\beta_v}\sigma_d(\beta)$ 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 13, the qth moment of $\max_{j,t}\|\Phi^*_{i,j,t}\tilde{q}_j^{-1}\varepsilon_{it}\|_q \lesssim J^{1-1/q}$. Thus we can prove that

$$\sup_{\beta\in\beta_v} |\sum_{t=1}^T \sqrt{T}^{-1/2}\sqrt{n_t J_t}^{-1}\sum_{j=1}^{J_t}\Phi^*_{i,j,t}\tilde{q}_j^{-1}\varepsilon_{it}\widehat{p}_{jt}(\beta) - Z_T(\beta)| = O_{\mathbb{P}}((n_u T)^{-\varepsilon'}), \tag{B.26}$$

for a constant $0 < \varepsilon' < 1/6$.

(3) Moreover, by Gaussian Maximal inequality, we have

$$\sup_{\beta_{1,2,3}} |G_T(\beta_1,\beta_2,\beta_3) - G_T(\beta_1^-,\beta_2^-,\beta_3^-)| \lesssim_{\mathbb{P}} \sqrt{\log J_a}T^{1/q}T^{-1/2}c_{n_u}. \tag{B.27}$$

□

B.11 Residual properties

We define the residual as $\widehat{\varepsilon}_{it} = R_{it} - \widehat{\mu}_t(\beta_{it})$. 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 B.10. *Under conditions of Theorem 3.1 and $T^{1/2q} \ll Th, T^{1/q} \ll \sqrt{Th}$, we have the residuals satisfying,*

$$\max_{t_0} |\widehat{\varepsilon}_{it_0} - \varepsilon_{it_0}| \lesssim_p \delta_T T^{1/q} = o(1). \tag{B.28}$$

Recall that we define $\mathcal{G}_{t-1}$ as a filtration and β_{it} is conditionally i.i.d. 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 $\tilde{q}_j = \mathbb{E}_{t-1}(\Phi^*_{i,j,t}) = \mathbb{E}(\Phi^*_{i,j,t}|\mathcal{G}_{t-1})$, and $\mathbb{E}(\Phi^*_{i,j,t}\varepsilon_{it}^2)|\mathcal{G}_{t-1}) = \sigma_t^2\tilde{q}_j$. 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 $\widehat{\sigma}_j^2 = n_t/J_t(\sum_{i=1}^{n_t}\widehat{\Phi}_{i,j,t}\widehat{\varepsilon}_{it}^2)(\sum_{i=1}^{n_t}\widehat{\Phi}_{i,j,t})^{-2}$. We see that $\sigma_j^2 = 1/J_t(\mathbb{E}_{t-1}(\Phi^*_{i,j,t}\varepsilon_{it}^2))(\mathbb{E}_{t-1}(\Phi^*_{i,j,t}))^{-2} = J_t^{-1}\sigma_t^2\tilde{q}_j^{-1}$. The variance estimator of $\widehat{L}_t(\beta)$ on the whole support of β is defined as $\widehat{\sigma}_t(\beta) = \sum_j \widehat{p}_j(\beta)\widehat{\sigma}_j^2$ targeting at $\sigma_t(\beta) = \sum_j \widehat{p}_j(\beta)\sigma_j^2$ in the population.

From Assumptions 7, we have $\min_j J_t\tilde{q}_j \geq c$, and $J_t\min_j \mathbb{E}_{t-1}(\Phi^*_{i,j,t}\varepsilon_{it}^2) \geq c$ 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 B.11. *Under the assumptions of Theorem B.6, $\sqrt{\delta_T}T^{1/(2q)}n_t^{-1/2}J_t^{1/2} \to 0$, we have,*

$$\max_t \sup_\beta |\sigma_t(\beta) - \widehat{\sigma}_t(\beta)| \lesssim_{\mathbb{P}} \delta_{1T} + \delta_{2T} = o(1/\sqrt{\log J}).$$

B.12 Strong approximation 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 $\widehat{\sigma}_t(\beta_{1,2,3}) = \sum_{j=1}^{J_t}[\widehat{p}_j(\beta_1) + \widehat{p}_j(\beta_3) - 2\widehat{p}_j(\beta_2)]^2\widehat{\sigma}_j^2$. Also $\sigma_t(\beta_{1,2,3}) = \sum_{j=1}^{J_t}[\widehat{p}_j(\beta_1) + \widehat{p}_j(\beta_3) - 2\widehat{p}_j(\beta_2)]^2\sigma_j^2$. We shall consider the test statistics

$$
\begin{aligned}
T_{n,t}(\beta_{1,2,3}) &= \sup_{\beta_1,\beta_2,\beta_3:\beta_1+\beta_3=2\beta_2} |\widehat{M}_t(\beta_1) - \widehat{M}_t(\beta_2) - \left[\widehat{M}_t(\beta_2) - \widehat{M}_t(\beta_3)\right] \\
&\quad -\{M_t(\beta_1) - M_t(\beta_2) - [M_t(\beta_2) - M_t(\beta_3)]\}|/[\widehat{\sigma}_t(\beta_{1,2,3})^{1/2}]|, \\
T_t(\beta_{1,2,3}) &= \sup_{\beta_1,\beta_2,\beta_3:\beta_1+\beta_3=2\beta_2} |(\widehat{p}_t^\top(\beta_1) - 2\widehat{p}_t^\top(\beta_2) + \widehat{p}_t^\top(\beta_3))\{\mathrm{diag}[\tilde{q}_{jt}]\}^{-1}\{n_t^{-1}\Phi_t^*\varepsilon_t\} \\
&\quad /[\sigma_t(\beta_{1,2,3})^{1/2}]|, \\
Z_t(\beta_{1,2,3}) &= \sup_{\beta_1,\beta_2,\beta_3:\beta_1+\beta_3=2\beta_2} |\sum_{j=1}^{J_t}(\widehat{p}_j^\top(\beta_1) - 2\widehat{p}_j^\top(\beta_2) + \widehat{p}_j^\top(\beta_3))Z_j\sigma_j \\
&\quad /[\sigma_t(\beta_{1,2,3})^{1/2}]|,
\end{aligned}
$$

where Z_j is a standard normal random variable.

Corollary B.11.1. *Under the conditions of Theorem 6.1 and Assumption 14, we have, conditional on $\mathcal{F}_{t-1}$,*

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

And

$$
\sup_{x\in\mathbb{R}} |\mathbb{P}(\sup_{\beta_1,\beta_2,\beta_3:\beta_1+\beta_3=2\beta_2} Z_{n_t}(\beta_{1,2,3}) \le x) - \mathbb{P}(\sup_{\beta_1,\beta_2,\beta_3:\beta_1+\beta_3=2\beta_2} Z_t(\beta_{1,2,3}) \le x)| \to 0. \tag{B.29}
$$

Now we discuss how to implement testing the hypothesis

$$
\mathsf{H}_0^{(u)} : \sup_{\beta_1,\beta_2,\beta_3:\beta_1+\beta_3=2\beta_2} |\mu_t(\beta_1) - \mu_t(\beta_2) - [\mu_t(\beta_2) - \mu_t(\beta_3)]| = 0.
$$

B.13 Test parametric hypothesis

Next we shall provide the theoretical justification of the above algorithm via a strong approximation theorem. Define $\delta_{lt} = \sqrt{n_t}(J_t l_{n,T}/n_t + 1/J_t)/\sqrt{J_t}$. We analyze the rate of the test statistics, and it is not hard to see that under the respective null hypothesis, we have

$$
\begin{aligned}
T_t^1 &= \max_\beta(\sqrt{n_t/J_t})|\widehat{M}_t(\beta) - M_t(\beta) - (\widehat{M}_t(\beta)^{H^1} - M_t(\beta))|/\widehat{\sigma}_t^{1/2}(\beta), \\
&= \max_\beta(\sqrt{n_t/J_t})|\widehat{p}_t^\top(\beta)\{\mathrm{diag}[\tilde{q}_{jt}]\}^{-1}\{n_t^{-1}\Phi_t^*\varepsilon_t\}/\sigma_t(\beta)^{1/2} + O_{\mathbb{P}}(\delta_{lt} \vee \sqrt{J_t}^{-1}) \\
&\quad -\beta\mathbb{E}(\beta_{it}^2)^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\beta_{it}\varepsilon_{it}/\sigma_t(\beta)^{1/2}| + o_{\mathbb{P}}(1/\sqrt{n_t}), \\
&= \sup_\beta|\sqrt{n_t/J_t}\sum_{j=1}^{J_t}\widehat{p}_{t,j}(\beta)(\tilde{q}_{jt}\sigma_j)^{-1}\{n_t^{-1}\Phi_{i,j,t}^*\varepsilon_{it}\} + O_{\mathbb{P}}(\delta_{lt} \vee \sqrt{J_t}^{-1})|, \\
T_t^2 &= \sup_\beta(\sqrt{n_t/J_t})|\widehat{M}_t(\beta) - M_t(\beta) - (\widehat{M}_t(\beta)^{H^2} - M_t(\beta))|/\widehat{\sigma}_t^{1/2}(\beta), \\
&= \max_\beta(\sqrt{n_t/J_t})|\widehat{p}_t^\top(\beta)\{\mathrm{diag}[\tilde{q}_{jt}\sigma_j]\}^{-1}\{n_t^{-1}\Phi_t^*\varepsilon_t\}/\sigma_t^{1/2}(\beta) + O_{\mathbb{P}}(\delta_{lt} \vee \sqrt{J_t}^{-1}) \\
&\quad -(1,\beta)(\mathcal{B}_t^\top\mathcal{B}_t)^{-1}\mathcal{B}_t\varepsilon_t/\sigma_t^{1/2}(\beta) + o_{\mathbb{P}}(1/\sqrt{n_t})|, \\
&= \sup_\beta|\sqrt{n_t/J_t}\sum_{j=1}^{J_t}\widehat{p}_{t,j}(\beta)\tilde{q}_{jt}^{-1}\{n_t^{-1}\Phi_{i,j,t}^*\varepsilon_{it}\}/\sigma_t^{1/2}(\beta) + O_{\mathbb{P}}(\delta_{lt} \vee \sqrt{J_t}^{-1})|.
\end{aligned}
$$

Thus the two terms both have leading term as $|\sqrt{n_t/J_t}\sum_{j=1}^{J_t}\widehat{p}_{t,j}(\beta)\tilde{q}_{jt}^{-1}\{n_t^{-1}\Phi_{i,j,t}^*\varepsilon_{it}\}/\sigma_t^{1/2}(\beta)|$. Then we shall apply strong approximation on the leading term. We can use the same the same strong approximation results as in Theorem 6.1. We now show the detailed rate for the estimator of $M_t(\beta)$. The model is

$$R_{it} = M_t(\beta_{it}) + \varepsilon_{it}. \tag{B.30}$$

Let

$$M_t(\beta_{it}) = \sum_{j=1}^{J_t}\mathbf{1}(\beta_{it} \in P_{jt})L_{t,j}^* + c_{it} + \varepsilon_{it}, \tag{B.31}$$

where c_{it} is the bias term. The bias term can be shown to be of order $|C_t|_\infty$. Let $C_t = [c_{it}]_i$. So we can write

$$R_t = \Phi_t^* M_t^* + C_t + \varepsilon_t, \tag{B.32}$$

where M_t^* is a $J_t \times 1$ dimensional vector of $\mathcal{M}_{t,j}^*$. $\mathcal{M}_{t,j}^* = \{\sum_{i=1}^{n_t}\mathbb{E}(\Phi_{i,j,t}|\mathcal{G}_{t-1})\}\{\sum_{i=1}^{n_t}\mathbb{E}(\Phi_{i,j,t}^* M(\beta_{it})|\mathcal{G}_{t-1})\}$. Now we let

$$
\begin{aligned}
\widehat{\mu}_t - M_t^* &= \{\widehat{p}_t^\top(\beta)\widehat{\Phi}_t\widehat{\Phi}_t^\top\}^{-1}\{\widehat{\Phi}_t R_t\} - M_t^*, && \text{(B.33)} \\
&= \{\widehat{p}_t^\top(\beta)\widehat{\Phi}_t\widehat{\Phi}_t^\top\}^{-1}\{\widehat{\Phi}_t(\Phi_t^{*\top}M_t^* + C_t)\} + \{\widehat{p}_t^\top(\beta)\widehat{\Phi}_t\widehat{\Phi}_t^\top\}^{-1}\{\widehat{\Phi}_t\varepsilon_t\} - M_t^*. && \text{(B.34)}
\end{aligned}
$$

Adopting the conditions and proofs as in Theorem B.6. It is not hard to derive that, $|\{\widehat{\Phi}_t\widehat{\Phi}_t^\top\}^{-1}\{\widehat{\Phi}_t(\Phi_t^{*\top}M_t^*)\} - M_t^*|_{\max} \lesssim_{\mathbb{P}} J_t l_{n,T}/n_t$, $|\{\widehat{\Phi}_t\widehat{\Phi}_t^\top\}^{-1}\{\widehat{\Phi}_t C_t|_{\max} \lesssim_{\mathbb{P}} 1/J_t$, and $|\{\widehat{\Phi}_t\widehat{\Phi}_t^\top\}^{-1}[\{\widehat{\Phi}_t\varepsilon_t\} - \{\Phi_t^*\varepsilon_t\}]|_{\max} \lesssim_{\mathbb{P}} J_t l_{n,T}/n_t$. $|(\{n_t^{-1}\widehat{\Phi}_t\widehat{\Phi}_t^\top\}^{-1} - \mathrm{diag}\{\tilde{q}_{jt}\}^{-1})\{n_t^{-1}\Phi_t^*\varepsilon_t\}|_{\max} \lesssim_{\mathbb{P}} |n_t^{-1}\Phi_t^*\varepsilon_t|_{\max}\sqrt{\log J_t}J_t^{3/2}/\sqrt{n_t}$. The following corollary provides the theoretical support of the uniform confidence band. Recall that we define $Z_t(\beta) = \sum_j \widehat{p}_{j,t}(\beta)Z_j$, where Z_js are standard normal random variables.

Corollary B.11.2. *Under conditions of Theorem B.6, and Assumption 14,*

$$\mathbb{P}(\sup_\beta|\sqrt{n_t/J_t}T_{c,t}(\beta) - Z_t(\beta)| \geq \delta_{lt} \vee \sqrt{J_t}^{-1}) \to 0, \tag{B.35}$$

$$\mathbb{P}(\sup_\beta|\sqrt{n_t/J_t}(\widehat{M}_t(\beta) - \widehat{M}_t(\beta)^{H^1})/\widehat{\sigma}_t^{1/2}(\beta) - Z_t(\beta)| \geq \delta_{lt} \vee \sqrt{J_t}^{-1}) \to 0, \tag{B.36}$$

$$\mathbb{P}(\sup_\beta|\sqrt{n_t/J_t}(\widehat{M}_t(\beta) - \widehat{M}_t(\beta)^{H^2})/\widehat{\sigma}_t^{1/2}(\beta) - Z_t(\beta)| \geq \delta_{lt} \vee \sqrt{J_t}^{-1}) \to 0. \tag{B.37}$$

Moreover, the above the results implies that,

$$\sup_x \mathbb{P}(\sup_\beta(\sqrt{n_t/J_t}T_{c,t}(\beta) \geq x) - \mathbb{P}(\sup_\beta Z_t(\beta) \geq x) \to 0, \tag{B.38}$$

$$\sup_x \mathbb{P}(\sup_\beta \sqrt{n_t/J_t}(\widehat{M}_t(\beta) - \widehat{M}_t(\beta)^{H^1})/\widehat{\sigma}_t(\beta)^{1/2} \geq x) - \mathbb{P}(\sup_\beta Z_t(\beta) \geq x) \to 0, \tag{B.39}$$

$$\sup_x \mathbb{P}(\sup_\beta \sqrt{n_t/J_t}(\widehat{M}_t(\beta) - \widehat{M}_t(\beta)^{H^2})/\widehat{\sigma}_t(\beta)^{1/2} \geq x) - \mathbb{P}(\sup_\beta Z_t(\beta) \geq x) \to 0. \tag{B.40}$$

Remark B.12 (Choice of bandwidth h for fixed t)**.** . *We shall note that selection of h favoring the performance of $\widehat{\mu}_t(\beta)$ shall be different from the choice of h in the first step. To see this, we look at the leading bias term. Recall that under certain restrictions,*

$$\delta_T + a_{nT} = \sqrt{(\log q_n)/n} \vee h \vee \sqrt{\log(q_n)/Th},$$

where $q_n = max(T, n_t)$.

$$\begin{aligned}
&J(\delta_T + a_{nT})^{1/2}\sqrt{\log q_n}/\sqrt{nT} + (\delta_T + a_{nT}) \\
&= h^{1/2}J/\sqrt{n} + J(\sqrt{\log(q_n)/Th})^{1/2}/\sqrt{n} + J((\log q_n)/n)^{1/4}(n)^{-1/2} + h + \sqrt{\log(q_n)/nTh}.
\end{aligned}$$

One admissible rate of h (up to a $\log q_n$ term) shall be order $T^{-1/2}$ instead of $T^{-1/3}$ (optimal rate as in step 1). Therefore undersmoothing is necessary for the good performance which is confirmed by the simulation results. J is of order $n_t^{-1/3}$. We make $n_t^2 \lesssim T^3$.

B.14 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

$$\delta_{j,k,q} = \sup_t \|X_{t,j} - X^{k*}_{t,j}\|. \tag{B.41}$$

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 $\Theta_{q,\alpha} = \||X_.|_\infty\|_{q,\alpha}(\log p)^{3/2} \wedge \Gamma_{\alpha,q}$. $\Psi_{2,\alpha} = \max_j \sup_{l\geq 0} l^{-\alpha}(\sum_{k\geq l}\delta_{j,k,q})$.

Theorem B.13. *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) $\Sigma_z = \lim_{n\to\infty} n^{-1}\sum_{t=1}^T \sum_{l\geq 0}\mathbb{E}(X_t X_{t-l}^\top)$, and $Z \sim$ 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 $\Phi_{\psi_1,\alpha'} = \max_q q^{-\alpha'}\max_j \Theta^z_{\alpha',q,j}$. 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,*

$$\mathbb{P}(|\sqrt{n}^{-1/2}\sum_{t=1}^T X_t - Z|_\infty \geq \delta(p, L, m, q, \alpha)) \to 0. \tag{B.42}$$

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 B.14. *(discussion of rates) We see that q, m, M, L, p interplays with each other. Compared to the i.i.d. 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 i.i.d. 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 $L^{-1/2}\sqrt{\log p^3}\log(p \vee L) \ll \delta(p, L, m, q, \alpha) \asymp L^{-1/6}(\log p)^{3/2} \vee (\log p)^{3/2}\sqrt{m/M}$.*