STAFF
REPORTS

NO. 1068
JULY 2023

REVISED
SEPTEMBER 2026

Beta-Sorted Portfolios

Matias Cattaneo | Richard K. Crump | Yingjie Feng | Weining Wang

FEDERAL RESERVE BANK *of* NEW YORK

Beta-Sorted Portfolios

Matias Cattaneo, Richard K. Crump, Yingjie Feng, and Weining Wang

Federal Reserve Bank of New York Staff Reports, no. 1068
July 2023; revised September 2026
https://doi.org/10.59576/sr.1068

Abstract

Beta-sorted portfolios—portfolios comprised of assets with similar covariation with 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 econometric 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 portfolio construction. Our framework rationalizes the well-known estimation algorithm with precise economic and statistical assumptions on the general data-generating process. We provide conditions which ensure valid estimation and inference allowing for a range of hypotheses of interest in financial applications. We show that the rate of convergence of the estimator changes depending on the value of beta. We demonstrate that valid inference depends critically on the object of interest and discuss drawbacks of the widely used Fama-MacBeth variance estimator. To address these limitations, we propose a new variance estimator. We demonstrate the usefulness of our theoretical results in two empirical applications, including one in which we introduce a novel risk factor that captures the business credit cycle and show that it predicts both the cross-sectional and time-series behavior of U.S. stock returns.

JEL classification: C12, C14, G12
Key words: beta pricing models, portfolio sorting, nonparametric estimation, partitioning, kernel regression, smoothly varying coefficients, Fama-MacBeth variance estimator

---

Crump: Federal Reserve Bank of New York (email: richard.crump@ny.frb.org). Cattaneo: Princeton University (email: cattaneo@princeton.edu). Feng: Tsinghua University (email: fengyj@sem.tsinghua.edu.cn). Wang: University of Groningen (email: ifs@uni-bonn.de). The authors thank Nina Boyarchenko, Mehmet Caner, Fernando Duarte, David Lucca, Olivier Scaillet, and participants at various seminars, workshops, and conferences for helpful comments and discussions. They also thank the editor and three anonymous referees for helpful comments and suggestions that substantially improved the paper. Dennis Kongoli provided excellent research assistance. Cattaneo gratefully acknowledges financial support from the National Science Foundation (SES-1947662, DMS-2210561, and SES-2241575), and from the John Simon Guggenheim Memorial Foundation through a 2026 Guggenheim Fellowship. Feng gratefully acknowledges financial support from the National Natural Science Foundation of China (NSFC) through grants 72522001 and 72203122. Wang gratefully acknowledges financial support from the ESRC (Grant Reference: ES/T01573X/1).

This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in this paper are those of the author(s) and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the author(s)

To view the authors' disclosure statements, visit
https://www.newyorkfed.org/research/staff_reports/sr1068.html.

1 Introduction

Deconstructing expected returns into asset-specific 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, has 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]

The implementation of beta-sorted portfolios entails the following two-step procedure, which incorporates a beta-adaptive portfolio construction (see, e.g., Bali et al. 2016). In a first step, time-varying risk factor exposures (betas) are estimated through (backward-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 exposure to the risk factors, and the magnitude of the 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 long/short 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 econometric model. We show that the model has key features which are important to consider for valid interpretation of the empirical results. Notably, we do not impose no-arbitrage conditions. Instead, the restrictions implied by no arbitrage become testable

[1] See, for example, Jagannathan and Wang (1998), Shanken and Zhou (2007), Kleibergen (2009), Ang et al. (2020), Gospodinov et al. (2014), Adrian et al. (2015), Gagliardini et al. (2016), Raponi et al. (2020), Giglio and Xiu (2021) and many others. For a recent survey, see Gagliardini et al. (2020).

[2] The empirical literature using beta-sorted portfolios is extensive. For a textbook treatment, see Bali et al. (2016), and for a few recent papers see, for example, Boons et al. (2020), Chen et al. (2021), Eisdorfer et al. (2022), Goldberg and Nozawa (2021), Fan et al. (2022), Chen et al. (2023), and Chen et al. (2025).

hypotheses. Furthermore, our framework precisely clarifies how the dynamics of the risk factors relate to the functional form of conditional expected returns. Within this framework, we introduce general sampling assumptions allowing for time-varying factor loadings, serially-correlated factors, and conditional heteroskedasticity across time and assets. We then study the asymptotic properties of the beta-sorted portfolio estimator and associated inference procedures in settings with large cross-sectional, n, and time-series, T, sample sizes (i.e., $n, T \to \infty$).

Our analysis yields new methodological and theoretical results for this popular procedure. At any fixed date, the beta-sorted portfolio estimator is centered on the sum of two terms: conditional expected returns and a linear combination of the common risk-factor innovations. By a law of large numbers, the time average of the second term converges to zero, so the estimator targets the sample average of conditional expected returns (SACER), which we argue is the natural object of interest. We establish uniform consistency for this time-averaged estimator under transparent high-level conditions. The key requirements are that the first-step beta estimates become uniformly accurate, the portfolios become sufficiently homogeneous in beta values, and each portfolio contains enough assets. Importantly, this consistency result does not impose a particular dynamic model for the betas and can accommodate commonly used parametric and nonparametric specifications (e.g., Adrian et al. 2015, Gagliardini et al. 2016, Gu et al. 2020).

These findings have direct implications for how beta-sorted portfolios should be constructed and reported. To estimate conditional expected returns for a fixed value of β, a researcher must track the portfolio containing that value in each period. Mechanically averaging portfolios with the same rank over time, as is commonly done, can combine different beta ranges as the cross-sectional distribution of these risk-factor exposures generally changes. The first-step window size, H, and the number of portfolios, J in the single-factor case, should both adapt to the available time-series and cross-sectional sample sizes rather than be fixed at conventional choices such as five years and ten portfolios. We develop data-driven choices for H and J and provide practical guidance for their use in estimation and inference.

Our distribution theory uncovers a sharp difference in the estimator's behavior at nonzero and at zero risk-factor exposures. For any fixed nonzero evaluation point, the estimator retains a common risk-factor component that is unaffected by the number of assets in each cross-section, so that the time-averaged estimator converges at only a $\sqrt{T}$ rate. When evaluated at the origin, this

common component disappears and idiosyncratic variation—including the additional uncertainty from estimating the risk-factor exposures—becomes the dominant source of uncertainty. Under our stated conditions, the estimator can attain the faster $\sqrt{nT/J^d}$ rate. This special role of zero follows from the model's factor structure and extends beyond the beta-sorted portfolio estimator.

We propose a variance estimator that separately estimates the common risk-factor and idiosyncratic components and accounts for the dependence created by overlapping first-step windows. We show that it provides valid, although potentially conservative, inference. To address the changing behavior near zero, we develop a new inference procedure that remains valid across different values of β, including at and near the origin. The procedure uses a data-driven rule to identify the region near zero and a conservative variance adjustment to ensure that the associated confidence intervals attain at least their nominal coverage asymptotically. We also show that the ubiquitous Fama-MacBeth (FM) variance estimator (Fama and MacBeth, 1973) is generally conservative for inference on the SACER at nonzero evaluation points but may not be valid at the origin, where it can either overestimate or underestimate uncertainty.

Our framework accommodates multiple risk factors and clarifies the trade-offs among common ways of constructing portfolios. Independent sorts form J groups based on each of the d estimated risk-factor exposures and then intersect them, producing J^d portfolios that may contain too few assets as d increases. Estimating several betas but sorting on only one avoids this issue. However, this approach changes the object being estimated. We characterize this alternative estimand and explain how our framework can be adapted to conduct inference on it. Conditional sorts first group assets by one estimated risk-factor exposure and then divide each group further using additional exposures. This approach can also help avoid empty portfolios, but does not eliminate the dimensionality problem. In this case, recovering the SACER at a fixed beta requires tracking the exposure ranges represented by each portfolio. Averaging returns or spreads by sequential rank across groups, as is commonly done in practice, generally combines different exposure regions, complicating interpretation.

Certain linear combinations of beta-sorted portfolio returns eliminate the common risk-factor component and, when constructed appropriately, can attain the same faster rate. A butterfly portfolio is one such "factor-cancelling contrast" and measures a discrete second difference in the SACER, thereby capturing curvature in the relationship between expected returns and risk-factor

exposures. Separately, we develop a test of the affine restriction implied by no arbitrage for raw returns and the corresponding linear restriction for excess returns.

From a theoretical perspective, beta-sorted portfolios present a number of technical challenges since the estimator relies on two nested nonparametric estimation steps together with a portfolio construction based on a first-step nonparametric generated regressor. More precisely, risk-factor exposures estimated nonparametrically in the first step enter directly into the nonsmooth partitioning scheme which complicates 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-sectional and time-series dimensions, this paper is most closely related to Gagliardini et al. (2016) (see also Gagliardini et al., 2020). The authors introduce a general framework and econometric methodology for inference in large-dimensional conditional factor models under no-arbitrage restrictions. They model risk-factor exposures as parametric functions of observable variables and establish conditions for consistent estimation and inference on the prices of risk. Although the statistical model under study shares important similarities with the setup of Gagliardini et al. (2016), there are substantial differences, and the models explored previously in the literature do not nest our setup. For example, our framework for the classical beta-sorted portfolio estimator accommodates data-generating processes that do not necessarily exclude arbitrage opportunities and allow risk-factor exposures to vary smoothly over time. Furthermore, unlike the existing literature (e.g., Adrian et al. 2015, Gagliardini et al. 2016), we establish valid estimation and inference without specifying the functional form of the conditional mean of the risk factors.

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

[3]For analysis of partitioning-based nonparametric estimators see Cattaneo et al. (2020b) and references therein. Partitioning-based estimators with random basis functions have been recently studied in Cattaneo et al. (2020a) and Cattaneo et al. (2024), but in those papers the conditioning variables are observed, while here the conditioning variable is generated using a preliminary time-series (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 et al. (2020) for surveys. A related literature endeavors to jointly estimate factor loadings and *latent* risk factors. See, for example, Connor and Linton (2007), Connor et al. (2012), Kelly et al. (2019), and Borri et al. (2024).

betas, and Cattaneo et al. (2020a), who study portfolio sorting estimators given observed individual characteristic variables. However, the individual results from each of these papers cannot be applied in our multi-step setting. Furthermore, the linkage between the two steps, through the role of the generated (nonparametrically estimated) regressor in the second-stage nonparametric partitioning estimator, represents a substantial technical challenge and has not been studied before.

We demonstrate the practical usefulness of our estimation and inference results in two empirical applications. First, we introduce a new risk factor—a measure of the business credit cycle—based on the Federal Reserve's Senior Loan Officer Opinion Survey and the ISM Manufacturing Index. We show that this factor is predictive of both the cross-sectional and time-series behavior of U.S. stock returns. Second, we revisit the capital share factor from Lettau et al. (2019). Using our new inference procedures, we find no evidence that the capital share risk factor is priced in the cross-section of U.S. stock returns.

We collect additional information and technical details in two Supplemental Appendices. The Supplemental Appendix: Theory (SA-Theory) develops the technical foundations for our results, including primitive conditions, additional theoretical results, and proofs. The Supplemental Appendix: Applied (SA-Applied) provides implementation details, data-driven guidance for tuning-parameter selection, simulation evidence, and the second empirical application. Unless otherwise stated, asymptotic limits are taken as $n, T \to \infty$, with H and J varying as specified below. We use $O_{\mathbb{P}}(1)$ and $o_{\mathbb{P}}(1)$ for sequences that are bounded in probability and converge to zero in probability, respectively. For positive random sequences a_n and b_n, $a_n \asymp_{\mathbb{P}} b_n$ means that both a_n/b_n and b_n/a_n are $O_{\mathbb{P}}(1)$. For positive deterministic sequences $a_n \asymp b_n$ means both a_n/b_n and b_n/a_n are $O(1)$. Finally, for a set $S \subseteq \mathbb{R}^d$, define its diameter as $\operatorname{diam}(S) = \sup\{\|s - s'\| : s, s' \in S\}$.

2 Model setup

We introduce a general econometric model of asset returns and show how it aligns with the two steps that comprise the beta-sorted portfolio algorithm. We explain the rationale behind the form of the estimator and discuss relevant properties of the model.

2.1 Modeling returns

Let R_{it} denote the return of asset i at time t, and f_t a vector of observable risk factors with $f_t \in \mathbb{R}^d$. We assume that asset returns are generated by the linear stochastic coefficient model,

$$R_{it} = \alpha_{it} + \beta_{it}^\top f_t + \varepsilon_{it}, \qquad \mathbb{E}\left[\varepsilon_{it} \middle| \mathcal{F}_{t-1}, f_t\right] = 0, \quad i = 1, \cdots, n_t, \quad t = 1, \cdots, T, \tag{2.1}$$

where α_{it} and β_{it} are random coefficients, ε_{it} is an idiosyncratic error term, and $\mathcal{F}_t$ is the information set up to time t.[5] The sigma field $\mathcal{F}_t$ will, in general, depend on T and $\{n_s : s < t\}$ but we suppress this dependence for simplicity of notation. We assume that α_{it} and β_{it} are both $\mathcal{F}_{t-1}$-measurable so we have that $\mathbb{E}[R_{it} | \mathcal{F}_{t-1}, f_t] = \alpha_{it} + \beta_{it}^\top f_t$. Finally, notice that equation (2.1) accommodates an unbalanced panel. Although each n_t may be different, we assume that they all grow at the same rate which ensures that each cross-section contributes to the asymptotic properties of the estimator (i.e., $n_t \asymp n$ for $t = 1, \cdots, T$).

To obtain the structural form of our model, we assume that there exists a non-random function $\mu_0(\cdot)$ such that

$$\mathbb{E}(R_{it} | \mathcal{F}_{t-1}) = \mu_0(\beta_{it}; \mathcal{G}_{t-1}) = \mu_t(\beta_{it}), \tag{2.2}$$

where $\mathcal{G}_t$ is the sub-sigma field of $\mathcal{F}_t$ generated only by i-invariant random variables such as time-series like f_t. Equation (2.2) restricts conditional expected returns to be a function of β_{it} only, but it allows the functional form to be random and to vary with past realizations of economy-wide random variables.[6] This restriction requires assets with the same risk-factor exposures (as measured by β_{it}) to have the same conditional expected return in a given period. For the discussion in this subsection, we take R_{it} to denote excess returns for simplicity of presentation. If R_{it} instead denotes raw returns, $\mu_t(0)$ may be interpreted as the zero-beta rate at time t.

Since the economy's state variables will reside in $\mathcal{G}_t$, a natural restriction is to impose that

$$\mathbb{E}[f_t | \mathcal{F}_{t-1}] = \mathbb{E}[f_t | \mathcal{G}_{t-1}], \qquad t = 1, 2, \ldots, T. \tag{2.3}$$

[5] For an alternative example of a random coefficient model tailored to a financial application, see Barras et al. (2022).

[6] This allows considerable flexibility in conditional expected returns, including time variation in the importance of economy-wide variables and in the strength of return predictability (e.g., Farmer et al., 2023).

This economic restriction also ensures that the systematic component of R_{it} is determined only by β_{it} and economy-wide variables. Using equation (2.3) along with equations (2.1) and (2.2), we obtain,

$$\alpha_{it} = \alpha_t(\beta_{it}), \qquad \alpha_t(\beta) = \mu_t(\beta) - \beta^\top \mathbb{E}(f_t|\mathcal{G}_{t-1}), \tag{2.4}$$

which clarifies the restriction on α_{it} implied by equation (2.2). Equation (2.4) is the nonparametric analog to a reduced-rank restriction in parametric models as it imposes cross-equation restrictions between the linear coefficients in equation (2.1). In fact, equation (2.4) nests as a special case the assumption of no arbitrage which directly implies a parametric reduced-rank regression model (e.g., Adrian et al., 2015). As shown in Gagliardini et al. (2016), in the absence of arbitrage opportunities, there exists a λ_t such that $\mu_t(\beta) = \beta^\top \lambda_t$ for all β which implies $\alpha_t(\beta) = \beta^\top (\lambda_t - \mathbb{E}(f_t|\mathcal{G}_{t-1}))$.

Finally, combining equations (2.1) and (2.4), we arrive at the structural form

$$R_{it} = \mu_t(\beta_{it}) + \beta_{it}^\top (f_t - \mathbb{E}[f_t|\mathcal{G}_{t-1}]) + \varepsilon_{it}. \tag{2.5}$$

Equation (2.5) may be compared to the standard beta pricing model (e.g., Cochrane, 2005, Chapter 12) and generalizations thereof (e.g., Cochrane, 1996; Adrian et al., 2015; Gagliardini et al., 2016). The most noteworthy difference is the presence of the (possibly) nonlinear, time-varying function $\mu_t(\beta_{it})$ in equation (2.5). When $\mu_t(\cdot)$ is nonlinear, the model departs from the no-arbitrage pricing restriction discussed above. Such departures can be motivated by structural models featuring violations of the law of one price. For example, equation (2.5) could arise in the margin-constraints model of Gârleanu and Pedersen (2011) under the assumption that the security's margin is a nonlinear function of its (past) beta.

2.2 Relation to the two-step estimator

To understand the rationale behind the beta-sorted portfolios algorithm, it is instructive to (temporarily) make the simplifying assumption that β_{it} is observable. Let us define systematic realized returns as

$$M_t(\beta) = \mu_t(\beta) + \beta^\top (f_t - \mathbb{E}[f_t|\mathcal{G}_{t-1}]), \tag{2.6}$$

which implies that $M_t(\beta_{it}) = \alpha_{it} + \beta_{it}^\top f_t$ and

$$R_{it} = M_t(\beta_{it}) + \varepsilon_{it}. \tag{2.7}$$

This equation makes clear that, for a fixed time period, we can only consistently (nonparametrically as $n_t \to \infty$) estimate the unknown function $M_t(\cdot)$ rather than the direct object of interest $\mu_t(\cdot)$. This comes about because both terms in $M_t(\beta)$ are an unknown transformation of β as they involve $\mu_t(\cdot)$ and $\mathbb{E}[f_t|\mathcal{G}_{t-1}]$ neither of which are observed. However, its sample average may be written as,

$$\frac{1}{T}\sum_{t=1}^{T} M_t(\beta) = \frac{1}{T}\sum_{t=1}^{T} \mu_t(\beta) + \frac{1}{T}\sum_{t=1}^{T} \beta^\top (f_t - \mathbb{E}[f_t|\mathcal{G}_{t-1}]). \tag{2.8}$$

Importantly, even though we do not observe $\mathbb{E}[f_t|\mathcal{G}_{t-1}]$, the second term has summands, $\beta^\top(f_t - \mathbb{E}[f_t|\mathcal{G}_{t-1}])$, which are a martingale difference sequence with respect to $\mathcal{G}_t$. Under mild conditions a law of large numbers will hold, and this sample average will converge to zero in probability so that

$$\frac{1}{T}\sum_{t=1}^{T} M_t(\beta) = \frac{1}{T}\sum_{t=1}^{T} \mu_t(\beta) + o_{\mathbb{P}}(1). \tag{2.9}$$

This demonstrates how $T^{-1}\sum_{t=1}^{T}\mu_t(\beta)$, the sample average of conditional expected returns, can be consistently estimable even if the individual $\mu_t(\beta)$ are not. A notable exception occurs when $\beta = 0$ as the linear term in $M_t(\beta)$ is eliminated (see Remark 2.2 for further discussion).

In practice, we must estimate $M_t(\cdot)$ from the data (which requires estimates of the β_{it}) to obtain a feasible estimator of $T^{-1}\sum_{t=1}^{T}\mu_t(\beta)$. In this setting the choice of estimator is informed by additional considerations beyond just statistical properties. Specifically, empirical practice in finance has revolved around two considerations. First, the estimator should have the interpretation of a feasible investment strategy, so that estimates of expected returns and their statistical significance can be directly assessed in economic terms. Second, the estimator should have a "real-time" formulation, i.e., when an additional time period of data is added, prior estimates are unaffected.

Taking these two considerations in mind translates to the following two-step procedure which is ubiquitous in the empirical finance literature:

Step One: Using only information up to time $t-1$, construct $\widehat{\beta}_{it}$ for $i = 1, \ldots, n_t$. The classical implementation uses rolling window OLS regressions,

$$\begin{pmatrix} \widehat{\alpha}_{it} \\ \widehat{\beta}_{it} \end{pmatrix} = \underset{a \in \mathbb{R},\, b \in \mathbb{R}^d}{\arg\min} \sum_{s=1}^{H} \big(R_{i(t-s)} - a - b^\top f_{t-s}\big)^2, \tag{2.10}$$

where H is the window size. This approach can be formally justified under the assumption that α_{it} and β_{it} are smoothly varying (random) coefficients which we discuss further in Section 4.

Step Two: Based on step one, group assets together with similar values of $\widehat{\beta}_{it}$ and form portfolios. The classical implementation uses marginal quantiles from the cross-sectional distribution of $\widehat{\beta}_{it}$ to separate the assets. Next, using the individual returns for time t, calculate the return for each portfolio. Finally, average the portfolio returns over time to assess the unconditional risk properties of the investment strategy.

Portfolios have been established as the fundamental investment object since at least Markowitz (1952, 1959) and Tobin (1958) and so are the natural choice to link returns, R_{it}, to risk exposures, β_{it}. However, since the β_{it} are unobserved they must first be estimated in a preliminary step. Moreover, to ensure the portfolio returns could have been earned in real time (without look ahead bias) each β_{it} is estimated using only information up to time $t-1$.

To formally write the beta-sorted portfolio estimator, let $\widehat{\Delta}_t = \{\widehat{P}_{jt} : 1 \le j \le \bar{J}_t\}$ be a collection of nonoverlapping cells covering the support $\mathcal{B}$ and all estimated betas, based on $\{\widehat{\beta}_{it} : i = 1, \ldots, n_t\}$. The $\widehat{P}_{jt}$ are portfolio designations; specifically, each asset i is allocated to a portfolio, $\widehat{P}_{jt}$, depending on its value for $\widehat{\beta}_{it}$. For example, in the classical implementation, portfolio designation is governed by the marginal quantiles of the (cross-sectional) distribution of $\widehat{\beta}_{it}$, i.e., when $d = 1$,

$$\begin{aligned}
\widehat{P}_{1t} &= \Big[\min_{1 \le i \le n_t} \widehat{\beta}_{it},\ \widehat{F}^{-1}_{\widehat{\beta},t}\big(1/\bar{J}_t\big)\Big), \\
\widehat{P}_{jt} &= \Big[\widehat{F}^{-1}_{\widehat{\beta},t}\big((j-1)/\bar{J}_t\big),\ \widehat{F}^{-1}_{\widehat{\beta},t}\big(j/\bar{J}_t\big)\Big), \qquad\qquad j = 2, \ldots \bar{J}_t - 1, \\
\widehat{P}_{\bar{J}_t t} &= \Big[\widehat{F}^{-1}_{\widehat{\beta},t}\big((\bar{J}_t - 1)/\bar{J}_t\big),\ \max_{1 \le i \le n_t} \widehat{\beta}_{it}\Big],
\end{aligned} \tag{2.11}$$

where $\widehat{F}^{-1}_{\widehat{\beta},t}(\cdot)$ is the empirical quantile function. In practice, the first and last cells are extended

outward, as needed, so that the partition covers the full support of the betas. When $d > 1$, the classical implementation uses the tensor (Cartesian) product of the partition in equation (2.11) across the d dimensions. It is important to emphasize that the portfolio that any specific β resides in may change over time. For example, when $d = 1$, 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.

To formalize the beta-sorted portfolio estimator we need to keep track of which portfolio any specific value of β resides in. We do this via

$$\widehat{p}_t(\beta) = \Big(\mathbb{1}(\beta \in \widehat{P}_{1t}), \cdots, \mathbb{1}(\beta \in \widehat{P}_{\bar{J}_t t})\Big)^\top. \tag{2.12}$$

The vector, $\widehat{p}_t(\beta)$, is comprised of a single element equal to one and all other elements equal to zero. For example, if the value of $\widehat{\beta}_{it}$ resides in the first portfolio, $\widehat{P}_{1t}$, then $\widehat{p}_t(\widehat{\beta}_{it}) = (1, 0, \ldots, 0)^\top$, and the return for the first portfolio at time t is

$$\frac{1}{\widehat{n}_{1t}} \sum_{i=1}^{n_t} \mathbb{1}\big\{\widehat{\beta}_{it} \in \widehat{P}_{1t}\big\} R_{it}, \qquad \widehat{n}_{1t} = \sum_{i=1}^{n_t} \mathbb{1}\big\{\widehat{\beta}_{it} \in \widehat{P}_{1t}\big\}. \tag{2.13}$$

More generally, the beta-sorted portfolio estimator at time t is

$$\widehat{M}_t(\beta) = \widehat{p}_t(\beta)^\top \left(\sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it}) \widehat{p}_t\big(\widehat{\beta}_{it}\big)^\top \right)^{-1} \sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it}) R_{it} \tag{2.14}$$

and the corresponding sample average estimator is

$$\widehat{\mu}(\beta) = \frac{1}{T - H} \sum_{t=H+1}^{T} \widehat{M}_t(\beta). \tag{2.15}$$

The estimator is comprised of $T - H$ individual estimates, $\widehat{M}_t(\beta)$, motivated by equation (2.8). Notice that $\widehat{M}_t(\beta)$ can be interpreted as the fitted value from a cross-sectional OLS regression of R_{it} on $\widehat{p}_t(\widehat{\beta}_{it})$ and evaluated at $\widehat{p}_t(\beta)$. Furthermore, it is straightforward to show that if $\beta \in \widehat{P}_{jt}$, then $\widehat{M}_t(\beta)$ may be equivalently written as,

$$\widehat{M}_t(\beta) = \frac{1}{\widehat{n}_{jt}} \sum_{i=1}^{n_t} \mathbb{1}\big\{\widehat{\beta}_{it} \in \widehat{P}_{jt}\big\} R_{it}, \qquad \widehat{n}_{jt} = \sum_{i=1}^{n_t} \mathbb{1}\big\{\widehat{\beta}_{it} \in \widehat{P}_{jt}\big\}. \tag{2.16}$$

The sample average in the definition of $\widehat{\mu}(\beta)$ begins at time $H+1$ to reflect the need for an initial sample period (of size H) to enable estimation of $\beta_{i(H+1)}$.[7]

We will provide conditions under which $\widehat{\mu}(\beta)$, given by equation (2.15), is uniformly consistent for the sample average of conditional expected returns (SACER) defined as

$$\bar{\mu}_T(\beta; H) = \frac{1}{T-H} \sum_{t=H+1}^{T} \mu_t(\beta). \tag{2.17}$$

The estimator, $\widehat{\mu}(\beta)$, thus provides an estimate of the SACER for a generic asset with risk exposures given by the evaluation point, β.

Remark 2.1 (Comparison to the typical implementation approach)**.** There is one subtle but important difference between equation (2.15) and the typical implementation of beta-sorted portfolios.[8] In existing practice, $\bar{J}_t = \bar{J}$ for all t, and the reported $\widehat{\mu}(\cdot)$ takes on only $\bar{J}$ values as the specific evaluation point β is not considered. For example, the first element of the reported $\widehat{\mu}(\cdot)$ is the average return across time associated with the first portfolio, $\widehat{P}_{1t}$. However, if the cross-sectional distribution of the β_{it} is changing over time then there is no guarantee that a chosen portfolio (e.g., the fifth portfolio) always represents assets with sufficiently similar betas.[9] Therefore, the typical implementation does not generally estimate the SACER at a fixed value of β; rather, it averages returns over beta regions that may change over time. This distinction is particularly important when we are interested in expected returns at intermediate values of β or in tests of monotonicity or other shape restrictions.

Remark 2.2 (Evaluation at $\beta = 0$)**.** As discussed above, the case where $\beta = 0$ is distinctive in this model. Observe that

$$\mathbb{E}[R_{it}|\mathcal{G}_{t-1}, \beta_{it} = 0] = M_t(0) = \mu_t(0), \tag{2.18}$$

so we no longer require the time averaging described by equation (2.8) for $\mu_t(0)$ to be consistently estimable. The disappearance of the common risk-factor component at zero is a property of the

[7] We use H as the length of the initial sample for convenience of notation. A longer burn-in period could also be used.

[8] We distinguish between "classical" beta-sorted portfolios and the "typical" implementation. By the former we mean the use of rolling window regressions to estimate β_{it} and marginal quantiles of the distribution of $\widehat{\beta}_{it}$ to form portfolios. By the latter we mean choosing a fixed number of portfolios (often ten) and reporting only the average returns of those portfolios.

[9] In fact, there may be no value of β that resides in the same ordered portfolio across all time periods.

data-generating process and is not specific to the choice of nonparametric estimator. This distinction will be made throughout the presentation of the theoretical results in the subsequent sections. However, it is important to emphasize that primary interest is generally in values of β which are of larger magnitude.

3 Consistency of beta-sorted portfolios

We first prove consistency of the beta-sorted portfolio estimator under minimal conditions. It is important to emphasize the challenging structure of the estimator as the (generally nonparametrically) generated regressors, $\widehat{\beta}_{it}$, enter the second stage through non-smooth indicator functions, $\widehat{p}_t(\widehat{\beta}_{it})$. Further, the first step and the second step share the same underlying data used to estimate β_{it} and managing this dependence presents a number of technical challenges.

We first provide conditions on the DGP.

Assumption 1 (DGP)**.** *The data satisfy equations* (2.1)–(2.3) *and* (i) β_{it} *has the compact support* $\mathcal{B}$ *for each* $i = 1, \ldots, n_t$ *and* $t = 1, \ldots, T$*;* (ii) *there exists a finite constant* $C > 0$ *such that* $|\mu_t(\beta) - \mu_t(\beta')| \leq C\|\beta - \beta'\|$ *for all* $\beta, \beta' \in \mathcal{B}$ *and* $H + 1 \leq t \leq T$*;* (iii) ε_{it} *is independent over* i *for each* t *conditional on* $\mathcal{F}_{t-1}$*;* (iv) $\max_{1\leq t\leq T} \max_{1\leq i\leq n_t} \mathbb{E}[|\varepsilon_{it}|^q|\mathcal{F}_{t-1}]$ *and* $\max_{1\leq t\leq T} \mathbb{E}[\|\tilde{f}_t\|^q|\mathcal{F}_{t-1}]$ *are bounded for some* $q > 2$ *where* $\tilde{f}_t = f_t - \mathbb{E}[f_t|\mathcal{G}_{t-1}]$.

Assumption 1(i)–(ii) are standard assumptions in the nonparametric literature which rule out discontinuities and other pathologies that would invalidate standard nonparametric estimation approaches. Assumption 1(iii) restricts the cross-sectional dependence properties of ε_{it}. This type of sampling assumption was introduced in Andrews (2005) and has been utilized in the financial econometrics literature by Gagliardini et al. (2016) and Cattaneo et al. (2020a). The assumption is quite general and allows for higher-order dependence in the distribution of ε_{it}; for example, it accommodates conditional heteroskedasticity both cross sectionally and over time. In Remark 4.3 of Section 4, we discuss weakening this assumption. Finally, Assumption 1(iv) comprises standard moment conditions on the factors and idiosyncratic errors.

Since the beta-sorted portfolio estimator consists of two steps, we must impose some conditions on the first-step estimator. However, we only impose minimal assumptions on the estimators of β_{it}.

Assumption 2 (First step estimator). *$\widehat{\beta}_{it}$ is $\mathcal{F}_{t-1}$-measurable, $H \leq cT$ for a constant $c \in (0,1)$, and there exists some sequence of positive constants $\mathsf{R}_{\mathtt{FS}}$ such that*

$$\max_{H+1\leq t\leq T} \max_{1\leq i\leq n_t} \|\widehat{\beta}_{it} - \beta_{it}\| = O_{\mathbb{P}}\big(\mathsf{R}_{\mathtt{FS}}\big).$$

Assumption 2 allows for an arbitrary collection of first-step estimators. It requires each $\widehat{\beta}_{it}$ to only use information available before R_{it} is realized and defines their uniform convergence rate, $\mathsf{R}_{\mathtt{FS}}$. We will impose that $\mathsf{R}_{\mathtt{FS}} = o(1)$ which is a natural (and generally necessary) assumption when working with generated regressors. For the rolling window OLS estimator in equation (2.10), Lemma SA-II.2.3 of SA-Theory provides primitive conditions under which

$$\mathsf{R}_{\mathtt{FS}} = \sqrt{\frac{\log(nT)}{H}} + \frac{(nT)^{1/q}\log(nT)}{H} + \mathsf{B}_{\mathtt{FS}}, \tag{3.1}$$

where the first two terms capture the variability of the estimator and $\mathsf{B}_{\mathtt{FS}}$ captures the bias from approximating the time-varying coefficients by a constant function in the window. Thus, we require H to diverge with T in such a way that ensures that all of these terms vanish asymptotically.

Importantly, we only require that uniformly consistent estimators of the β_{it} are available but they are not required to be of the form given by equation (2.10). Assumption 2 covers, for example, cases where a parametric or nonparametric form for the β_{it} is specified in terms of additional variables (e.g. Connor et al., 2012; Gagliardini et al., 2016; Gu et al., 2020) so long as suitable estimators are available.

We now introduce the following assumption which ensures that the probability of an empty portfolio vanishes as the sample size grows.

Assumption 3 (Partition). *Let $\widehat{n}_{\min} = \min_{H+1\leq t\leq T} \min_{1\leq j\leq \bar{J}_t} \widehat{n}_{jt}$. There exists some positive sequence $n_{\min} > 0$ such that $\widehat{n}_{\min} \geq n_{\min}$ with probability approaching one.*

Our first main result is the following:

Theorem 3.1. *Suppose Assumptions 1–3 hold and the following conditions are satisfied:*

(a) $\mathsf{R}_{\mathtt{FS}} = o(1)$,

(b) $\widehat{\mathsf{R}}_{\mathtt{D}} = o_{\mathbb{P}}(1)$ *where* $\widehat{\mathsf{R}}_{\mathtt{D}} = \max_{H+1\leq t\leq T} \max_{1\leq j\leq \bar{J}_t} \operatorname{diam}(\widehat{P}_{jt})$,

(c) $\frac{\log(nT)}{n_{\min}} = o(1)$,

(d) $\limsup_{n,T\to\infty} \frac{(nT)^{2/q}\log(nT)}{n_{\min}} \leq C$ *for some constant* C.

Then,

$$\sup_{\beta\in\mathcal{B}} \left|\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H)\right| = o_{\mathbb{P}}(1).$$

Theorem 3.1 shows that the beta-sorted portfolio estimator is uniformly consistent under general conditions that follow naturally from the structure of the estimator: Condition (a) requires that the first-step estimators of the β_{it} approximate their targets sufficiently well. Combining this requirement with condition (b) ensures that assets within a portfolio have increasingly similar true betas. Condition (c) ensures that all portfolios are comprised of a sufficient number of assets. Together, conditions (c) and (d) ensure that the contribution of the idiosyncratic errors vanishes asymptotically.

If the partition is "quasi-uniform" (so that no portfolio covers an arbitrarily large or small portion of the support) and uses a number of intervals of order J along each of the d coordinates, and each portfolio contains roughly the same number of assets, then it can be shown that

$$\max_{H+1\leq t\leq T} \max_{1\leq j\leq \bar{J}_t} \operatorname{diam}(\widehat{P}_{jt}) = O_{\mathbb{P}}(J^{-1}), \qquad n_{\min} \asymp \frac{n}{J^d}, \tag{3.2}$$

where $\bar{J}_t \asymp J^d$ uniformly in t. In this case, conditions (b)–(d) require that $J \to \infty$, $J^d \log(nT)/n \to 0$, and $J^d(nT)^{2/q}\log(nT)/n = O(1)$. Consequently, J performs a similar balancing act as H; specifically, J must grow fast enough to eliminate the approximation bias from grouping assets with different betas together, but slowly enough to ensure that each portfolio contains a sufficient number of assets. For the marginal-quantile portfolio construction introduced in equation (2.11), Lemma SA-II.3.3 of SA-Theory provides primitive conditions under which the resulting partition is quasi-uniform over appropriately defined interior portfolios.

Apart from assuming a compact support and the availability of uniformly consistent first-step estimators, consistency of $\widehat{\mu}(\cdot)$ requires no direct restrictions on the time-series or cross-sectional properties of β_{it}. For example, the β_{it} can be random, can be correlated in the cross-section and the time-series, and could depend on additional variables in a (possibly) nonlinear fashion. Furthermore, the factors, f_t, are only required to satisfy the (conditional) moment restrictions given

by Assumption 1(iv), but otherwise have no restrictions on their serial dependence properties.

Remark 3.2 (Technical details on Theorem 3.1)**.** Under the conditions of Theorem 3.1, Theorem SA-I.2.1 of SA-Theory establishes the stronger result that, for any arbitrarily slowly diverging sequence ℓ_n (e.g., $\ell_n = \log\log(nT)$), there exists a fixed, finite constant $C > 0$ such that, with probability approaching one, simultaneously for all $\beta \in \mathcal{B}$,

$$\left|\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H)\right| \leq C\ell_n \left[\mathsf{R}_{\mathsf{FS}} + \widehat{\mathsf{R}}_{\mathsf{D}} + \sqrt{\frac{\log(nT)}{n_{\min}}} + \frac{\|\beta\|}{\sqrt{T}}\right]. \tag{3.3}$$

Unlike the uniform consistency result given in Theorem 3.1, this bound identifies the components that can determine the accuracy of the estimator. It is noteworthy that the result does not require a specific partitioning rule (i.e., portfolio construction) since only the maximum diameter and the minimum number of assets in any one portfolio enter the bound. The final term arises from the unforecastable component of the risk factors and is proportional to the norm of the evaluation point β. For a fixed nonzero β, this term is unaffected by the cross-sectional sample size, whereas it vanishes at $\beta = 0$. The disappearance of this term at $\beta = 0$ is the source of the differential convergence rates of the estimator that we establish in Section 4.

3.1 Object of interest: SACER vs. PACER

In Theorem 3.1 we showed that $\widehat{\mu}(\beta)$ is a consistent estimator of the SACER. Following the arguments in Section 2, the SACER is the natural candidate for our object of interest. However, it is important to point out that the SACER is a sample average of (generally) random quantities and, hence, random itself. The randomness arises from the presence of $\mu_t(\cdot)$ which depends on the i-invariant sigma field $\mathcal{G}_{t-1}$. We can contrast SACER to its population counterpart, the population average of conditional expected returns (PACER). To do so we need the following assumption:

Assumption P (PACER)**.** *There exists a non-random function* $\mu(\cdot)$ *such that* $\sup_{\beta\in\mathcal{B}} \left|\bar{\mu}_T(\beta; H) - \mu(\beta)\right| = o_{\mathbb{P}}(1)$.

Assumption P imposes a uniform convergence condition on the SACER and defines the PACER as its non-random (probability) limit. In the special case where $\mu_t(\cdot) = \mu(\cdot)\ \forall t$, the SACER and PACER are equal.

The SACER and the PACER represent two distinct objects of interest. To make this distinction concrete, suppose that there exists a finite-dimensional state vector S_{t-1} that is $\mathcal{G}_{t-1}$-measurable and $\mu_t(\beta) = \mu_0(\beta; \mathsf{S}_{t-1})$. Further, assume that $\{\mathsf{S}_t\}$ is stationary and ergodic and $\mu_0(\cdot;\cdot)$ satisfies the regularity conditions required for a uniform law of large numbers.[10] Assumption P then holds with $\mu(\beta) = \mathbb{E}[\mu_0(\beta; \mathsf{S}_0)]$. Thus, the PACER averages conditional expected returns over the stationary distribution of the state variables, whereas the SACER averages conditional expected returns over the particular realizations of those state variables observed during the sample period. Consequently, the SACER and PACER may differ meaningfully in any particular sample.

We recommend that empirical researchers focus on the SACER rather than the PACER for a few reasons. First, even if Assumption P fails to hold, the SACER is well-defined under our remaining assumptions. The PACER may not exist if, for example, some state variables driving $\mu_t(\cdot)$ are not stationary. Second, and more importantly, the SACER is a more intuitive and interpretable object of interest as we have a wealth of data summarizing what occurred over our sample. For example, we observe the behavior of macroeconomic aggregates, changes in the legal and regulatory landscape, or the presence of unusual events (e.g., financial crises or natural disasters). Conversely, interpretation of the PACER will generally hinge on the stationary distribution of the economy-wide state variables and the mapping $\mu_0(\cdot;\cdot)$ from those states to conditional expected returns. Characterizing the properties of this distribution would be challenging (e.g., assessing the probability of recessions or financial crises) and the appeal of the approach herein is that we can avoid making specific assumptions about the functional form of $\mu_t(\cdot)$. Third, inference on the PACER requires additional assumptions beyond those needed for inference on the SACER. In addition to the estimation error $\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H)$, inference on the PACER must characterize the time-series variation in $\bar{\mu}_T(\beta; H) - \mu(\beta)$. Despite our preference for the SACER, Section 4.3.1 discusses this additional source of uncertainty and the resulting challenges for inference on the PACER.

4 Inference for beta sorted portfolios

In the previous section we established consistency under minimal conditions, requiring only that the first-step estimators are well behaved and the portfolio-formation rule, using these estimators,

[10] Equation (2.2) allows for more generality; however, we form our discussion around a finite set of state variables governing the economy for simplicity of exposition and to cement intuition.

results in portfolios with sufficiently similar β values and sufficiently many assets. To move the theory beyond consistency, however, we will now specialize to the family of first-step estimators and portfolio-formation rules that are in widespread practice.

4.1 Classical beta-sorted portfolio estimator

The classical beta-sorted portfolio estimator is comprised of (i) rolling window OLS regressions to estimate β_{it} as given by equation (2.10) and (ii) forming portfolios based on the marginal quantiles of the cross-sectional distribution of $\widehat{\beta}_{it}$ as given by equation (2.11) in the special case of $d = 1$. Taking these mainstays of empirical practice as our starting point, we will obtain an asymptotic linearization of $\widehat{\mu}(\cdot)$ that underpins our results on asymptotic normality and feasible inference.

The first ingredient of beta sorts is an estimate of β_{it}. It will be convenient to introduce the following generalization of the conventional rolling window estimator. Let $1 \leq s \leq H$ and define

$$r_{\mathfrak{p}}(u) = (1, u, \ldots, u^{\mathfrak{p}})^{\top}, \qquad X_s(t) = r_{\mathfrak{p}}\left(s/H\right) \otimes (1, f_{t-s}^{\top})^{\top}$$

where $\otimes$ denotes the Kronecker product and $\mathfrak{p}$ is a fixed integer. The corresponding estimator of β_{it} is formed from the following local polynomial regression estimator,

$$\widehat{\beta}_{it} = S\,\widehat{\theta}_{it}, \qquad \widehat{\theta}_{it} = \left(\sum_{s=1}^{H} X_s(t)X_s(t)^{\top}\right)^{-1} \sum_{s=1}^{H} X_s(t)R_{i(t-s)}, \tag{4.1}$$

where S selects elements $2, \ldots, d+1$ of $\widehat{\theta}_{it}$ which correspond to the estimator of β_{it}. This is still a rolling-window least squares estimator except with higher-order polynomial terms. In SA-Theory we also allow for kernel weights in equation (4.1) but we suppress this generalization since we focus on equal weights to match empirical practice. When $\mathfrak{p} = 0$, we recover the rolling-window estimator in equation (2.10). For the SACER estimation and inference procedures discussed in this section, we use $\mathfrak{p} = 0$ and $\mathfrak{p} = 1$. The general formulation also accommodates the local quadratic regressions, $\mathfrak{p} = 2$, used in the empirical affine specification test.

We require that $H \to \infty$ and $H/T \to 0$ to ensure that the number of observations in each window grows but the size of the window itself, relative to the full sample, shrinks. The polynomial order, $\mathfrak{p}$, controls the flexibility of the fit within the window. A more flexible fit can reduce the

(first-step) approximation bias. Indeed, we will use the first-step estimator with $\mathfrak{p} = 1$ to conduct valid pointwise inference at fixed nonzero evaluation points based on a data-driven choice of H (see Section SA-III.1 of SA-Applied for details).[11]

Just as the first step of the estimation procedure can be generalized to allow for more flexible fits (i.e., $\mathfrak{p} > 0$), so too can the second step. For example, rather than use a constant fit within the portfolio we can use the least-squares fit from a regression of R_{it} on a constant and $\widehat{\beta}_{it}$ using *only* the assets placed in that portfolio. To distinguish the order of this second-step fit from the first-step polynomial order, let $\mathfrak{q}$ denote the polynomial order within each portfolio and write the corresponding estimators as $\widehat{M}_t(\beta;\mathfrak{q})$ and $\widehat{\mu}(\beta;\mathfrak{q}) = (T-H)^{-1}\sum_{t=H+1}^{T}\widehat{M}_t(\beta;\mathfrak{q})$. The case $\mathfrak{q} = 0$ is exactly the classical portfolio-mean estimator whereas $\mathfrak{q} = 1$ corresponds to running linear regressions within the portfolio. This more flexible fit can reduce the magnitude of the second-step approximation bias. In particular, we will need to use $\widehat{\mu}(\beta;1)$ to conduct inference when $\beta = 0$. Further, even when $\beta \neq 0$, using $\mathfrak{q} = 1$ can reduce the magnitude of the approximation bias. It also eliminates a higher-order remainder term exactly.[12] Throughout, we use $(\mathfrak{p},\mathfrak{q}) = (0,0)$ for point estimation and $(\mathfrak{p},\mathfrak{q}) = (1,1)$ for inference on the SACER.

4.2 Asymptotic normality

In this section, we establish asymptotic normality of the estimator under the additional conditions collected in Assumption SA-DGP of SA-Theory. Most are standard regularity conditions, so we discuss only the more noteworthy restrictions here. First, we impose more structure on the idiosyncratic errors by writing $\varepsilon_{it} = \varsigma_{it}\eta_{it}$, requiring the primitive innovations, η_{it}, to be independent across assets and time while allowing for cross-sectional and time-series dependence in the scale factors, ς_{it} (Assumption SA-DGP(ii)). Second, we impose conditions on α_{it} and β_{it} to ensure a well-behaved time-varying coefficient model and assume that the histories of these coefficients through date t are conditionally independent across assets, while allowing their conditional distributions to differ across assets (Assumption SA-DGP(iii)). This permits systematic differences in factor exposures across groups such as industries. Third, we impose a near-epoch dependence assumption on the risk factors and require the time average of the conditional variance to converge to a deterministic

[11] Section SA-II.2 of SA-Theory also provides additional results, including primitive smoothness conditions and uniform convergence rates for the first-step estimator with arbitrary $\mathfrak{p}$.

[12] Specifically, $\mathfrak{q} \geq 1$ implies that $\mathscr{R}_3(\beta) = 0$; see Lemma SA-I.3.1 of SA-Theory.

positive-definite matrix (Assumption SA-DGP(iv)).

The starting point for the distributional analysis is the stochastic linearization. For fixed $\mathfrak{q} \in \{0, 1\}$, we have that

$$\widehat{\mu}(\beta; \mathfrak{q}) - \bar{\mu}_T(\beta; H) = \bar{\mathscr{S}}_f(\beta) + \bar{\mathscr{S}}_\varepsilon(\beta; \mathfrak{q}) + \mathscr{R}(\beta; \mathfrak{q}), \tag{4.2}$$

where $\bar{\mathscr{S}}_f(\beta)$ is the risk factor component,

$$\bar{\mathscr{S}}_f(\beta) = \frac{1}{T - H} \sum_{t=H+1}^{T} \beta^\top \big(f_t - \mathbb{E}[f_t \mid \mathcal{G}_{t-1}]\big), \tag{4.3}$$

$\bar{\mathscr{S}}_\varepsilon(\beta; \mathfrak{q})$ is the idiosyncratic component, and $\mathscr{R}(\beta; \mathfrak{q})$ collects all remaining approximation errors. The idiosyncratic component captures two distinct sources of estimation error. The first source is the direct contribution of the idiosyncratic shocks, ε_{it}, to the second-step portfolio fit. This term also arises in standard nonparametric regression problems. The second source is the indirect contribution of past idiosyncratic shocks through the estimated betas used in the second step. Because the rolling estimation windows overlap, the same shock can affect beta estimates in more than one period. Formal expressions are provided in Section SA-II.4 of SA-Theory.

We can use equation (4.2) to introduce some key properties of $\widehat{\mu}(\beta; \mathfrak{q})$. First, under the rate conditions in Section SA-II.6 of SA-Theory, it can be shown that $\mathscr{R}(\beta; \mathfrak{q})$ is negligible relative to the other terms in the linearization. Second, the remaining terms $\bar{\mathscr{S}}_f(\beta)$ and $\bar{\mathscr{S}}_\varepsilon(\beta; \mathfrak{q})$ have different rates of convergence. Specifically, $\bar{\mathscr{S}}_f(0) = 0$ and

$$\bar{\mathscr{S}}_f(\beta) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right), \qquad \beta \neq 0, \tag{4.4}$$

whereas

$$\bar{\mathscr{S}}_\varepsilon(0; \mathfrak{q}) = O_{\mathbb{P}}\left(\sqrt{\frac{J^d}{nT}}\right), \tag{4.5}$$

and, importantly, $\bar{\mathscr{S}}_\varepsilon(\beta; \mathfrak{q}) = o_{\mathbb{P}}(T^{-1/2})$ when $\beta \neq 0$. Thus, $\bar{\mathscr{S}}_f(\beta)$ is the dominant first-order term when $\beta \neq 0$. This implies that, at any fixed nonzero evaluation point, the estimator achieves only a $\sqrt{T}$ rate of convergence even though it uses approximately nT observations. However, the pointwise convergence rate changes discontinuously at $\beta = 0$. At the origin, $\bar{\mathscr{S}}_f(0) = 0$, so, to first

order, only $\bar{\mathscr{I}}_\varepsilon(0;\mathfrak{q})$ contributes, yielding a faster convergence rate.

The $\sqrt{T}$ convergence rate is a consequence of the common factor structure imposed in equation (2.1). At any fixed $\beta \neq 0$, each period's return contains the innovation to the risk factor, $\beta^\top(f_t - \mathbb{E}[f_t \mid \mathcal{G}_{t-1}])$. Increasing the cross-sectional sample size averages away the effect of idiosyncratic shocks but has no effect on this common component, whose influence can only be reduced by averaging over time. It is this feature that explains why $\bar{\mathscr{I}}_f(\beta)$ dominates the other terms in the linearization. This slower rate of convergence is the cost of remaining agnostic about the functional form of $\mathbb{E}[f_t \mid \mathcal{G}_{t-1}]$. If the conditional mean of the risk factors were known, the factor innovation could be removed from each portfolio return and a faster convergence rate could then be achieved. However, this would require a functional-form assumption on the factor dynamics and would lead to an entirely different estimator. As our focus is on the classical beta-sorted portfolio estimator, we do not discuss such an approach further.

The following theorem formalizes the distinction between the pointwise asymptotic distributions at zero and at nonzero evaluation points. Let $\mathcal{B}^\circ \subset \operatorname{int}(\mathcal{B})$ denote an interior evaluation region containing zero.

Theorem 4.1 (Pointwise asymptotic normality)**.** *Fix $\beta \in \mathcal{B}^\circ$ and suppose the conditions of Theorem SA-II.6.2 of SA-Theory hold. If $\beta \neq 0$, then, for each $\mathfrak{q} \in \{0,1\}$,*

$$\frac{\widehat{\mu}(\beta;\mathfrak{q}) - \bar{\mu}_T(\beta;H)}{\sqrt{\Sigma_f(\beta)}} \rightsquigarrow \mathcal{N}(0,1), \qquad \Sigma_f(\beta) \asymp_{\mathbb{P}} T^{-1},$$

where $\rightsquigarrow$ denotes convergence in distribution and

$$\Sigma_f(\beta) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \beta^\top \mathbb{V}\mathrm{ar}(f_t \mid \mathcal{G}_{t-1})\beta.$$

If $\beta = 0$, then

$$\frac{\widehat{\mu}(0;1) - \bar{\mu}_T(0;H)}{\sqrt{\Sigma_\varepsilon(0;1)}} \rightsquigarrow \mathcal{N}(0,1), \qquad \Sigma_\varepsilon(0;1) \asymp_{\mathbb{P}} \frac{J^d}{nT},$$

where $\Sigma_\varepsilon(0;\mathfrak{q})$ is the conditional variance of $\bar{\mathscr{I}}_\varepsilon(0;\mathfrak{q})$ and is defined in Section SA-II.6 of SA-Theory.

Theorem 4.1 establishes pointwise asymptotic normality of the beta-sorted portfolio estimator centered at the SACER and formalizes the distinction between the convergence rates described

above. Although the leading asymptotic variance at any fixed nonzero evaluation point, $\Sigma_f(\beta)$, does not depend on the order $\mathfrak{q}$ of the second-step fit, the choice of $\mathfrak{q}$ affects the second-step approximation bias, and $\mathfrak{q} = 1$ is required to conduct inference at the origin. For all values of β, asymptotic normality also requires the approximation errors from estimating the β_{it} in the first step and fitting $\mu_t(\cdot)$ in the second step to be negligible relative to the corresponding standard error.

Remark 4.2 (Technical details on Theorem 4.1)**.** Section SA-I.3 of SA-Theory develops a stochastic linearization based on generic first-step estimators, and Section SA-II.4 specializes it to the classical estimator outlined in Section 4.1. Further, SA-Theory gives conditions under which $\mathscr{R}(\beta; \mathfrak{q})$ is negligible relative to $T^{-1/2}$ when $\beta \neq 0$ and $\sqrt{J^d/(nT)}$ when $\beta = 0$. To establish our pointwise inference result at $\beta = 0$, we modify the marginal-quantile partition in equation (2.11) so that zero is a fixed breakpoint in every marginal partition. This ensures that zero occupies the same boundary position in its selected portfolio at every date, providing the sharper control of the estimated partition needed at the origin. At a fixed $\beta \neq 0$, the leading common risk-factor component, $\bar{\mathscr{S}}_f(\beta)$, does not depend on this adjustment. The partition anchored at zero can nevertheless be used at all evaluation points to provide a common implementation. Importantly, the marginal quantile partition of equation (2.11) remains valid for consistent estimation of the SACER.

Remark 4.3 (Cross-sectional dependence)**.** Equation (4.2) clarifies the scope for weakening the cross-sectional independence assumptions. For a fixed $\beta \neq 0$, the leading term is $\bar{\mathscr{S}}_f(\beta)$, whose behavior is solely governed by the evaluation point and the time-series properties of the risk factors. Consequently, sufficiently weak cross-sectional dependence that preserves the orders of $\bar{\mathscr{S}}_\varepsilon(\beta; \mathfrak{q})$ and $\mathscr{R}(\beta; \mathfrak{q})$ would leave the first-order distribution unchanged. This suggests that the result could accommodate, for example, additional weak cross-sectional dependence in the idiosyncratic errors of the type considered by Chamberlain and Rothschild (1983).

Remark 4.4 (Time series properties of conditional expected returns)**.** One way to satisfy the first-step bias requirement in Assumption SA-FSB of SA-Theory is to require α_{it} and β_{it} to vary sufficiently smoothly within each rolling estimation window. However, equation (2.4) shows that the relevant restriction cannot be interpreted independently of the functional form for conditional expected returns because $\alpha_t(\beta) = \mu_t(\beta) - \beta^\top \mathbb{E}[\, f_t |\, \mathcal{G}_{t-1}]$. To illustrate, suppose that S_t is a vector of $\mathcal{G}_t$-measurable state variables which follows a stationary linear time-series process (with nonde-

generate innovations) and $\mathbb{E}[f_t \mid \mathcal{G}_{t-1}] = \zeta + \Lambda \mathsf{S}_{t-1}$, with $\Lambda \neq 0$ (as in, e.g., Adrian et al. 2015 or Gagliardini et al. 2016). Then, in order for $\alpha_{it} = \alpha_t(\beta_{it})$ to vary smoothly, $\mu_t(\cdot)$ *must* vary over time in a way that offsets the nonsmooth variation induced by $\beta^\top \Lambda \mathsf{S}_{t-1}$. In that case, the SACER and PACER are distinct objects and need not coincide in a given sample, which directly affects the interpretation of estimation and inference in empirical applications.

4.3 Feasible inference on the SACER

We have established conditions which ensure asymptotic normality of the beta-sorted portfolio estimator when centered at the SACER. In this section, we investigate the properties of two feasible inference procedures. The first is the so-called Fama-MacBeth (FM) variance estimator which is ubiquitous in existing applications of beta-sorted portfolios. The FM variance estimator can be motivated by the classical sample variance estimator and constructed as

$$\widehat{\Sigma}_{\mathtt{FM}}(\beta; \mathfrak{q}) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \left(\widehat{M}_t(\beta; \mathfrak{q}) - \widehat{\mu}(\beta; \mathfrak{q}) \right)^2. \tag{4.6}$$

Here, $\widehat{M}_t(\beta; \mathfrak{q})$ for $t = H+1, \ldots, T$ serves as the sample "observations" and $\widehat{\mu}(\beta; \mathfrak{q})$ serves as the sample "mean." Consequently, apart from a degrees-of-freedom correction, the FM variance estimator also has an interpretation as a leave-t out jackknife variance estimator. $\widehat{\Sigma}_{\mathtt{FM}}(\beta; 0)$ is ubiquitous in the empirical finance literature. We will discuss the theoretical properties of this conventional FM variance estimator, and its generalization given by equation (4.6), later in this section.

The second feasible inference procedure is based on the theoretical results we have presented. From equation (4.2) and Theorem 4.1, sampling uncertainty has a common risk-factor component, $\Sigma_f(\beta)$, and an idiosyncratic component, $\Sigma_\varepsilon(\beta; 1)$, which we will estimate separately. At any fixed $\beta \neq 0$, $\Sigma_f(\beta)$ is the leading variance component, which we estimate using the following plug-in estimator:

$$\widehat{\Sigma}_f(\beta) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \left[\beta^\top \left(f_t - \widehat{f}_{t|t-1} \right) \right]^2. \tag{4.7}$$

Here, $\widehat{f}_{t|t-1}$ is an estimate of $\mathbb{E}[f_t | \mathcal{G}_{t-1}]$ using only information contained in $\mathcal{G}_{t-1}$. For example, a parametric approach would lead to $\widehat{f}_{t|t-1} = h_{t-1}(\widehat{\vartheta})$ for some $h_{t-1}(\vartheta)$ with $h_{t-1}(\widehat{\vartheta})$ a corresponding

feasible estimate. As a simple example, in the case of an AR(1) we would obtain $h_{t-1}(\vartheta) = \vartheta_0 + \vartheta_1 f_{t-1}$, and we would estimate the parameters, $(\vartheta_0, \vartheta_1)$, recursively. Importantly, the estimator, $\widehat{f}_{t|t-1}$, need not be based on the correct specification. If the conditional mean of the factors is misspecified, then Theorem 4.5 below shows that

$$\widehat{\Sigma}_f(\beta) - \Sigma_f(\beta) = \mathscr{B}_{\mathtt{F}}(\beta) + o_{\mathbb{P}}(T^{-1}), \tag{4.8}$$

where

$$\mathscr{B}_{\mathtt{F}}(\beta) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \left[\beta^\top \left(\widehat{f}_{t|t-1} - \mathbb{E}[f_t|\mathcal{G}_{t-1}]\right)\right]^2. \tag{4.9}$$

Clearly, $\mathscr{B}_{\mathtt{F}}(\beta)$ is non-negative; thus, we can still conduct valid, albeit possibly conservative, inference based on $\widehat{\Sigma}_f(\beta)$.

We estimate the variance contribution from the idiosyncratic component using a bin-restricted delete-i jackknife. Let $\widehat{M}_t^{(-i)}(\beta; \mathfrak{q})$ be the period-t portfolio estimator computed after deleting unit i from the second-step cross-section while holding the portfolio formation rule represented by the partition $\widehat{\Delta}_t$ fixed, and define

$$\widehat{d}_{it}(\beta; \mathfrak{q}) = \widehat{M}_t^{(-i)}(\beta; \mathfrak{q}) - \widehat{M}_t(\beta; \mathfrak{q}). \tag{4.10}$$

Set $\widehat{d}_{it}(\beta; \mathfrak{q}) = 0$ when unit i is not observed at date t. The jackknife variance estimator is

$$\widehat{\Sigma}_\varepsilon(\beta; \mathfrak{q}) = \frac{1}{(T-H)^2} \sum_i \sum_{t=H+1}^{T} \sum_{s=H+1}^{T} \widehat{d}_{it}(\beta; \mathfrak{q}) \widehat{d}_{is}(\beta; \mathfrak{q}) \mathbb{1}\{|t-s| \leq H\}. \tag{4.11}$$

The truncation at lag H captures the serial dependence induced by the overlapping first-step estimation windows. Pairs farther than H periods apart do not share primitive idiosyncratic shocks after the first-step correction is isolated. Importantly, calculating $\widehat{\Sigma}_\varepsilon(\beta; \mathfrak{q})$ does not require any re-estimation steps in practice. Using properties of leave-one-out OLS estimation, we can express $\widehat{d}_{it}(\beta; \mathfrak{q})$ as a function of the fitted residuals from the second-step regression and the leverage of unit i in that regression. Section SA-II.7 of SA-Theory gives equivalent computationally simple formulations of this jackknife variance estimator.

Combining the two components gives our proposed variance estimator

$$\widehat{\Sigma}(\beta) = \widehat{\Sigma}_f(\beta) + \widehat{\Sigma}_\varepsilon(\beta; 1). \tag{4.12}$$

Theorem 4.5 (Plug-in variance estimator)**.** *Fix $\beta \in \mathcal{B}^\circ$ and suppose the conditions of Theorem SA-II.7.3 of SA-Theory hold. If $\beta \neq 0$, then*

$$\widehat{\Sigma}(\beta) - \Sigma_f(\beta) - \Sigma_\varepsilon(\beta; 1) - \mathscr{B}_{\mathtt{F}}(\beta) = o_{\mathbb{P}}(T^{-1}).$$

If $\beta = 0$, then

$$\widehat{\Sigma}(0) - \Sigma_\varepsilon(0; 1) = o_{\mathbb{P}}\left(\frac{J^d}{nT}\right).$$

Theorem 4.5 shows that the newly-proposed variance estimator satisfies the required rate restrictions to ensure feasible inference. When the conditional mean of the risk factors is correctly specified and estimated sufficiently accurately, the plug-in variance estimator is consistent for the first-order variance. When the conditional mean is misspecified, the plug-in variance estimator is conservative and overestimates the first-order variance by $\mathscr{B}_{\mathtt{F}}(\beta)$. The jackknife variance estimator is consistent for the idiosyncratic component of the first-order variance at $\beta = 0$. Thus, Theorem 4.5 establishes that $\widehat{\Sigma}(\beta)$ for $\beta \neq 0$ and $\widehat{\Sigma}(0)$ at the origin can be used to conduct valid, possibly conservative, pointwise inference.

We now discuss the properties of the Fama-MacBeth variance estimator.

Theorem 4.6 (Fama–MacBeth variance estimator)**.** *Fix $\beta \in \mathcal{B}^\circ$ and $\mathfrak{q} \in \{0, 1\}$, and suppose the conditions of Theorem SA-II.7.4 of SA-Theory hold. Define*

$$\Sigma_\mu(\beta) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \big(\mu_t(\beta) - \bar{\mu}_T(\beta; H)\big)^2.$$

If $\beta \neq 0$, then

$$\widehat{\Sigma}_{\mathtt{FM}}(\beta; \mathfrak{q}) = \Sigma_f(\beta) + \Sigma_\mu(\beta) + o_{\mathbb{P}}(T^{-1}).$$

Theorem 4.6 shows that the FM variance estimator is generally conservative for inference on the SACER at a fixed nonzero β as it includes the time-series variation in conditional expected returns,

$\Sigma_\mu(\beta)$, which is not part of the first-order variance. This additional term is nonnegative and can be strictly positive when $\mu_t(\beta)$ varies over time. The intuition for this result can be gleaned from the following decomposition of the summands of the estimator:

$$\widehat{M}_t(\beta;\mathfrak{q}) - \widehat{\mu}(\beta;\mathfrak{q}) = [\widehat{M}_t(\beta;\mathfrak{q}) - \mu_t(\beta)] + [\mu_t(\beta) - \bar{\mu}_T(\beta;H)] - [\widehat{\mu}(\beta;\mathfrak{q}) - \bar{\mu}_T(\beta;H)]. \tag{4.13}$$

After squaring and averaging this decomposition, the first component generates the leading sampling variance, $\Sigma_f(\beta)$, while the second generates $\Sigma_\mu(\beta)$. The final component and the relevant cross-products are asymptotically negligible.

The properties of the FM variance estimator at $\beta = 0$ are very different. As we saw in Section 4.2, the uncertainty of the estimator depends importantly on the fact that the second-step procedure relies on first-step estimation. The FM estimator uses the period-by-period dispersion of the portfolio estimates but does not account for the dependence induced by the overlapping windows in the first step. As such, it need not estimate the appropriate variance at the origin. This is the case even when the no-arbitrage condition holds and the conditional expected return function is linear in β. Depending on the DGP, $\widehat{\Sigma}_{\mathtt{FM}}(0;\mathfrak{q})$ may either overestimate or underestimate the true variance.

Corollary 4.6.1 (Pointwise feasible inference)**.** *Suppose the conditions of Theorems 4.1, 4.5, and 4.6 hold. Let $\mathsf{z}_{1-\alpha/2}$ be the $1-\alpha/2$ quantile of a standard Gaussian random variable. For any nonnegative variance estimator $\widehat{V}$, define*

$$\mathcal{I}_{1-\alpha}(\beta;\mathfrak{q},\widehat{V}) = \left[\widehat{\mu}(\beta;\mathfrak{q}) - \mathsf{z}_{1-\alpha/2}\widehat{V}^{1/2},\ \widehat{\mu}(\beta;\mathfrak{q}) + \mathsf{z}_{1-\alpha/2}\widehat{V}^{1/2}\right]$$

and $\widehat{\Sigma}(\beta)_+ = \max\left\{\widehat{\Sigma}(\beta), 0\right\}$. Then:

(i) For each fixed $\beta \neq 0$ and $\mathfrak{q} \in \{0,1\}$,

$$\liminf_{n,T\to\infty} \mathbb{P}\left(\bar{\mu}_T(\beta;H) \in \mathcal{I}_{1-\alpha}\big(\beta;\mathfrak{q},\widehat{\Sigma}(\beta)_+\big)\right) \geq 1-\alpha.$$

The coverage probability converges to $1-\alpha$ if $T\mathscr{B}_{\mathtt{F}}(\beta) = o_{\mathbb{P}}(1)$.

(ii) For each fixed $\beta \neq 0$ *and* $\mathfrak{q} \in \{0,1\}$,

$$\liminf_{n,T\to\infty} \mathbb{P}\left(\bar{\mu}_T(\beta;H) \in \mathcal{I}_{1-\alpha}\big(\beta;\mathfrak{q},\widehat{\Sigma}_{\mathtt{FM}}(\beta;\mathfrak{q})\big)\right) \geq 1-\alpha,$$

The coverage probability converges to $1-\alpha$ *if* $T\Sigma_\mu(\beta) = o_{\mathbb{P}}(1)$.

(iii) For $\beta = 0$,

$$\mathbb{P}\left(\bar{\mu}_T(0;H) \in \mathcal{I}_{1-\alpha}\big(0;1,\widehat{\Sigma}(0)_+\big)\right) \to 1-\alpha.$$

Corollary 4.6.1 shows that for $\beta \neq 0$ confidence intervals based on either the newly-proposed variance estimator or the FM variance estimator attain at least their nominal coverage probability. The former variance estimator is asymptotically exact when $T\mathscr{B}_{\mathtt{F}}(\beta)$ is asymptotically negligible, whereas the FM interval is asymptotically exact when the time-series variation in conditional expected returns is asymptotically negligible. When $\beta = 0$, the newly-proposed variance estimator is asymptotically exact. The use of $\widehat{\Sigma}(\beta)_+$ in Corollary 4.6.1 arises because the jackknife component of the variance estimator, $\widehat{\Sigma}_\varepsilon(\beta;\mathfrak{q})$, can be negative in finite samples.[13] This ensures that the standard error is well defined. In the limit, this adjustment is unnecessary with probability approaching one.

Remark 4.7 (Uniformity over β)**.** Corollary 4.6.1 establishes coverage at each fixed evaluation point, but it does not guarantee that coverage remains valid uniformly as β approaches zero. This distinction matters because of the differential behavior of the limiting distribution at the origin that we have discussed. Section SA-II.8 of SA-Theory develops a uniformly valid procedure that uses the newly-proposed variance estimator away from zero and a conservative variance adjustment near zero. Let $\mathcal{I}^{\mathtt{UI}}_{1-\alpha}(\beta)$ denote the resulting pointwise confidence interval. These confidence intervals attain at least their nominal coverage probability uniformly over $\beta \in \mathcal{B}^\circ$, i.e.,

$$\liminf_{n,T\to\infty} \inf_{\beta\in\mathcal{B}^\circ} \mathbb{P}\left(\bar{\mu}_T(\beta;H) \in \mathcal{I}^{\mathtt{UI}}_{1-\alpha}(\beta)\right) \geq 1-\alpha.$$

Remark 4.8. Theorems 4.5 and 4.6 imply that we can construct an asymptotic lower bound on

[13] Alternatively, we could add weights to the jackknife variance estimator to ensure non-negativity as is done in the HAC literature (e.g., Newey and West, 1987). However, in empirically-relevant settings we do not observe negative values of $\widehat{\Sigma}_\varepsilon(\beta;\mathfrak{q})$.

the time-series variation in conditional expected returns, as measured by $\Sigma_\mu(\beta)$, by taking the difference between $\widehat{\Sigma}_{\mathtt{FM}}(\beta;\mathfrak{q})$ and $\widehat{\Sigma}(\beta)$. Under the assumptions of these theorems, for each fixed $\beta \neq 0$ and $\mathfrak{q} \in \{0,1\}$,

$$\widehat{\Sigma}_{\mathtt{FM}}(\beta;\mathfrak{q}) - \widehat{\Sigma}(\beta) = \Sigma_\mu(\beta) - \mathscr{B}_{\mathtt{F}}(\beta) + o_{\mathbb{P}}(T^{-1}). \tag{4.14}$$

Because $\mathscr{B}_{\mathtt{F}}(\beta) \geq 0$, equation (4.14) shows that this difference provides an asymptotic lower bound on $\Sigma_\mu(\beta)$. When $T\mathscr{B}_{\mathtt{F}}(\beta) = o_{\mathbb{P}}(1)$, the bound is asymptotically sharp. Thus, the contrast between the two variance estimators provides information about the time-series variation in conditional expected returns. In both our empirical applications, we find that this difference is relatively large, making the bound informative.

Remark 4.9. Theorem 4.5 and Corollary 4.6.1 imply that valid, although potentially conservative, inference on the SACER does not require correct specification of $\mathbb{E}[f_t \mid \mathcal{G}_{t-1}]$. A misspecified predictor increases the estimated variance through $\mathscr{B}_{\mathtt{F}}(\beta)$ but does not invalidate coverage. This contrasts with, for example, Adrian et al. (2015) and Gagliardini et al. (2016), which impose a first-order Markov structure on the factor dynamics. In practice, specifying the correct functional form and the appropriate conditioning variables for the factor dynamics can be challenging. This robustness is an advantage of the beta-sorted portfolio estimator.

The choices of H and J govern the amount of smoothing in the two estimation steps. Longer first-step windows reduce the variability of the estimated betas but can increase approximation bias when risk-factor exposures vary over time, while forming more portfolios reduces within-portfolio approximation bias but leaves fewer assets in each portfolio. Part III of SA-Applied develops data-driven choices, based on mean squared error criteria, at fixed nonzero evaluation points, together with a near-origin adjustment for uniform-in-β inference. It also explains why factor-cancelling inference requires dedicated undersmoothing. Part IV reports Monte Carlo evidence on the finite-sample performance of the SACER estimators, the proposed and FM variance estimators, and the data-driven tuning rules. The simulation results are consistent with our theoretical predictions. We find that confidence intervals based on our proposed variance estimator perform well when the evaluation point is away from the origin. Near zero, the uniform-in-β adjustment substantially improves coverage relative to the pointwise procedure.

4.3.1 Estimation and Inference on the PACER

Although we recommend focusing on the SACER, it is useful to compare the requirements for estimation and inference on the PACER. For $\mathfrak{q} \in \{0, 1\}$, decompose

$$\widehat{\mu}(\beta; \mathfrak{q}) - \mu(\beta) = \big[\widehat{\mu}(\beta; \mathfrak{q}) - \bar{\mu}_T(\beta; H)\big] + \big[\bar{\mu}_T(\beta; H) - \mu(\beta)\big]. \tag{4.15}$$

The first term is $o_{\mathbb{P}}(1)$ under the conditions in Section 3 and Theorem SA-I.2.1 of SA-Theory. At any fixed β, consistency for the PACER therefore requires only a law of large numbers for $\bar{\mu}_T(\beta; H)$. Assumption P imposes this convergence uniformly over $\beta \in \mathcal{B}$.

Inference on the PACER, however, requires substantially more structure. When the estimator is centered at the SACER, its leading component has a martingale-difference structure that underlies Theorem 4.1. Centering at the PACER introduces the additional term $\bar{\mu}_T(\beta; H) - \mu(\beta)$, which reflects the time-series variation in conditional expected returns and is not generally a martingale difference. A PACER central limit theorem must, therefore, characterize the asymptotic distribution of this time-series average, including its long-run variance, as well as its covariance with the SACER-centered estimation error (the first term in equation (4.15)). Assumption P alone is insufficient for such results.

Consequently, the current variance results do not by themselves provide feasible inference on the PACER. At a fixed nonzero β, Theorem 4.6 shows that the FM variance estimator captures the period-by-period variation in conditional expected returns through $\Sigma_\mu(\beta)$. It does not, in general, estimate the long-run variance of this process when conditional expected returns are serially dependent, nor does it account for their possible covariance with the SACER-centered estimation error. Establishing a PACER central limit theorem and constructing a feasible variance estimator would therefore require substantially more structure on the time-series dynamics of the DGP.

5 Extensions and further discussion

5.1 Multiple factors and portfolio construction

The beta-sorted portfolio estimator that we have studied throughout the paper is the one based on so-called independent sorts, i.e., forming portfolios based on tensor products of the marginal

partitions of each factor exposure (using marginal quantiles as breakpoints) as described in earlier sections. This is the multivariate analog of the estimator when $d = 1$. A drawback of this approach is that, as we showed, the number of portfolios grows exponentially in the number of factors (at the rate J^d), which can lead to sparsely populated portfolios and even empty portfolios. In practice, researchers often use a small number of factors to avoid this problem; in particular, applications with $d > 2$ are rare (one exception is Chang et al. (2013), who used $d = 4$).

An alternative approach that avoids multidimensional sorting is to form portfolios based on a single target risk-factor exposure, while controlling for the other factors in the first-step regression.[14] However, such an approach targets a different object than the SACER and the distinction is analogous to the difference between the "short" and "long" regression when studying omitted variables in OLS regressions. Suppose that the first risk factor is the target and is used for portfolio construction whereas the other $d - 1$ factors serve only as control variables in the first-step regression. Partition the evaluation point β as $\beta = (\beta_1, \beta_{-1}^\top)^\top$. For simplicity of presentation, suppose the β_{it} are conditionally *i.i.d.* across assets rather than just conditionally independent. Let $\mathbb{E}_t^\beta[\cdot \mid \beta_{it,1} = b_1]$ denote the expectation with respect to their common period-t conditional distribution, further conditioning on the first component being equal to b_1.[15] The resultant estimator targets the following object

$$\bar{\mu}_{T|1}(b_1; H) = \frac{1}{T-H} \sum_{t=H+1}^{T} \mu_{t|1}(b_1), \qquad \mu_{t|1}(b_1) = \mathbb{E}_t^\beta[\mu_t(\beta_{it}) \mid \beta_{it,1} = b_1]\,. \tag{5.1}$$

Like the SACER, $\bar{\mu}_{T|1}(b_1; H)$ is generally random because both conditional expected returns and the cross-sectional distribution of the remaining betas, $\beta_{it,-1}$, may vary over time. Sorting on a single target variable, therefore, averages over the remaining exposures rather than holding them fixed and does not generally recover the SACER at any specified value of $(b_1, b_{-1}^\top)^\top$.

We can use the theoretical structure introduced in earlier sections to describe the asymptotic distribution of this marginal sorting estimator which we label as $\widehat{\mu}_{|1}(b_1; \mathfrak{q})$. Define the average beta of assets with target risk factor $\beta_1 = b_1$ at time t as $\vartheta_{t|1}(b_1) = \mathbb{E}_t^\beta[\beta_{it} \mid \beta_{it,1} = b_1]$. Under corresponding regularity and rate conditions, the leading expansion of the marginal beta-sorted

[14] For recent examples, see Bali et al. (2017), Chen et al. (2021), Fan et al. (2022), and Chen et al. (2023).

[15] Formally, $\mathbb{E}_t^\beta[\cdot \mid \beta_{it,1} = b_1] = \mathbb{E}[\cdot \mid \mathcal{D}_{t-1}, \beta_{it,1} = b_1]$ where $\mathcal{D}_{t-1}$ is the design sigma field which we suppress for simplicity. The design sigma field and the related filtrations are defined in Section SA-II.1.2 of SA-Theory.

portfolio estimator is

$$\widehat{\mu}_{|1}(b_1;\mathfrak{q}) - \bar{\mu}_{T|1}(b_1;H) = \frac{1}{T-H}\sum_{t=H+1}^{T} \vartheta_{t|1}(b_1)^\top \big(f_t - \mathbb{E}[f_t \mid \mathcal{G}_{t-1}]\big) + \bar{\mathscr{I}}_{|1}(b_1) + \mathscr{R}_{|1}(b_1), \qquad (5.2)$$

where $\bar{\mathscr{I}}_{|1}(b_1) = O_{\mathbb{P}}\big(\sqrt{J/(nT)}\big)$ represents the contribution from the idiosyncratic component and $\mathscr{R}_{|1}(b_1)$ collects all other remainder terms. As before, when the risk factor component is nondegenerate (and the remainder is negligible) the estimator possesses a $\sqrt{T}$ rate. Importantly, setting $b_1 = 0$ does not generally eliminate the factor component because the average exposures to the remaining factors may be nonzero. The stronger condition $\vartheta_{t|1}(0) = 0$ for all t eliminates the common-factor component at $b_1 = 0$. Our theoretical framework could be extended to conduct inference on $\bar{\mu}_{T|1}(b_1;H)$ using a variance estimator analogous to $\widehat{\Sigma}(\beta)$. However, this marginal target depends on the joint distribution of risk-factor exposures and can change even when the conditional expected return function remains unchanged. Thus, care must be taken when translating empirical results based on marginal sorting to economic conclusions.

A more principled alternative for mitigating the sparse portfolio problem is to impose additive separability of conditional expected returns across the components of β. Specifically, suppose that

$$\mu_t(\beta) = \mu_{0t} + \sum_{k=1}^{d} \mu_{kt}(\beta_k), \qquad \mu_{kt}(0) = 0, \quad k = 1,\dots,d. \qquad (5.3)$$

Additive separability rules out interactions among factor exposures because the effect of changing β_k cannot depend on the remaining components of β. However, it still retains substantial flexibility as each $\mu_{kt}(\cdot)$ may be nonlinear and time varying. Notice that the no-arbitrage restriction is nested as the special case where $\mu_{kt}(\beta_k) = \beta_k \lambda_{kt}$.

Under additive separability, a lower-dimensional estimator is available. To provide intuition, consider the case where $\mathfrak{q} = 0$. Without additive separability, $\widehat{M}_t(\cdot;0)$ can be obtained, at each t, via a cross-sectional regression of R_{it} on approximately J^d indicator functions. In contrast, additive separability of $\mu_t(\cdot)$ directly implies additive separability of the realized systematic return function $M_t(\cdot)$. We can therefore rely on a cross-sectional regression of R_{it} on approximately dJ indicator functions instead.[16]

[16] A fully additively separable specification is not necessary. Partially additive specifications can allow interactions within low-dimensional groups of risk factor exposures while retaining much of the reduction in dimensionality. In

Importantly, additive separability alone does not imply that the marginal sorting estimator discussed above recovers the component $\mu_{1t}(b_1)$. Under the additive model in equation (5.3),

$$\mu_{t|1}(b_1) = \mu_{0t} + \mu_{1t}(b_1) + \sum_{k=2}^{d} \mathbb{E}_t^{\beta}[\mu_{kt}(\beta_{it,k}) \mid \beta_{it,1} = b_1]\,. \tag{5.4}$$

When the risk-factor exposures are dependent, each summand in the final term can vary with b_1, which can alter both the level and the shape of $\mu_{t|1}(b_1)$. In the special case in which each summand is (almost surely) constant in b_1, marginal sorting recovers the shape and contrasts of $\mu_{1t}(b_1)$. However, neither zero correlation nor zero rank correlation across the dimensions of risk-factor exposures is sufficient for this conclusion. Joint additive estimation avoids this restriction by controlling flexibly for the other risk-factor exposure components.

A further useful feature of the additive estimator is that it retains a well-defined interpretation when additive separability is misspecified. At each t, the conditional expected return target becomes the best mean-square approximation to $\mu_t(\cdot)$ within the space of additively separable functions, where mean-square error is computed using $\mathbb{E}_t^{\beta}[\cdot]$. This interpretation follows because the factor component of $M_t(\cdot)$ is already linear in β (and therefore already additively separable). The time-averaged estimator can thus be understood as targeting the sample average of these additive approximations rather than the unrestricted SACER.

Remark 5.1 (Conditional sorts)**.** Another common approach to portfolio construction is a conditional, or sequential, sort.[17] For example, to study exposure β_1 while controlling for exposure β_2, researchers first sort assets into portfolios using the marginal quantiles of $\widehat{\beta}_{it,2}$ and then, within each of those portfolios, sort assets using *conditional* quantiles of $\widehat{\beta}_{it,1}$. The procedure can be repeated recursively with additional risk-factor exposures. Because the breakpoints at each stage depend on portfolio membership at the preceding stages, the resulting partition is not a tensor product of marginal-quantile partitions and generally depends on the ordering of the procedure. However, the complete collection of final portfolios generated by the sequential sorts partitions the

that case, the dimensionality cost is governed by the largest group of exposures allowed to interact rather than by the full dimension d. For example, we could have $\mu_t(\beta) = \mu_{0t} + \mu_{12t}(\beta_1, \beta_2) + \sum_{k=3}^{d} \mu_{kt}(\beta_k)$ so the dimension would be of the order $J^2 + (d-2)J$.

[17] For examples of conditional sorts involving estimated risk-factor exposures, see Herskovic et al. (2016), Bollerslev et al. (2016), and Bali et al. (2017).

observed cross-section.[18] Thus, provided that the resulting cells shrink appropriately and their boundaries within the support are recorded, an estimator evaluated at a fixed β continues to target the SACER. Importantly, this statement requires keeping track of the part of the support each portfolio represents. For example, the fifth β_1 portfolio within a low-β_2 group can cover a very different range of β_1 from the fifth β_1 portfolio within a high-β_2 group. If instead portfolio returns or high-minus-low spreads are averaged across parent portfolios based only on their sequential ranks, without explicitly keeping track of the β ranges represented by each portfolio, as is common in empirical applications, the target becomes a weighted average of conditional expected returns or return differences, respectively, and does not generally identify the SACER or a SACER contrast at a fixed beta.

Although conditional sorts can ensure nonempty portfolios when the cross-section is sufficiently large, they do not eliminate the curse of dimensionality. With J portfolios formed for each of d exposures, there are still J^d final portfolios containing approximately n_t/J^d assets each. Thus, conditional sorts alleviate empty portfolios caused by dependence among the risk-factor exposures, but not the sparsity caused by increasing the number of exposures. Our high-level framework can accommodate conditional sorts, but providing primitive conditions under which our results apply is challenging because the recursively estimated breakpoints require, throughout the sequential sorting procedure, uniform control over both the β ranges represented by the portfolios and the number of assets they contain.

5.2 Factor-cancelling portfolios and linear-pricing restrictions

The stochastic expansion in Section 4.2 shows that, for $\beta \neq 0$, the common risk-factor component is the leading source of uncertainty for the beta-sorted portfolio estimator. Inspecting equation (4.3), we observe that this term is linear in β; consequently, it can be eliminated for judiciously chosen linear contrasts. Let $\mathsf{b}_1, \ldots, \mathsf{b}_{K_\Psi}$ be fixed exposure vectors, let $a_1, \ldots, a_{K_\Psi}$ be fixed weights, and define $\widehat{\Psi}(\mathsf{q}) = \sum_{k=1}^{K_\Psi} a_k \widehat{\mu}(\mathsf{b}_k; \mathsf{q})$ and its target $\bar{\Psi}_T = \sum_{k=1}^{K_\Psi} a_k \bar{\mu}_T(\mathsf{b}_k; H)$. If the weights and

[18] Although the final portfolios partition the observed cross-section, covering the entire support $\mathcal{B}$ requires extending the outer conditional intervals within every parent portfolio to the corresponding boundaries of $\mathcal{B}$.

evaluation points satisfy

$$\sum_{k=1}^{K_\Psi} a_k \mathsf{b}_k = 0, \tag{5.5}$$

then the leading common risk-factor component in the stochastic expansion cancels exactly.

A specific example of such a linear contrast is a "butterfly" trade where one goes long one unit of each of two assets, one with risk-factor exposure b_1 and one with risk-factor exposure b_3, and short two units of an asset with risk-factor exposure b_2. Then, if $\mathsf{b}_1 - \mathsf{b}_2 = \mathsf{b}_2 - \mathsf{b}_3$ and we take weights as $(1, -2, 1)$, equation (5.5) is satisfied. The estimator for the sample average of the conditional expected returns of such a trade is then

$$\widehat{\Psi}_{\mathtt{BF}}(\mathsf{q}) = \widehat{\mu}(\mathsf{b}_1; \mathsf{q}) - 2\widehat{\mu}(\mathsf{b}_2; \mathsf{q}) + \widehat{\mu}(\mathsf{b}_3; \mathsf{q}). \tag{5.6}$$

This contrast is a discrete second difference and therefore a measure of curvature in the sample-average conditional expected return function.

Consistency of $\widehat{\Psi}(\mathsf{q})$ follows directly under the conditions in Section 3 and Theorem SA-I.2.1 of SA-Theory because the contrast is just a fixed finite linear combination of consistent estimators. In fact, consistency holds regardless of whether equation (5.5) is satisfied. Inference, on the other hand, does not follow directly from the pointwise results in Section 4.2. Because the factor-cancelling restriction eliminates the leading common risk-factor component, the distributional properties of $\widehat{\Psi}(\mathsf{q}) - \bar{\Psi}_T$ in the limit are akin to those of $\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H)$ when $\beta = 0$.[19] The rate of convergence is now governed by the idiosyncratic component in equation (4.2). Under the conditions in Theorem SA-II.9.1 of SA-Theory, it is shown that the estimator enjoys the faster $\sqrt{nT/J^d}$ rate. In order to derive these properties, however, we require a partition anchored at the prespecified evaluation points, as detailed in Section SA-II.9 of SA-Theory. This construction is the analogue, for factor-cancelling linear contrasts, of the partition anchored at zero (see Remark 4.2).

A natural hypothesis of interest is whether conditional expected returns satisfy the restrictions implied by the absence of arbitrage opportunities. As we have discussed, no-arbitrage conditions

[19] Ordinary high-minus-low portfolios do not generally satisfy condition (5.5) and therefore retain the common risk-factor component.

restrict the form of $\mu_t(\beta)$ to be linear (for excess returns) or affine (for raw returns). One approach would be to test whether the time-averaged conditional expected return of a butterfly trade equals zero. A broader approach tests the linear or affine specification jointly across a collection of values for the risk-factor exposures, which is the approach we propose.

Here, we will focus on the affine case. Consider the following null hypothesis,

$$\mathsf{H}_0^{\mathtt{LIN}} : \mu_t(b) = a_t + c_t^\top b, \qquad b \in \mathcal{B}, \quad t = 1, \ldots, T, \tag{5.7}$$

where the intercept and slopes may vary over time. Under the null, realized returns satisfy

$$R_{it} = a_t + \varpi_t^\top \beta_{it} + \varepsilon_{it}, \tag{5.8}$$

where $\varpi_t = c_t + f_t - \mathbb{E}[f_t \mid \mathcal{G}_{t-1}]$. Thus, the realized cross-sectional slope is ϖ_t rather than c_t because it absorbs the contemporaneous risk-factor innovation. If the β_{it} were known, then a_t and ϖ_t could be estimated by a cross-sectional regression at each t.

The procedure evaluates equation (5.7) at a fixed collection of values for the risk-factor exposures: $\mathsf{b}_1, \ldots, \mathsf{b}_{K_{\mathtt{LIN}}}$. At each date t, we construct two beta estimates using adjacent, nonoverlapping first-step windows. The earlier estimate, $\widehat{\beta}_{it}$, uses observations dated $t-H, \ldots, t-1$, and the later estimate, $\widehat{\beta}_{i(t+H)}$, uses observations dated $t, \ldots, t+H-1$ and is paired with the outcome $R_{i(t+H)}$.

Because the later beta estimate contains first-step estimation error, an OLS regression of $R_{i(t+H)}$ on $\widehat{\beta}_{i(t+H)}$ would generally exhibit errors-in-variables bias. To address this bias in the linear fit used for the test, we estimate the relation by instrumental variables, using $\widehat{\beta}_{it}$ as an instrument for $\widehat{\beta}_{i(t+H)}$. Since the two windows do not overlap, the instrument is predetermined with respect to the innovations entering the later beta estimate and asset return. Relevance additionally requires that the earlier beta estimate retain sufficient cross-sectional information about the later beta. This condition is natural in our setting because risk-factor exposures would be expected to evolve gradually over short horizons, preserving the bulk of the cross-sectional information across adjacent estimation windows.

Let $\widehat{a}_{t+H}$ and $\widehat{\varpi}_{t+H}$ denote the resulting IV coefficient estimates and define

$$\widehat{e}_{i(t+H)} = R_{i(t+H)} - \widehat{a}_{t+H} - \widehat{\varpi}_{t+H}^{\top}\widehat{\beta}_{i(t+H)}. \tag{5.9}$$

For an evaluation point b, let $\widehat{w}_{it}(b)$ denote the weight assigned to asset i by the beta-sorted portfolio estimator using the beta estimates from the earlier window. The test averages the residual returns as follows:

$$\widehat{\nu}(b) = \frac{1}{T-2H}\sum_{t=H+1}^{T-H}\sum_{i\in\mathcal{I}_t^{\mathtt{LIN}}}\widehat{w}_{it}(b)\widehat{e}_{i(t+H)}, \tag{5.10}$$

where $\mathcal{I}_t^{\mathtt{LIN}}$ contains the assets observed in both windows and whose return, $R_{i(t+H)}$, is non-missing. The IV regression, portfolio partition, and normalization of portfolio weights all use this matched sample. Under the null, these residual portfolio returns should be close to zero.

Stacking the moments over the fixed grid gives $\widehat{\boldsymbol{\nu}}_{\mathtt{LIN}} = (\widehat{\nu}(\mathsf{b}_1), \ldots, \widehat{\nu}(\mathsf{b}_{K_{\mathtt{LIN}}}))^{\top}$. The Wald statistic is $\mathsf{W}_{\mathtt{LIN}} = \widehat{\boldsymbol{\nu}}_{\mathtt{LIN}}^{\top}\widehat{\Sigma}_{\mathtt{LIN}}^{-1}\widehat{\boldsymbol{\nu}}_{\mathtt{LIN}}$, where $\widehat{\Sigma}_{\mathtt{LIN}}$ accounts for dependence across the grid points and the serial dependence from overlapping observations. Under the null, $\mathsf{W}_{\mathtt{LIN}}$ converges to a chi-squared distribution with $K_{\mathtt{LIN}}$ degrees of freedom, so large values provide evidence against the affine specification at one or more of the selected exposure points. Because the grid is fixed and the moments are averaged over time, this is not an omnibus test over the entire support and nonlinear departures from the affine specification that offset across dates need not be detected. Technical conditions and the variance estimator are provided in Section SA-II.10 of SA-Theory.

6 Empirical application

In this section, we introduce a novel risk factor and show that it predicts both the cross-sectional and time-series behavior of U.S. stock returns. In Section SA-IV.2.2 of SA-Applied, we present a second empirical application that revisits the capital share factor of Lettau et al. (2019). Together, these applications illustrate the practical value of the new theoretical and inferential results developed earlier in the paper.

Our risk factor is a new measure of the business credit cycle. Standard measures of the business credit cycle use ratios of credit aggregates to measures of output. Although theoretically

appealing, these ratios can be difficult to interpret because their movements may reflect changes in the composition of the underlying aggregates rather than changes in credit conditions relative to economic activity. Here we take a different approach. We rely on the Federal Reserve's Senior Loan Officer Opinion Survey[20] (SLOOS) to measure credit conditions and the ISM Manufacturing Index to measure economic activity. This construction has three distinct advantages. First, as the SLOOS and ISM can both be expressed as diffusion indices, their behavior is less sensitive to changes in the structure of the economy. Second, they are much more timely than credit aggregates and national accounts data which tend to be released with a substantial lag. Third, they are not subject to revision outside of updated seasonal factors for the ISM index. Thus, we have a timely factor which we can evaluate in real time with no look-ahead bias.

Our factor is constructed as

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

where SLOOS_t is the net percentage of large domestic banks tightening standards for commercial and industrial loans to all firms and ISM_t is the ISM Manufacturing Index. For both series we date each observation to the month in which it became publicly available.[21] Although both the SLOOS and the ISM are based on questions with a similar format, they are scaled differently and so the affine transformation of the SLOOS is implemented so that they are both on the same scale (between 0 and 100). To understand why this is (the inverse of) a credit-to-output type measure, note that a fall in the SLOOS corresponds to easier lending standards and a fall in the ISM to weaker economic activity. Thus, a low value of the business-credit factor is intended to capture greater credit availability relative to economic activity. The reverse interpretation applies when the business-credit factor is high. Our factor series begins in January 1965 representing the first SLOOS observation.

As a preliminary check for the validity of our factor, we assess its ability to predict future market

[20] For background information on the SLOOS and its relationship to lending and economic activity, see Lown et al. (2000) and Lown and Morgan (2006).

[21] The Senior Loan Officer Opinion Survey is currently conducted on a quarterly basis. To construct a monthly series we keep the SLOOS value constant until a new value is available. For the period from 1984m1 through 1990m1, the credit standards question was not included in the SLOOS, and we use the net willingness to make consumer installment loans by large domestic banks, with its sign reversed to match the tightening convention. For the initial period from 1965m1 through 1967m9, we use the credit-standards series for all banks and all firms.

returns. Motivated by equation (2.1), we ask whether the persistence of the factor translates into predictability of future equity returns. We estimate predictive regressions of the form

$$r_{t,t+\ell} = a^{(\ell)} + b^{(\ell)} z_t + v^{(\ell)}_{t+\ell}, \tag{6.2}$$

where $r_{t,t+\ell}$ is the cumulative market excess return over the following ℓ months and z_t denotes the predictor under consideration. We compare our risk factor with the benchmark predictors studied by Welch and Goyal (2008). In Table 1 we present in-sample R^2 from predictive regressions for forecast horizons of 1, 3, 6, and 12 months ahead. The first fourteen rows report results for the benchmark predictors. The row labeled "CGP" reports results using only the SLOOS component of our factor, as in Chava et al. (2015), while the row labeled "BCF" reports results for our business-credit factor. The results are stark. The in-sample R^2 from our new risk factor far outstrips that of the other predictors considered. The out-of-sample results, reported in Table SA-IV.2.1 of SA-Applied, lead to the same conclusion, namely, that our risk factor outperforms each of the other predictors by a wide margin.

We now investigate how the business-credit factor performs in explaining the cross-section of equity returns. We use monthly raw returns from the Center for Research in Security Prices (CRSP) from January 1965 through December 2019. We retain common shares listed on the New York Stock Exchange, American Stock Exchange, or Nasdaq and follow Bali et al. (2016) in incorporating delisting returns. The first 120 months are reserved for initial first-step estimation, so the reported expected-return estimates cover January 1975 through December 2019. We select the first-step window H_{it} and second-step number of portfolios J_t using the data-driven approach described in Section SA-III.1 of SA-Applied. Within each window, we use all available nonmissing observations for an asset and require at least 36 usable prior observations to estimate β_{it} and include the asset in that month's portfolio sort. Further implementation details are provided in Section SA-III.4 of SA-Applied.

Figure 1 reports the estimated SACER, $\widehat{\mu}(\beta)$, over a grid ranging from -1.25 to 1.25 in increments of 0.25. The black line is the $(\mathfrak{p}, \mathfrak{q}) = (0, 0)$ point estimate. The estimated SACER exhibits a clear negative relationship with beta. It is around 1.5 percent per month at negative beta values and around 1 percent at positive beta values. The overall pattern indicates that greater exposure

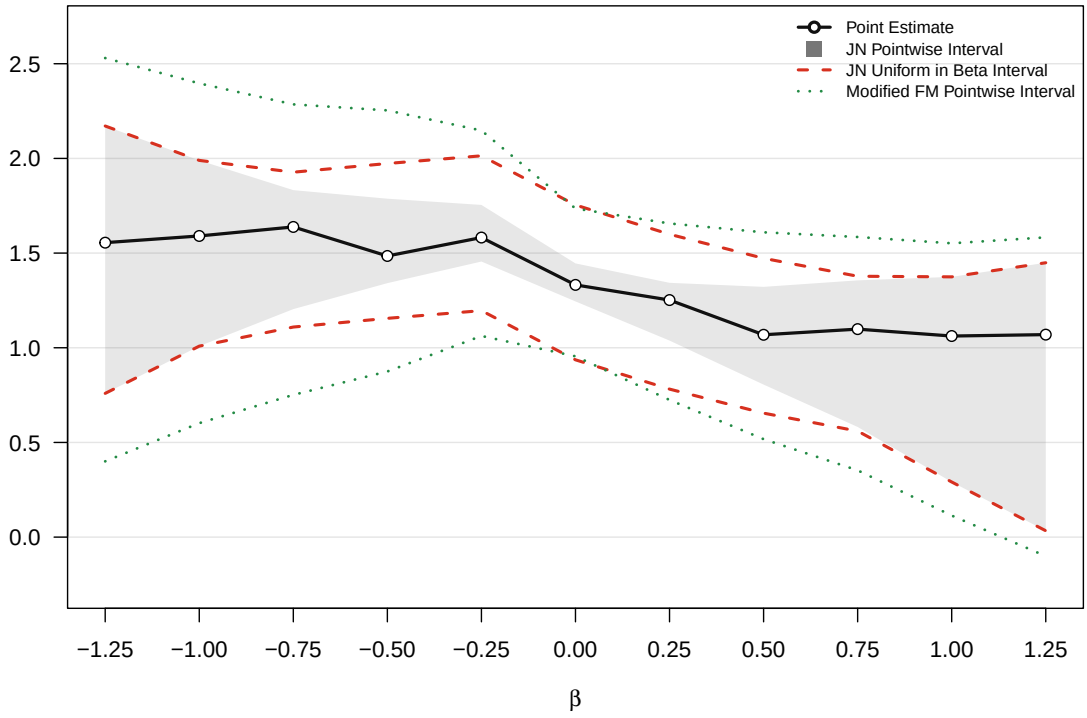
Table 1: **In-sample Predictive Regressions** This table reports R^2 (in percent) from predictive regressions of cumulative market excess returns on an individual predictor from Welch and Goyal (2008) for horizons of 1, 3, 6, and 12 months ahead. The row labeled "CGP" reports results for the SLOOS only portion of our risk factor as studied in Chava et al. (2015). The row labeled "BCF" reports results for our proposed risk factor. The sample period is 1965m1–2019m12.

	$\ell = 1$	$\ell = 3$	$\ell = 6$	$\ell = 12$
(log) Dividend Price Ratio	0.09	0.31	0.71	1.45
(log) Dividend Yield	0.11	0.32	0.74	1.48
(log) Earnings Price Ratio	0.03	0.04	0.06	0.25
(log) Dividend Payout Ratio	0.02	0.20	0.56	0.74
Stock Variance	1.01	0.11	0.15	0.65
Book-to-Market Ratio	0.00	0.01	0.07	0.12
Net Equity Expansion	0.15	0.22	0.48	1.25
Treasury Bill Yield	0.40	0.79	1.09	1.57
Long Term Treasury Yield	0.13	0.17	0.14	0.00
Long Term Treasury Return	1.16	0.71	1.88	1.34
Term Spread	0.52	1.38	2.50	6.99
Default Yield Spread	0.26	0.82	2.31	2.88
Default Return Spread	0.30	0.24	0.29	0.02
(lagged) Inflation	0.01	0.46	1.84	2.29
CGP	1.26	3.04	3.65	4.06
BCF	2.89	7.87	10.82	12.68

to the business-credit factor is associated with lower expected returns.

We use our new theoretical results to conduct inference on the SACER. In the figure, the grey region is the 95% pointwise interval based on our full variance estimator, $\widehat{\Sigma}(\beta)$ (denoted JN), while the green dotted lines delimit the 95% pointwise interval based on the modified FM variance estimator, $\widehat{\Sigma}_{\mathtt{FM}}(\beta; 1)$. Both pointwise intervals use $(\mathfrak{p}, \mathfrak{q}) = (1, 1)$, and therefore need not be centered on the black line. Using our data-driven choices of tuning parameters, these intervals are asymptotically valid at fixed nonzero evaluation points (see Lemma SA-III.1.1 of SA-Applied). The red dashed lines report the 95% JN uniform-in-beta interval discussed in Remark 4.7. In this application, the estimated boundaries are approximately $\beta = \pm 0.79$, so the uniform adjustment applies to the seven reported grid points from -0.75 through 0.75. Consistent with the theory, Figure 1 shows that the uniform-in-beta adjustment widens the interval near zero, where the risk-factor component of the variance becomes small, while leaving the JN pointwise interval unchanged away from zero. The figure also shows that the modified FM pointwise intervals are wider than the JN pointwise intervals at every reported value of β.

Figure 1: **Inference on the SACER.** The black line reports the $(\mathfrak{p}, \mathfrak{q}) = (0, 0)$ estimate of the SACER. The grey region is the 95% JN pointwise interval. The red dashed lines are the 95% JN uniform-in-beta intervals, and the green dotted lines are the 95% modified FM pointwise intervals. All three intervals use $(\mathfrak{p}, \mathfrak{q}) = (1, 1)$. Results are reported for monthly returns (in percent). The sample period is 1965m1–2019m12 with a 120-month burn-in period.



We quantify the economic magnitude of the downward slope in Figure 1 using a high-minus-low contrast. Table 2 provides results for the contrast between $\beta = -0.75$ and 0.75. We choose these two values as the contrast endpoints since they reside in the realized support in almost every month (both endpoints lie within the realized range in 94.6% of the months). The point estimate of this contrast, based on the $(\mathfrak{p}, \mathfrak{q}) = (0, 0)$ estimator, is -0.540 percentage points per month corresponding to the difference between the two points on the black line in Figure 1. Table 2 presents the results from our new inference procedures. The JN standard error is 0.304, giving a (two-sided) p-value of 0.070. The modified FM variance estimator produces a slightly smaller standard error, giving a two-sided p-value of 0.054. Thus, both the JN and modified FM procedures reject a zero contrast at the 10% level. As a reference, using $\widehat{\Sigma}_f$ alone gives a smaller standard error and a p-value of 0.019. Although both endpoints are away from zero, so the idiosyncratic component is asymptotically of smaller order, we recommend the full JN variance estimator because it retains

this source of uncertainty. Finally, applying the affine specification test from Section 5.2 to the seven central grid points yields a Wald statistic of 7.871, with an associated chi-squared p-value of 0.344. Thus, we do not reject the null of an affine specification at conventional significance levels.

Table 2: **High Minus Low Contrast.** The table reports inference for $\widehat{\mu}(0.75) - \widehat{\mu}(-0.75)$ in monthly percentage points. Each row reports the standard error, 90% confidence interval, and two-sided normal-approximation p-value under the JN, modified FM, or $\widehat{\Sigma}_f$-only variance estimator. The confidence intervals are centered at the $(\mathfrak{p}, \mathfrak{q}) = (1, 1)$ estimate.

Variance estimator	SE	90% CI	p-value
JN	0.304	$[-1.049, -0.050]$	0.070
Modified FM	0.285	$[-1.019, -0.080]$	0.054
$\widehat{\Sigma}_f$ only	0.234	$[-0.934, -0.165]$	0.019

7 Conclusion

Beta-sorted portfolios are a commonly used empirical tool in asset pricing. In a first step, time-varying factor exposures are estimated by rolling regressions of asset returns on observable risk factors to measure how returns co-vary with those factors. In a second step, assets with similar risk-factor exposures are grouped into portfolios, whose returns are compared across different exposure levels. Yet the simple and intuitively appealing algorithm belies a more complicated statistical setting involving a two-step procedure in which both stages involve nonparametric estimation.

We provide a comprehensive econometric framework that rationalizes this commonly used estimator and links its properties to the first-step estimation window and the number of portfolios formed in the second step. We establish uniform consistency and pointwise asymptotic normality for the time-averaged beta-sorted portfolio estimator, showing that its rate of convergence is $\sqrt{T}$ at fixed nonzero risk-factor exposures but a faster rate of $\sqrt{nT/J^d}$ at the origin. We also show that certain weighted linear combinations can attain this faster rate of convergence. We introduce a new variance estimator, characterize the properties of the ubiquitous Fama-MacBeth variance estimator, and provide inference procedures for hypotheses of interest. Finally, we identify limitations of current practice and provide guidance on implementation and interpretation.

We apply our new theoretical results in two substantive empirical applications. First, we introduce a new risk factor designed to capture the business-credit cycle. We find a downward-sloping relationship between expected returns and exposure to this factor, with a statistically significant

high-minus-low contrast using our proposed inference procedures. Second, we revisit the capital share risk factor proposed by Lettau et al. (2019). Using our new inference procedures, we do not find a statistically significant relation in the cross-section of U.S. equity returns.

References

Adrian, T., Crump, R. K., Moench, E., 2015. Regression-based estimation of dynamic asset pricing models. Journal of Financial Economics 118, 211–244.

Andrews, D. W. K., 2005. Cross-section regression with common shocks. Econometrica 73, 1551–1585.

Ang, A., Kristensen, D., 2012. Testing conditional factor models. Journal of Financial Economics 106, 132–156.

Ang, A., Liu, J., Schwarz, K., 2020. Using stocks or portfolios in tests of factor models. Journal of Financial and Quantitative Analysis 55, 709–750.

Bali, T. G., Brown, S. J., Tang, Y., 2017. Is economic uncertainty priced in the cross-section of stock returns? Journal of Financial Economics 126, 471–489.

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

Barras, L., Gagliardini, P., Scaillet, O., 2022. Skill, scale, and value creation in the mutual fund industry. Journal of Finance 77, 601–638.

Bollerslev, T., Li, S. Z., Todorov, V., 2016. Roughing up beta: Continuous versus discontinuous betas and the cross section of expected stock returns. Journal of Financial Economics 120, 464–490.

Boons, M., Duarte, F., de Roon, F., Szymanowska, M., 2020. Time-varying inflation risk and stock returns. Journal of Financial Economics 136, 444–470.

Borri, N., Chetverikov, D., Liu, Y., Tsyvinski, A., 2024. One factor to bind the cross-section of returns. Working Paper 32365, NBER.

Cattaneo, M. D., Crump, R. K., Farrell, M. H., Feng, Y., 2024. On binscatter. American Economic Review 114, 1488–1514.

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

Cattaneo, M. D., Farrell, M. H., Feng, Y., 2020b. Large sample properties of partitioning-based series estimators. Annals of Statistics 48, 1718–1741.

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

Chang, B. Y., Christoffersen, P., Jacobs, K., 2013. Market skewness risk and the cross section of stock returns. Journal of Financial Economics 107, 46–68.

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

Chen, Y., Han, B., Pan, J., 2021. Sentiment trading and hedge fund returns. Journal of Finance 76, 2001–2033.

Chen, Z., Da, Z., Huang, D., Wang, L., 2023. Presidential economic approval rating and the cross-section of stock returns. Journal of Financial Economics 147, 106–131.

Chen, Z., Lu, A., Zhu, X., 2025. Investor sentiment and the pricing of macro risks for hedge funds. Management Science 71, 1623–1645.

Cochrane, J. H., 1996. A cross-sectional test of an investment-based asset pricing model. Journal of Political Economy 104, 572–621.

Cochrane, J. H., 2005. Asset Pricing. Princeton University Press, Princeton, revised ed.

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

Connor, G., Linton, O., 2007. Semiparametric estimation of a characteristic-based factor model of common stock returns. Journal of Empirical Finance 14, 694–717.

Eisdorfer, A., Froot, K., Ozik, G., Sadka, R., 2022. Competition links and stock returns. Review of Financial Studies 35, 4300–4340.

Fama, E. F., MacBeth, J. D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81, 607–636.

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

Farmer, L. E., Schmidt, L., Timmermann, A., 2023. Pockets of predictability. Journal of Finance 78, 1279–1341.

Gagliardini, P., Ossola, E., Scaillet, O., 2016. Time-varying risk premium in large cross-sectional equity data sets. Econometrica 84, 985–1046.

Gagliardini, P., Ossola, E., Scaillet, O., 2020. Estimation of large dimensional conditional factor models in finance. In: Durlauf, S. N., Hansen, L. P., Heckman, J. J., Matzkin, R. L. (eds.), *Handbook of Econometrics*, Elsevier, vol. 7A, chap. 3, pp. 219–282.

Gârleanu, N., Pedersen, L. H., 2011. Margin-based asset pricing and deviations from the law of one price. Review of Financial Studies 24, 1980–2022.

Giglio, S., Xiu, D., 2021. Asset pricing with omitted factors. Journal of Political Economy 129, 1947–1990.

Goldberg, J., Nozawa, Y., 2021. Liquidity supply in the corporate bond market. Journal of Finance 76, 755–796.

Gospodinov, N., Kan, R., Robotti, C., 2014. Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors. Review of Financial Studies 27, 2139–2170.

Gospodinov, N., Robotti, C., 2013. Asset pricing theories, models, and tests. In: Baker, H. K., Filbeck, G. (eds.), *Portfolio Theory and Management*, Oxford University Press, chap. 3, pp. 46–72.

Goyal, A., 2012. Empirical cross-sectional asset pricing: A survey. Financial Markets and Portfolio Management 26, 3–38.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Review of Financial Studies 33, 2223–2273.

Herskovic, B., Kelly, B. T., Lustig, H., Van Nieuwerburgh, S., 2016. The common factor in idiosyncratic volatility: Quantitative asset pricing implications. Journal of Financial Economics 119, 249–283.

Jagannathan, R., Wang, Z., 1998. An asymptotic theory for estimating beta-pricing models using cross-sectional regression. Journal of Finance 53, 1285–1309.

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

Kleibergen, F., 2009. Tests of risk premia in linear factor models. Journal of Econometrics 149, 149–173.

Lettau, M., Ludvigson, S. C., Ma, S., 2019. Capital share risk in U.S. asset pricing. Journal of Finance 74, 1753–1792.

Lown, C. S., Morgan, D. P., 2006. The credit cycle and the business cycle: New findings using the loan officer opinion survey. Journal of Money, Credit and Banking 38, 1575–1597.

Lown, C. S., Morgan, D. P., Rohatgi, S., 2000. Listening to loan officers: The impact of commercial credit standards on lending and output. FRBNY Economic Policy Review 6, 1–16.

Markowitz, H., 1952. Portfolio selection. Journal of Finance 7, 77–91.

Markowitz, H., 1959. Portfolio Selection: Efficient Diversification of Investments. John Wiley & Sons, New York.

Nagel, S., 2013. Empirical cross-sectional asset pricing. Annual Review of Financial Economics 5, 167–199.

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Raponi, V., Robotti, C., Zaffaroni, P., 2020. Testing beta-pricing models using large cross-sections. Review of Financial Studies 33, 2796–2842.

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

Tobin, J., 1958. Liquidity preference as behavior towards risk. Review of Economic Studies 25, 65–86.

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

Beta-Sorted Portfolios
Supplemental Appendix: Theory

Matias D. Cattaneo* Richard K. Crump† Yingjie Feng‡ Weining Wang§

September 22, 2026

Abstract

This supplemental appendix gives the theoretical details for beta-sorted portfolio estimation and inference. Part I develops general high-level consistency and stochastic linearization results for beta-sorted estimators based on generated regressors and data-dependent portfolios. Part II specializes these results to the classical beta-sorted portfolio construction based on rolling-window first-step regressions and marginal-quantile portfolios, including convergence rates, distributional approximations, feasible variance estimation, uniform-in-β inference, and factor-cancelling contrasts. The companion Supplemental Appendix: Applied provides implementation guidance, simulation evidence, and empirical applications.

*Department of Economics, Princeton University.
†Capital Markets, Federal Reserve Bank of New York.
‡School of Economics and Management, Tsinghua University.
§School of Economics, University of Bristol.

Contents

Part I: General Beta-Sorted Portfolios 1

SA-I.1 Setup 1
SA-I.1.1 Estimator . . . 2
SA-I.1.2 Notation . . . 3

SA-I.2 Convergence Rate 4
SA-I.2.1 Result . . . 4
SA-I.2.2 Discussion . . . 5
SA-I.2.3 Proof of Theorem SA-I.2.1 . . . 5

SA-I.3 Stochastic Linearization 8
SA-I.3.1 Decomposition . . . 9
SA-I.3.2 Result . . . 10
SA-I.3.3 Discussion . . . 11
SA-I.3.4 Proof of Lemma SA-I.3.1 . . . 12

Part II: Classical Beta-Sorted Portfolios 14

SA-II.1 Assumptions 14
SA-II.1.1 Primitive Conditions . . . 14
SA-II.1.2 Filtrations . . . 15
SA-II.1.3 Discussion . . . 16

SA-II.2 First-Step Beta Estimates 17
SA-II.2.1 Decomposition . . . 17
SA-II.2.2 Results . . . 18
SA-II.2.3 Discussion . . . 20
SA-II.2.4 Proof of Lemma SA-II.2.2 . . . 22
SA-II.2.5 Proof of Lemma SA-II.2.3 . . . 23
SA-II.2.6 Proof of Lemma SA-II.2.4 . . . 26

SA-II.3 Second-Step Partitioning 28
SA-II.3.1 Results . . . 30
SA-II.3.2 Discussion . . . 32
SA-II.3.3 Proof of Lemma SA-II.3.1 . . . 33
SA-II.3.4 Proof of Lemma SA-II.3.2 . . . 36
SA-II.3.5 Proof of Lemma SA-II.3.3 . . . 37

SA-II.4 Stochastic Linearization 39
SA-II.4.1 Decomposition . . . 39
SA-II.4.2 Results . . . 40
SA-II.4.3 Discussion . . . 41

SA-II.4.4 Proof of Lemma SA-II.4.1(a) . . . 42
SA-II.4.5 Proof of Lemma SA-II.4.1(b) . . . 52
SA-II.4.6 Proof of Lemma SA-II.4.2 . . . 57

SA-II.5 Convergence Rate **62**
SA-II.5.1 Results . . . 62
SA-II.5.2 Proof of Corollary SA-II.5.1 . . . 63
SA-II.5.3 Proof of Corollary SA-II.5.2 . . . 64

SA-II.6 Pointwise-in-β Distributional Approximation **64**
SA-II.6.1 Results . . . 65
SA-II.6.2 Discussion . . . 66
SA-II.6.3 Proof of Lemma SA-II.6.1 . . . 67
SA-II.6.4 Proof of Theorem SA-II.6.2 . . . 69

SA-II.7 Pointwise-in-β Variance Estimation **70**
SA-II.7.1 Results . . . 71
SA-II.7.2 Fama-MacBeth Variance Estimator . . . 73
SA-II.7.3 Discussion . . . 74
SA-II.7.4 Proof of Lemma SA-II.7.1 . . . 75
SA-II.7.5 Preliminary Results for Proof of Lemma SA-II.7.2 . . . 76
SA-II.7.6 Proof of Lemma SA-II.7.2 . . . 88
SA-II.7.7 Proof of Theorem SA-II.7.3 . . . 90
SA-II.7.8 Proof of Theorem SA-II.7.4 . . . 90

SA-II.8 Uniform-in-β Inference **95**
SA-II.8.1 Result . . . 95
SA-II.8.2 Proof of Theorem SA-II.8.1 . . . 97

SA-II.9 Factor-Cancelling Linear Contrasts **101**
SA-II.9.1 Setup and Linearization . . . 101
SA-II.9.2 Evaluation-Point-Anchored Partitioning . . . 102
SA-II.9.3 Result . . . 104
SA-II.9.4 Discussion . . . 105
SA-II.9.5 Proof of Theorem SA-II.9.1 . . . 105

SA-II.10 Fixed-Grid Test of Linearity **106**
SA-II.10.1 Testing Problem . . . 106
SA-II.10.2 IV Fit and Test Statistic . . . 108
SA-II.10.3 Decomposition . . . 109
SA-II.10.4 Results . . . 110
SA-II.10.5 Discussion . . . 113
SA-II.10.6 Proof of Lemma SA-II.10.1 . . . 115
SA-II.10.7 Proof of Lemma SA-II.10.2 . . . 119
SA-II.10.8 Proof of Theorem SA-II.10.3 . . . 128

SA-II.10.9 Proof of Lemma SA-II.10.4 128

Part I: General Beta-Sorted Portfolios

This part of the supplemental appendix considers a general portfolio scheme and establishes two main results under high-level conditions for the beta-sorted portfolio estimator: (i) uniform consistency, and (ii) stochastic linearization. Part II specializes this setup to the classical beta-sorted portfolio estimator, where portfolios are constructed using first-step local rolling-window regressions and second-step quantile-spaced partitions. Part II also provides primitive conditions for the high-level conditions imposed here.

SA-I.1 Setup

We consider the following asset return model throughout:

$$R_{it} = \alpha_{it} + \beta_{it}^\top f_t + \varepsilon_{it}, \qquad \mathbb{E}[\varepsilon_{it}|\mathcal{F}_{t-1}, f_t] = 0, \qquad i = 1, \dots, n_t, \quad t = 1, \dots, T,$$

where α_{it} and β_{it} are $\mathcal{F}_{t-1}$-measurable, $f_t \in \mathbb{R}^d$ is a column vector of observable factors, and β_{it} takes values in a common compact support $\mathcal{B} \subset \mathbb{R}^d$ containing 0 as an interior point. For each finite sample, let n be a cross-sectional index bound such that $[n_t] \subseteq [n]$ for all $t \le T$, where for any positive integer m, $[m] := \{1, \dots, m\}$. We define

$$\begin{aligned}
\mathcal{G}_t &= \sigma\big(\{f_s\}_{s=1}^t, \{\aleph_{2s}\}_{s=1}^t\big), \\
\mathcal{F}_t &= \sigma\big(\{f_s\}_{s=1}^t, \{\{\varepsilon_{is}\}_{i=1}^{n_s}\}_{s=1}^t, \{\aleph_{1i}\}_{i=1}^n, \{\aleph_{2s}\}_{s=1}^t, \{\{\aleph_{3is}\}_{i=1}^{n_s}\}_{s=1}^t\big).
\end{aligned}$$

Here $\aleph_{1i}$, $\aleph_{2t}$, and $\aleph_{3it}$ are auxiliary random vectors that may enter the data-generating process. The predictable coefficients α_{it} and β_{it} are $\mathcal{F}_{t-1}$-measurable. Part II gives a primitive innovation-free representation in terms of factor, scale, and auxiliary histories dated at most $t-1$. Thus $\{\mathcal{F}_t\}$ and $\{\mathcal{G}_t\}$ are filtrations with $\mathcal{G}_t \subseteq \mathcal{F}_t$ and $\sigma(f_s : 1 \le s \le t) \subseteq \mathcal{G}_t$. We interpret $\mathcal{F}_t$ as the full information set and $\mathcal{G}_t$ as the factor-state information set. We assume that there is a fixed constant $c > 0$ such that $cn \le n_t \le n$ for all $t = 1, \dots, T$, and asymptotic approximations are taken as both n and T diverge to infinity.

We assume that there exists a nonrandom function $\mu_0(\cdot\,;\cdot)$ such that

$$\mathbb{E}[R_{it}|\mathcal{F}_{t-1}] = \mu_0(\beta_{it}; \mathcal{G}_{t-1}) =: \mu_t(\beta_{it}),$$

and the structural form is

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

We impose $\mathbb{E}[f_t|\mathcal{F}_{t-1}] = \mathbb{E}[f_t|\mathcal{G}_{t-1}]$ and therefore write

$$R_{it} = \mu_t(\beta_{it}) + \beta_{it}^\top \widetilde{f}_t + \varepsilon_{it}, \qquad \widetilde{f}_t = f_t - \mathbb{E}[f_t|\mathcal{G}_{t-1}].$$

It follows that

$$\mu_t(\cdot) = \mu_0(\cdot\,; \mathcal{G}_{t-1}), \qquad \alpha_t(b) := \mu_t(b) - b^\top \mathbb{E}[f_t \mid \mathcal{G}_{t-1}], \quad b \in \mathcal{B}, \qquad \alpha_{it} = \alpha_t(\beta_{it}).$$

Let $H = H_{nT}$ be an integer satisfying $1 \le H < T$. In this general part, H records the first-step

initialization period, so second-step estimation uses dates $t = H+1, \ldots, T$. Part II specializes H to the rolling-window length and imposes its asymptotic rate conditions.

At a fixed date, a cross-sectional regression of returns on risk exposures recovers the realized systematic return function

$$M_t(\beta) = \mu_t(\beta) + \beta^\top \widetilde{f}_t,$$

rather than $\mu_t(\beta)$ itself. The time average of the second term, $(T-H)^{-1}\sum_{t=H+1}^{T} \beta^\top \widetilde{f}_t$, has martingale-difference summands with respect to $\{\mathcal{G}_t\}$ because $\mathbb{E}[\beta^\top \widetilde{f}_t | \mathcal{G}_{t-1}] = 0$. Hence this average is asymptotically negligible for consistency under the conditions below, although it contributes to the first-order distribution when $\beta \neq 0$.

The goal is to estimate and conduct inference on the *sample average of conditional expected returns (SACER)*:

$$\bar{\mu}_T(\beta; H) = \frac{1}{T-H}\sum_{t=H+1}^{T} \mu_t(\beta).$$

SA-I.1.1 Estimator

The beta-sorted portfolio procedure targets $\bar{\mu}_T(\cdot)$ by averaging over time cross-sectional piecewise polynomial regression fits constructed using estimated partitions (portfolios). More precisely, the estimator is constructed in two steps:

(i) obtain an estimator $\widehat{\beta}_{it}$ of β_{it} for each $i = 1, \ldots, n_t$ and $t = H+1, \ldots, T$; and

(ii) for each $t = H+1, \ldots, T$, construct a collection of $\bar{J}_t$ disjoint cells $\widehat{\Delta}_t = \{\widehat{P}_{jt} : j \in [\bar{J}_t]\}$, and then define the following estimator of $\bar{\mu}_T(\beta; H)$:

$$\widehat{\mu}(\beta) = \frac{1}{T-H}\sum_{t=H+1}^{T} \widehat{M}_t(\beta), \qquad \widehat{M}_t(\beta) = \widehat{p}_t(\beta)^\top \left(\sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it})\widehat{p}_t(\widehat{\beta}_{it})^\top\right)^{-1} \sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it}) R_{it},$$

where $\widehat{p}_t(b)$ denotes a piecewise polynomial expansion based on the partition $\widehat{\Delta}_t$.

Without loss of generality, we write this piecewise polynomial expansion in cell-normalized coordinates. For each cell $\widehat{P}_{jt}$, choose a center $\widehat{c}_{jt}$ and a nonsingular scaling/rotation matrix $\widehat{D}_{jt}$ such that $\widehat{D}_{jt}^{-1}(b - \widehat{c}_{jt})$ is uniformly bounded over $b \in \widehat{P}_{jt}$. Let $\widehat{\jmath}_t(b)$ denote the index of the cell containing b. For a nonnegative integer $\mathfrak{q}$, let $\check{p}_{\mathfrak{q}}(z) = (z^s : |s| \leq \mathfrak{q})^\top$ denote the column vector collecting, in a fixed order, all d-variate monomials in z with total degree no greater than $\mathfrak{q}$, using the multi-index notation collected in Section SA-I.1.2. For any cell P equipped with a center c_P and nonsingular scaling/rotation matrix D_P, define the localized degree-$\mathfrak{q}$ block basis by

$$p_{\mathfrak{q}}(b, P) = \mathbb{1}(b \in P)\check{p}_{\mathfrak{q}}\big(D_P^{-1}(b - c_P)\big).$$

Cell indicators are understood with the fixed boundary tie-breaking convention for the corresponding partition, so a point contributes to at most one block. For any admissible partition $\Delta = \{P_j : j \in [\bar{J}]\}$ equipped

with its cell normalizations, define the associated stacked basis by

$$p(b;\Delta) = \big(p_{\mathfrak{q}}(b,P_1)^\top,\ldots,p_{\mathfrak{q}}(b,P_{\bar{J}})^\top\big)^\top.$$

We then take $\widehat{p}_t(b) = p(b;\widehat{\Delta}_t)$. This stacked-block form emphasizes that $\widehat{p}_t(b)$ is the product of a cell-indicator and the polynomial basis used within the selected cell. When $\mathfrak{q}=0$, $\check{p}_{\mathfrak{q}}$ is the scalar 1, so $p_{\mathfrak{q}}(b,P)=\mathbb{1}(b\in P)$. If $K_{\mathfrak{q}}$ denotes the number of monomials, then $\widehat{p}_t(\beta)\in\mathbb{R}^{\bar{J}_t K_{\mathfrak{q}}}$.

At this level of generality, we allow for generic choices of the first-step estimator $\widehat{\beta}_{it}$ of β_{it} and the partitioning scheme $\{\widehat{\Delta}_t\}_{t=H+1}^T$, while in the second step, for each $t=H+1,\ldots,T$, we allow for a piecewise polynomial regression of total degree $\mathfrak{q}$ to obtain $\widehat{M}_t(\cdot)$.

Although β_{it} is assumed to have bounded support $\mathcal{B}$, the first-step estimator $\widehat{\beta}_{it}$ may lie outside $\mathcal{B}$. Thus, we fix $\delta_{\mathcal{B}}>0$ and introduce a compact convex set $\mathcal{B}^+$ satisfying

$$\left\{b\in\mathbb{R}^d : \inf_{a\in\mathcal{B}}\|b-a\|\le\delta_{\mathcal{B}}\right\}\subseteq\mathcal{B}^+.$$

When $\widehat{\beta}_{it}$ is uniformly consistent for β_{it} in the sense that

$$\max_{H+1\le t\le T}\max_{i\in[n_t]}\|\widehat{\beta}_{it}-\beta_{it}\| = o_{\mathbb{P}}(1),$$

the estimated $\widehat{\beta}_{it}$ will belong to $\mathcal{B}^+$ for all $i\in[n_t]$ and $t=H+1,\ldots,T$, with probability approaching one. In the theoretical analysis below, μ_t is understood as a bounded continuous extension from $\mathcal{B}$ to $\mathcal{B}^+$ whenever it is evaluated at $\widehat{\beta}_{it}$ outside $\mathcal{B}$. Thus, $\mathcal{B}^+$ is only an extension device for evaluating functions at first-step estimates and along Taylor paths; it does not change the support $\mathcal{B}$.

For the interior results below, fix an evaluation region $\mathcal{B}^\circ\subset\mathrm{int}(\mathcal{B})$. For each t, define

$$\mathcal{J}_t = \{j\in[\bar{J}_t] : \widehat{P}_{jt}\cap\mathcal{B}^\circ\neq\emptyset\}.$$

Thus, $\mathcal{J}_t$ collects precisely the estimated cells that may contain evaluation points in $\mathcal{B}^\circ$.

SA-I.1.2 Notation

For a positive integer m, let $[m]=\{1,\ldots,m\}$. All vectors are column vectors unless explicitly transposed. For a generic vector $v\in\mathbb{R}^d$, v_j denotes its j-th element and $\|v\|$ denotes its Euclidean norm. For a matrix $A\in\mathbb{R}^{m\times n}$, $\|A\|$ denotes its operator norm; for a square symmetric matrix A, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote its minimum and maximum eigenvalues. For a scalar- or vector-valued random variable Y and $r\ge1$, write $\|Y\|_r=(\mathbb{E}\|Y\|^r)^{1/r}$, where the inner norm is absolute value for scalars and Euclidean norm for vectors. For a set $S\subseteq\mathbb{R}^d$, define its diameter as $\mathrm{diam}(S)=\sup\{\|s-s'\| : s,s'\in S\}$. We write $\otimes$ for the Kronecker product.

For $\ell,u\in\mathbb{R}^d$ such that $\ell\le u$ elementwise, define the rectangle

$$[\ell,u] = \{b=(b_1,\ldots,b_d)^\top : b_l\in[\ell_l,u_l],\ l\in[d]\},$$

where ℓ_l and u_l denote the l-th elements of ℓ and u. Thus, $\mathrm{diam}([\ell,u])=(\sum_{l\in[d]}(u_l-\ell_l)^2)^{1/2}$.

For a multi-index $m\in\mathbb{Z}_+^d$, let $|m|=\sum_{j=1}^d m_j$, and for $z\in\mathbb{R}^d$, write $z^m=z_1^{m_1}\cdots z_d^{m_d}$. For a nonnegative integer $\mathfrak{q}$, $(z^m : |m|\le\mathfrak{q})^\top$ denotes the column vector collecting, in a fixed order, all d-variate monomials in z

with total degree no greater than $\mathfrak{q}$. For a multivariate function $g : \mathbb{R}^d \to \mathbb{R}$, $\nabla g = \nabla^1 g$ denotes its gradient, $\nabla^2 g$ its Hessian, and

$$g^{(m)}(z) = \frac{\partial^{|m|}}{\partial z_1^{m_1} \cdots \partial z_d^{m_d}} g(z).$$

For a univariate function g, we write $g \in L^1(\mathbb{R})$ if $\int_{\mathbb{R}} |g(z)| \, \mathrm{d}z < \infty$. For a function $g : \mathcal{X} \to \mathbb{R}$ on a compact set $\mathcal{X} \subseteq \mathbb{R}^d$ and some $s > 0$, we say g is Hölder-s on $\mathcal{X}$ if all partial derivatives $g^{(m)}$ with $|m| \leq \underline{s}$ exist and are continuous on $\mathcal{X}$, and all partial derivatives with $|m| = \underline{s}$ are Hölder continuous with exponent $s - \underline{s}$ uniformly on $\mathcal{X}$, where $\underline{s} = \max\{k \in \mathbb{Z}_+ : k < s\}$. Whenever a function enters the analysis through its Hessian, we additionally require it to be twice continuously differentiable with a uniformly bounded Hessian on its domain. This requirement is additional to Hölder-s smoothness when $s = 2$ under the preceding convention.

For sequences, $a_n \lesssim b_n$ denotes that $\limsup_n |a_n/b_n|$ is finite; $a_n = O_{\mathbb{P}}(b_n)$ or $a_n \lesssim_{\mathbb{P}} b_n$ means that $\lim_{\epsilon \to \infty} \limsup_n \mathbb{P}[|a_n/b_n| \geq \epsilon] = 0$; $a_n = o(b_n)$ or $a_n \ll b_n$ denotes $a_n/b_n \to 0$; and $a_n = o_{\mathbb{P}}(b_n)$ denotes that a_n/b_n converges to zero in probability. We write $a_n \asymp b_n$ when $a_n \lesssim b_n$ and $b_n \lesssim a_n$.

SA-I.2 Convergence Rate

For the remaining results in this part, define

$$\widehat{n}_{jt} = \sum_{i=1}^{n_t} \mathbb{1}(\widehat{\beta}_{it} \in \widehat{P}_{jt}), \qquad \widehat{n}_{\max} = \max_{H+1 \leq t \leq T} \max_{j \in [\bar{J}_t]} \widehat{n}_{jt}, \qquad \widehat{n}_{\min} = \min_{H+1 \leq t \leq T} \min_{j \in [\bar{J}_t]} \widehat{n}_{jt}.$$

For each cell, define the estimated Gram block

$$\widehat{Q}_{jt} = \frac{1}{n_t} \sum_{i=1}^{n_t} \mathbb{1}(\widehat{\beta}_{it} \in \widehat{P}_{jt}) \check{p}_{\mathfrak{q}}\big(\widehat{D}_{jt}^{-1}(\widehat{\beta}_{it} - \widehat{c}_{jt})\big) \check{p}_{\mathfrak{q}}\big(\widehat{D}_{jt}^{-1}(\widehat{\beta}_{it} - \widehat{c}_{jt})\big)^\top.$$

When $\mathfrak{q} = 0$, $\check{p}_{\mathfrak{q}}$ is the scalar 1, and $\widehat{Q}_{jt} = \widehat{n}_{jt}/n_t$ on every nonempty cell.

SA-I.2.1 Result

Theorem SA-I.2.1 (Convergence Rate)**.** *Consider the second-step estimator with polynomial degree $\mathfrak{q} \geq 0$. Suppose that the following conditions hold:*

(i) ***DGP****. $\{\beta_{it} : i \in [n_t], t \in [T]\}$ take values in compact support $\mathcal{B}$; μ_t admits a bounded extension to the neighborhood $\mathcal{B}^+$ of $\mathcal{B}$ and this extension is uniformly Hölder-s_μ on $\mathcal{B}^+$ for some $s_\mu \geq 1$; $\mathbb{E}[f_t|\mathcal{F}_{t-1}] = \mathbb{E}[f_t|\mathcal{G}_{t-1}]$, $\mathbb{E}[\varepsilon_{it}|\mathcal{F}_{t-1}, f_t] = 0$, and ε_{it} is independent over i for each t conditional on $\mathcal{F}_{t-1}$. For some $q > 2$, $\max_{t \in [T]} \max_{i \in [n_t]} \mathbb{E}[|\varepsilon_{it}|^q|\mathcal{F}_{t-1}] \lesssim 1$ and $\max_{t \in [T]} \mathbb{E}[\|\widetilde{f}_t\|^q|\mathcal{F}_{t-1}] \lesssim 1$.*

(ii) ***First-Step****. $\widehat{\beta}_{it}$ is $\mathcal{F}_{t-1}$-measurable, and $H \leq cT$ for a constant $c \in (0, 1)$. With probability approaching one, $\widehat{\beta}_{it} \in \mathcal{B}^+$ for all $i \in [n_t]$ and $t = H + 1, \ldots, T$. Define the realized first-step error*

$$\widehat{\mathsf{R}}_{\mathsf{FS}} = \max_{H+1 \leq t \leq T} \max_{i \in [n_t]} \|\widehat{\beta}_{it} - \beta_{it}\|.$$

(iii) ***Partition and Design.*** *The estimated partition $\widehat{\Delta}_t$ and its associated centers and scaling matrices are $\mathcal{F}_{t-1}$-measurable for each t. With probability approaching one, $\mathcal{B} \cup \{\widehat{\beta}_{it} : i \in [n_t]\}$ is contained in $\cup_{j\in[\bar{J}_t]}\widehat{P}_{jt}$ for each $t = H+1,\dots,T$, $\widehat{n}_{\min} > 0$, and, for all $t = H+1,\dots,T$ and $j \in [\bar{J}_t]$,*

$$1 \lesssim \frac{n_t}{\widehat{n}_{jt}}\lambda_{\min}(\widehat{Q}_{jt}) \le \frac{n_t}{\widehat{n}_{jt}}\lambda_{\max}(\widehat{Q}_{jt}) \lesssim 1.$$

Define the realized partition diameter

$$\widehat{\mathsf{R}}_{\mathtt{D}} = \max_{H+1\le t\le T}\max_{j\in[\bar{J}_t]} \operatorname{diam}(\widehat{P}_{jt}).$$

If $\frac{(nT)^{2/\mathfrak{q}}\log(nT)}{\widehat{n}_{\min}} \lesssim_{\mathbb{P}} 1$, then for any $\ell_n \to \infty$, with probability approaching one, the following bound holds for all $\beta \in \mathcal{B}$:

$$\left|\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H)\right| \lesssim \ell_n \left[\widehat{\mathsf{R}}_{\mathtt{FS}} + \widehat{\mathsf{R}}_{\mathtt{D}}^{\min\{s_\mu,\mathfrak{q}+1\}} + \sqrt{\frac{\log(nT)}{\widehat{n}_{\min}}} + \frac{\|\beta\|}{\sqrt{T}}\right].$$

All implicit constants in this theorem are independent of the evaluation point β and of n, T, but may depend on the Hölder and design constants.

SA-I.2.2 Discussion

(i) The sequence $\ell_n \to \infty$ is an arbitrary high-probability slack sequence. Equivalently, the same display can be read as the corresponding $\lesssim_{\mathbb{P}}$ rate, while retaining the explicit dependence of the factor term on the evaluation point β.

(ii) When $\mathfrak{q} = 0$, the within-cell polynomial basis $\check{p}_{\mathfrak{q}}$ is the scalar 1 and $\widehat{Q}_{jt} = \widehat{n}_{jt}/n_t$ on every nonempty cell. Thus, the design condition in (iii) reduces to the nonempty-cell condition $\widehat{n}_{\min} > 0$, and since $s_\mu \ge 1$, the second-step approximation rate is governed by the cell-diameter term $\widehat{\mathsf{R}}_{\mathtt{D}}$.

(iii) If the partition is quasi-uniform and each cell roughly contains the same number of observations, then $\widehat{\mathsf{R}}_{\mathtt{D}} \lesssim_{\mathbb{P}} J^{-1}$ and $n/J^d \lesssim_{\mathbb{P}} \widehat{n}_{\min} \le \widehat{n}_{\max} \lesssim_{\mathbb{P}} n/J^d$. In that case, the theorem gives the rate

$$\ell_n \left[\widehat{\mathsf{R}}_{\mathtt{FS}} + J^{-\min\{s_\mu,\mathfrak{q}+1\}} + \sqrt{\frac{J^d\log(nT)}{n}} + \frac{\|\beta\|}{\sqrt{T}}\right].$$

(iv) With a boundary-specific modification of the assumptions and proof, the same argument can allow boundary cells that are large outside the support but have a small intersection with the feasible support $\mathcal{B}$. For example, if $\beta = \underline{\beta}$ is on the lower boundary and the first cell intersects $\mathcal{B}$ only on a shrinking neighborhood of $\underline{\beta}$, then any $\widehat{\beta}_{it}$ in that cell is still close to β with high probability.

SA-I.2.3 Proof of Theorem SA-I.2.1

Work on the event where the first-step estimates belong to $\mathcal{B}^+$, the estimated partition covers $\mathcal{B}$ and the first-step estimates, and the design conditions in the theorem hold. This event has probability approaching one.

For each $\beta \in \mathcal{B}$, let $\widehat{\jmath}_t(\beta)$ denote the index of the cell containing β. The fitted value can be written as a weighted average,

$$\widehat{M}_t(\beta) = \sum_{i=1}^{n_t} \widehat{w}_{it}(\beta) R_{it},$$

where

$$\widehat{w}_{it}(\beta) = \frac{\mathbb{1}(\widehat{\beta}_{it} \in \widehat{P}_{\widehat{\jmath}_t(\beta)t})}{n_t} \check{p}_{\mathfrak{q}}\big(\widehat{D}^{-1}_{\widehat{\jmath}_t(\beta)t}(\beta - \widehat{c}_{\widehat{\jmath}_t(\beta)t})\big)^\top \widehat{Q}^{-1}_{\widehat{\jmath}_t(\beta)t} \check{p}_{\mathfrak{q}}\big(\widehat{D}^{-1}_{\widehat{\jmath}_t(\beta)t}(\widehat{\beta}_{it} - \widehat{c}_{\widehat{\jmath}_t(\beta)t})\big).$$

The boundedness of the normalized polynomial basis and the scaled eigenvalue condition imply that with probability approaching one, uniformly over t and β,

$$\sum_{i=1}^{n_t} |\widehat{w}_{it}(\beta)| \lesssim 1, \qquad \sum_{i=1}^{n_t} \widehat{w}_{it}(\beta)^2 \lesssim \frac{1}{\widehat{n}_{\widehat{\jmath}_t(\beta)t}}.$$

Applying the same algebra to the constant polynomial gives $\sum_{i=1}^{n_t} \widehat{w}_{it}(\beta) = 1$. Moreover, the weights reproduce all polynomials of total degree no greater than $\mathfrak{q}$ on the selected cell:

$$\sum_{i=1}^{n_t} \widehat{w}_{it}(\beta)\, \mathrm{poly}(\widehat{\beta}_{it}) = \mathrm{poly}(\beta)$$

for every such polynomial poly. This includes constant functions when $\mathfrak{q} = 0$ and coordinate functions when $\mathfrak{q} \geq 1$.

Using $R_{it} = \mu_t(\beta_{it}) + \beta_{it}^\top \widetilde{f}_t + \varepsilon_{it}$,

$$\begin{aligned}
\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H) = {} & \frac{1}{T-H} \sum_{t=H+1}^{T} \sum_{i=1}^{n_t} \widehat{w}_{it}(\beta) \varepsilon_{it} \\
& + \frac{1}{T-H} \sum_{t=H+1}^{T} \sum_{i=1}^{n_t} \widehat{w}_{it}(\beta) \big(\mu_t(\beta_{it}) - \mu_t(\beta)\big) \\
& + \frac{1}{T-H} \sum_{t=H+1}^{T} \left(\sum_{i=1}^{n_t} \widehat{w}_{it}(\beta) \beta_{it} \right)^\top \widetilde{f}_t.
\end{aligned}$$

We first bound the approximation term. Since $s_\mu \geq 1$, the uniform Hölder condition implies a uniform Lipschitz bound on $\mathcal{B}^+$. Therefore,

$$\max_{H+1 \leq t \leq T} \max_{i \in [n_t]} \big|\mu_t(\beta_{it}) - \mu_t(\widehat{\beta}_{it})\big| \lesssim \widehat{\mathsf{R}}_{\mathtt{FS}}.$$

Together with the absolute-weight-sum bound, this controls the contribution from replacing β_{it} by $\widehat{\beta}_{it}$. For the remaining within-cell term, local polynomial approximation gives, uniformly over t and cells, a polynomial poly_{jt} of degree no greater than $\mathfrak{q}$ such that

$$\sup_{b \in \widehat{P}_{jt} \cap \mathcal{B}^+} \big|\mu_t(b) - \mathrm{poly}_{jt}(b)\big| \lesssim \mathrm{diam}(\widehat{P}_{jt})^{\min\{s_\mu, \mathfrak{q}+1\}}.$$

Using polynomial reproduction,

$$\left|\sum_{i=1}^{n_t} \widehat{w}_{it}(\beta)\big(\mu_t(\widehat{\beta}_{it}) - \mu_t(\beta)\big)\right| \lesssim \widehat{\mathsf{R}}_{\mathtt{D}}^{\min\{s_\mu, \mathfrak{q}+1\}}.$$

Thus, uniformly over $\beta \in \mathcal{B}$,

$$\left|\frac{1}{T-H}\sum_{t=H+1}^{T}\sum_{i=1}^{n_t} \widehat{w}_{it}(\beta)\big(\mu_t(\beta_{it}) - \mu_t(\beta)\big)\right| \lesssim \widehat{\mathsf{R}}_{\mathtt{FS}} + \widehat{\mathsf{R}}_{\mathtt{D}}^{\min\{s_\mu, \mathfrak{q}+1\}}.$$

Next, consider the idiosyncratic-error term. For each component $k \in [K_{\mathfrak{q}}]$ of $\check{p}_{\mathfrak{q}}\big(\widehat{D}_{jt}^{-1}(\widehat{\beta}_{it} - \widehat{c}_{jt})\big)$,

$$\frac{1}{\widehat{n}_{jt}}\sum_{i=1}^{n_t} \mathbb{1}(\widehat{\beta}_{it} \in \widehat{P}_{jt}) \left[\check{p}_{\mathfrak{q}}\big(\widehat{D}_{jt}^{-1}(\widehat{\beta}_{it} - \widehat{c}_{jt})\big)\right]_k \varepsilon_{it}$$

is a conditionally mean-zero average, independent over i given $\mathcal{F}_{t-1}$, with conditional variance bounded by a constant multiple of $1/\widehat{n}_{jt}$. On the event $\widehat{n}_{\min} > 0$, the number of cell-component averages is of order $\sum_{t=H+1}^{T} \bar{J}_t \le \sum_{t=H+1}^{T} n_t \lesssim nT$. Since the normalized regressors are bounded, a standard truncation and union-bound argument for conditionally independent arrays yields

$$\max_{H+1\le t\le T} \max_{j\in[\bar{J}_t]} \max_{k\in[K_{\mathfrak{q}}]} \left|\frac{1}{\widehat{n}_{jt}}\sum_{i=1}^{n_t} \mathbb{1}(\widehat{\beta}_{it} \in \widehat{P}_{jt}) \left[\check{p}_{\mathfrak{q}}\big(\widehat{D}_{jt}^{-1}(\widehat{\beta}_{it} - \widehat{c}_{jt})\big)\right]_k \varepsilon_{it}\right| \lesssim \ell_n\sqrt{\frac{\log(nT)}{\widehat{n}_{\min}}}$$

with probability approaching one; the rate restriction involving $\widehat{n}_{\min}$ ensures that the truncation contribution is absorbed into the displayed bound. For any evaluation point β, the weighted error is a bounded linear combination of these component averages in the selected cell. The boundedness of the normalized polynomial basis and the scaled eigenvalue bound for $\widehat{Q}_{jt}$ then gives, uniformly over $\beta \in \mathcal{B}$,

$$\left|\frac{1}{T-H}\sum_{t=H+1}^{T}\sum_{i=1}^{n_t} \widehat{w}_{it}(\beta)\varepsilon_{it}\right| \lesssim \ell_n\sqrt{\frac{\log(nT)}{\widehat{n}_{\min}}}.$$

Finally, decompose the factor term as

$$\begin{aligned}
&\frac{1}{T-H}\sum_{t=H+1}^{T}\left(\sum_{i=1}^{n_t} \widehat{w}_{it}(\beta)\beta_{it}\right)^{\top}\widetilde{f}_t \\
&= \frac{1}{T-H}\sum_{t=H+1}^{T} \beta^{\top}\widetilde{f}_t \\
&\quad + \frac{1}{T-H}\sum_{t=H+1}^{T}\left(\sum_{i=1}^{n_t} \widehat{w}_{it}(\beta)(\beta_{it} - \widehat{\beta}_{it})\right)^{\top}\widetilde{f}_t \\
&\quad + \frac{1}{T-H}\sum_{t=H+1}^{T}\left(\sum_{i=1}^{n_t} \widehat{w}_{it}(\beta)\widehat{\beta}_{it} - \beta\right)^{\top}\widetilde{f}_t.
\end{aligned}$$

By the martingale-difference property of $\widetilde{f}_t$ with respect to $\mathcal{G}_t$ and the conditional moment bound,

$$\left\| \frac{1}{T-H} \sum_{t=H+1}^{T} \widetilde{f}_t \right\| \lesssim_{\mathbb{P}} T^{-1/2},$$

so the first term is $\lesssim_{\mathbb{P}} \|\beta\| T^{-1/2}$ uniformly over β. Also,

$$\frac{1}{T-H} \sum_{t=H+1}^{T} \|\widetilde{f}_t\| \lesssim_{\mathbb{P}} 1$$

by Markov's inequality. The second term is $\lesssim_{\mathbb{P}} \widehat{\mathsf{R}}_{\mathsf{FS}}$ because the weights have uniformly bounded absolute sum. For the third term, polynomial reproduction gives

$$\sum_{i=1}^{n_t} \widehat{w}_{it}(\beta) \widehat{\beta}_{it} = \beta$$

when $\mathfrak{q} \geq 1$. When $\mathfrak{q} = 0$, the same expression is the ordinary cell average of the $\widehat{\beta}_{it}$'s in the selected cell, and hence its distance from β is bounded by the diameter of that cell. Therefore, uniformly over $\beta \in \mathcal{B}$,

$$\left| \frac{1}{T-H} \sum_{t=H+1}^{T} \left(\sum_{i=1}^{n_t} \widehat{w}_{it}(\beta) \beta_{it} \right)^{\top} \widetilde{f}_t \right| \lesssim_{\mathbb{P}} \frac{\|\beta\|}{\sqrt{T}} + \widehat{\mathsf{R}}_{\mathsf{FS}} + \widehat{\mathsf{R}}_{\mathsf{D}}^{\min\{s_\mu, \mathfrak{q}+1\}}.$$

Combining the three bounds and absorbing the stochastic constants into the arbitrary high-probability slack sequence ℓ_n gives the claimed result, uniformly over $\beta \in \mathcal{B}$. □

SA-I.3 Stochastic Linearization

This section derives a high-level stochastic linearization for second-step polynomial degree $\mathfrak{q} \geq 0$. The result is stated in terms of generic first-step estimators and generic partitions, so it does not rely on the specific local-polynomial first step or marginal-quantile-spaced partition introduced later. Let J denote a generic partition-resolution sequence; the conditions below encode cell mass of order J^{-d} and cell diameter of order J^{-1}.

Let $\{\mathcal{D}_t\}$ denote an auxiliary design filtration with $\mathcal{D}_{t-1} \subseteq \mathcal{F}_{t-1}$. In this high-level section, $\mathcal{D}_{t-1}$ records the information used to construct the conditionally nonrandom reference partition; Section SA-II.1.2 gives its concrete definition for the classical beta-sorted estimator. To state the linearization, introduce a reference partition,

$$\Delta_t = \{P_{jt} : j \in [\bar{J}_t]\}$$

that is $\mathcal{D}_{t-1}$-measurable and has the same number of cells as the estimated partition $\widehat{\Delta}_t$ from Section SA-I.1. For each reference cell paired with an estimated cell intersecting $\mathcal{B}^\circ$, choose a $\mathcal{D}_{t-1}$-measurable center c_{jt} and nonsingular scaling/rotation matrix D_{jt} so that the normalized polynomial basis is uniformly bounded on the cell. Tail reference cells may instead be assigned any $\mathcal{D}_{t-1}$-measurable nonsingular normalization; no uniform boundedness or design condition is imposed on them because they are never selected by an interior evaluation point. Thus, $\mathcal{D}_{t-1}$ governs the reference partition and reference objects, whereas the realized

estimated partition $\widehat{\Delta}_t$ may depend on first-step estimates and is only required below to be $\mathcal{F}_{t-1}$-measurable. The reference cells and estimated cells are paired by their common index j.

Define the reference basis maps by

$$p_t(b) = p(b; \Delta_t), \qquad p_{jt}(b) = p_{\mathfrak{q}}(b, P_{jt}).$$

Both basis maps are evaluated at the raw argument b; in particular, the classical construction does not restrict the generated regressor to reside in $\mathcal{B}$. The estimated marginal partition in Section SA-II.3 uses the coordinatewise sample extrema as its outer endpoints and therefore contains every realized $\widehat{\beta}_{it}$. Reference cells are paired with estimated cells by their split-quantile ranks. Only pairs whose estimated cell intersects $\mathcal{B}^\circ$ are normalized and used below; no reference design condition is imposed on the cells outside $\mathcal{B}^\circ$. Write $\widehat{p}_{jt}(b)$ for the j-th $K_{\mathfrak{q}}$-dimensional block of the estimated basis $\widehat{p}_t(b)$ defined in Section SA-I.1. Thus $p_t(\widehat{\beta}_{it})$ denotes the reference basis evaluated at the first-step estimate, while $p_t(\beta_{it})$ denotes the reference basis evaluated at the true exposure. The reference Gram matrix below remains centered at the generated-regressor distribution, but the leading idiosyncratic term uses $p_t(\beta_{it})$ so that its coefficients are fixed by the innovation-free conditioning field in the classical verification. The estimated and reference Gram matrices are

$$\widehat{Q}_t = \frac{1}{n_t}\sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it})\widehat{p}_t(\widehat{\beta}_{it})^\top \qquad \text{and} \qquad Q_t = \frac{1}{n_t}\sum_{i=1}^{n_t} \mathbb{E}\big[p_t(\widehat{\beta}_{it})p_t(\widehat{\beta}_{it})^\top \mid \mathcal{D}_{t-1}\big].$$

Both matrices are block diagonal. Write $\widehat{Q}_{jt}$ and Q_{jt} for their j-th $K_{\mathfrak{q}} \times K_{\mathfrak{q}}$ diagonal blocks. Thus Q_t is the conditional reference Gram matrix for the generated regressor evaluated on the conditionally nonrandom reference partition. When $\mathfrak{q} = 0$, $Q_{jt} = n_t^{-1}\sum_{i=1}^{n_t} \mathbb{P}(\widehat{\beta}_{it} \in P_{jt} \mid \mathcal{D}_{t-1})$. In every expression involving Q_t^{-1} or $\widehat{Q}_t^{-1}$ and an evaluation point in $\mathcal{B}^\circ$, the inverse is blockwise shorthand for the inverse of the selected block. Thus the algebra does not require invertibility or a polynomial normalization for tail blocks outside $\mathcal{J}_t$.

SA-I.3.1 Decomposition

Using the block Gram matrices defined above, the second-step estimator admits the decomposition

$$\begin{aligned}
\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H) &= \frac{1}{T-H}\sum_{t=H+1}^{T} (\widehat{M}_t(\beta) - \mu_t(\beta)) \\
&= \mathscr{G}(\beta) + \mathscr{E}(\beta) - \mathscr{L}(\beta) + \mathscr{B}(\beta) \\
&\quad + \mathscr{R}_1(\beta) + \mathscr{R}_2(\beta) + \mathscr{R}_3(\beta) + \mathscr{R}_4(\beta) - \mathscr{R}_5(\beta),
\end{aligned}$$

where the potentially leading terms are

$$\begin{aligned}
\mathscr{G}(\beta) &= \frac{1}{T-H}\sum_{t=H+1}^{T} \beta^\top \widetilde{f}_t, \\
\mathscr{E}(\beta) &= \frac{1}{T-H}\sum_{t=H+1}^{T} \widehat{p}_t(\beta)^\top Q_t^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t} p_t(\beta_{it})\varepsilon_{it}, \\
\mathscr{L}(\beta) &= \frac{1}{T-H}\sum_{t=H+1}^{T} \widehat{p}_t(\beta)^\top \widehat{Q}_t^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it})[\nabla\mu_t(\beta_{it})]^\top(\widehat{\beta}_{it} - \beta_{it}),
\end{aligned}$$

$$\mathscr{B}(\beta) = \frac{1}{T-H}\sum_{t=H+1}^{T} \widehat{p}_t(\beta)^\top \widehat{Q}_t^{-1} \frac{1}{n_t}\sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it})\big(\mu_t(\widehat{\beta}_{it}) - \mu_t(\beta)\big),$$

and the higher-order remainder terms are

$$\mathscr{R}_1(\beta) = \frac{1}{T-H}\sum_{t=H+1}^{T} \widehat{p}_t(\beta)^\top \left(\widehat{Q}_t^{-1} - Q_t^{-1}\right) \frac{1}{n_t}\sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it})\varepsilon_{it},$$

$$\mathscr{R}_2(\beta) = \frac{1}{T-H}\sum_{t=H+1}^{T} \widehat{p}_t(\beta)^\top Q_t^{-1} \frac{1}{n_t}\sum_{i=1}^{n_t} \big(\widehat{p}_t(\widehat{\beta}_{it}) - p_t(\beta_{it})\big)\varepsilon_{it},$$

$$\mathscr{R}_3(\beta) = \frac{1}{T-H}\sum_{t=H+1}^{T} \left[\widehat{p}_t(\beta)^\top \widehat{Q}_t^{-1} \left(\frac{1}{n_t}\sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it})\widehat{\beta}_{it}^\top\right) - \beta^\top\right] \widetilde{f}_t,$$

$$\mathscr{R}_4(\beta) = \frac{1}{T-H}\sum_{t=H+1}^{T} \widehat{p}_t(\beta)^\top \widehat{Q}_t^{-1} \left(\frac{1}{n_t}\sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it})(\beta_{it} - \widehat{\beta}_{it})^\top\right) \widetilde{f}_t,$$

$$\mathscr{R}_5(\beta) = \frac{1}{2(T-H)}\sum_{t=H+1}^{T} \widehat{p}_t(\beta)^\top \widehat{Q}_t^{-1} \frac{1}{n_t}\sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it})\big(\widehat{\beta}_{it} - \beta_{it}\big)^\top \nabla^2\mu_t(\tilde{\beta}_{it})\big(\widehat{\beta}_{it} - \beta_{it}\big),$$

with $\tilde{\beta}_{it} = a_{it}\widehat{\beta}_{it} + (1-a_{it})\beta_{it}$ for some $a_{it} \in [0,1]$.

The expanded set $\mathcal{B}^+$ from Section SA-I.1 handles first-step estimates and Taylor intermediate points that may lie outside the support $\mathcal{B}$. Recall that $\mathcal{B}^\circ$ is the interior evaluation region and $\mathcal{J}_t$ is the set of estimated cells intersecting it. For each $j \in \mathcal{J}_t$, the reference block Q_{jt} below is formed from the reference cell paired with the estimated cell $\widehat{P}_{jt}$; the partition-design condition requires these paired reference cells to satisfy the same interior design bounds. Recall the previously defined quantity $\widehat{\mathsf{R}}_{\mathtt{FS}}$, and also define

$$\widehat{\mathsf{R}}_{\mathtt{D}}^{\circ} = \max_{H+1\le t\le T}\max_{j\in\mathcal{J}_t} \operatorname{diam}(\widehat{P}_{jt}),$$

$$\widehat{\mathsf{R}}_{\mathtt{Q}} = \max_{H+1\le t\le T}\max_{j\in\mathcal{J}_t} \|\widehat{Q}_{jt} - Q_{jt}\|,$$

$$\widehat{\mathsf{R}}_{\mathtt{P}} = \max_{H+1\le t\le T}\max_{j\in\mathcal{J}_t} \frac{1}{n_t}\sum_{i=1}^{n_t} \|\widehat{p}_{jt}(\widehat{\beta}_{it}) - p_{jt}(\beta_{it})\|^2.$$

SA-I.3.2 Result

The following lemma establishes a general stochastic linearization for generic first-step estimators and partitioning schemes. Part II then verifies these high-level quantities for the classical beta-sorted portfolio estimator.

Lemma SA-I.3.1 (Stochastic Linearization)**.** *Maintain the setup and notation of Section SA-I.1 and the definitions above. Suppose that conditions (i) and (ii) in Theorem SA-I.2.1 hold, and that μ_t is uniformly Hölder-s_μ on $\mathcal{B}^+$ and twice continuously differentiable on $\mathcal{B}^+$ with uniformly bounded first and second derivatives. Assume further that the following conditions hold:*

(i) (Partition and Design) the estimated partition $\widehat{\Delta}_t$ and its associated centers and scaling matrices are $\mathcal{F}_{t-1}$-measurable for each t, $\mathcal{B}^\circ \cup \{\widehat{\beta}_{it} : i \in [n_t]\}$ is contained in $\cup_{j\in[\bar{J}_t]}\widehat{P}_{jt}$ for each $t = H+1,\dots,T$ w.p.a.1, $\widehat{\mathsf{R}}_{\mathtt{D}}^{\circ} \lesssim_{\mathbb{P}} J^{-1}$, and $J^{-d} \lesssim_{\mathbb{P}} \lambda_{\min}(Q_{jt}) \le \lambda_{\max}(Q_{jt}) \lesssim_{\mathbb{P}} J^{-d}$ uniformly over $t = H+1,\dots,T$ and

$j \in \mathcal{J}_t$. *Also, uniformly over the same indices,*

$$\lambda_{\max}\left(\frac{1}{n_t}\sum_{i=1}^{n_t} p_{jt}(\beta_{it})p_{jt}(\beta_{it})^\top\right) \lesssim_{\mathbb{P}} J^{-d}.$$

(ii) (Gram Convergence) $\widehat{\mathsf{R}}_{\mathtt{Q}} = o_{\mathbb{P}}(J^{-d})$.

Then, for each fixed $\beta \in \mathcal{B}^\circ$,

$$|\mathscr{G}(\beta)| \lesssim_{\mathbb{P}} \frac{\|\beta\|}{\sqrt{T}}, \qquad |\mathscr{E}(\beta)| \lesssim_{\mathbb{P}} \sqrt{\frac{J^d}{nT}},$$
$$|\mathscr{R}_1(\beta)| \lesssim_{\mathbb{P}} \sqrt{\frac{J^{3d}(\widehat{\mathsf{R}}_{\mathtt{Q}})^2}{nT}}, \qquad |\mathscr{R}_2(\beta)| \lesssim_{\mathbb{P}} \sqrt{\frac{J^{2d}\widehat{\mathsf{R}}_{\mathtt{P}}}{nT}},$$
$$|\mathscr{R}_3(\beta)| \lesssim_{\mathbb{P}} \mathbb{1}(\mathfrak{q}=0)\frac{\widehat{\mathsf{R}}^\circ_{\mathtt{D}}}{\sqrt{T}}, \qquad |\mathscr{R}_4(\beta)| \lesssim_{\mathbb{P}} \frac{\widehat{\mathsf{R}}_{\mathtt{FS}}}{\sqrt{T}}.$$

Moreover,

$$\sup_{\beta\in\mathcal{B}^\circ}|\mathscr{L}(\beta)| \lesssim_{\mathbb{P}} \widehat{\mathsf{R}}_{\mathtt{FS}}, \qquad \sup_{\beta\in\mathcal{B}^\circ}|\mathscr{R}_5(\beta)| \lesssim_{\mathbb{P}} (\widehat{\mathsf{R}}_{\mathtt{FS}})^2,$$
$$\sup_{\beta\in\mathcal{B}^\circ}|\mathscr{B}(\beta)| \lesssim_{\mathbb{P}} (\widehat{\mathsf{R}}^\circ_{\mathtt{D}})^{\min\{s_\mu,\mathfrak{q}+1\}}, \qquad \sup_{\beta\in\mathcal{B}^\circ}|\mathscr{R}_3(\beta)| \lesssim_{\mathbb{P}} \mathbb{1}(\mathfrak{q}=0)\widehat{\mathsf{R}}^\circ_{\mathtt{D}}.$$

The constants implicit in $\lesssim_{\mathbb{P}}$ *do not depend on the evaluation point* $\beta \in \mathcal{B}^\circ$.

SA-I.3.3 Discussion

(i) The condition $\widehat{\mathsf{R}}_{\mathtt{Q}} = o_{\mathbb{P}}(J^{-d})$ is a relative Gram-convergence condition, because each reference block Q_{jt} has eigenvalues of order J^{-d}. Together with the reference design condition, it implies $\widehat{Q}_{jt} \asymp_{\mathbb{P}} J^{-d}$ uniformly over $j \in \mathcal{J}_t$ and controls the inverse perturbation $\widehat{Q}_{jt}^{-1} - Q_{jt}^{-1}$ used in the bound for $\mathscr{R}_1(\beta)$. The realized size of $\widehat{\mathsf{R}}_{\mathtt{P}}$ controls the generated-regressor basis-switching term $\mathscr{R}_2(\beta)$.

(ii) Since the variance bounds used in the proof do not depend on β, the pointwise stochastic bounds for $\mathscr{E}(\beta)$ and $\mathscr{R}_j(\beta)$ can be read uniformly in the following sense: if $G(\beta)$ denotes one of these terms and r_n is the displayed rate, then for any deterministic sequence $\ell_n \to \infty$,

$$\limsup_{n,T\to\infty}\sup_{\beta\in\mathcal{B}^\circ}\mathbb{P}\Big(|G(\beta)| > \ell_n r_n\Big) = 0.$$

(iii) The interior partition and design conditions may not hold for all bins. The index set $\mathcal{J}_t$ restricts attention to cells that can contain evaluation points in $\mathcal{B}^\circ$. Given the results in Section SA-II.3, this can be guaranteed if we focus on interior points β that are separated from the boundary of $\mathcal{B}$ by more than the cell width and the first-step smoothing scale.

(iv) When $\mathfrak{q} = 0$, the block basis is the scalar 1. In that special case, $\widehat{\mathsf{R}}_{\mathtt{Q}}$ is the cell-average discrepancy and $\widehat{\mathsf{R}}_{\mathtt{P}}$ is the cell-switching discrepancy between the generated-regressor basis $\widehat{p}_t(\widehat{\beta}_{it})$ and the true-exposure reference basis $p_t(\beta_{it})$. The latter enters only through $\mathscr{R}_2(\beta)$.

(v) For the classical marginal-quantile partition, Part II verifies the high-level quantities appearing in this lemma under primitive conditions. In particular, Section SA-II.3 establishes the partition, Gram, and basis-switching bounds, while Section SA-II.4 uses the structure of the rolling-window first step to control the generated-regressor terms. The rate substitutions are therefore postponed until the classical setup and its notation have been introduced.

SA-I.3.4 Proof of Lemma SA-I.3.1

Work on an event whose probability approaches one and on which the conditions in the lemma hold. By Weyl's inequality, for every $t = H+1, \dots, T$ and $j \in \mathcal{J}_t$,

$$J^{-d} \lesssim \lambda_{\min}(\widehat{Q}_{jt}) \le \lambda_{\max}(\widehat{Q}_{jt}) \lesssim J^{-d}.$$

Because $\mathbb{E}[f_t|\mathcal{F}_{t-1}] = \mathbb{E}[f_t|\mathcal{G}_{t-1}]$ and $\mathcal{G}_{t-1} \subseteq \mathcal{F}_{t-1}$, $\widetilde{f}_t$ has mean zero conditional on $\mathcal{F}_{t-1}$. Thus any $\mathcal{F}_{t-1}$-measurable coefficient multiplying $\widetilde{f}_t$ generates a martingale-difference summand with respect to $\{\mathcal{F}_t\}$. The bound on $\mathscr{G}(\beta)$ is immediate from the martingale-difference property of $\widetilde{f}_t$ and the conditional moment bound.

For $\mathscr{E}(\beta)$, let $\widehat{\jmath}_t(\beta)$ denote the cell containing β in $\widehat{\Delta}_t$. Since $\beta \in \mathcal{B}^\circ$, $\widehat{\jmath}_t(\beta) \in \mathcal{J}_t$, and the paired indexing identifies the corresponding reference block. By the martingale structure and $\mathbb{E}[\varepsilon_{it}^2|\mathcal{F}_{t-1}] \lesssim 1$, for each fixed $\beta \in \mathcal{B}^\circ$,

$$\begin{aligned}
&\frac{1}{(T-H)^2}\sum_{t=H+1}^{T} \mathbb{V}\mathrm{ar}\left[\widehat{p}_t(\beta)^\top Q_t^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t} p_t(\beta_{it})\varepsilon_{it}\,\middle|\,\mathcal{F}_{t-1}\right] \\
&\lesssim \frac{1}{T^2}\sum_{t=H+1}^{T}\frac{1}{n_t}\widehat{p}_{\widehat{\jmath}_t(\beta)t}(\beta)^\top Q_{\widehat{\jmath}_t(\beta)t}^{-1}\left(\frac{1}{n_t}\sum_{i=1}^{n_t} p_{\widehat{\jmath}_t(\beta)t}(\beta_{it})p_{\widehat{\jmath}_t(\beta)t}(\beta_{it})^\top\right)Q_{\widehat{\jmath}_t(\beta)t}^{-1}\widehat{p}_{\widehat{\jmath}_t(\beta)t}(\beta) \lesssim \frac{1}{T^2}\frac{J^d T}{n} = \frac{J^d}{nT}.
\end{aligned}$$

The bound on $\mathscr{E}(\beta)$ follows by the Lenglart–Rebolledo inequality.

For $\mathscr{L}(\beta)$ and $\mathscr{R}_5(\beta)$, bounded derivatives, bounded basis functions, and the eigenvalue bounds for $\widehat{Q}_{jt}$ give

$$|\mathscr{R}_5(\beta)| \le \frac{1}{2(T-H)}\sum_{t=H+1}^{T}\left[\frac{1}{n_t}\sum_{i=1}^{n_t}\left|\widehat{p}_t(\beta)^\top\widehat{Q}_t^{-1}\widehat{p}_t(\widehat{\beta}_{it})\right|\|\nabla^2\mu_t(\tilde{\beta}_{it})\|\right]\widehat{\mathsf{R}}_{\mathtt{FS}}^2 \lesssim \widehat{\mathsf{R}}_{\mathtt{FS}}^2,$$

and

$$|\mathscr{L}(\beta)| \le \frac{1}{T-H}\sum_{t=H+1}^{T}\left[\frac{1}{n_t}\sum_{i=1}^{n_t}\left|\widehat{p}_t(\beta)^\top\widehat{Q}_t^{-1}\widehat{p}_t(\widehat{\beta}_{it})\right|\|\nabla\mu_t(\beta_{it})\|\right]\widehat{\mathsf{R}}_{\mathtt{FS}} \lesssim \widehat{\mathsf{R}}_{\mathtt{FS}}.$$

The same bounds hold uniformly over $\beta \in \mathcal{B}^\circ$.

Next, consider $\mathscr{B}(\beta)$. On each interior cell, let poly_{jt} denote a polynomial of degree at most $\mathtt{q}$ satisfying the usual approximation bound

$$\sup_{b \in \widehat{P}_{jt}\cap\mathcal{B}^+}|\mu_t(b) - \mathrm{poly}_{jt}(b)| \lesssim \mathrm{diam}(\widehat{P}_{jt})^{\min\{s_\mu,\mathtt{q}+1\}}.$$

Because the second-step basis reproduces every polynomial of degree at most $\mathtt{q}$ on the estimated cell con-

taining β,

$$\widehat{p}_t(\beta)^\top \widehat{Q}_t^{-1} \frac{1}{n_t} \sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it}) \operatorname{poly}_{\widehat{\jmath}_t(\beta)t}(\widehat{\beta}_{it}) = \operatorname{poly}_{\widehat{\jmath}_t(\beta)t}(\beta).$$

The boundedness of the absolute sum of the local polynomial weights then gives

$$\sup_{\beta \in \mathcal{B}^\circ} |\mathscr{B}(\beta)| \lesssim (\widehat{\mathsf{R}}_{\mathtt{D}}^\circ)^{\min\{s_\mu, \mathfrak{q}+1\}}.$$

For $\mathscr{R}_3(\beta)$, if $\mathfrak{q} \geq 1$, the global linear function $g(b) = b$ is reproduced exactly by the local polynomial basis, and therefore $\mathscr{R}_3(\beta) = 0$. If $\mathfrak{q} = 0$,

$$\left\| \widehat{p}_t(\beta)^\top \widehat{Q}_t^{-1} \left(\frac{1}{n_t} \sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it}) \widehat{\beta}_{it}^\top \right) - \beta^\top \right\| \leq \widehat{\mathsf{R}}_{\mathtt{D}}^\circ,$$

because the left-hand side is the distance between β and the average of the estimated betas in the same cell as β. This gives

$$|\mathscr{R}_3(\beta)| \lesssim_{\mathbb{P}} \mathbb{1}(\mathfrak{q} = 0) \frac{\widehat{\mathsf{R}}_{\mathtt{D}}^\circ}{\sqrt{T}} \qquad \text{and} \qquad \sup_{\beta \in \mathcal{B}^\circ} |\mathscr{R}_3(\beta)| \lesssim_{\mathbb{P}} \mathbb{1}(\mathfrak{q} = 0) \widehat{\mathsf{R}}_{\mathtt{D}}^\circ$$

because the term is identically zero when $\mathfrak{q} \geq 1$.

Finally, consider $\mathscr{R}_j(\beta)$, $j = 1, 2, 4$. The block-diagonal structure gives

$$\|\widehat{Q}_{jt}^{-1} - Q_{jt}^{-1}\| \lesssim J^{2d} \widehat{\mathsf{R}}_{\mathtt{Q}}$$

uniformly over $t = H+1, \ldots, T$ and $j \in \mathcal{J}_t$. Boundedness of the estimated basis and the eigenvalue bounds for $\widehat{Q}_{jt}$ control the conditional second moment of the selected cell average. Therefore,

$$\begin{aligned} &\frac{1}{T} \max_{H+1 \leq t \leq T} \mathbb{V}\mathrm{ar}\left[\widehat{p}_t(\beta)^\top \left(\widehat{Q}_t^{-1} - Q_t^{-1} \right) \frac{1}{n_t} \sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it}) \varepsilon_{it} \,\middle|\, \mathcal{F}_{t-1} \right] \\ &\lesssim \frac{1}{nTJ^d} \left(J^{2d} \widehat{\mathsf{R}}_{\mathtt{Q}} \right)^2 = \frac{J^{3d} \widehat{\mathsf{R}}_{\mathtt{Q}}^2}{nT}. \end{aligned}$$

For $\mathscr{R}_2(\beta)$,

$$\begin{aligned} &\frac{1}{T} \max_{H+1 \leq t \leq T} \mathbb{V}\mathrm{ar}\left[\widehat{p}_t(\beta)^\top Q_t^{-1} \frac{1}{n_t} \sum_{i=1}^{n_t} \left(\widehat{p}_t(\widehat{\beta}_{it}) - p_t(\beta_{it}) \right) \varepsilon_{it} \,\middle|\, \mathcal{F}_{t-1} \right] \\ &\lesssim \frac{J^{2d}}{T} \max_{H+1 \leq t \leq T} \max_{j \in \mathcal{J}_t} \frac{1}{n_t^2} \sum_{i=1}^{n_t} \|\widehat{p}_{jt}(\widehat{\beta}_{it}) - p_{jt}(\beta_{it})\|^2 \lesssim \frac{J^{2d} \widehat{\mathsf{R}}_{\mathtt{P}}}{nT}. \end{aligned}$$

For $\mathscr{R}_4(\beta)$,

$$\frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \mathbb{V}\mathrm{ar}\left[\widehat{p}_t(\beta)^\top \widehat{Q}_t^{-1} \left(\frac{1}{n_t} \sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it}) (\beta_{it} - \widehat{\beta}_{it})^\top \right) \widetilde{f}_t \,\middle|\, \mathcal{F}_{t-1} \right] \lesssim \frac{\widehat{\mathsf{R}}_{\mathtt{FS}}^2}{T}.$$

The desired pointwise bounds follow again by the Lenglart–Rebolledo inequality. □

Part II: Classical Beta-Sorted Portfolios

This part specializes the general setup in Part I to the classical beta-sorted portfolio estimator, where the portfolios are constructed using local rolling-window least-squares estimates of β_{it}.

We keep the notation from Part I, including the estimator $\widehat{\mu}(\beta)$, the estimated and reference partitions $\widehat{\Delta}_t$ and Δ_t, and the basis maps $\widehat{p}_t(\cdot)$ and $p_t(\cdot)$. The specialization has two ingredients: the first-step betas are estimated by rolling-window least-squares regressions, and the second-step portfolios are formed by constructing tensor products based on marginal empirical quantiles of the estimated betas. Throughout Part II, we take $\mathcal{B} = [\underline{\beta}, \bar{\beta}]$ to be a rectangle with $\underline{\beta} = (\underline{\beta}_1, \ldots, \underline{\beta}_d)^\top$, $\bar{\beta} = (\bar{\beta}_1, \ldots, \bar{\beta}_d)^\top$, and 0 in its interior.

SA-II.1 Assumptions

All notation from Part I remains in force.

We employ the following simple notion of time-series dependence. For $r \geq 1$, a sequence of random variables $\{Y_t\}_{t\in\mathbb{Z}}$ with finite r-th moments is L_r-near-epoch dependent on a sequence $\{\xi_t\}_{t\in\mathbb{Z}}$ with approximation constants a_m if, for $\mathcal{H}_{t,m} = \sigma(\xi_{t-m}, \ldots, \xi_{t+m})$,

$$\|Y_t - \mathbb{E}[Y_t|\mathcal{H}_{t,m}]\|_r \leq a_m,$$

where $a_m \to 0$ as $m \to \infty$.

SA-II.1.1 Primitive Conditions

The following primitive conditions specialize the high-level conditions from Part I to the classical beta-sorted estimator. They govern the data-generating process, the predictability and cross-sectional distribution of the time-varying coefficients, and the dependence and moment conditions for the factors and idiosyncratic errors. The precise definitions of the filtrations and limit sigma-fields appearing in the assumption are collected in Section SA-II.1.2.

Assumption SA-DGP (DGP). *The following conditions hold:*

(i) ***Smoothness.*** *μ_t admits a bounded extension to the neighborhood $\mathcal{B}^+$ of $\mathcal{B}$, and this extension is uniformly Hölder-s_μ on $\mathcal{B}^+$ for some $s_\mu > 0$.*

(ii) ***Idiosyncratic Error.*** *$\varepsilon_{it} = \varsigma_{it}\eta_{it}$, where (a) η_{it} is i.i.d. across i and over t with mean 0 and variance 1, is independent of $\mathcal{E}_\infty$, and has a Lebesgue density g_η with weak derivative $g_\eta^{(1)} \in L^1(\mathbb{R})$; moreover, there exists a moment exponent $q > 2(3+d)$ such that $\mathbb{E}[|\eta_{it}|^q] < \infty$; (b) ς_{it} is $\mathcal{E}_t$-measurable, bounded, and bounded away from zero for all $i \in [n_t]$ and $t \in [T]$.*

(iii) ***Varying Coefficients.***

(a) ***Predictability and Boundedness.*** *The coefficients α_{it} and β_{it} are $\mathcal{E}_{t-1}$-measurable, $\beta_{it} \in \mathcal{B}$, and $|\alpha_{it}| + \|\beta_{it}\|$ is bounded a.s. by a deterministic finite bound uniformly over i, t, and T.*

(b) ***Independence.*** *Conditional on $\mathcal{D}_{t-1}$, the unit-specific predictable histories*

$$\big((\alpha_{i1}, \beta_{i1}^\top), \ldots, (\alpha_{it}, \beta_{it}^\top)\big)$$

are independent (but not necessarily identically distributed) over $i \in [n_t]$, and β_{it} has a density $f_{\beta_{it}}(b|\mathcal{D}_{t-1})$ conditional on $\mathcal{D}_{t-1}$, which is bounded, bounded away from zero, and Lipschitz continuous on $\mathcal{B}$.

(iv) ***Factors.*** *$\mathbb{E}[f_t|\mathcal{F}_{t-1}] = \mathbb{E}[f_t|\mathcal{G}_{t-1}]$, $\lambda_{\min}(\mathbb{E}[\bar{f}_t\bar{f}_t^\top]) \asymp 1$ with $\bar{f}_t = (1, f_t^\top)^\top$, for all $t \in [T]$, and $\max_{1\le t\le T}\mathbb{E}[\|\widetilde{f}_t\|^4|\mathcal{F}_{t-1}] \lesssim_{\mathbb{P}} 1$. There exists a deterministic positive-definite matrix $\Omega_f \in \mathbb{R}^{d\times d}$ with $0 < \lambda_{\min}(\Omega_f) \le \lambda_{\max}(\Omega_f) < \infty$ such that*

$$\left\| \frac{1}{T-H}\sum_{t=H+1}^{T} \mathbb{V}\mathrm{ar}(f_t \mid \mathcal{G}_{t-1}) - \Omega_f \right\| = o_{\mathbb{P}}(1). \tag{SA-II.1.1}$$

For the moment exponent q in part (ii)*, there exists a fixed $\upsilon > 0$ such that, for each coordinate $k \in [d]$, $\sup_{t\in\mathbb{Z}}\mathbb{E}[|f_{t,k}|^{q+\upsilon}] < \infty$ and the process $f_{t,k}$ is $L_{q+\upsilon}$-near-epoch dependent on a strictly stationary strongly mixing process $\{\xi_t\}$ with mixing coefficients $\alpha(m)$ and approximation constants $a_m \downarrow 0$. Assume there exist constants $C_a, C_\alpha > 0$ and $\rho_a, \rho_\alpha \in (0,1)$ such that*

$$a_m \le C_a\rho_a^m, \qquad \alpha(m) \le C_\alpha\rho_\alpha^m, \qquad m \ge 1.$$

Remark SA-II.1.1 (Primitive verification of factor variance stabilization)**.** *Condition* (SA-II.1.1) *holds, for example, if $\{\widetilde{f}_t\}$ is strictly stationary, its coordinates are L_q-near-epoch dependent on the same strongly mixing process with the exponential rates in Assumption SA-DGP(iv), and $\Omega_f = \mathbb{E}[\widetilde{f}_0\widetilde{f}_0^\top]$ is positive definite. Indeed, because $q > 4$, the NED law of large numbers applies to $\widetilde{f}_t\widetilde{f}_t^\top$. Moreover, $\widetilde{f}_t\widetilde{f}_t^\top - \mathbb{E}[\widetilde{f}_t\widetilde{f}_t^\top \mid \mathcal{G}_{t-1}]$ is a matrix-valued martingale difference whose average is $O_{\mathbb{P}}((T-H)^{-1/2})$ by the fourth-moment bound and martingale orthogonality. Hence the average conditional variance converges to Ω_f whenever $T - H \to \infty$.*

Assumption SA-SEP (Interior separation). *Let $J = J_{nT} \to \infty$ denote the common order of the coordinatewise portfolio counts used to construct the second-step partition. Writing $b = (b_1, \dots, b_d)^\top$, the interior evaluation region $\mathcal{B}^\circ$ introduced in Section SA-I.1 satisfies*

$$\inf_{b\in\mathcal{B}^\circ}\min_{l\in[d]}\min\{|b_l - \underline{\beta}_l|, |\bar{\beta}_l - b_l|\} \gg J^{-1} + \frac{\log(nT)}{\sqrt{H}}.$$

SA-II.1.2 Filtrations

The idiosyncratic-error condition in Assumption SA-DGP separates predictable model information from the i.i.d. innovations. In Part II we use four nested sigma-fields, ordered from common factor information to full information:

$$\begin{aligned}
\mathcal{G}_t &= \sigma\big(\{f_s\}_{s=1}^t, \{\aleph_{2s}\}_{s=1}^t\big),\\
\mathcal{D}_t &= \sigma\big(\{f_s\}_{s=1}^t, \{\{\varsigma_{is}\}_{i=1}^{n_s}\}_{s=1}^t, \{\aleph_{2s}\}_{s=1}^t\big),\\
\mathcal{E}_t &= \sigma\big(\{f_s\}_{s=1}^t, \{\{\varsigma_{is}\}_{i=1}^{n_s}\}_{s=1}^t, \{\aleph_{1i}\}_{i=1}^n, \{\aleph_{2s}\}_{s=1}^t, \{\{\aleph_{3is}\}_{i=1}^{n_s}\}_{s=1}^t\big),\\
\mathcal{F}_t &= \sigma\big(\{f_s\}_{s=1}^t, \{\{\varsigma_{is}\}_{i=1}^{n_s}\}_{s=1}^t, \{\{\eta_{is}\}_{i=1}^{n_s}\}_{s=1}^t, \{\aleph_{1i}\}_{i=1}^n, \{\aleph_{2s}\}_{s=1}^t, \{\{\aleph_{3is}\}_{i=1}^{n_s}\}_{s=1}^t\big).
\end{aligned}$$

Thus

$$\mathcal{G}_t \subseteq \mathcal{D}_t \subseteq \mathcal{E}_t \subseteq \mathcal{F}_t.$$

Here $\mathcal{G}_t$ is the common information used to form conditional factor means and variances. The design field $\mathcal{D}_t$ adds realized volatility scales and is the conditioning set for reference portfolios, conditional beta densities, and reference cell objects. It excludes the unit-specific auxiliary variables $\aleph_{1i}$ and $\aleph_{3is}$, so after conditioning on $\mathcal{D}_{t-1}$ the predictable unit histories in Assumption SA-DGP(iii)(b) remain independent across i. The environment field $\mathcal{E}_t$ adds those unit-specific auxiliary variables and therefore contains the realized predictable paths α_{is} and β_{is}, but still excludes the i.i.d. innovations. The full field $\mathcal{F}_t$ adds the realized innovations η_{is}; under the bounded-away-from-zero condition on ς_{it}, this is equivalently the field obtained by augmenting $\mathcal{E}_t$ with the idiosyncratic errors ε_{is}.

The corresponding limit sigma-fields are

$$\mathcal{G}_\infty = \sigma(\cup_{t\geq 1}\mathcal{G}_t)\,, \qquad \mathcal{D}_\infty = \sigma(\cup_{t\geq 1}\mathcal{D}_t)\,, \qquad \mathcal{E}_\infty = \sigma(\cup_{t\geq 1}\mathcal{E}_t)\,, \qquad \mathcal{F}_\infty = \sigma(\cup_{t\geq 1}\mathcal{F}_t)\,.$$

Remark SA-II.1.2 (Interpretation of the auxiliary variables)**.** *The variables $\aleph_{1i}$, $\aleph_{2t}$, and $\aleph_{3it}$ are bookkeeping devices for features of the data-generating process not otherwise specified. Their inclusion in the sigma-fields above does not require that they be observed by the econometrician or used to construct the estimator. The variable $\aleph_{1i}$ collects time-invariant unit-specific information, such as a fixed industry classification or a permanent firm type. The variable $\aleph_{2t}$ collects time-varying common information not already summarized by f_t, such as an aggregate volatility or liquidity state. Finally, $\aleph_{3it}$ collects time-varying unit-specific information, such as a predetermined exogenous firm characteristic (e.g., firm age) or an unobserved firm-specific state governing the evolution of its risk factor exposure. Thus, β_{it} may depend on $\aleph_{1i}$ and on the histories of $\aleph_{2s}$, $\aleph_{3is}$, and f_s through date $t-1$, with α_{it} then restricted to satisfy $\alpha_{it} = \alpha_t(\beta_{it})$.*

All of the examples remain subject to Assumption SA-DGP. In particular, Assumption SA-DGP(iii)(b) separately requires the resulting predictable coefficient histories to be independent, although not necessarily identically distributed, across assets conditional on $\mathcal{D}_{t-1}$. For example, fixed industry membership may be included in $\aleph_{1i}$, whereas a time-varying state shared by firms in an industry should be included as a component of the common vector $\aleph_{2t}$, and hence in the conditioning information. Moreover, because the auxiliary variables belong to $\mathcal{E}_\infty$, Assumption SA-DGP(ii)(a) requires the primitive idiosyncratic innovations η_{it} to be independent of them.

SA-II.1.3 Discussion

The filtration notation in this part should be read as the primitive information structure for the classical beta-sorted estimator. Although the same letters $\mathcal{F}_t$ and $\mathcal{G}_t$ are used in Part I, the high-level fields in Part I and the primitive fields in this part are local to their respective arguments. Part I uses a compact two-field notation to state abstract measurability and mean-zero conditions. This part refines that structure by separating common factor information, design information, innovation-free environment information, and full information.

This distinction matters for verifying the high-level idiosyncratic-error arguments under Assumption SA-DGP. The contemporaneous scale ς_{it} is $\mathcal{E}_t$-measurable and need not be measurable with respect to the past full field $\mathcal{F}_{t-1}$ as defined in this part. Thus, if the high-level Part I arguments were applied literally under

the primitive assumptions in this part, the conditioning field used for period-t idiosyncratic cross-sectional arrays would have to fix the current innovation-free scales and design objects while still excluding the current primitive innovations. Equivalently, one may introduce the intermediate field

$$\mathcal{C}_t = \mathcal{E}_t \vee \sigma\big(\{\{\eta_{is}\}_{i=1}^{n_s}\}_{s=1}^{t-1}\big),$$

which contains the past full information and the current innovation-free environment, but not the current shocks $\{\eta_{it} : i \in [n_t]\}$. Conditional on $\mathcal{C}_t$, the variables $\varepsilon_{it} = \varsigma_{it}\eta_{it}$ are independent over i and have conditional mean zero. The first-step estimates, partitions, and reference weights are measurable with respect to subfields of $\mathcal{C}_t$ in the arguments where they are held fixed, and martingale or mean-zero statements relative to coarser past information follow by iterated expectations. The specialized proofs below use this conditioning logic, sometimes through proof-specific stabilized design fields, without changing the rates or substantive conclusions.

SA-II.2 First-Step Beta Estimates

Our main inference results focus on a local-polynomial least-squares estimator of β_{it}, which is commonly used in practice. When the panel is unbalanced, the notation $i \in [n_t]$ in the rolling-window formulas below refers to units observed throughout the window ending at t, so that every lagged observation $R_{i(t-s)}$ with $s \in [H]$ is available. Specifically, let $\mathfrak{p} \geq 0$, let $K(\cdot)$ be a kernel function supported on $[0,1]$, and define

$$r_{\mathfrak{p}}(u) = (1, u, \ldots, u^{\mathfrak{p}})^\top, \qquad X_s(t) = r_{\mathfrak{p}}\left(s/H\right) \otimes (1, f_{t-s}^\top)^\top \in \mathbb{R}^{(d+1)(\mathfrak{p}+1)}.$$

Let $S = (e_2^\top, e_3^\top, \ldots, e_{d+1}^\top)^\top$, where each $e_j \in \mathbb{R}^{(\mathfrak{p}+1)(d+1)}$ is the j-th standard basis vector. Define

$$\widehat{\beta}_{it} = S\,\widehat{\theta}_{it}, \qquad \widehat{\theta}_{it} = \underset{\theta \in \mathbb{R}^{(d+1)(\mathfrak{p}+1)}}{\arg\min} \sum_{s=1}^{H} \Big(R_{i(t-s)} - X_s(t)^\top \theta\Big)^2 K\Big(\frac{s}{H}\Big). \tag{SA-II.2.1}$$

In this case, we can write

$$\widehat{\beta}_{it} = \frac{1}{H}\sum_{s=1}^{H} w_s(t) R_{i(t-s)}, \qquad w_s(t) = S\widehat{A}(t)^{-1} X_s(t) K\Big(\frac{s}{H}\Big), \qquad \widehat{A}(t) = \frac{1}{H}\sum_{s=1}^{H} K\Big(\frac{s}{H}\Big) X_s(t) X_s(t)^\top.$$

Assumption SA-FS (First step: least squares). *$\widehat{\beta}_{it}$ is the least-squares estimator defined in (SA-II.2.1), where $H \to \infty$ and $h = H/T = o(1)$. The kernel $K(\cdot) : [0,1] \to \mathbb{R}_+$ is bounded and Lipschitz continuous, and $\int_0^1 r_{\mathfrak{p}}(s) r_{\mathfrak{p}}(s)^\top K(s)\, \mathrm{d}s$ is positive definite.*

SA-II.2.1 Decomposition

The estimation error admits the decomposition

$$\widehat{\beta}_{it} - \beta_{it} = \big(\breve{\beta}_{it} - \beta_{it}\big) + \frac{1}{H}\sum_{s=1}^{H} w_s(t)\varepsilon_{i(t-s)}, \tag{SA-II.2.2}$$

where

$$\breve{\beta}_{it} = \frac{1}{H}\sum_{s=1}^{H} w_s(t)(\alpha_{i(t-s)} + \beta_{i(t-s)}^{\top} f_{t-s}).$$

The term $\breve{\beta}_{it} - \beta_{it}$ is the first-step bias component, while the second term is the first-step variance component. To characterize the bias component, define

$$\widehat{\mathsf{B}}_{\mathsf{FS}} = \max_{H+1\le t\le T} \max_{i\in[n_t]} \|\breve{\beta}_{it} - \beta_{it}\|.$$

Assumption SA-FSB (First step: bias). *There exists a deterministic sequence* $\mathsf{B}_{\mathsf{FS}} > 0$ *such that* $\mathsf{B}_{\mathsf{FS}} = o(H^{-1/2})$ *and*

$$H^{(d-1)/2} \max_{H+1\le t\le T} \mathbb{P}\Big(\widehat{\mathsf{B}}_{\mathsf{FS}} > \mathsf{B}_{\mathsf{FS}} \,\Big|\, \mathcal{D}_{t-1}\Big) \lesssim_{\mathbb{P}} \mathsf{B}_{\mathsf{FS}} + H^{-1}.$$

The uniform rate is summarized by

$$\mathsf{R}_{\mathsf{FS}} = \sqrt{\frac{\log(nT)}{H}} + \frac{(nT)^{1/q}\log(nT)}{H} + \mathsf{B}_{\mathsf{FS}}, \qquad h = \frac{H}{T}.$$

Assumption SA-FSB implies

$$\widehat{\mathsf{B}}_{\mathsf{FS}} \le \mathsf{B}_{\mathsf{FS}} \qquad \text{w.p.a.1.}$$

Consequently, R_{FS} controls the uniform first-step error. If $(nT)^{2/q}\log(nT)/H \to 0$, the middle term is dominated by $\sqrt{\log(nT)/H}$. Lemma SA-II.2.4 gives primitive bounds for B_{FS} under several smoothness models.

Remark SA-II.2.1 (Stronger and weaker first-step bias conditions)**.** *A stronger sufficient condition for Assumption SA-FSB is* $\widehat{\mathsf{B}}_{\mathsf{FS}} \le \mathsf{B}_{\mathsf{FS}}$ *a.s. for all sufficiently large* n, T*. In addition, inspection of our proofs shows that the weaker conditions* $\widehat{\mathsf{B}}_{\mathsf{FS}} \lesssim_{\mathbb{P}} \mathsf{B}_{\mathsf{FS}}$ *and* $H^{(d-1)/2}\max_{H+1\le t\le T}\max_{i\in[n_t]}\mathbb{P}\big(\|\breve{\beta}_{it} - \beta_{it}\| > \mathsf{B}_{\mathsf{FS}} \,|\, \mathcal{D}_{t-1}\big) \lesssim_{\mathbb{P}} \mathsf{B}_{\mathsf{FS}} + H^{-1}$ *are sufficient. We impose the stronger condition in Assumption SA-FSB for clarity of presentation.*

SA-II.2.2 Results

The first result verifies the uniform concentration of the local-polynomial Gram matrix. This regularity condition is used below to control the inverse Gram matrix uniformly over the rolling windows.

Lemma SA-II.2.2 (First-Step Gram Matrix)**.** *Suppose that Assumptions SA-DGP(iv) and SA-FS hold. Then,*

$$\max_{H+1\le t\le T} \big\|\widehat{A}(t) - \mathbb{E}[\widehat{A}(t)]\big\| \lesssim_{\mathbb{P}} \frac{T^{2/q}}{H} + \sqrt{\frac{\log T}{H}}.$$

Together with Assumptions SA-DGP(iv) and SA-FS, this implies that, if $T^{2/q}/H + \sqrt{\log T/H} = o(1)$, then $\sup_{H+1\le t\le T}\|\widehat{A}(t)^{-1}\| \lesssim_{\mathbb{P}} 1$. The lower bound follows from the uniform positive definiteness of $\mathbb{E}[\bar{f}_t\bar{f}_t^{\top}]$ and the positive definiteness of the kernel-weighted polynomial Gram matrix in Assumption SA-FS.

The next result gives the uniform first-step estimation error rate R_{FS} in the least-squares case, which provides the high-level bound $\widehat{\mathsf{R}}_{\mathsf{FS}} \lesssim_{\mathbb{P}} \mathsf{R}_{\mathsf{FS}}$ for Theorem SA-I.2.1.

Lemma SA-II.2.3 (First-Step Convergence Rate)**.** *Suppose Assumptions SA-DGP(ii) and (iv) and SA-FS and SA-FSB hold, and that* $T^{2/q}/H + \sqrt{\log T/H} = o(1)$*. Then,*

$$\widehat{\mathsf{R}}_{\mathsf{FS}} = \max_{H+1\le t\le T} \max_{i\in[n_t]} \|\widehat{\beta}_{it} - \beta_{it}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(nT)}{H}} + \frac{(nT)^{1/q}\log(nT)}{H} + \mathsf{B}_{\mathsf{FS}} = \mathsf{R}_{\mathsf{FS}}.$$

Assumption SA-DGP does not impose temporal smoothness on the realized coefficient paths. The next result gives primitive bounds on the deterministic threshold B_{FS} under several smoothness models.

Lemma SA-II.2.4 (First-Step Bias Threshold)**.**
Suppose Assumptions SA-DGP(iv) and SA-FS hold, and

$$\frac{T^{2/(q-d-1)}}{H} = o(1).$$

(i) ***Nonrandom*** $\beta_i(\cdot)$***: Hölder-s.*** *Suppose that* $\alpha_{it} = \alpha_i(t/T)$ *and* $\beta_{it} = \beta_i(t/T)$ *for all* i *and* t*, where* $\alpha_i(\cdot)$ *and* $\beta_i(\cdot)$ *are nonrandom, bounded, and Hölder-*s *on* $[0,1]$ *uniformly in* $i\in[n]$ *for some* $s>0$*. If*

$$\sqrt{H}\, h^{\min\{s,\mathfrak{p}+1\}} \to 0,$$

then Assumption SA-FSB holds with

$$\mathsf{B}_{\mathsf{FS}} \lesssim h^{\min\{s,\mathfrak{p}+1\}}.$$

(ii) ***Stochastic*** $\beta_i(\cdot)$***: Brownian Motion.*** *Suppose that* $\beta_{it} = \beta_i^0 + \sigma_i W_i(t/T)$ *for all* i *and* t*, where* $\beta_i^0 \in \mathbb{R}^d$ *is nonrandom,* $W_i(\cdot)$ *is a* d*-dimensional standard Brownian motion on* $[0,1]$*, independent across* i*, and* $0 < \underline{\sigma} \le \sigma_i \le \overline{\sigma} < \infty$ *uniformly in* i*. Suppose also that the conditional intercept function satisfies*

$$\alpha_t(b) = \alpha(t/T, b), \qquad b\in\mathcal{B},$$

where the possibly random map $\alpha(\cdot,\cdot)$ *is bounded, Hölder-*s *in its first argument uniformly over its second argument for some* $s > 1/2$*, and Lipschitz in its second argument uniformly over its first argument, a.s. with deterministic uniform bounds. If*

$$\sqrt{Hh\log(nT)} \to 0,$$

then Assumption SA-FSB holds with

$$\mathsf{B}_{\mathsf{FS}} \lesssim h^{1/2}\big(\log(nT)\big)^{1/2}.$$

(iii) ***Stochastic*** $\beta_i(\cdot)$***: Nonrandom Trend Plus Brownian Motion.*** *Suppose that the conditional intercept function satisfies*

$$\alpha_t(b) = \alpha(t/T, b), \qquad b\in\mathcal{B},$$

where the possibly random map $\alpha(\cdot,\cdot)$ is bounded, Hölder-s in its first argument uniformly over its second argument, and Lipschitz in its second argument uniformly over its first argument, a.s. with deterministic uniform bounds. Suppose also that

$$\beta_{it} = \bar{\beta}_i(t/T) + \sigma_i W_i(t/T),$$

where $\bar{\beta}_i(\cdot)$ is nonrandom and Hölder-s as in case (i) with $s \in (0, \mathfrak{p}+1]$, $W_i(\cdot)$ is a d-dimensional standard Brownian motion on $[0,1]$, independent across i, and the scale condition in case (ii) holds. If

$$\sqrt{H}\,\max\left\{h^s, h^{1/2}\big(\log(nT)\big)^{1/2}\right\} \to 0,$$

then Assumption SA-FSB holds with

$$\mathsf{B}_{\mathtt{FS}} \lesssim \max\left\{h^s, h^{1/2}\big(\log(nT)\big)^{1/2}\right\}.$$

(iv) ***Stochastic $\beta_i(\cdot)$: Integrated Brownian Motion.*** *Suppose that*

$$\beta_{it} = \beta_i^0 + \sigma_i \int_0^{t/T} W_i(v)\,\mathrm{d}v,$$

where $\beta_i^0 \in \mathbb{R}^d$ is nonrandom, $W_i(\cdot)$ is a d-dimensional standard Brownian motion on $[0,1]$, independent across i, and the scale condition in case (ii) holds. Suppose also that

$$\alpha_t(b) = \alpha(t/T, b), \qquad b \in \mathcal{B},$$

where the possibly random map $\alpha(\cdot,\cdot)$ is twice continuously differentiable jointly in its arguments, a.s., and the function and all its derivatives through order two are uniformly bounded by deterministic constants. If $h\log(nT) = o(1)$ and

$$\sqrt{H}\,h^{\min\{\mathfrak{p}+1,\,3/2\}}\big(\log(nT)\big)^{1/2} \to 0,$$

then Assumption SA-FSB holds with

$$\mathsf{B}_{\mathtt{FS}} \lesssim h^{\min\{\mathfrak{p}+1,\,3/2\}}\big(\log(nT)\big)^{1/2}.$$

SA-II.2.3 Discussion

(i) *First-step rate under smooth trends.* Lemma SA-II.2.4 translates the path restrictions stated directly in that lemma into primitive bounds on the deterministic threshold $\mathsf{B}_{\mathtt{FS}}$. If $(nT)^{2/q}\log(nT)/H \to 0$, then the term $(nT)^{1/q}\log(nT)/H$ in $\mathsf{R}_{\mathtt{FS}}$ is dominated by $\sqrt{\log(nT)/H}$, and Lemma SA-II.2.3 gives

$$\mathsf{R}_{\mathtt{FS}} \lesssim_{\mathbb{P}} \sqrt{\frac{\log(nT)}{H}} + \mathsf{B}_{\mathtt{FS}}.$$

In the nonrandom Hölder-s case, writing $h = H/T$, Lemma SA-II.2.4(i) gives $\mathsf{B}_{\mathtt{FS}} \lesssim h^{\min\{s,\mathfrak{p}+1\}}$. Thus the local-polynomial order reduces the first-step bias only up to the smoothness ceiling s.

In the integrated-Brownian case (iv), the analogous distinction is

$$\mathsf{B}_{\mathtt{FS}} \lesssim \begin{cases} h\big(\log(nT)\big)^{1/2}, & \mathfrak{p} = 0, \\ h^{3/2}\big(\log(nT)\big)^{1/2}, & \mathfrak{p} \geq 1. \end{cases}$$

A degree-one fit absorbs the current random slope $\sigma_i W_i(t/T)$ of the integrated path. Higher polynomial orders do not further improve this component because that slope itself has Brownian local variation. In particular, the displayed primitive $\mathsf{B}_{\mathtt{FS}}$ bound is $o(H^{-1/2})$ whenever

$$\sqrt{H}\, h^{\min\{\mathfrak{p}+1,\, 3/2\}}\big(\log(nT)\big)^{1/2} \to 0.$$

The restriction $T^{2/(q-d-1)}/H = o(1)$ in Lemma SA-II.2.4 is stronger than the concentration rate $T^{2/q}/H = o(1)$ used for the first-step Gram matrix because the lemma verifies the conditional-tail requirement in Assumption SA-FSB through the stronger $H^{(d+1)/2}$-weighted global-tail bound. It is implied by $T^{(3+d)/q}/H = o(1)$, the rate used in the partition results, because $q > 2(3+d)$ implies $2/(q-d-1) < (3+d)/q$.

(ii) *Conditional intercepts in the Brownian cases.* The conditions in cases (ii)–(iii) are imposed on the conditional intercept function $\alpha_t(b) = \mu_t(b) - b^\top \mathbb{E}[f_t \mid \mathcal{G}_{t-1}]$, rather than on a separate nonrandom path for α_{it}. Because $\alpha_{it} = \alpha_t(\beta_{it})$, temporal smoothness of $\alpha_t(b)$ at fixed b controls variation in the conditional intercept function, while Lipschitz continuity in b transfers the local modulus of β_{it} to α_{it}. In case (iv), the stronger joint twice-continuous differentiability condition permits a first-order expansion of $\tau \mapsto \alpha(\tau, \beta_i(\tau))$. The local polynomial can therefore absorb both its time derivative and the contribution of the current random beta slope, leaving a remainder of integrated-Brownian order.

(iii) *Compact support in the Brownian cases.* Cases (ii)–(iii) use the Brownian component only through its local modulus of continuity. The proof requires

$$\max_{i\in[n]} \sup_{|u-v|\leq h} \big\|W_i(u) - W_i(v)\big\| \lesssim_{\mathbb{P}} h^{1/2}\big(\log(nT)\big)^{1/2}$$

and does not otherwise use the global level of $W_i(\cdot)$. Since Assumption SA-DGP(iii)(a) imposes a fixed compact support $\mathcal{B}$ on β_{it}, the Brownian specifications should be read as a convenient shorthand for any compact-state process with this same local modulus, such as reflected Brownian motion on $\mathcal{B}$ or a diffusion on a compact state space. For such a compact-state process, the threshold bound $\mathsf{B}_{\mathtt{FS}} \lesssim h^{1/2}(\log(nT))^{1/2}$ is unchanged, provided the remaining requirements of Assumption SA-DGP(iii) are maintained.

The same interpretation applies to case (iv). There the proof uses

$$\max_{i\in[n]} \sup_{|u-v|\leq h} \left\| \int_v^u \big(W_i(z) - W_i(v)\big)\, \mathrm{d}z \right\| \lesssim_{\mathbb{P}} h^{3/2}\big(\log(nT)\big)^{1/2}.$$

Thus the displayed integrated-Brownian specification may be replaced by a compact-state process with continuously differentiable paths whose derivative has the same Brownian local modulus; the rates in case (iv) are unchanged.

SA-II.2.4 Proof of Lemma SA-II.2.2

Recall $X_s(t) = r_{\mathfrak{p}}(s/H) \otimes (1, f_{t-s}^\top)^\top$ and $K_{\max} = \sup_{u\in[0,1]} |K(u)| < \infty$.

Define

$$V_{s,t} = \frac{1}{H} X_s(t) X_s(t)^\top K\Big(\frac{s}{H}\Big),$$

so that $\widehat{A}(t) = \sum_{s=1}^{H} V_{s,t}$.

Put $r_* = (q+\upsilon)/2 > q/2 > 2$.

Step 1: NED property of $V_{s,t}$. Since $X_s(t) = r_{\mathfrak{p}}(s/H) \otimes (1, f_{t-s}^\top)^\top$ is a function of f_{t-s}, Assumption SA-DGP(iv) implies that $X_s(t)X_s(t)^\top$ is L_{r_*}-NED on $\{\xi_t\}$ with geometric approximation constants. To see this, let $\mathcal{H}_{t-s,k} = \sigma(\xi_{t-s-k}, \dots, \xi_{t-s+k})$, $V_{s,t}^{(k)} = \mathbb{E}[V_{s,t} \mid \mathcal{H}_{t-s,k}]$, and $X_s^{(k)}(t) = \mathbb{E}[X_s(t) \mid \mathcal{H}_{t-s,k}]$. The contraction property of conditional expectation and the product inequality give

$$\|V_{s,t} - V_{s,t}^{(k)}\|_{r_*} \le \frac{2K_{\max}}{H} \big(\|X_s(t)\|_{q+\upsilon} + \|X_s^{(k)}(t)\|_{q+\upsilon}\big) \|X_s(t) - X_s^{(k)}(t)\|_{q+\upsilon} \lesssim H^{-1}\rho_a^k.$$

Step 2: Moment bound. Uniformly in s and t, $\mathbb{E}[\|V_{s,t}\|^{r_*}] \lesssim H^{-r_*}\mathbb{E}[\|X_s(t)\|^{q+\upsilon}] \lesssim H^{-r_*}$.

Step 3: Rolling-sum bound. Let Y_u be any centered scalar entry of $(1, f_u^\top)^\top(1, f_u^\top)$. It has a uniformly bounded r_*-th moment and is geometrically L_{r_*}-NED. Set $m = C_0 \log T$ and $Y_u^{(m)} = \mathbb{E}[Y_u \mid \mathcal{H}_{u,m}]$. Because Y_u is centered, so is $Y_u^{(m)}$, and conditional-expectation contraction gives $\|Y_u^{(m)}\|_{r_*} \le \|Y_u\|_{r_*} \lesssim 1$. Moreover, $\{Y_u^{(m)}\}$ is strongly mixing: observations separated by more than $2m$ have mixing coefficient at most $\alpha(j-2m)$. Therefore, at probability level v, its inverse mixing coefficient is at most $2m + C\log(e/v)$, while Markov's inequality bounds the quantile of $|Y_u^{(m)}|$ by Cv^{-1/r_*}. For sufficiently large x, these two bounds give

$$\int_{\{0<v<1:\, C\{m+\log(e/v)\}v^{-1/r_*}>x\}} Cv^{-1/r_*}\, \mathrm{d}v \lesssim x^{-(r_*-1)}\{m + \log(e+x)\}^{r_*-1}.$$

This verifies the tail-integral condition in Rio (2017, Theorem 6.2).

For two entries Y_u and Y_{u+j}, approximate each one using $\mathcal{H}_{u,\lfloor j/3\rfloor}$ and $\mathcal{H}_{u+j,\lfloor j/3\rfloor}$. The covariance inequality in Rio (2017, Theorem 1.1(a)) and the geometric NED bound give, for some $\rho \in (0,1)$,

$$|\mathbb{C}\mathrm{ov}(Y_u, Y_{u+j})| \lesssim \rho_a^{\lfloor j/3\rfloor} + \alpha(\lfloor j/3\rfloor)^{1-2/r_*} \lesssim \rho^j.$$

Thus the sum of the absolute covariances is uniformly bounded. Since $\|Y_u - Y_u^{(m)}\|_2 \lesssim \rho_a^m$, on every interval of length at most $2H$ the sum of the absolute covariances of $Y_u^{(m)}$ is at most $CH + CH^2\rho_a^m \lesssim CH$ when C_0 is sufficiently large.

The variables $Y_u^{(m)}$ have now been verified to be centered and strongly mixing, with the moment, tail-integral, and covariance bounds required by Rio (2017, Theorem 6.2). Apply that theorem on consecutive blocks of length H. An interval crossing a block boundary splits into at most two partial sums, and there are $O(T/H)$ blocks. Also, for any fixed $A > 0$, C_0 can be chosen so that, uniformly for $\lambda \ge 1$,

$$\mathbb{P}\left(\sum_{u=1}^{T} |Y_u - Y_u^{(m)}| > \lambda\right) \le \lambda^{-1} \sum_{u=1}^{T} \mathbb{E}|Y_u - Y_u^{(m)}| \le \frac{CT\rho_a^m}{\lambda} = o(T^{-A}).$$

Choose C_0 large enough to absorb this approximation error. Consequently, for any $a \ge 1$ and $1 \le \lambda \le T$

with λ/a sufficiently large, we have the following bound, where the maximum is over time intervals I:

$$\mathbb{P}\left(\max_{|I|\le H}\left|\sum_{u\in I} Y_u\right| > C\lambda\right) \lesssim \frac{T}{H}\left(1+\frac{\lambda^2}{aH}\right)^{-a/2} + Ta^{r_*-1}\lambda^{-r_*}(\log T)^{r_*-1}. \tag{SA-II.2.3}$$

Indeed, the tail-integral term on one block is the preceding integral multiplied by H/λ and evaluated at $x = \lambda/a$.

For each matrix entry, the deterministic weights formed from $r_{\mathfrak{p}}(s/H)$ and $K(s/H)$ are uniformly bounded and have bounded total change over $s = 1,\dots,H$. Expanding a weighted sum into partial sums and successive changes in these weights therefore bounds it by a constant times the largest unweighted partial sum in (SA-II.2.3). Take $a = M\log T$ and $\lambda = M\{T^{2/q} + \sqrt{H\log T}\}$. The first term in (SA-II.2.3) tends to zero for sufficiently large M. For the second term, the moment gap gives

$$T\lambda^{-r_*}(\log T)^{2(r_*-1)} \lesssim T^{1-2r_*/q}(\log T)^{2(r_*-1)} = T^{-\upsilon/q}(\log T)^{2(r_*-1)} = o(1).$$

A finite union over the matrix entries now yields

$$\max_{H+1\le t\le T}\left\|\widehat{A}(t) - \mathbb{E}[\widehat{A}(t)]\right\| \lesssim_{\mathbb{P}} \frac{T^{2/q}}{H} + \sqrt{\frac{\log T}{H}}.$$

The covariance bound above also gives $(T-H)^{-1}\sum_{t=H+1}^{T}\|\widehat{A}(t) - \mathbb{E}[\widehat{A}(t)]\|^2 = O_{\mathbb{P}}(H^{-1})$ by taking expectations and applying Markov's inequality. The same rolling-sum argument applies to the fixed powers of $\|f_t\|$ used below because their available moment is strictly above the working moment used in the corresponding rate. □

SA-II.2.5 Proof of Lemma SA-II.2.3

Define

$$N_i(t) = \frac{1}{H}\sum_{s=1}^{H} X_s(t)R_{i(t-s)}K\Big(\frac{s}{H}\Big), \qquad \text{and} \qquad U_{it} = \frac{1}{H}\sum_{s=1}^{H} X_s(t)\varepsilon_{i(t-s)}K\Big(\frac{s}{H}\Big).$$

The normal equations give $\widehat{\theta}_{it} = \widehat{A}(t)^{-1}N_i(t)$. Since

$$N_i(t) = \frac{1}{H}\sum_{s=1}^{H} X_s(t)\big(\alpha_{i(t-s)} + \beta_{i(t-s)}^{\top}f_{t-s}\big)K\Big(\frac{s}{H}\Big) + U_{it},$$

the definition of $\breve{\beta}_{it}$ implies the exact identity

$$\widehat{\beta}_{it} - \breve{\beta}_{it} = S\widehat{A}(t)^{-1}U_{it}.$$

Step 1: Bounding $\widehat{A}(t)^{-1}$. By Assumptions SA-DGP(iv) and SA-FS, $\lambda_{\min}(\mathbb{E}[\widehat{A}(t)]) \ge c > 0$ uniformly in $t = H+1,\dots,T$. Indeed, writing $\bar{f}_r = (1, f_r^{\top})^{\top}$, any vector a conformable with $X_s(t)$ can be written as polynomial coefficients $a_0,\dots,a_{\mathfrak{p}} \in \mathbb{R}^{d+1}$, so that $(X_s(t)^{\top}a)^2 = (\bar{f}_{t-s}^{\top}a(s/H))^2$, where $a(u) = \sum_{k=0}^{\mathfrak{p}} a_k u^k$. The uniform lower bound on $\mathbb{E}[\bar{f}_r\bar{f}_r^{\top}]$ gives a constant multiple of $H^{-1}\sum_{s=1}^{H}K(s/H)\|a(s/H)\|^2$, and this discrete polynomial Gram matrix is bounded away from zero for all large H by the positive definite integral

in Assumption SA-FS. By Lemma SA-II.2.2, $\|\widehat{A}(t) - \mathbb{E}[\widehat{A}(t)]\| \lesssim_{\mathbb{P}} T^{2/q}/H + \sqrt{\log T/H} = o_{\mathbb{P}}(1)$ uniformly in t. Weyl's inequality then gives

$$\lambda_{\min}(\widehat{A}(t)) \geq \lambda_{\min}(\mathbb{E}[\widehat{A}(t)]) - \|\widehat{A}(t) - \mathbb{E}[\widehat{A}(t)]\| \geq c - o_{\mathbb{P}}(1) > 0$$

uniformly over $t \in [H+1, T]$ with probability approaching one. Hence

$$\sup_{H+1 \leq t \leq T} \|\widehat{A}(t)^{-1}\| \lesssim_{\mathbb{P}} 1. \tag{SA-II.2.4}$$

Step 2: Bounding U_{it}. Condition on $\mathcal{E}_\infty$. Then the factor history and volatility scales are fixed, and $\varepsilon_{i(t-s)} = \varsigma_{i(t-s)}\eta_{i(t-s)}$ with independent mean-zero innovations. Hence, for fixed (i, t_0), the array $\{X_s(t_0)\varepsilon_{i(t_0-s)} : s \in [H]\}$ is a martingale difference array in the index s with respect to the filtration $\mathcal{H}_{s,it_0} = \mathcal{E}_\infty \vee \sigma(\eta_{i(t_0-r)} : 1 \leq r \leq s)$. The probability bounds below can therefore be read conditionally on $\mathcal{E}_\infty$ and then integrated. Abel's summation formula gives, for any $t_0 \in [H+1, T]$,

$$\begin{aligned}
&\left\| \sum_{s=1}^{H} K\Big(\frac{s}{H}\Big) X_s(t_0)\varepsilon_{i(t_0-s)} \right\| \\
&\quad \leq \max_{\ell \in [H]} \left\| \sum_{s=1}^{\ell} X_s(t_0)\varepsilon_{i(t_0-s)} \right\| \sum_{s=1}^{H-1} \left| K\Big(\frac{s+1}{H}\Big) - K\Big(\frac{s}{H}\Big) \right| + \left\| \sum_{s=1}^{H} X_s(t_0)\varepsilon_{i(t_0-s)} \right\| K(1).
\end{aligned}$$

Since K is Lipschitz continuous, $\sum_{s=1}^{H-1} |K(\frac{s+1}{H}) - K(\frac{s}{H})| \leq L_K$ for a finite constant L_K, and $K(1) \leq \|K\|_\infty < \infty$. Therefore,

$$\max_{H+1 \leq t_0 \leq T} \max_{i \in [n_{t_0}]} \|U_{it_0}\| \lesssim \frac{1}{H} \max_{H+1 \leq t_0 \leq T} \max_{i \in [n_{t_0}]} \max_{\ell \in [H]} \left\| \sum_{s=1}^{\ell} X_s(t_0)\varepsilon_{i(t_0-s)} \right\|.$$

Fix t_0 and i, put $\tau = (nT)^{1/q}$, and write

$$Y_{s,it_0} = X_s(t_0)\varepsilon_{i(t_0-s)}, \qquad s \in [H].$$

For a fixed constant $M > 0$, define the conditionally recentered truncated increment

$$Y^{(M)}_{s,it_0} = Y_{s,it_0}\mathbb{1}(\|Y_{s,it_0}\| \leq M\tau) - \mathbb{E}[Y_{s,it_0}\mathbb{1}(\|Y_{s,it_0}\| \leq M\tau) \,|\, \mathcal{H}_{s-1,it_0}].$$

By construction,

$$\mathbb{E}\Big[Y^{(M)}_{s,it_0} \,\Big|\, \mathcal{H}_{s-1,it_0}\Big] = 0, \qquad \|Y^{(M)}_{s,it_0}\| \leq 2M\tau.$$

Thus the truncated array remains a martingale-difference array with respect to the original filtration; no conditioning on a future-shock event is involved.

Its predictable quadratic variation satisfies, conditional on $\mathcal{E}_\infty$,

$$\sum_{s=1}^{H} \mathbb{E}\Big[\|Y^{(M)}_{s,it_0}\|^2 \,\Big|\, \mathcal{H}_{s-1,it_0}\Big] \lesssim \sum_{s=1}^{H} \|X_s(t_0)\|^2,$$

because the volatility scales are bounded and $\mathbb{E}[\eta_{it}^2] = 1$. Moreover,

$$\max_{H+1\le t\le T} \frac{1}{H}\sum_{s=1}^{H} \|X_s(t)\|^2 \lesssim_{\mathbb{P}} 1,$$

by applying the same maximal Fuk–Nagaev inequality used in Lemma SA-II.2.2 to the $L_{q/2}$-NED sequence $\|\bar{f}_t\|^2$. Hence, on an $\mathcal{E}_\infty$-measurable event whose probability can be made arbitrarily close to one, the coordinatewise maximal martingale Bernstein inequality (Freedman, 1975) gives

$$\mathbb{P}\left(\max_{\ell\in[H]}\left\|\sum_{s=1}^{\ell} Y_{s,it_0}^{(M)}\right\| > x \,\middle|\, \mathcal{E}_\infty\right) \lesssim \exp\left\{-\frac{cx^2}{H + M\tau x}\right\}. \tag{SA-II.2.5}$$

The dimension of Y_{s,it_0} is fixed, so the coordinatewise inequality and a finite union bound imply the displayed norm bound.

It remains to transfer this inequality to the original increments. Let

$$\mathcal{A}_M = \left\{\max_{H+1\le t_0\le T}\max_{i\in[n_{t_0}]}\max_{s\in[H]} \|Y_{s,it_0}\| \le M\tau\right\}.$$

Since $r_{\mathfrak{p}}(\cdot)$ is bounded, the maximum over rolling-window summands is controlled by the maximum over the nT calendar shock pairs. Markov's inequality, the finite q-th moments of $(1+\|f_r\|)\varepsilon_{ir}$, and a union bound give

$$\mathbb{P}(\mathcal{A}_M^c) \lesssim M^{-q}. \tag{SA-II.2.6}$$

Because $\mathbb{E}[Y_{s,it_0} \mid \mathcal{H}_{s-1,it_0}] = 0$, on $\mathcal{A}_M$,

$$\sum_{s=1}^{\ell} Y_{s,it_0} = \sum_{s=1}^{\ell} Y_{s,it_0}^{(M)} - \sum_{s=1}^{\ell} \mathbb{E}[Y_{s,it_0}\mathbb{1}(\|Y_{s,it_0}\| > M\tau) \,|\, \mathcal{H}_{s-1,it_0}], \qquad \ell \in [H].$$

The recentering term is negligible at the target scale. Indeed, uniformly over (i, t_0), conditional Markov's inequality and the bounded volatility scales give

$$\begin{aligned}\sum_{s=1}^{H} \|\mathbb{E}[Y_{s,it_0}\mathbb{1}(\|Y_{s,it_0}\| > M\tau) \,|\, \mathcal{H}_{s-1,it_0}]\| &\le (M\tau)^{1-q}\sum_{s=1}^{H}\mathbb{E}[\|Y_{s,it_0}\|^q \,|\, \mathcal{E}_\infty] \\ &\lesssim_{\mathbb{P}} (M\tau)^{1-q}\sum_{r=1}^{T}(1+\|f_r\|)^q \lesssim_{\mathbb{P}} M^{1-q}\frac{\tau}{n} \lesssim_{\mathbb{P}} M^{1-q}\tau,\end{aligned}$$

where $\tau^q = nT$ and $T^{-1}\sum_{r=1}^{T}(1+\|f_r\|)^q = O_{\mathbb{P}}(1)$ by the uniform q-moment bound in Assumption SA-DGP(iv).

For fixed M, taking x to be a sufficiently large multiple of $\sqrt{H\log(nT)} + \tau\log(nT)$ in (SA-II.2.5) and applying a union bound over at most nT pairs (i, t_0) gives a probability tending to zero. The recentering term is $o_{\mathbb{P}}\big(\sqrt{H\log(nT)} + \tau\log(nT)\big)$, while (SA-II.2.6) can be made arbitrarily small by choosing M large.

Therefore,

$$\max_{H+1\le t_0\le T}\max_{i\in[n_{t_0}]}\max_{\ell\in[H]}\left\|\sum_{s=1}^{\ell}Y_{s,it_0}\right\|\lesssim_{\mathbb{P}}\sqrt{H\log(nT)}+(nT)^{1/q}\log(nT).$$

Dividing by H in the definition of U_{it} yields

$$\max_{H+1\le t\le T}\max_{i\in[n_t]}\|U_{it}\|\lesssim_{\mathbb{P}}\sqrt{\frac{\log(nT)}{H}}+\frac{(nT)^{1/q}\log(nT)}{H}.$$

Step 3: Combining with the realized bias. By the exact identity above, (SA-II.2.4), and the preceding rate bound,

$$\max_{H+1\le t\le T}\max_{i\in[n_t]}\|\widehat{\beta}_{it}-\breve{\beta}_{it}\|\lesssim_{\mathbb{P}}\sqrt{\frac{\log(nT)}{H}}+\frac{(nT)^{1/q}\log(nT)}{H}.$$

The triangle inequality and Assumption SA-FSB, which gives $\widehat{\mathsf{B}}_{\mathtt{FS}}\le\mathsf{B}_{\mathtt{FS}}$ w.p.a.1, yield

$$\max_{H+1\le t\le T}\max_{i\in[n_t]}\|\widehat{\beta}_{it}-\beta_{it}\|\lesssim_{\mathbb{P}}\sqrt{\frac{\log(nT)}{H}}+\frac{(nT)^{1/q}\log(nT)}{H}+\mathsf{B}_{\mathtt{FS}}.$$

□

SA-II.2.6 Proof of Lemma SA-II.2.4

We first give an argument common to all four cases. Apply (SA-II.2.3) with the fixed threshold $\lambda=cH$ and $a=q+2$, both to the rolling Gram matrices and to the rolling averages of $(1+\|f_t\|)^2$. Because $T^{2/(q-d-1)}/H=o(1)$, H grows at a positive power of T and hence $H^{-\upsilon/2}(\log T)^{(q+\upsilon)/2-1}=o(1)$. The two terms in (SA-II.2.3) are therefore bounded by $CTH^{-q/2}$. It follows that, for a sufficiently large C,

$$\begin{aligned}&\mathbb{P}\left(\max_{H+1\le t\le T}\|\widehat{A}(t)^{-1}\|>C\right)\\&\quad+\mathbb{P}\left(\max_{H+1\le t\le T}\frac{1}{H}\sum_{s=1}^{H}(1+\|f_{t-s}\|)^2>C\right)\\&\lesssim TH^{-q/2}=o(H^{-(d+1)/2}).\end{aligned}\tag{SA-II.2.7}$$

Indeed,

$$H^{(d+1)/2}TH^{-q/2}=\left(\frac{T^{2/(q-d-1)}}{H}\right)^{(q-d-1)/2}\to 0.$$

The inverse bound follows from the uniform lower eigenvalue bound for $\mathbb{E}[\widehat{A}(t)]$ and the same fixed-threshold inequality.

Now let $a_h>0$ be deterministic. Suppose that, outside an event with probability $o(H^{-(d+1)/2})$, for every (i,t) there exists a coefficient vector $\gamma_{it}=(\gamma_{it,0}^\top,\dots,\gamma_{it,\mathfrak{p}}^\top)^\top$ satisfying $S\gamma_{it}=\beta_{it}$ and, uniformly over $s\in[H]$,

$$\alpha_{i(t-s)}+\beta_{i(t-s)}^\top f_{t-s}=X_s(t)^\top\gamma_{it}+e_{its},\qquad |e_{its}|\le Ca_h(1+\|f_{t-s}\|).\tag{SA-II.2.8}$$

Using the definition of $\breve{\beta}_{it}$ and the normal equations,

$$\breve{\beta}_{it} - \beta_{it} = S\widehat{A}(t)^{-1}\frac{1}{H}\sum_{s=1}^{H} X_s(t)e_{its}K\Big(\frac{s}{H}\Big).$$

Consequently, (SA-II.2.7) and boundedness of the kernel imply, for a sufficiently large fixed constant C,

$$\mathbb{P}\Big(\widehat{\mathsf{B}}_{\mathsf{FS}} > Ca_h\Big) = o(H^{-(d+1)/2}). \tag{SA-II.2.9}$$

Set $\mathsf{B}_{\mathsf{FS}} = Ca_h$. If $\sqrt{H}\,a_h \to 0$, then $\mathsf{B}_{\mathsf{FS}} = o(H^{-1/2})$. Moreover, because

$$\mathbb{E}\Big[\mathbb{1}\Big(\widehat{\mathsf{B}}_{\mathsf{FS}} > \mathsf{B}_{\mathsf{FS}}\Big)\,\Big|\,\mathcal{D}_{t-1}\Big], \qquad t = H+1,\dots,T,$$

is a nonnegative martingale sampled along the nested design filtration, Doob's maximal inequality and (SA-II.2.9) give, for every $\epsilon > 0$,

$$\mathbb{P}\Big(H^{(d+1)/2}\max_{H+1\le t\le T}\mathbb{P}\Big(\widehat{\mathsf{B}}_{\mathsf{FS}} > \mathsf{B}_{\mathsf{FS}}\,\Big|\,\mathcal{D}_{t-1}\Big) > \epsilon\Big) \le \frac{H^{(d+1)/2}}{\epsilon}\mathbb{P}\Big(\widehat{\mathsf{B}}_{\mathsf{FS}} > \mathsf{B}_{\mathsf{FS}}\Big) \to 0.$$

Consequently,

$$\frac{H^{(d-1)/2}}{\mathsf{B}_{\mathsf{FS}} + H^{-1}}\max_{H+1\le t\le T}\mathbb{P}\Big(\widehat{\mathsf{B}}_{\mathsf{FS}} > \mathsf{B}_{\mathsf{FS}}\,\Big|\,\mathcal{D}_{t-1}\Big) \le H^{(d+1)/2}\max_{H+1\le t\le T}\mathbb{P}\Big(\widehat{\mathsf{B}}_{\mathsf{FS}} > \mathsf{B}_{\mathsf{FS}}\,\Big|\,\mathcal{D}_{t-1}\Big) = o_{\mathbb{P}}(1),$$

which verifies Assumption SA-FSB. Thus it remains only to verify (SA-II.2.8) with the stated a_h and tail probability.

Case (i). The standard Taylor–Hölder expansion of $\alpha_i(\cdot)$ and each coordinate of $\beta_i(\cdot)$ around t/T gives a local polynomial in s/H of degree at most $\mathfrak{p}$, with a deterministic remainder bounded by $a_h = h^{\min\{s,\mathfrak{p}+1\}}$ uniformly over i, t, and $s \in [H]$. The common argument therefore proves part (i).

Case (ii). For every fixed $M > 0$, the Gaussian maximal inequality and the Brownian modulus of continuity give a fixed constant C_M such that

$$\mathbb{P}\left(\max_{i\in[n]}\sup_{|u-v|\le h}\|W_i(u) - W_i(v)\| > C_M h^{1/2}\big(\log(nT)\big)^{1/2}\right) \lesssim (nT)^{-M}. \tag{SA-II.2.10}$$

Take $M > (d+1)/2$. Put the current Brownian level in the constant beta coefficient, so that $S\gamma_{it} = \beta_{it}$. For the intercept component, use the degree-$\mathfrak{p}$ local Taylor–Hölder polynomial for $\tau \mapsto \alpha(\tau, \beta_{it})$ around t/T. Its temporal remainder is $O(h^{\min\{s,\mathfrak{p}+1\}})$. Lipschitz continuity in the beta argument adds at most a constant multiple of the Brownian increment in (SA-II.2.10). Since $\min\{s, \mathfrak{p}+1\} > 1/2$, (SA-II.2.8) holds outside an $o(H^{-(d+1)/2})$ event with $a_h = h^{1/2}\big(\log(nT)\big)^{1/2}$. The common argument proves part (ii).

Case (iii). Choose the local polynomial coefficients for the nonrandom beta trend as in case (i), and put the current Brownian level in the constant beta coefficient. The temporal intercept remainder is bounded by h^s, while the Lipschitz contribution from the beta argument is bounded by the beta increment. Its nonrandom component is of order $h^{\min\{s,1\}}$, which is bounded by the claimed rate, and its Brownian component satisfies (SA-II.2.10). Therefore (SA-II.2.8) holds outside an $o(H^{-(d+1)/2})$ event with $a_h = \max\{h^s, h^{1/2}\big(\log(nT)\big)^{1/2}\}$. The common argument proves part (iii). The Brownian component is

nowhere differentiable, so increasing $\mathfrak{p}$ does not improve this component of the threshold.

Case (iv). Write $I_i(u) = \int_0^u W_i(v)\,\mathrm{d}v$. For every fixed $M > 0$, the Brownian maximal and modulus inequalities give a fixed constant C_M such that the probability that either

$$\max_{i\in[n]} \sup_{u\in[0,1]} \|W_i(u)\| > C_M\big(\log(nT)\big)^{1/2},$$
$$\max_{i\in[n]} \sup_{|u-v|\le h} \|I_i(u) - I_i(v) - W_i(v)(u-v)\| > C_M h^{3/2}\big(\log(nT)\big)^{1/2}$$

is bounded by a constant multiple of $(nT)^{-M}$. Take $M > (d+1)/2$.

If $\mathfrak{p} = 0$, use the current levels of the beta and conditional intercept paths as the constant local-polynomial coefficients. Boundedness of the first derivatives of α gives (SA-II.2.8) with $a_h = h\big(\log(nT)\big)^{1/2}$. If $\mathfrak{p} \ge 1$, include the current slope $\sigma_i W_i(t/T)$ in the degree-one beta coefficient. For $g_i(u) = \alpha(u, \beta_i(u))$, also include

$$g_i'(u) = \partial_1\alpha(u,\beta_i(u)) + \nabla_b\alpha(u,\beta_i(u))^\top \sigma_i W_i(u)$$

at $u = t/T$ in the degree-one intercept coefficient. The joint second-derivative bound on α, the Brownian bounds above, and $h\log(nT) = o(1)$ imply (SA-II.2.8) with $a_h = h^{3/2}\big(\log(nT)\big)^{1/2}$. Combining the two polynomial orders gives the threshold in part (iv), and the common argument completes the proof. □

SA-II.3 Second-Step Partitioning

Using the generic partition (portfolios) notation from Section SA-I.3 and the concrete design filtration $\{\mathcal{D}_t\}$ from Section SA-II.1.2, we now specialize the estimated partition (portfolio) to the marginal-quantile scheme used by the classical beta-sorted estimator. Recall that $\mathcal{B} = [\underline{\beta}, \bar{\beta}]$ is a rectangle. The partition is formed by tensor products across coordinates:

$$\widehat{\Delta}_t = \otimes_{l\in[d]}\widehat{\Delta}_{t,l}, \qquad \widehat{\Delta}_{t,l} = \Big\{[\hat{\tau}_{0t,l}, \hat{\tau}_{1t,l}), [\hat{\tau}_{1t,l}, \hat{\tau}_{2t,l}), \ldots, [\hat{\tau}_{(J_{t,l}-1)t,l}, \hat{\tau}_{J_{t,l}t,l}]\Big\}.$$

Here $\hat{\tau}_{jt,l}$ denotes a possibly random breakpoint and $J_{t,l}$ denotes the number of intervals for coordinate $l \in [d]$, so that $\bar{J}_t = \prod_{l\in[d]} J_{t,l}$.

For any first-step estimator $\widehat{\beta}_{it} = (\widehat{\beta}_{it,1}, \ldots, \widehat{\beta}_{it,d})^\top$, coordinate $l \in [d]$, and time $t = H+1, \ldots, T$, define

$$\widehat{F}_{\widehat{\beta},t,l}(z) = \frac{1}{n_t}\sum_{i=1}^{n_t} \mathbb{1}(\widehat{\beta}_{it,l} \le z) \qquad \text{and} \qquad \bar{F}_{\widehat{\beta},t,l}(z) = \frac{1}{n_t}\sum_{i=1}^{n_t} \mathbb{P}(\widehat{\beta}_{it,l} \le z \mid \mathcal{D}_{t-1}).$$

Their generalized inverse functions are

$$\widehat{F}^{-1}_{\widehat{\beta},t,l}(\kappa) = \inf\{z\in\mathbb{R} : \widehat{F}_{\widehat{\beta},t,l}(z) \ge \kappa\} \qquad \text{and} \qquad \bar{F}^{-1}_{\widehat{\beta},t,l}(\kappa) = \inf\{z\in\mathbb{R} : \bar{F}_{\widehat{\beta},t,l}(z) \ge \kappa\},$$

for $\kappa \in (0,1)$.

Our main inference results focus on partitions formed by tensor products of marginal empirical quantiles of the components of $\widehat{\beta}_{it}$, separately on the negative and nonnegative sides of zero. For each $l \in [d]$ and

$t = H+1, \ldots, T$, define the split sample sizes

$$\widehat{N}^-_{t,l} = \sum_{i=1}^{n_t} \mathbb{1}(\widehat{\beta}_{it,l} < 0) \qquad \text{and} \qquad \widehat{N}^+_{t,l} = \sum_{i=1}^{n_t} \mathbb{1}(\widehat{\beta}_{it,l} \geq 0).$$

On the event $\min\{\widehat{N}^-_{t,l}, \widehat{N}^+_{t,l}\} > 0$, define the split empirical distribution functions by

$$\widehat{F}_{\widehat{\beta},t,l,-}(z) = \frac{1}{\widehat{N}^-_{t,l}} \sum_{i=1}^{n_t} \mathbb{1}(\widehat{\beta}_{it,l} < 0)\mathbb{1}(\widehat{\beta}_{it,l} \leq z) \qquad \text{and} \qquad \widehat{F}_{\widehat{\beta},t,l,+}(z) = \frac{1}{\widehat{N}^+_{t,l}} \sum_{i=1}^{n_t} \mathbb{1}(\widehat{\beta}_{it,l} \geq 0)\mathbb{1}(\widehat{\beta}_{it,l} \leq z),$$

with generalized inverse functions

$$\widehat{F}^{-1}_{\widehat{\beta},t,l,-}(\kappa) = \inf\{z < 0 : \widehat{F}_{\widehat{\beta},t,l,-}(z) \geq \kappa\} \qquad \text{and} \qquad \widehat{F}^{-1}_{\widehat{\beta},t,l,+}(\kappa) = \inf\{z \geq 0 : \widehat{F}_{\widehat{\beta},t,l,+}(z) \geq \kappa\},$$

for $\kappa \in (0,1)$. If either split sample is empty, define its empirical distribution and quantiles arbitrarily. The uniform empirical sign-share bound in the proof of Lemma SA-II.3.2 shows that this event has probability approaching zero.

Assumption SA-PA (Marginal-quantile-spaced partition). *For each $t = H+1, \ldots, T$, the partition $\widehat{\Delta}_t = \otimes_{l\in[d]} \widehat{\Delta}_{t,l}$ with*

$$\widehat{\Delta}_{t,l} = \Big\{[\hat{\tau}^-_{0t,l}, \hat{\tau}^-_{1t,l}), [\hat{\tau}^-_{1t,l}, \hat{\tau}^-_{2t,l}), \ldots, [\hat{\tau}^-_{(J^-_{t,l}-1)t,l}, 0), [0, \hat{\tau}^+_{1t,l}), [\hat{\tau}^+_{1t,l}, \hat{\tau}^+_{2t,l}), \ldots, [\hat{\tau}^+_{(J^+_{t,l}-1)t,l}, \hat{\tau}^+_{J^+_{t,l}t,l}]\Big\}$$

where $\hat{\tau}^-_{0t,l} = \min_{i:\widehat{\beta}_{it,l}<0} \widehat{\beta}_{it,l}$ and $\hat{\tau}^+_{J^+_{t,l}t,l} = \max_{i:\widehat{\beta}_{it,l}\geq 0} \widehat{\beta}_{it,l}$, $\hat{\tau}^-_{jt,l} = \widehat{F}^{-1}_{\widehat{\beta},t,l,-}(j/J^-_{t,l})$ for $j \in [J^-_{t,l}-1]$, $\hat{\tau}^+_{jt,l} = \widehat{F}^{-1}_{\widehat{\beta},t,l,+}(j/J^+_{t,l})$ for $j \in [J^+_{t,l}-1]$, and $J^-_{t,l} + J^+_{t,l} = J_{t,l}$. Assume that $J^-_{t,l}, J^+_{t,l} \asymp J$ for all $l \in [d]$ and some $J \to \infty$ as $n \to \infty$, and $J^d/n = o(1)$.

The corresponding reference partition is denoted by $\Delta_t = \otimes_{l\in[d]} \Delta_{t,l}$ and is nonrandom conditionally on $\mathcal{D}_{t-1}$. It uses the conditional analogs of the same split quantiles, and should be read as the conditional reference counterpart of the estimated quantile partition, rather than as a partition formed from the distribution of the true β_{it}. For each coordinate $l \in [d]$,

$$\Delta_{t,l} = \Big\{(\tau^-_{0t,l}, \tau^-_{1t,l}), [\tau^-_{1t,l}, \tau^-_{2t,l}), \ldots, [\tau^-_{(J^-_{t,l}-1)t,l}, 0), [0, \tau^+_{1t,l}), [\tau^+_{1t,l}, \tau^+_{2t,l}), \ldots, [\tau^+_{(J^+_{t,l}-1)t,l}, \tau^+_{J^+_{t,l}t,l})\Big\},$$

where $\tau^-_{0t,l} = -\infty$ and $\tau^+_{J^+_{t,l}t,l} = +\infty$, $\tau^-_{jt,l} = \bar{F}^{-1}_{\widehat{\beta},t,l,-}(j/J^-_{t,l})$ for $j \in [J^-_{t,l}-1]$, and $\tau^+_{jt,l} = \bar{F}^{-1}_{\widehat{\beta},t,l,+}(j/J^+_{t,l})$ for $j \in [J^+_{t,l}-1]$, with

$$\bar{\pi}_{t,l,-} = \frac{1}{n_t} \sum_{i=1}^{n_t} \mathbb{P}(\widehat{\beta}_{it,l} < 0 \mid \mathcal{D}_{t-1}), \qquad \bar{\pi}_{t,l,+} = \frac{1}{n_t} \sum_{i=1}^{n_t} \mathbb{P}(\widehat{\beta}_{it,l} \geq 0 \mid \mathcal{D}_{t-1}),$$

$$\bar{F}_{\widehat{\beta},t,l,-}(z) = \frac{1}{n_t \bar{\pi}_{t,l,-}} \sum_{i=1}^{n_t} \mathbb{P}(\widehat{\beta}_{it,l} \leq z, \widehat{\beta}_{it,l} < 0 \mid \mathcal{D}_{t-1}), \qquad z < 0,$$

$$\bar{F}_{\widehat{\beta},t,l,+}(z) = \frac{1}{n_t \bar{\pi}_{t,l,+}} \sum_{i=1}^{n_t} \mathbb{P}(0 \leq \widehat{\beta}_{it,l} \leq z \mid \mathcal{D}_{t-1}), \qquad z \geq 0.$$

Thus $\bar{F}_{\widehat{\beta},t,l,-}$ and $\bar{F}_{\widehat{\beta},t,l,+}$ are distribution functions on the negative and nonnegative sides, respectively. Their generalized inverses are defined over $z < 0$ and $z \geq 0$, respectively, as for the split empirical distribution

functions above. Accordingly, define marginal split-mixture densities

$$\bar{f}_{\widehat{\beta},t,l,-}(z) = \frac{1}{n_t \bar{\pi}_{t,l,-}} \sum_{i=1}^{n_t} f_{\widehat{\beta}_{it,l}|\mathcal{D}_{t-1}}(z|\mathcal{D}_{t-1}) \mathbb{1}(z < 0) \quad \text{and}$$
$$\bar{f}_{\widehat{\beta},t,l,+}(z) = \frac{1}{n_t \bar{\pi}_{t,l,+}} \sum_{i=1}^{n_t} f_{\widehat{\beta}_{it,l}|\mathcal{D}_{t-1}}(z|\mathcal{D}_{t-1}) \mathbb{1}(z \geq 0).$$

These ratios are defined as zero if the corresponding pooled sign probability vanishes. Lemma SA-II.3.1(i) implies that $\bar{\pi}_{t,l,-}$ and $\bar{\pi}_{t,l,+}$ are uniformly bounded away from zero with probability approaching one, so this convention is asymptotically irrelevant. Cell membership, counts, and polynomial basis evaluation all use the raw generated regressor $\widehat{\beta}_{it}$. The estimated partition therefore covers the coordinatewise range of the realized generated regressors, while the reference partition uses the corresponding complete conditional split distributions. The cells outside $\mathcal{B}^\circ$ may be unbounded, but they do not intersect the interior evaluation region with probability approaching one and are not used in any design or inverse-Gram claim below. The reference basis $p_t(\cdot)$ is defined from this Δ_t, and the reference Gram matrix Q_t is the conditional expectation of $p_t(\widehat{\beta}_{it})p_t(\widehat{\beta}_{it})^\top$ as in Section SA-I.3. For the verification with polynomial degree $\mathfrak{q} \geq 0$, the basis normalization is cell-specific. For $j \in \mathcal{J}_t$, write the paired reference and estimated rectangles as $P_{jt} = \prod_{l\in[d]}[a_{jtl}, b_{jtl})$ and $\widehat{P}_{jt} = \prod_{l\in[d]}[\widehat{a}_{jtl}, \widehat{b}_{jtl})$, with the endpoint and tie-breaking conventions inherited from the corresponding partitions. We set

$$c_{jt,l} = \frac{a_{jtl} + b_{jtl}}{2}, \qquad D_{jt} = \operatorname{diag}\{b_{jt1} - a_{jt1}, \ldots, b_{jtd} - a_{jtd}\},$$
$$\widehat{c}_{jt,l} = \frac{\widehat{a}_{jtl} + \widehat{b}_{jtl}}{2}, \qquad \widehat{D}_{jt} = \operatorname{diag}\{\widehat{b}_{jt1} - \widehat{a}_{jt1}, \ldots, \widehat{b}_{jtd} - \widehat{a}_{jtd}\}.$$

Thus, for b in the reference cell P_{jt}, the polynomial block is $\check{p}_{\mathfrak{q}}(D_{jt}^{-1}(b - c_{jt}))$, while for b in the estimated cell $\widehat{P}_{jt}$, the polynomial block is $\check{p}_{\mathfrak{q}}(\widehat{D}_{jt}^{-1}(b - \widehat{c}_{jt}))$. The block entering the stacked basis is always the block of the cell containing the argument b. When $\mathfrak{q} = 0$, this reduces to the cell indicator. When $\mathfrak{q} \geq 1$, the monomials are therefore formed from the normalized coordinates of b within its own rectangle. For the cell containing the origin, the split construction sets the lower endpoint equal to zero in every coordinate, both for the reference and estimated partitions. Hence $D_{jt}^{-1}(0 - c_{jt}) = \widehat{D}_{jt}^{-1}(0 - \widehat{c}_{jt}) = (-1/2, \ldots, -1/2)^\top$ on that cell, and therefore $\widehat{p}_t(0) = p_t(0)$ for every $\mathfrak{q} \geq 0$. This is the specialization of the affine scaling/rotation convention from Part I to the axis-aligned quantile rectangles used here. On the event where all side lengths are of order J^{-1}, the normalized basis is uniformly bounded and has derivative of order J in the original beta coordinates. The rest of this section collects the verification results for this random partitioning scheme and basis function.

Under Assumption SA-SEP, $\mathcal{J}_t$ contains precisely the estimated cells intersecting $\mathcal{B}^\circ$. Define the corresponding interior cell counts

$$\widehat{n}^\circ_{\min} = \min_{H+1\leq t\leq T} \min_{j\in\mathcal{J}_t} \widehat{n}_{jt}, \qquad \widehat{n}^\circ_{\max} = \max_{H+1\leq t\leq T} \max_{j\in\mathcal{J}_t} \widehat{n}_{jt}.$$

SA-II.3.1 Results

We present three results that establish the main properties of the quantile-based portfolio construction.

The first result establishes the density regularity of the generated regressor and translates it into density and spacing properties of the split reference quantile portfolios.

Lemma SA-II.3.1 (Generated-Regressor Density and Reference Quantile Spacing)**.** *Suppose that Assumptions SA-DGP, SA-FS, SA-PA, SA-SEP, and SA-FSB hold. Assume further that* $T^{(3+d)/q}/H + \sqrt{\log T/H} = o(1)$*. Then, with probability approaching one, the following statements hold:*

(i) *The conditional density* $f_{\widehat{\beta}_{it}|\mathcal{D}_{t-1}}(\cdot|\mathcal{D}_{t-1})$ *is uniformly bounded and bounded away from zero on* $\mathcal{B}$*, uniformly over* $i \in [n_t]$ *and* $t = H+1, \dots, T$*.*

(ii) *The marginal split-mixture densities* $\bar{f}_{\widehat{\beta},t,l,-}$ *and* $\bar{f}_{\widehat{\beta},t,l,+}$ *are uniformly bounded and bounded away from zero on* $[\underline{\beta}_l, 0)$ *and* $[0, \bar{\beta}_l]$*, respectively, uniformly over* $l \in [d]$ *and* $t = H+1, \dots, T$*.*

Consequently, the finite reference cells paired with cells in $\mathcal{J}_t$ *satisfy*

$$J^{-1} \lesssim_{\mathbb{P}} \min_{H+1\le t\le T} \min_{j\in\mathcal{J}_t^\circ} \min_{l\in[d]} (b_{jtl} - a_{jtl}) \le \max_{H+1\le t\le T} \max_{j\in\mathcal{J}_t^\circ} \max_{l\in[d]} (b_{jtl} - a_{jtl}) \lesssim_{\mathbb{P}} J^{-1},$$

where $\mathcal{J}_t^\circ = \{j \in \mathcal{J}_t : P_{jt} \text{ is bounded}\}$*.*

The next result compares the empirical breakpoints in $\widehat{\Delta}_t$ with their conditional reference counterparts in Δ_t. The first two bounds control levels of the split quantiles, while the last two control adjacent spacings, which are needed when comparing cell boundaries.

Lemma SA-II.3.2 (Quantile Convergence)**.** *Suppose that Assumptions SA-DGP, SA-FS, SA-PA, SA-SEP, and SA-FSB hold. Assume further that* $T^{(3+d)/q}/H + \sqrt{\log T/H} = o(1)$ *and* $\log T/n = o(1)$*. Then, uniformly over the empirical cells intersecting* $\mathcal{B}^\circ$ *and their rank-paired reference cells,*

$$\max_{H+1\le t\le T} \max_{j\in\mathcal{J}_t} \max_{l\in[d]} \left(|\widehat{a}_{jtl} - a_{jtl}| + |\widehat{b}_{jtl} - b_{jtl}| \right) \lesssim_{\mathbb{P}} \sqrt{\frac{\log T}{n}},$$
$$\max_{H+1\le t\le T} \max_{j\in\mathcal{J}_t} \max_{l\in[d]} \left| (\widehat{b}_{jtl} - \widehat{a}_{jtl}) - (b_{jtl} - a_{jtl}) \right| \lesssim_{\mathbb{P}} \sqrt{\frac{\log T}{n}}.$$

If, in addition, $J^2 \log T/n = o(1)$*, then the estimated cell widths satisfy*

$$J^{-1} \lesssim_{\mathbb{P}} \min_{H+1\le t\le T} \min_{j\in\mathcal{J}_t} \min_{l\in[d]} (\widehat{b}_{jtl} - \widehat{a}_{jtl}) \le \max_{H+1\le t\le T} \max_{j\in\mathcal{J}_t} \max_{l\in[d]} (\widehat{b}_{jtl} - \widehat{a}_{jtl}) \lesssim_{\mathbb{P}} J^{-1}.$$

With the quantile approximation in hand, we can verify the portfolio geometry, portfolio counts, and Gram behavior used by the subsequent results.

Lemma SA-II.3.3 (Generated Partition and Basis Moments)**.** *Suppose that Assumptions SA-DGP, SA-FS, SA-PA, SA-SEP, and SA-FSB hold. Assume further that*

$$\frac{T^{(3+d)/q}}{H} + \sqrt{\frac{\log T}{H}} = o(1), \qquad \frac{J^d \log(nT)}{n} = o(1).$$

Then the following results hold.

1. *Generated partition coverage and geometry:* $\mathcal{B}^\circ \cup \{\widehat{\beta}_{it} : i \in [n_t]\}$ *is contained in* $\cup_{j\in[\bar{J}_t]} \widehat{P}_{jt}$ *uniformly over* t *w.p.a.1, and* $\widehat{\mathsf{R}}_{\mathsf{D}}^\circ \lesssim_{\mathbb{P}} J^{-1}$*.*

2. *Cell counts:*

$$\frac{n}{J^d} \lesssim_{\mathbb{P}} \widehat{n}_{\min}^\circ \le \widehat{n}_{\max}^\circ \lesssim_{\mathbb{P}} \frac{n}{J^d}.$$

3. *The generated Gram blocks:*

$$J^{-d} \lesssim_{\mathbb{P}} \lambda_{\min}(\widehat{Q}_{jt}) \leq \lambda_{\max}(\widehat{Q}_{jt}) \lesssim_{\mathbb{P}} J^{-d}$$

uniformly over $t = H+1, \ldots, T$ *and interior cells* $j \in \mathcal{J}_t$.

4. *The reference Gram blocks:*

$$J^{-d} \lesssim_{\mathbb{P}} \lambda_{\min}(Q_{jt}) \leq \lambda_{\max}(Q_{jt}) \lesssim_{\mathbb{P}} J^{-d}$$

uniformly over $t = H+1, \ldots, T$ *and* $j \in \mathcal{J}_t$.

5. *The empirical true-exposure reference second moments:*

$$\lambda_{\max}\left(\frac{1}{n_t}\sum_{i=1}^{n_t} p_{jt}(\beta_{it})p_{jt}(\beta_{it})^\top\right) \lesssim_{\mathbb{P}} J^{-d}$$

uniformly over $t = H+1, \ldots, T$ *and* $j \in \mathcal{J}_t$.

6. *The comparison quantities from Lemma SA-I.3.1:*

$$\widehat{\mathsf{R}}_{\mathtt{Q}} \lesssim_{\mathbb{P}} \mathsf{R}_{\mathtt{Q}} = J^{-d}\sqrt{\frac{J^d \log(nT)}{n}} + J^{-(d-1)}\sqrt{\frac{\log T}{n}} + \frac{\log(nT)}{n},$$

and

$$\widehat{\mathsf{R}}_{\mathtt{P}} \lesssim_{\mathbb{P}} \mathsf{R}_{\mathtt{P}} = J^{-d}\min\left\{J\left(\mathsf{R}_{\mathtt{FS}} + \sqrt{\frac{\log T}{n}}\right), 1\right\} + \sqrt{\frac{J^{-d}\min\left\{J\left(\mathsf{R}_{\mathtt{FS}} + \sqrt{\frac{\log T}{n}}\right), 1\right\}\log(nT)}{n}}$$
$$+ \frac{\log(nT)}{n} + \mathbb{1}(\mathfrak{q} \geq 1)J^{-d}\min\left\{J^2\left(\mathsf{R}_{\mathtt{FS}} + \sqrt{\frac{\log T}{n}}\right)^2, 1\right\}.$$

SA-II.3.2 Discussion

(i) The rate $\mathsf{R}_{\mathtt{Q}}$ controls the inverse-Gram perturbation. It compares the empirical generated Gram matrix with the conditional reference Gram matrix of $\widehat{\beta}_{it}$ evaluated on the reference partition. Its terms correspond to ordinary cross-sectional sampling variation within cells, movement of the empirical quantile endpoints, and empirical discretization. The condition $\mathsf{R}_{\mathtt{Q}} = o(J^{-d})$ is implied by

$$\frac{J^d \log(nT)}{n} = o(1) \qquad \text{and} \qquad \frac{J^2 \log T}{n} = o(1).$$

If these conditions hold, then the construction gives the measurability and coverage components of the partition-design condition (i), while Lemma SA-II.3.3 verifies the diameter, design, and Gram-convergence components in conditions (i) and (ii) of Lemma SA-I.3.1. The stochastic linearization for $\mathfrak{q} = 0$ can therefore be obtained through the comparison below.

(ii) The rate $\mathsf{R}_{\mathtt{P}}$ controls the basis-switching variance term $\mathscr{R}_2(\beta)$. It compares $\widehat{p}_t(\widehat{\beta}_{it})$ with $p_t(\beta_{it})$. The switch set is therefore a boundary strip between paired estimated and reference cells, with width

determined by both the first-step displacement $\widehat{\beta}_{it}-\beta_{it}$ and the endpoint quantile error. The condition $\mathsf{R}_{\mathtt{P}}=o(J^{-d})$ is implied by the same two displays above and by $J\mathsf{R}_{\mathtt{FS}}=o(1)$.

(iii) For $\mathfrak{q}\geq 1$, $\mathsf{R}_{\mathtt{P}}$ also contains a nonswitch polynomial-coordinate term. On observations that remain in paired cells, the indicators agree but the normalized coordinates differ because the estimated and reference cell endpoints are not identical and because $\widehat{\beta}_{it}$ differs from β_{it}. Bounded derivatives of the normalized basis convert the combined displacement into the final term in $\mathsf{R}_{\mathtt{P}}$.

SA-II.3.3 Proof of Lemma SA-II.3.1

Because $3+d>2$, the condition $T^{(3+d)/q}/H+\sqrt{\log T/H}=o(1)$ implies the rate condition in Lemma SA-II.2.2. Hence, on an event with probability approaching one, $\widehat{A}(t)$ is invertible and

$$1\lesssim\lambda_{\min}\left(\frac{1}{H}\sum_{l=1}^{H}c_{itl}c_{itl}^{\top}\right)\leq\lambda_{\max}\left(\frac{1}{H}\sum_{l=1}^{H}c_{itl}c_{itl}^{\top}\right)\lesssim 1,$$

uniformly over $i\in[n_t]$ and $t=H+1,\ldots,T$, where

$$c_{itl}=S\widehat{A}(t)^{-1}X_l(t)K\Big(\frac{l}{H}\Big)\varsigma_{i(t-l)}.$$

It remains to verify the higher-order coefficient-moment bound used below. The uniform inverse bound for $\widehat{A}(t)$ and the boundedness of the kernel and $\varsigma_{i(t-l)}$ give

$$\max_{i\in[n_t]}\frac{1}{H}\sum_{l=1}^{H}\|c_{itl}\|^{3+d}\lesssim_{\mathbb{P}}\frac{1}{H}\sum_{l=1}^{H}\|X_l(t)\|^{3+d}\lesssim 1+\frac{1}{H}\sum_{l=1}^{H}\|f_{t-l}\|^{3+d}.$$

The process $\{\|f_t\|^{3+d}\}$ has a finite moment of order $q/(3+d)>2$ and inherits the geometric near-epoch dependence in Assumption SA-DGP(iv). Applying the same maximal Fuk–Nagaev argument as in Lemma SA-II.2.2, now with moment exponent $q/(3+d)$, yields

$$\max_{H+1\leq t\leq T}\left|\frac{1}{H}\sum_{l=1}^{H}\left(\|f_{t-l}\|^{3+d}-\mathbb{E}[\|f_{t-l}\|^{3+d}]\right)\right|\lesssim_{\mathbb{P}}\frac{T^{(3+d)/q}}{H}+\sqrt{\frac{\log T}{H}}=o_{\mathbb{P}}(1).$$

Since $\sup_t\mathbb{E}[\|f_t\|^{3+d}]<\infty$, intersecting these events gives an event $\mathcal{A}$ with probability approaching one on which

$$\max_{H+1\leq t\leq T}\max_{i\in[n_t]}\frac{1}{H}\sum_{l=1}^{H}\|c_{itl}\|^{3+d}\lesssim 1$$

and the preceding inverse and eigenvalue bounds all hold. On this event, let

$$\mathcal{H}_{it}^{\circ}=\sigma\Big((\alpha_{i(t-s)},\beta_{i(t-s)}^{\top},\varsigma_{i(t-s)}):s\in[H]\Big)\vee\mathcal{D}_{t-1}.$$

Conditional on $\mathcal{H}^\circ_{it}$, the only remaining randomness in the first-step variance component comes from the i.i.d. innovations $\{\eta_{i(t-s)} : s \in [H]\}$. Let $f_{\widehat{\beta}_{it}|\mathcal{H}^\circ_{it}}(\cdot|\mathcal{H}^\circ_{it})$ be the corresponding conditional density. Then

$$f_{\widehat{\beta}_{it}|\mathcal{D}_{t-1}}(z|\mathcal{D}_{t-1}) = \mathbb{E}\Big[f_{\widehat{\beta}_{it}|\mathcal{H}^\circ_{it}}(z|\mathcal{H}^\circ_{it})\big|\mathcal{D}_{t-1}\Big] = H^{d/2}\mathbb{E}\Big[f_{G_{it}}\big(\sqrt{H}(z-\breve{\beta}_{it})\big)\big|\mathcal{D}_{t-1}\Big],$$

where $G_{it} = H^{-1/2}\sum_{l=1}^{H} c_{itl}\eta_{i(t-l)}$ and $\breve{\beta}_{it}$ is defined in (SA-II.2.2). Given $\mathcal{H}^\circ_{it}$, the coefficients c_{itl} and the signal $\breve{\beta}_{it}$ are fixed, and the independence of the innovations in Assumption SA-DGP(ii) gives the displayed convolution representation. We record the needed properties of $f_{G_{it}}$ directly. Let φ_η denote the characteristic function of η_{it}. Conditional on $\mathcal{H}^\circ_{it}$,

$$\varphi_{G_{it}}(u) = \prod_{l=1}^{H}\varphi_\eta\left(\frac{c_{itl}^\top u}{\sqrt{H}}\right).$$

The condition $g_\eta^{(1)} \in L^1(\mathbb{R})$ implies polynomial decay of φ_η away from the origin, while the moment condition on η_{it} gives the usual expansion $\varphi_\eta(v) = 1 - v^2/2 + O(|v|^{2+\delta})$ for some $\delta > 0$ as $v \to 0$. Together with the eigenvalue bounds on $H^{-1}\sum_l c_{itl}c_{itl}^\top$ and the moment average bound on the c_{itl}'s, the standard three-region argument for products of characteristic functions applies: near the origin the product has Gaussian decay, on intermediate shells a fixed fraction of the projected arguments is bounded away from zero, and in the tails the decay from $g_\eta^{(1)} \in L^1(\mathbb{R})$ is applied to enough projected factors. Hence

$$\int_{\mathbb{R}^d}(1+\|u\|)|\varphi_{G_{it}}(u)|\,\mathrm{d}u \lesssim 1.$$

Fourier inversion therefore gives a continuously differentiable density

$$f_{G_{it}}(x) = (2\pi)^{-d}\int_{\mathbb{R}^d}\exp(-\mathrm{i}u^\top x)\varphi_{G_{it}}(u)\,\mathrm{d}u$$

with $\sup_x\{f_{G_{it}}(x) + \|\nabla f_{G_{it}}(x)\|\} \lesssim 1$ uniformly.

The same Fourier calculation also gives a uniform lower bound near the origin. Indeed, for fixed r and $\|u\| \le r$, the expansion of φ_η gives

$$\varphi_{G_{it}}(u) = \exp\left\{-\frac{1}{2H}\sum_{l=1}^{H}(c_{itl}^\top u)^2\right\} + o(1),$$

uniformly on $\|u\| \le r$. The matrix inside the quadratic form has eigenvalues bounded and bounded away from zero on $\mathcal{A}$, so the real part of the Fourier integral over a sufficiently large fixed ball is close to the corresponding nondegenerate Gaussian integral. The tail outside that ball can be made uniformly small by the preceding integrability bound. Hence $f_{G_{it}}(0) \gtrsim 1$ uniformly, and boundedness of $\nabla f_{G_{it}}$ implies that $f_{G_{it}}(x) \gtrsim 1$ on a fixed neighborhood of zero. The same Fourier argument, with integration by parts, also gives a polynomial tail bound $f_{G_{it}}(x) \lesssim (1+\|x\|)^{-d-1}$ uniformly. Let

$$\mathcal{A}_{it,\mathsf{FS}} = \Big\{\|\breve{\beta}_{it} - \beta_{it}\| \le \mathsf{B}_{\mathsf{FS}}\Big\}.$$

By Assumption SA-FSB and the inclusion $\mathcal{A}^c_{it,\mathsf{FS}} \subseteq \{\widehat{\mathsf{B}}_{\mathsf{FS}} > \mathsf{B}_{\mathsf{FS}}\}$,

$$\sqrt{H}\,\mathsf{B}_{\mathsf{FS}} = o(1), \qquad H^{d/2} \max_{H+1\le t\le T} \max_{i\in[n_t]} \mathbb{P}\big(\mathcal{A}^c_{it,\mathsf{FS}} \,\big|\, \mathcal{D}_{t-1}\big) \lesssim_{\mathbb{P}} \sqrt{H}\,\mathsf{B}_{\mathsf{FS}} + H^{-1/2} = o_{\mathbb{P}}(1).$$

We split the conditional convolution into its contributions on $\mathcal{A}_{it,\mathsf{FS}}$ and $\mathcal{A}^c_{it,\mathsf{FS}}$. Since $\sup_x f_{G_{it}}(x) \lesssim 1$, the latter contribution satisfies, uniformly over i, t, and $z \in \mathcal{B}$,

$$H^{d/2}\mathbb{E}\Big[f_{G_{it}}\Big(\sqrt{H}(z-\breve{\beta}_{it})\Big)\,\mathbb{1}(\mathcal{A}^c_{it,\mathsf{FS}})\,\Big|\,\mathcal{D}_{t-1}\Big] \lesssim H^{d/2}\mathbb{P}\big(\mathcal{A}^c_{it,\mathsf{FS}} \,\big|\, \mathcal{D}_{t-1}\big) = o_{\mathbb{P}}(1). \tag{SA-II.3.1}$$

On $\mathcal{A}_{it,\mathsf{FS}}$, $\|\breve{\beta}_{it} - \beta_{it}\| \le \mathsf{B}_{\mathsf{FS}}$. Hence, for every $r > 0$,

$$\mathbb{P}\Big(\|\breve{\beta}_{it} - z\| \le r,\, \mathcal{A}_{it,\mathsf{FS}} \,\Big|\, \mathcal{D}_{t-1}\Big) \le \mathbb{P}(\|\beta_{it} - z\| \le r + \mathsf{B}_{\mathsf{FS}} \,|\, \mathcal{D}_{t-1}) \lesssim \big(r + \mathsf{B}_{\mathsf{FS}}\big)^d,$$

where the last inequality follows from the uniform upper bound on $f_{\beta_{it}}(\cdot \mid \mathcal{D}_{t-1})$. Decomposing $\mathcal{B}$ into annuli of width $H^{-1/2}$ around z, using $\sqrt{H}\,\mathsf{B}_{\mathsf{FS}} = o(1)$ and the polynomial tail of $f_{G_{it}}$, shows that the contribution on $\mathcal{A}_{it,\mathsf{FS}}$ is uniformly bounded. Together with (SA-II.3.1), this yields

$$\sup_{z\in\mathcal{B}} f_{\widehat{\beta}_{it}|\mathcal{D}_{t-1}}(z \mid \mathcal{D}_{t-1}) \lesssim_{\mathbb{P}} 1.$$

For the lower bound, fix $r_0 > 0$ small enough that $f_{G_{it}}(x)$ is uniformly bounded away from zero on $\{x : \|x\| \le 2r_0\}$. Since $\sqrt{H}\,\mathsf{B}_{\mathsf{FS}} = o(1)$, uniformly in $z \in \mathcal{B}$,

$$\begin{aligned} f_{\widehat{\beta}_{it}|\mathcal{D}_{t-1}}(z \mid \mathcal{D}_{t-1}) &\gtrsim H^{d/2}\mathbb{P}\Big(\|\beta_{it} - z\| \le r_0 H^{-1/2},\, \mathcal{A}_{it,\mathsf{FS}} \,\Big|\, \mathcal{D}_{t-1}\Big) \\ &\gtrsim H^{d/2}\left[\mathbb{P}\Big(\|\beta_{it} - z\| \le r_0 H^{-1/2} \,\Big|\, \mathcal{D}_{t-1}\Big) - \mathbb{P}\big(\mathcal{A}^c_{it,\mathsf{FS}} \,\big|\, \mathcal{D}_{t-1}\big)\right]. \end{aligned}$$

Because $\mathcal{B}$ is a rectangle, its intersection with a ball of radius $r_0 H^{-1/2}$ centered at any $z \in \mathcal{B}$ contains a cone whose volume is bounded below by a constant multiple of $H^{-d/2}$. The uniform lower bound on $f_{\beta_{it}}(\cdot \mid \mathcal{D}_{t-1})$ and Assumption SA-FSB therefore imply

$$\inf_{z\in\mathcal{B}} f_{\widehat{\beta}_{it}|\mathcal{D}_{t-1}}(z \mid \mathcal{D}_{t-1}) \gtrsim_{\mathbb{P}} 1.$$

This proves the full-vector density statement without requiring a density for the signal $\breve{\beta}_{it}$ itself.

The marginal density statement follows by applying the preceding argument to each coordinate. The eigenvalue bounds for $H^{-1}\sum_{s=1}^{H} c_{its}c_{its}^\top$ bound each coordinate variance above and away from zero, and the exceptional-event bound above is stronger than needed for one coordinate. For the split mixtures,

$$\begin{aligned} \bar{f}_{\widehat{\beta},t,l,-}(z) &= \frac{1}{n_t \bar{\pi}_{t,l,-}} \sum_{i=1}^{n_t} f_{\widehat{\beta}_{it,l}|\mathcal{D}_{t-1}}(z|\mathcal{D}_{t-1})\mathbb{1}(z<0), \\ \bar{f}_{\widehat{\beta},t,l,+}(z) &= \frac{1}{n_t \bar{\pi}_{t,l,+}} \sum_{i=1}^{n_t} f_{\widehat{\beta}_{it,l}|\mathcal{D}_{t-1}}(z|\mathcal{D}_{t-1})\mathbb{1}(z\ge 0). \end{aligned}$$

Because 0 is an interior point of $\mathcal{B}$ and the marginal densities are bounded and bounded away from zero on $[\underline{\beta}_l, \bar{\beta}_l]$, the denominators are bounded away from zero w.p.a.1. Thus the split densities are bounded and bounded away from zero on $[\underline{\beta}_l, 0)$ and $[0, \bar{\beta}_l]$, respectively. The separation of $\mathcal{B}^\circ$ from $\partial\mathcal{B}$ and $J^-_{t,l}, J^+_{t,l} \asymp J$

imply that every finite reference endpoint paired with a cell in $\mathcal{J}_t$ lies in this density-regular region w.p.a.1. The displayed interior spacing bound therefore follows from the mean value theorem. □

SA-II.3.4 Proof of Lemma SA-II.3.2

For $a \in \{-,+\}$, write $\mathcal{S}_- = (-\infty, 0)$ and $\mathcal{S}_+ = [0, \infty)$. Define

$$\widehat{A}_{tla}(z) = \frac{1}{n_t}\sum_{i=1}^{n_t} \mathbb{1}(\widehat{\beta}_{it,l} \in \mathcal{S}_a)\mathbb{1}(\widehat{\beta}_{it,l} \le z),$$
$$\bar{A}_{tla}(z) = \frac{1}{n_t}\sum_{i=1}^{n_t} \mathbb{P}(\widehat{\beta}_{it,l} \in \mathcal{S}_a, \widehat{\beta}_{it,l} \le z|\mathcal{D}_{t-1}),$$

and let $\widehat{\pi}_{tla} = n_t^{-1}\sum_{i=1}^{n_t} \mathbb{1}(\widehat{\beta}_{it,l} \in \mathcal{S}_a)$ and $\bar{\pi}_{tla} = n_t^{-1}\sum_{i=1}^{n_t} \mathbb{P}(\widehat{\beta}_{it,l} \in \mathcal{S}_a|\mathcal{D}_{t-1})$. Then $\widehat{F}_{\widehat{\beta},t,l,a} = \widehat{A}_{tla}/\widehat{\pi}_{tla}$ and $\bar{F}_{\widehat{\beta},t,l,a} = \bar{A}_{tla}/\bar{\pi}_{tla}$ on the corresponding side of zero.

Conditional on $\mathcal{D}_{t-1}$, $\widehat{\beta}_{it}$ is independent over i. The classes of half-lines intersected with $\mathcal{S}_-$ and $\mathcal{S}_+$ have finite VC index and envelope one. Therefore the i.n.i.d. maximal inequality in Cattaneo et al. (2024, Lemma SA24), combined with conditional bounded-difference concentration and followed by a union bound over t, coordinates, and signs, gives

$$\max_{H+1\le t\le T}\max_{l\in[d]}\max_{a\in\{-,+\}}\sup_{z\in\mathcal{S}_a}|\widehat{A}_{tla}(z) - \bar{A}_{tla}(z)| \lesssim_{\mathbb{P}} \sqrt{\frac{\log T}{n}},$$
$$\max_{H+1\le t\le T}\max_{l\in[d]}\max_{a\in\{-,+\}}|\widehat{\pi}_{tla} - \bar{\pi}_{tla}| \lesssim_{\mathbb{P}} \sqrt{\frac{\log T}{n}}.$$

Lemma SA-II.3.1 and the fact that 0 lies in the interior of $\mathcal{B}$ imply $\bar{\pi}_{tla} \gtrsim_{\mathbb{P}} 1$ uniformly. Hence $\widehat{\pi}_{tla} \gtrsim_{\mathbb{P}} 1$ uniformly and the ratio map gives

$$\max_{H+1\le t\le T}\max_{l\in[d]}\sup_{z<0}|\widehat{F}_{\widehat{\beta},t,l,-}(z) - \bar{F}_{\widehat{\beta},t,l,-}(z)| \lesssim_{\mathbb{P}} \sqrt{\frac{\log T}{n}},$$
$$\max_{H+1\le t\le T}\max_{l\in[d]}\sup_{z\ge 0}|\widehat{F}_{\widehat{\beta},t,l,+}(z) - \bar{F}_{\widehat{\beta},t,l,+}(z)| \lesssim_{\mathbb{P}} \sqrt{\frac{\log T}{n}}.$$

Since $\log T/n = o(1)$, the uniform split-CDF bounds above, Assumption SA-SEP, and $J \to \infty$ imply that the first negative and last positive empirical intervals do not intersect $\mathcal{B}^\circ$ in any coordinate w.p.a.1. Hence all rank-paired reference cells indexed by $\mathcal{J}_t$ have finite endpoints on this event. For the probability indices defining cells in $\mathcal{J}_t$, the associated reference quantiles lie in the density-regular region identified in Lemma SA-II.3.1. The generalized inverse relation and the empirical quantile overshoot therefore give the first displayed bound in the lemma. The side-length bound follows from

$$\left|(\widehat{b}_{jtl} - \widehat{a}_{jtl}) - (b_{jtl} - a_{jtl})\right| \le |\widehat{b}_{jtl} - b_{jtl}| + |\widehat{a}_{jtl} - a_{jtl}|,$$

uniformly over t, $j \in \mathcal{J}_t$, and $l \in [d]$. Under $J^2 \log T/n = o(1)$, the right-hand side is $o_{\mathbb{P}}(J^{-1})$. Combining this result with the reference width bounds in Lemma SA-II.3.1 gives the asserted estimated width bounds. □

SA-II.3.5 Proof of Lemma SA-II.3.3

Let $\widehat{P}_{jt} = [\widehat{\ell}_{jt}, \widehat{u}_{jt}]$ denote a generic rectangle in $\widehat{\Delta}_t$ and $P_{jt} = [\ell_{jt}, u_{jt}]$ the corresponding rectangle in the reference partition Δ_t. Let $\widehat{\psi}_{jt}(b) = \widehat{p}_{jt}(b)\widehat{p}_{jt}(b)^\top$ and $\psi_{jt}(b) = p_{jt}(b)p_{jt}(b)^\top$. The condition $J^d \log(nT)/n = o(1)$ implies $\log T/n = o(1)$, so the argument in the proof of Lemma SA-II.3.2 ensures that all reference cells paired with $j \in \mathcal{J}_t$ have finite endpoints w.p.a.1. Lemma SA-II.3.1 gives reference side lengths of order J^{-1}. For cells indexed by $j \in \mathcal{J}_t$, the estimated side lengths satisfy the same order. To see this, apply the localized VC/Bernstein inequality to split intervals on each side of zero. Since the split densities are bounded and bounded away from zero, an interval with side length larger than a sufficiently large multiple of J^{-1} has conditional probability larger than a sufficiently large multiple of J^{-1}, while an interval with side length smaller than a sufficiently small multiple of J^{-1} has conditional probability smaller than a sufficiently small multiple of J^{-1}. The empirical split quantile intervals have empirical mass of order J^{-1} on their corresponding side, up to a discretization error of order n^{-1}, and $J^d \log(nT)/n = o(1)$ implies the localized stochastic error is $o_{\mathbb{P}}(J^{-1})$ uniformly. Hence the estimated and reference cells indexed by $\mathcal{J}_t$ belong w.p.a.1 to a uniformly quasi-uniform class of rectangles. Moreover, the first and last empirical order statistics lie outside $\mathcal{B}^\circ$ coordinatewise w.p.a.1: Assumption SA-SEP places at least one complete quantile cell between $\mathcal{B}^\circ$ and each support boundary, and every sample extremum lies weakly beyond the corresponding first or last interior empirical quantile. Thus the empirical partition covers $\mathcal{B}^\circ$ and, by construction, every realized $\widehat{\beta}_{it}$. Assumption SA-SEP implies that every cell indexed by $j \in \mathcal{J}_t$ is separated from the outer boundary of $\mathcal{B}$ by more than the first-step smoothing scale, uniformly over t. The side-length bounds give

$$\widehat{\mathsf{R}}^\circ_{\mathtt{D}} \lesssim_{\mathbb{P}} J^{-1}.$$

The classes generated by the entries of $\widehat{\psi}_{jt}$ and ψ_{jt} are polynomial transformations of rectangle indicators. Hence they are VC-type with bounded envelopes and conditional variances of order J^{-d}. All empirical-process bounds below are first taken uniformly over deterministic quasi-uniform rectangles equipped with the midpoint and side-length normalization described above. The empirical quantile cells are then substituted into those uniform bounds.

The same maximal inequality applied only to the cell indicators indexed by $\mathcal{J}_t$ gives

$$\frac{n}{J^d} \lesssim_{\mathbb{P}} \widehat{n}^\circ_{\min} \leq \widehat{n}^\circ_{\max} \lesssim_{\mathbb{P}} \frac{n}{J^d}.$$

For a deterministic admissible rectangle P with its midpoint and side-length normalization, write $\psi_P(b) = p_{\mathfrak{q}}(b, P)p_{\mathfrak{q}}(b, P)^\top$. Conditional on $\mathcal{D}_{t-1}$, the maximal inequality gives

$$\max_{H+1\leq t\leq T} \sup_P \left\| \frac{1}{n_t} \sum_{i=1}^{n_t} \left(\psi_P(\widehat{\beta}_{it}) - \mathbb{E}[\psi_P(\widehat{\beta}_{it})|\mathcal{D}_{t-1}] \right) \right\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(nT)}{nJ^d}}.$$

The supremum is over the deterministic quasi-uniform rectangle class containing the realized estimated cells w.p.a.1, so evaluating the display at $P = \widehat{P}_{jt}$ controls the empirical part of $\widehat{Q}_{jt}$. The conditional density of $\widehat{\beta}_{it}$ is bounded and bounded away from zero on interior cells by Lemma SA-II.3.1. After the affine change of variables to cell-normalized coordinates, the corresponding population generated Gram block has eigenvalues of order J^{-d} because the monomials in $\check{p}_{\mathfrak{q}}$ are linearly independent on a fixed normalized rectangle. Since

$J^d \log(nT)/n = o(1)$, Weyl's inequality gives

$$J^{-d} \lesssim_{\mathbb{P}} \lambda_{\min}(\widehat{Q}_{jt}) \leq \lambda_{\max}(\widehat{Q}_{jt}) \lesssim_{\mathbb{P}} J^{-d}$$

uniformly over $t = H+1, \ldots, T$ and $j \in \mathcal{J}_t$. Multiplying by $n_t/\widehat{n}_{jt} \asymp_{\mathbb{P}} J^d$ gives the scaled Gram bound

$$1 \lesssim_{\mathbb{P}} \frac{n_t}{\widehat{n}_{jt}} \lambda_{\min}(\widehat{Q}_{jt}) \leq \frac{n_t}{\widehat{n}_{jt}} \lambda_{\max}(\widehat{Q}_{jt}) \lesssim_{\mathbb{P}} 1.$$

For the reference Gram blocks, recall that

$$Q_{jt} = \frac{1}{n_t} \sum_{i=1}^{n_t} \mathbb{E}[p_{jt}(\widehat{\beta}_{it}) p_{jt}(\widehat{\beta}_{it})^\top | \mathcal{D}_{t-1}].$$

After the change of variables $b = c_{jt} + D_{jt} z$, the normalized support for z is a rectangle with nonempty interior. The reference side-length bounds from Lemma SA-II.3.1, together with the conditional density bounds for $\widehat{\beta}_{it}$ in the same lemma, imply that the Jacobian is of order J^{-d} and that the weight attached to the monomial Gram matrix is bounded and bounded away from zero. Since the monomials up to degree $\mathfrak{q}$ are linearly independent on this rectangle,

$$J^{-d} \lesssim_{\mathbb{P}} \lambda_{\min}(Q_{jt}) \leq \lambda_{\max}(Q_{jt}) \lesssim_{\mathbb{P}} J^{-d}$$

uniformly over $t = H+1, \ldots, T$ and $j \in \mathcal{J}_t$.

For the empirical true-exposure reference second moments, apply the same VC-type maximal inequality to the class $\psi_{jt}(\beta_{it})$. Conditional on $\mathcal{D}_{t-1}$, Assumption SA-DGP(iii) gives independent unit histories and conditional densities for β_{it} that are bounded on $\mathcal{B}$. Since the reference cells have side lengths of order J^{-1}, the population block moments are of order J^{-d}. The empirical deviation is of order $\sqrt{\log(nT)/(nJ^d)}$, which is $o_{\mathbb{P}}(J^{-d})$ under $J^d \log(nT)/n = o(1)$. Hence

$$\lambda_{\max}\left(\frac{1}{n_t} \sum_{i=1}^{n_t} p_{jt}(\beta_{it}) p_{jt}(\beta_{it})^\top \right) \lesssim_{\mathbb{P}} J^{-d}$$

uniformly over $t = H+1, \ldots, T$ and $j \in \mathcal{J}_t$.

It remains to bound $\widehat{\mathsf{R}}_{\mathfrak{q}}$. Add and subtract the population counterpart of the estimated rectangle, obtained by evaluating the deterministic-class map $P \mapsto n_t^{-1} \sum_i \mathbb{E}[\psi_P(\widehat{\beta}_{it}) | \mathcal{D}_{t-1}]$ at $P = \widehat{P}_{jt}$. The resulting empirical-process term is already bounded above by $\sqrt{\log(nT)/(nJ^d)}$. For the population-comparison term, compare the estimated and reference rectangles while holding the generated regressor fixed. The endpoint quantile bounds in Lemma SA-II.3.2 imply that the symmetric difference between paired rectangles has conditional mass $J^{-(d-1)}\sqrt{\log T/n}$, up to the empirical discretization term $\log(nT)/n$. When $\mathfrak{q} \geq 1$, the nonswitch change in normalized polynomial coordinates is of the same order: bounded derivatives of $\check{p}_{\mathfrak{q}}$ give a pointwise perturbation of order $J\sqrt{\log T/n}$ on a cell of mass J^{-d}. Combining these bounds gives $\widehat{\mathsf{R}}_{\mathfrak{q}} \lesssim_{\mathbb{P}} \mathsf{R}_{\mathfrak{q}}$.

For $\widehat{\mathsf{R}}_{\mathfrak{p}}$, work on the high-probability event from Lemma SA-II.2.3 and Lemma SA-II.3.2 on which $\max_{i,t} \|\widehat{\beta}_{it} - \beta_{it}\| \lesssim \mathsf{R}_{\mathfrak{FS}}$ and all paired endpoints differ by at most a constant multiple of $\sqrt{\log T/n}$. Decompose the full basis discrepancy into the generated-regressor displacement and the endpoint displacement:

$$\widehat{p}_{jt}(\widehat{\beta}_{it}) - p_{jt}(\beta_{it}) = \left\{ \widehat{p}_{jt}(\widehat{\beta}_{it}) - p_{jt}(\widehat{\beta}_{it}) \right\} + \left\{ p_{jt}(\widehat{\beta}_{it}) - p_{jt}(\beta_{it}) \right\}.$$

Any observation for which the corresponding cell indicators differ in either component belongs to a neighborhood of ∂P_{jt} (the boundary of P_{jt}) with radius proportional to $\mathsf{R}_{\mathtt{FS}}+\sqrt{\log T/n}$. When $J(\mathsf{R}_{\mathtt{FS}}+\sqrt{\log T/n}) \leq 1$, the density bounds give the boundary-strip probability $J^{-d}J(\mathsf{R}_{\mathtt{FS}}+\sqrt{\log T/n})$. Otherwise, the squared basis difference is supported on the union of the paired cells, whose mass is $O_{\mathbb{P}}(J^{-d})$. A Bernstein inequality in the first case and the cell-mass bound in the second give the first three terms in $\mathsf{R}_{\mathtt{P}}$.

On the complement of the switch set, the paired cell indicators agree. Both basis blocks are then evaluated in corresponding cells, and only the normalized polynomial coordinates can differ. For $\mathfrak{q} \geq 1$, bounded derivatives of the normalized polynomial basis, the first-step error bound, and the endpoint quantile bounds in Lemma SA-II.3.2 imply

$$\max_{H+1\leq t\leq T}\max_{j\in\mathcal{J}_t}\max_{i\in[n_t]}\left\|\widehat{p}_{jt}(\widehat{\beta}_{it})-p_{jt}(\beta_{it})\right\|\mathbb{1}\left(\widehat{\beta}_{it}\in\widehat{P}_{jt},\ \beta_{it}\in P_{jt}\right)\lesssim_{\mathbb{P}}\min\left\{J\left(\mathsf{R}_{\mathtt{FS}}+\sqrt{\frac{\log T}{n}}\right),1\right\}.$$

The squared basis difference on the nonswitch set is therefore bounded by the corresponding squared displacement times the empirical mass of the paired cell, which is $\lesssim_{\mathbb{P}} J^{-d}$ uniformly. This gives the final term in $\mathsf{R}_{\mathtt{P}}$ when $\mathfrak{q} \geq 1$ and proves $\widehat{\mathsf{R}}_{\mathtt{P}} \lesssim_{\mathbb{P}} \mathsf{R}_{\mathtt{P}}$. □

SA-II.4 Stochastic Linearization

This section specializes the general stochastic linearization from Section SA-I.3 to the classical estimator defined in Sections SA-II.2 and SA-II.3. The first-step decomposition in (SA-II.2.2) separates the bias component $\breve{\beta}_{it}-\beta_{it}$ from the first-step variance component, which determines the generated-regressor term $\mathscr{L}(\beta)$.

SA-II.4.1 Decomposition

Using the first-step decomposition in (SA-II.2.2), decompose

$$\mathscr{L}(\beta)=\mathscr{L}_{\mathtt{B}}(\beta)+\mathscr{L}_{\mathtt{V}}(\beta),$$

where

$$\mathscr{L}_{\mathtt{B}}(\beta)=\frac{1}{T-H}\sum_{t=H+1}^{T}\widehat{p}_t(\beta)^\top\widehat{Q}_t^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\widehat{p}_t(\widehat{\beta}_{it})[\nabla\mu_t(\beta_{it})]^\top\left(\breve{\beta}_{it}-\beta_{it}\right),$$
$$\mathscr{L}_{\mathtt{V}}(\beta)=\frac{1}{T-H}\sum_{t=H+1}^{T}\widehat{p}_t(\beta)^\top\widehat{Q}_t^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\widehat{p}_t(\widehat{\beta}_{it})[\nabla\mu_t(\beta_{it})]^\top\left(\widehat{\beta}_{it}-\breve{\beta}_{it}\right).$$

$\mathscr{L}_{\mathtt{B}}(\beta)$ is a bias term related to the bias of the first-step estimation, while $\mathscr{L}_{\mathtt{V}}(\beta)$ contains both a bias component (due to the correlation between $\widehat{p}_t(\widehat{\beta}_{it})$ and $\varepsilon_{i(t-s)}$'s) and a variance component. Let $A(t)=\mathbb{E}[\widehat{A}(t)]$. The reference leading term for $\mathscr{L}_{\mathtt{V}}(\beta)$ is

$$\mathscr{A}(\beta)=\frac{1}{T-H}\sum_{t=1}^{T}\sum_{i=1}^{n_t}w_{it}(\beta)\varepsilon_{it}$$

with

$$w_{it}(\beta) = \frac{1}{H} \sum_{s=\max\{1,H+1-t\}}^{\min\{H,T-t\}} \frac{1}{n_{t+s}} \widehat{p}_{t+s}(\beta)^\top Q_{t+s}^{-1} p_{t+s}(\beta_{i(t+s)}) [\nabla \mu_{t+s}(\beta_{i(t+s)})]^\top \\ \times K\Big(\frac{s}{H}\Big) SA(t+s)^{-1} X_s(t+s) \mathbb{1}(i \le n_{t+s}).$$

Replacing β_{it} by $\widehat{\beta}_{it}$ can move the normalized polynomial coordinates even if both points remain in the same cell when $\mathfrak{q} \ge 1$. The proof below isolates the centered part of this smooth-interior movement and controls it by its uniform bound. We do not keep this component as a separate leading term: it is absorbed into the linear-approximation remainder. The remaining difference between $\mathscr{L}_{\mathtt{V}}(\beta)$ and its reference approximation is driven by the conditional mean of the same-cell contribution, boundary switching, endpoint movement, and Gram perturbations.

SA-II.4.2 Results

The contribution of the first-step bias $\breve{\beta}_{it} - \beta_{it}$ is a direct consequence of Assumption SA-FSB. The preliminary lemma below controls both the conditional mean and conditional variance of the localized block-basis perturbation generated by the first-step variance term. For $\mathfrak{q} = 0$, this perturbation is exactly the cell-switching component. For $\mathfrak{q} \ge 1$, the same lemma also covers the smooth same-cell movement of the normalized polynomial coordinates.

Lemma SA-II.4.1 (Generated-Regressor Basis-Perturbation Bias and Variance)**.** *Suppose that Assumptions SA-DGP, SA-FS, SA-PA, and SA-FSB hold and that $s_\mu \ge 2$. Assume further that*

$$\frac{\max\{J^{2/(q-2)}T^{2/q}, T^{2(3+d)/q}\}}{H} = o(1).$$

Let $P = [\ell, u] \subset \mathcal{B}$ satisfy

$$\min_{k\in[d]} \min\{|\ell_k - \underline{\beta}_k|, |\bar{\beta}_k - u_k|\} \gg \frac{\log(nT)}{\sqrt{H}}, \qquad J^{-1} \lesssim \min_{k\in[d]}(u_k - \ell_k) \le \max_{k\in[d]}(u_k - \ell_k) \lesssim J^{-1}.$$

Then, the following results hold.

(a) *Conditional mean:*

$$\left\| \frac{1}{H} \sum_{s=1}^{H} \mathbb{E}\left[\Big(p_{\mathfrak{q}}(\widehat{\beta}_{it}, P) - p_{\mathfrak{q}}(\beta_{it}, P)\Big) [\nabla \mu_t(\beta_{it})]^\top S\widehat{A}(t)^{-1} X_s(t) K\Big(\frac{s}{H}\Big) \varsigma_{i(t-s)} \eta_{i(t-s)} \middle| \mathcal{D}_{t-1} \right] \right\| \\ \lesssim_{\mathbb{P}} \frac{1}{HJ^d} + \frac{\mathsf{B}_{\mathtt{FS}}}{J^d}.$$

(b) *Conditional variance:*

$$\left\| \mathbb{V}\mathrm{ar}\left[\frac{1}{H} \sum_{s=1}^{H} \Big(p_{\mathfrak{q}}(\widehat{\beta}_{it}, P) - p_{\mathfrak{q}}(\beta_{it}, P)\Big) [\nabla \mu_t(\beta_{it})]^\top S\widehat{A}(t)^{-1} X_s(t) K\Big(\frac{s}{H}\Big) \varsigma_{i(t-s)} \eta_{i(t-s)} \middle| \mathcal{D}_{t-1} \right] \right\| \\ \lesssim_{\mathbb{P}} \frac{\log^d(nT)}{HJ^{d-1}\max\{J, \sqrt{H}\}} + \Big(\frac{1}{HJ^d} + \frac{\mathsf{B}_{\mathtt{FS}}}{J^d}\Big)^2 + \mathbb{1}(\mathfrak{q} \ge 1) \frac{J^{2-d}}{H^2}.$$

These results hold uniformly over $i \in [n_t]$, $t = H+1,\dots,T$, and rectangles P satisfying the conditions imposed above.

The next lemma refines only the generated-regressor component of Lemma SA-I.3.1. The remaining terms in the stochastic decomposition are already controlled by the general result in Part I. The refinement extracts the leading term $\mathscr{A}(\beta)$ and absorbs the smooth-interior polynomial movement that arises when $\mathfrak{q} \geq 1$ into the linear-approximation remainder.

Lemma SA-II.4.2 (Classical Generated-Regressor Linearization)**.** *Suppose that Assumptions SA-DGP, SA-FS, SA-PA, SA-SEP, and SA-FSB hold and that $s_\mu \geq 2$. Assume further that*

$$\frac{J^d \log(nT)}{n} = o(1),$$
$$\frac{J^{d-1}\max\{J,\sqrt{H}\}\log^2(nT)}{\min\{n^{1-2/q}T^{1-4/q}h, n\}} \lesssim 1, \qquad \frac{\max\{J^{2/(q-2)}T^{2/q}, T^{2(3+d)/q}\}}{H} = o(1). \tag{SA-II.4.1}$$

Then, the bias term satisfies

$$\sup_{\beta\in\mathcal{B}^\circ} |\mathscr{L}_{\mathtt{B}}(\beta)| \lesssim_{\mathbb{P}} \mathsf{B}_{\mathtt{FS}},$$

and the variance terms satisfy

$$\sup_{\beta\in\mathcal{B}^\circ} |\mathscr{A}(\beta)| \lesssim_{\mathbb{P}} \sqrt{\frac{J^d\log(nT)}{nH}}, \qquad |\mathscr{A}(0)| \lesssim_{\mathbb{P}} \sqrt{\frac{J^d}{nT}},$$

and

$$\sup_{\beta\in\mathcal{B}^\circ} \left|\mathscr{L}_{\mathtt{V}}(\beta) - \mathscr{A}(\beta)\right| \lesssim_{\mathbb{P}} \mathsf{R}_{\mathtt{LA}}$$

with

$$\begin{aligned}\mathsf{R}_{\mathtt{LA}} = \frac{1}{H} + \mathsf{B}_{\mathtt{FS}} &+ \sqrt{\frac{J^{2d}\log^{d+1}(nT)}{nHJ^{d-1}\max\{J,\sqrt{H}\}}} + \sqrt{\frac{J^{d+1}\sqrt{\log T/n}\log(nT)}{nH}} + \min\{J^d\mathsf{R}_{\mathtt{Q}},1\}\sqrt{\frac{J^d\log(nT)}{nH}} \\ &+ \mathbb{1}(\mathfrak{q}\geq 1)\sqrt{\frac{J^{d+2}\log(nT)}{nH^2}} + \mathbb{1}(\mathfrak{q}\geq 1)\sqrt{\frac{J^{d+2}\log T\log(nT)}{n^2H}}.\end{aligned}$$

SA-II.4.3 Discussion

(i) The comparison replaces the generated design in $\mathscr{L}_{\mathtt{V}}(\beta)$ by the reference design. For $\mathfrak{q} = 0$, the within-cell basis is constant, so only boundary switching matters. For $\mathfrak{q} \geq 1$, movement inside a cell also changes the normalized polynomial coordinates. The proof isolates the centered product of this same-cell movement and the first-step shock multiplier, but the lemma absorbs it into $\mathsf{R}_{\mathtt{LA}}$ using its uniform bound. This conservative treatment avoids relying on additional time-averaging for the same-cell component at the origin.

SA-II.4.4 Proof of Lemma SA-II.4.1(a)

The proof is separated into several steps. We first prove the switching-bias bound, which is the case $\mathfrak{q} = 0$. The final step extends the same mean calculation to each component of the localized basis $p_{\mathfrak{q}}(\cdot, P)$.

Step 1: Definitions and notation. For any generic rectangle of interest $P = [\ell, u]$ satisfying the boundary restrictions in the lemma, we write $\sqrt{H}(P - c) = [\sqrt{H}(\ell - c), \sqrt{H}(u - c)]$ to be the scaled shifted rectangle. For any vector v, v_{-j} denotes the subvector excluding the j-th entry; for any rectangle $\mathcal{R} = [\ell, u]$ in $\mathbb{R}^d$, $\mathcal{R}_{-j} = [\ell_{-j}, u_{-j}]$ in $\mathbb{R}^{d-1}$. Also write $\mathcal{B}_j = [\underline{\beta}_j, \bar{\beta}_j]$ for the j-th coordinate projection of $\mathcal{B}$.

Recall that

$$\breve{\beta}_{it} = \frac{1}{H}\sum_{l=1}^{H} S\widehat{A}(t)^{-1} X_l(t) K\Big(\frac{l}{H}\Big)\big(\alpha_{i(t-l)} + \beta_{i(t-l)}^{\top} f_{t-l}\big)$$

with $\max_{i,t} \|\breve{\beta}_{it} - \beta_{it}\| \le \mathsf{B}_{\mathtt{FS}}$ w.p.a.1 by Assumption SA-FSB, and that $\mathcal{E}_{t-1}$ contains the predictable paths, factor history, and volatility scales used in the conditioning arguments, but excludes the i.i.d. innovations η_{is}.

For each i and t, define the local regular event

$$\mathcal{A}_{it,\mathtt{FS}} = \Big\{\|\breve{\beta}_{it} - \beta_{it}\| \le \mathsf{B}_{\mathtt{FS}}\Big\}.$$

Assumption SA-FSB and $\mathcal{A}_{it,\mathtt{FS}}^c \subseteq \{\widehat{\mathsf{B}}_{\mathtt{FS}} > \mathsf{B}_{\mathtt{FS}}\}$ give

$$H^{(d-1)/2} \max_{H+1\le t\le T} \max_{i\in[n_t]} \mathbb{P}\big(\mathcal{A}_{it,\mathtt{FS}}^c \,\big|\, \mathcal{D}_{t-1}\big) \lesssim_{\mathbb{P}} \mathsf{B}_{\mathtt{FS}} + H^{-1}. \tag{SA-II.4.2}$$

Because $\sqrt{H}(\mathsf{B}_{\mathtt{FS}} + H^{-1}) = o(1)$, this also gives

$$H^{d/2} \max_{H+1\le t\le T} \max_{i\in[n_t]} \mathbb{P}\big(\mathcal{A}_{it,\mathtt{FS}}^c \,\big|\, \mathcal{D}_{t-1}\big) \lesssim_{\mathbb{P}} \sqrt{H}\,\mathsf{B}_{\mathtt{FS}} + H^{-1/2} = o_{\mathbb{P}}(1). \tag{SA-II.4.3}$$

Every $\mathcal{D}_{t-1}$-conditional calculation below is split over $\mathcal{A}_{it,\mathtt{FS}}$ and its complement. On the regular event, the signal displacement is bounded by the deterministic $\mathsf{B}_{\mathtt{FS}}$. On the complement, the conditional convolution representations below, including their leave-one-out and leave-two-out versions, are bounded by the corresponding uniformly integrable density kernels multiplied by $H^{d/2}\mathbb{P}(\mathcal{A}_{it,\mathtt{FS}}^c \mid \mathcal{D}_{t-1}) = o_{\mathbb{P}}(1)$.

More explicitly, the uniform conditional density bound for β_{it} gives

$$\mathbb{P}\big(\mathcal{A}_{it,\mathtt{FS}}^c,\, \beta_{it} \in P \,\big|\, \mathcal{D}_{t-1}\big) \lesssim_{\mathbb{P}} \min\big\{\mathbb{P}(\mathcal{A}_{it,\mathtt{FS}}^c \mid \mathcal{D}_{t-1}), J^{-d}\big\},$$

while conditional convolution on $\mathcal{A}_{it,\mathtt{FS}}^c$ gives

$$\mathbb{P}\Big(\mathcal{A}_{it,\mathtt{FS}}^c,\, \widehat{\beta}_{it} \in P \,\Big|\, \mathcal{D}_{t-1}\Big) \lesssim_{\mathbb{P}} \mathbb{P}(\mathcal{A}_{it,\mathtt{FS}}^c \mid \mathcal{D}_{t-1}) \min\Big\{1, \frac{H^{d/2}}{J^d}\Big\} = o_{\mathbb{P}}(J^{-d}).$$

These two bounds cover respectively $J < \sqrt{H}$ and $J \ge \sqrt{H}$. For the conditional mean, the leave-one-out Taylor coefficient and conditional convolution give the sharper exceptional-event bound

$$\frac{1}{\sqrt{H}}\mathbb{P}\big(\mathcal{A}_{it,\mathtt{FS}}^c \,\big|\, \mathcal{D}_{t-1}\big) \min\Big\{1, \frac{H^{d/2}}{J^d}\Big\}$$

$$\leq \frac{1}{J^d}\min\left\{\frac{J^d}{\sqrt{H}}, H^{(d-1)/2}\right\}\mathbb{P}\big(\mathcal{A}_{it,\mathsf{FS}}^c \,\big|\, \mathcal{D}_{t-1}\big) \lesssim_{\mathbb{P}} \frac{\mathsf{B}_{\mathsf{FS}}+H^{-1}}{J^d} = \frac{\mathsf{B}_{\mathsf{FS}}}{J^d}+\frac{1}{HJ^d},$$

by (SA-II.4.2). For the conditional variance, the two preceding cell-probability bounds imply that the exceptional-event contribution is absorbed by $\{J^{d-1}\max(J,\sqrt{H})\}^{-1}$. Thus the drift terms in the calculations below are evaluated on $\mathcal{A}_{it,\mathsf{FS}}$ and are controlled by B_{FS}, without treating $\widehat{\mathsf{B}}_{\mathsf{FS}}$ as $\mathcal{D}_{t-1}$-measurable. To avoid repeating event indicators, the $\mathcal{A}_{it,\mathsf{FS}}$ indicator is suppressed in the regular-event displacement bounds below; the complementary contribution is always controlled by the bounds just displayed.

Write

$$c_{itl} = S\widehat{A}(t)^{-1}X_l(t)K\Big(\frac{l}{H}\Big)\varsigma_{i(t-l)} \in \mathbb{R}^d,$$
$$c_{ts} = S\widehat{A}(t)^{-1}X_s(t)K\Big(\frac{s}{H}\Big) \in \mathbb{R}^d.$$

Define $G_{its} = \frac{1}{\sqrt{H}}\sum_{\substack{l=1\\l\neq s}}^{H} S\widehat{A}(t)^{-1}X_l(t)K\Big(\frac{l}{H}\Big)\varsigma_{i(t-l)}\eta_{i(t-l)}$ and $\delta_{its} = \frac{1}{\sqrt{H}}c_{its}\eta_{i(t-s)}$. Let

$$F_{G_{its}}(z) = \mathbb{P}\bigg(\frac{1}{\sqrt{H}}\sum_{\substack{l=1\\l\neq s}}^{H} S\widehat{A}(t)^{-1}X_l(t)K\Big(\frac{l}{H}\Big)\varsigma_{i(t-l)}\eta_{i(t-l)} \leq z \,\Big|\, \mathcal{E}_{t-1}, \eta_{i(t-s)}\bigg)$$

be the conditional CDF of G_{its}, with $f_{G_{its}}(z)$ the corresponding conditional density. Its existence follows from the same characteristic-function argument as in Lemma SA-II.3.1. For any rectangle $[\ell,u]$, $F_{G_{its}}([\ell,u]) = \mathbb{P}(G_{its}\in[\ell,u]|\mathcal{E}_{t-1},\eta_{i(t-s)}) = \mathbb{P}(G_{its}\in[\ell,u]|\mathcal{E}_{t-1})$ by independence of $\eta_{i(t-s)}$. Use $f_{\beta_{it}}(\cdot|\mathcal{D}_{t-1})$ to denote the conditional density of β_{it} given $\mathcal{D}_{t-1}$, which is possibly heterogeneous over i and t.

By independence, the (conditional) characteristic function of G_{its} can be written as

$$\varphi_{G_{its}}(z) = \prod_{\substack{l=1\\l\neq s}}^{H}\varphi_\eta(c_{itl}^\top z/\sqrt{H}), \quad \text{for } z\in\mathbb{R}^d$$

where $\varphi_\eta(z) = \mathbb{E}[e^{iz\eta}]$ is the characteristic function of a generic random variable η that follows the same distribution as η_{it}. By the inversion formula,

$$f_{G_{its}}(z) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-iu^\top z}\varphi_{G_{its}}(u)\,\mathrm{d}u, \quad \text{for } z\in\mathbb{R}^d.$$

Its derivative is

$$\nabla f_{G_{its}}(z) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}(-iu)e^{-iu^\top z}\varphi_{G_{its}}(u)\,\mathrm{d}u.$$

Before we prove the main results, we need some properties of the density $f_{G_{its}}$. We summarize them in the following claim: with probability approaching one, uniformly over $i\in[n_t]$, $t=H+1,\ldots,T$, $s\in[H]$, and $k=0,1,2,3$,

$$\sup_{z\in\mathbb{R}^d}\|\nabla^k f_{G_{its}}(z)\|_\infty \lesssim 1, \qquad \|\nabla^k f_{G_{its}}(z)\|_\infty \lesssim \|z\|^{-d-1} \quad \text{for } \|z\|\geq 1. \tag{SA-II.4.4}$$

We will prove this claim later in Step 5.

Step 2: Key expansion.

$$
\begin{aligned}
&\mathbb{E}\Big[\Big(\mathbb{1}(\widehat{\beta}_{it}\in P)-\mathbb{1}(\beta_{it}\in P)\Big)(\nabla\mu_t(\beta_{it}))^\top\Big(S\widehat{A}(t)^{-1}X_s(t)\Big)K\Big(\frac{s}{H}\Big)\varsigma_{i(t-s)}\eta_{i(t-s)}\Big|\mathcal{D}_{t-1}\Big]\\
&=\mathbb{E}\Big[(\nabla\mu_t(\beta_{it}))^\top c_{its}\mathbb{E}\Big[\mathbb{1}(\widehat{\beta}_{it}\in P)\eta_{i(t-s)}\Big|\mathcal{E}_{t-1}\Big]\Big|\mathcal{D}_{t-1}\Big]\\
&=\mathbb{E}\Big[(\nabla\mu_t(\beta_{it}))^\top c_{its}\mathbb{E}\Big[\mathbb{1}\Big(\frac{1}{\sqrt{H}}\sum_{\substack{l=1\\l\neq s}}^{H}c_{itl}\eta_{i(t-l)}\in\sqrt{H}\Big(P-\breve{\beta}_{it}-\frac{1}{H}c_{its}\eta_{i(t-s)}\Big)\Big)\eta_{i(t-s)}\\
&\qquad\qquad\Big|\mathcal{E}_{t-1},\eta_{i(t-s)}\Big]\Big|\mathcal{D}_{t-1}\Big]\\
&=\mathbb{E}\Big[(\nabla\mu_t(\beta_{it}))^\top c_{its}\eta_{i(t-s)}\Big(F_{G_{its}}\Big(\sqrt{H}(P-\breve{\beta}_{it})\Big)\\
&\qquad\qquad+F_{G_{its}}\Big(\sqrt{H}(P-\breve{\beta}_{it}-\frac{1}{H}c_{its}\eta_{i(t-s)})\Big)-F_{G_{its}}\Big(\sqrt{H}(P-\breve{\beta}_{it})\Big)\Big)\Big|\mathcal{D}_{t-1}\Big]\\
&=\mathbb{E}\Big[(\nabla\mu_t(\beta_{it}))^\top c_{its}\eta_{i(t-s)}\Big(F_{G_{its}}\Big(\sqrt{H}(P-\breve{\beta}_{it}-\frac{1}{H}c_{its}\eta_{i(t-s)})\Big)-F_{G_{its}}\Big(\sqrt{H}(P-\breve{\beta}_{it})\Big)\Big)\Big|\mathcal{D}_{t-1}\Big]
\end{aligned}
$$

by the independence between $\eta_{i(t-s)}$ and variables in $\mathcal{E}_{t-1}$.

Now, by the density bound (SA-II.4.4), we further have

$$
\begin{aligned}
&\mathbb{E}\Big[(\nabla\mu_t(\beta_{it}))^\top c_{its}\mathbb{E}\Big[\mathbb{1}(\widehat{\beta}_{it}\in P)\eta_{i(t-s)}\Big|\mathcal{E}_{t-1}\Big]\Big|\mathcal{D}_{t-1}\Big]\\
&=\mathbb{E}\Big[(\nabla\mu_t(\beta_{it}))^\top c_{its}\eta_{i(t-s)}\int_{\sqrt{H}(P-\breve{\beta}_{it})}(f_{G_{its}}(z-\delta_{its})-f_{G_{its}}(z))\,\mathrm{d}z\Big|\mathcal{D}_{t-1}\Big]\\
&=\mathbb{E}\Big[(\nabla\mu_t(\beta_{it}))^\top c_{its}\eta_{i(t-s)}\Big[-\delta_{its}^\top\int_{\sqrt{H}(P-\breve{\beta}_{it})}\nabla f_{G_{its}}(z)\,\mathrm{d}z\\
&\qquad\qquad+\delta_{its}^\top\int_{\sqrt{H}(P-\breve{\beta}_{it})}\int_0^1\Big(\nabla f_{G_{its}}(z)-\nabla f_{G_{its}}(z-v\delta_{its})\Big)\,\mathrm{d}v\,\mathrm{d}z\Big]\Big|\mathcal{D}_{t-1}\Big]\\
&=-\frac{1}{\sqrt{H}}\mathbb{E}\Big[(\nabla\mu_t(\beta_{it}))^\top c_{its}c_{its}^\top\int_{\sqrt{H}(P-\breve{\beta}_{it})}\nabla f_{G_{its}}(z)dz\Big|\mathcal{D}_{t-1}\Big]\\
&\quad+\frac{1}{H}\mathbb{E}\Big[(\nabla\mu_t(\beta_{it}))^\top c_{its}c_{its}^\top\mathbb{E}\Big[\eta_{i(t-s)}^3\int_{\sqrt{H}(P-\breve{\beta}_{it})}\int_0^1(1-v)\nabla^2 f_{G_{its}}(z-v\delta_{its})c_{its}\,\mathrm{d}v\,\mathrm{d}z\Big|\mathcal{E}_{t-1}\Big]\Big|\mathcal{D}_{t-1}\Big]\\
&=-\frac{1}{\sqrt{H}}I_{1,s}+\frac{1}{H}I_{2,s},
\end{aligned}
$$

where $I_{1,s}$ and $I_{2,s}$ denote the corresponding terms.

Step 3: Bound $I_{1,s}$.

$$
\begin{aligned}
&\Big|\mathbb{E}\Big[(\nabla\mu_t(\beta_{it}))^\top c_{its}c_{its}^\top\int_{\sqrt{H}(P-\breve{\beta}_{it})}\nabla f_{G_{its}}(z)\,\mathrm{d}z\Big|\mathcal{D}_{t-1}\Big]\Big|\\
&\quad=\Big|\mathbb{E}\Big[(\nabla\mu_t(\beta_{it}))^\top c_{ts}c_{ts}^\top\varsigma_{i(t-s)}^2\int_{\sqrt{H}(P-\breve{\beta}_{it})}\nabla f_{G_{its}}(z)\,\mathrm{d}z\Big|\mathcal{D}_{t-1}\Big]\Big|\\
&=|c_{ts}^\top Mc_{ts}|\\
&\le\|c_{ts}\|^2\|M\|.
\end{aligned}
$$

where

$$M = \mathbb{E}\Big[\varsigma^2_{i(t-s)} \int_{\sqrt{H}(P-\breve{\beta}_{it})} \nabla f_{G_{its}}(z)\,\mathrm{d}z\ (\nabla\mu_t(\beta_{it}))^\top \Big| \mathcal{D}_{t-1}\Big].$$

It is enough to consider each entry of M, since d is fixed.

A typical (j,k)-th entry of M is

$$\begin{aligned}
&\mathbb{E}\Big[\varsigma^2_{i(t-s)} \int_{\sqrt{H}(P-\breve{\beta}_{it})} \nabla_j f_{G_{its}}(z)\,\mathrm{d}z\ \nabla_k\mu_t(\beta_{it}) \Big| \mathcal{D}_{t-1}\Big] \\
&= \mathbb{E}\Big[\varsigma^2_{i(t-s)} \int_{\sqrt{H}(P-\beta_{it})} \nabla_j f_{G_{its}}(z)\,\mathrm{d}z\ \nabla_k\mu_t(\beta_{it}) \Big| \mathcal{D}_{t-1}\Big] \\
&\quad + \mathbb{E}\Big[\varsigma^2_{i(t-s)} \int_{\sqrt{H}(P-\breve{\beta}_{it})} \nabla_j f_{G_{its}}(z)\,\mathrm{d}z\ \nabla_k\mu_t(\beta_{it}) - \varsigma^2_{i(t-s)} \int_{\sqrt{H}(P-\beta_{it})} \nabla_j f_{G_{its}}(z)\,\mathrm{d}z\ \nabla_k\mu_t(\beta_{it}) \Big| \mathcal{D}_{t-1}\Big].
\end{aligned}$$

For the integration by parts in b_j, write $f_{G_{its}}(z) \equiv f_{G_{its}}(z_j; z_{-j})$ and define

$$m(b_j; z_{-j}, b_{-j}) = \int_0^{\sqrt{H}(u_j-\ell_j)} f_{G_{its}}(z + \sqrt{H}(\ell - b))\,\mathrm{d}z_j.$$

On the one hand,

$$\begin{aligned}
&\mathbb{E}\Big[\varsigma^2_{i(t-s)} \int_{\sqrt{H}(P-\beta_{it})} \nabla_j f_{G_{its}}(z)\,\mathrm{d}z\ \nabla_k\mu_t(\beta_{it}) \Big| \mathcal{D}_{t-1}\Big] \\
&= \varsigma^2_{i(t-s)} \int_{\mathcal{B}} \Big(\int_0^{\sqrt{H}(u-\ell)} \nabla_j f_{G_{its}}(z + \sqrt{H}(\ell - b))\,\mathrm{d}z\Big) \\
&\qquad\qquad \times \nabla_k\mu_t(b) f_{\beta_{it}}(b|\mathcal{D}_{t-1})\,\mathrm{d}b \\
&= \varsigma^2_{i(t-s)} \int_{\mathcal{B}_{-j}} \int_0^{\sqrt{H}(u_{-j}-\ell_{-j})} \int_{\mathcal{B}_j} \Big(\int_0^{\sqrt{H}(u_j-\ell_j)} \nabla_j f_{G_{its}}(z + \sqrt{H}(\ell - b))\ \mathrm{d}z_j\Big) \\
&\qquad\qquad \times \nabla_k\mu_t(b) f_{\beta_{it}}(b|\mathcal{D}_{t-1})\,\mathrm{d}b_j\,\mathrm{d}z_{-j}\,\mathrm{d}b_{-j} \\
&= -\frac{\varsigma^2_{i(t-s)}}{\sqrt{H}} \int_{\mathcal{B}_{-j}} \int_0^{\sqrt{H}(u_{-j}-\ell_{-j})} m(b_j; z_{-j}, b_{-j}) \nabla_k\mu_t(b) \\
&\qquad\qquad \times f_{\beta_{it}}(b|\mathcal{D}_{t-1})\Big|_{\underline{\beta}_j}^{\bar{\beta}_j} \mathrm{d}z_{-j}\,\mathrm{d}b_{-j} \\
&\quad + \frac{\varsigma^2_{i(t-s)}}{\sqrt{H}} \int_{\mathcal{B}_{-j}} \int_0^{\sqrt{H}(u_{-j}-\ell_{-j})} \int_{\mathcal{B}_j} \Big(\int_0^{\sqrt{H}(u_j-\ell_j)} f_{G_{its}}(z + \sqrt{H}(\ell - b))\ \mathrm{d}z_j\Big) \\
&\qquad\qquad \times \nabla_j\Big(\nabla_k\mu_t(b) f_{\beta_{it}}(b|\mathcal{D}_{t-1})\Big)\,\mathrm{d}b_j\,\mathrm{d}z_{-j}\,\mathrm{d}b_{-j} \\
&= O\Big(\frac{1}{\sqrt{H}J^d}\Big) \\
&\quad + \frac{\varsigma^2_{i(t-s)}}{\sqrt{H}} \int_{\mathcal{B}_{-j}} \int_0^{\sqrt{H}(u_{-j}-\ell_{-j})} \int_{\mathcal{B}_j} \Big(\int_0^{\sqrt{H}(u_j-\ell_j)} f_{G_{its}}(z + \sqrt{H}(\ell - b))\ \mathrm{d}z_j\Big) \\
&\qquad\qquad \times \nabla_j\Big(\nabla_k\mu_t(b) f_{\beta_{it}}(b|\mathcal{D}_{t-1})\Big)\,\mathrm{d}b_j\,\mathrm{d}z_{-j}\,\mathrm{d}b_{-j} \\
&= O\Big(\frac{1}{\sqrt{H}J^d}\Big) + \frac{\varsigma^2_{i(t-s)}}{\sqrt{H}} \int_{\sqrt{H}(P-\mathcal{B})} f_{G_{its}}(z) \Big(\int_{\mathcal{B}\cap(P-z/\sqrt{H})} \nabla_j\Big(\nabla_k\mu_t(b) f_{\beta_{it}}(b|\mathcal{D}_{t-1})\Big)\,\mathrm{d}b\Big)\,\mathrm{d}z
\end{aligned}$$

$$\lesssim_{\mathbb{P}} \frac{1}{\sqrt{H}J^d}$$

where the inner integral in the last display is $\lesssim_{\mathbb{P}} J^{-d}$ uniformly. The above calculation uses the bounded derivatives of μ_t and the bounded Lipschitz conditional density $f_{\beta_{it}}(\cdot|\mathcal{D}_{t-1})$; the integration-by-parts step is understood coordinatewise using the corresponding bounded weak derivative. We also use that $[\ell, u]$ is at least of order $\log(nT)/\sqrt{H}$ away from the boundary of $\mathcal{B}$. To see the boundary term, define also

$$\tilde{m}(b_j, z_{-j}) = \int_0^{\sqrt{H}(u_j-\ell_j)} f_{G_{its}}(z_j + \sqrt{H}(\ell_j - b_j); z_{-j})\,\mathrm{d}z_j.$$

$$\begin{aligned}
&- \frac{1}{\sqrt{H}} \int_{\mathcal{B}_{-j}} \int_0^{\sqrt{H}(u_{-j}-\ell_{-j})} m(b_j; z_{-j}, b_{-j}) \nabla_k \mu_t(b) f_{\beta_{it}}(b|\mathcal{D}_{t-1}) \Big|_{\underline{\beta}_j}^{\bar{\beta}_j} \mathrm{d}z_{-j}\,\mathrm{d}b_{-j} \\
&= \frac{1}{\sqrt{H}} \int_{\sqrt{H}(P_{-j}-\mathcal{B}_{-j})} \int_{\mathcal{B}_{-j}\cap(P_{-j}-z_{-j}/\sqrt{H})} \Big[\tilde{m}(\underline{\beta}_j, z_{-j}) \nabla_k \mu_t(\underline{\beta}_j, b_{-j}) f_{\beta_{it}}(\underline{\beta}_j, b_{-j}|\mathcal{D}_{t-1}) \\
&\qquad\qquad - \tilde{m}(\bar{\beta}_j, z_{-j}) \nabla_k \mu_t(\bar{\beta}_j, b_{-j}) f_{\beta_{it}}(\bar{\beta}_j, b_{-j}|\mathcal{D}_{t-1}) \Big] \mathrm{d}b_{-j}\,\mathrm{d}z_{-j} \\
&\lesssim \frac{1}{\sqrt{H}} \Big(\int_{\sqrt{H}(P_{-j}-\mathcal{B}_{-j})} \int_0^{\sqrt{H}(u_j-\ell_j)} f_{G_{its}}(z_j + \sqrt{H}(\ell_j - \underline{\beta}_j); z_{-j})\,\mathrm{d}z_j\,\mathrm{d}z_{-j} \\
&\qquad + \int_{\sqrt{H}(P_{-j}-\mathcal{B}_{-j})} \int_0^{\sqrt{H}(u_j-\ell_j)} f_{G_{its}}(z_j + \sqrt{H}(\ell_j - \bar{\beta}_j); z_{-j})\,\mathrm{d}z_j\,\mathrm{d}z_{-j} \Big) \frac{1}{J^{d-1}} \\
&\lesssim \frac{1}{\sqrt{H}J^{d-1}} \mathbb{P}(|G_{its,j}| > \tau_n|\mathcal{E}_{t-1}) \lesssim_{\mathbb{P}} \frac{1}{\sqrt{H}J^d},
\end{aligned}$$

where $G_{its,j}$ denotes the j-th coordinate of G_{its} and $\tau_n \gtrsim \log(nT)$ denotes the distance between $[\ell, u]$ and the boundary of the support of β_{it}. The last inequality follows from Bernstein's inequality and the usual truncation argument: on an $\mathcal{E}_{t-1}$-measurable event that happens w.p.a.1, each innovation is truncated at M and recentered before applying Bernstein, while Chebyshev's inequality is applied to the centered complementary tail sum.

$$\mathbb{P}(|G_{its,j}| \geq \tau_n|\mathcal{E}_{t-1}) \leq \exp\Big(- \frac{0.5\tau_n^2}{M_0 + \tau_n M/3\sqrt{H}} \Big) + \frac{1}{\tau_n^2 M^{q-2}} \lesssim J^{-1} \qquad \text{for } M \asymp J^{\frac{1}{q-2}}.$$

because $J^{1/(q-2)}T^{1/q}/\sqrt{H} = o(1)$ and $\tau_n \gtrsim \log(nT)$.

On the other hand, we will use a partitioning argument to bound the remainder in $I_{1,s}$. Specifically, if $J \gtrsim H^{1/2}$, we can introduce a multi-index $m \in \mathbb{Z}^d$ and $\hbar \asymp H^{-1/2}$ and define $\mathcal{B}_m = \{b \in \mathcal{B} : b_j \in [\ell_j + m_j\hbar, \ell_j + (m_j+1)\hbar), j \in [d]\}$. Denote this collection of subrectangles by $\{\mathcal{B}_m : m \in \mathcal{M}\}$ for an index set $\mathcal{M}$. Since $\mathcal{B}$ is a bounded rectangle, we need at most a constant multiple of $H^{d/2}$ such subrectangles to cover the entire $\mathcal{B}$, i.e., $|\mathcal{M}| \lesssim H^{d/2}$. Then, we have

$$\begin{aligned}
&\mathbb{E}\Big[\varsigma_{i(t-s)}^2 \int_{\sqrt{H}(P-\breve{\beta}_{it})} \nabla_j f_{G_{its}}(z)\,\mathrm{d}z\ \nabla_k \mu_t(\beta_{it}) - \varsigma_{i(t-s)}^2 \int_{\sqrt{H}(P-\beta_{it})} \nabla_j f_{G_{its}}(z)\,\mathrm{d}z\ \nabla_k \mu_t(\beta_{it}) \Big| \mathcal{D}_{t-1} \Big] \\
&= \varsigma_{i(t-s)}^2 \mathbb{E}\Big[\int_{\sqrt{H}(P-\beta_{it})} \Big(\nabla_j f_{G_{its}}(z + \sqrt{H}(\beta_{it} - \breve{\beta}_{it})) - \nabla_j f_{G_{its}}(z) \Big)\,\mathrm{d}z\ \nabla_k \mu_t(\beta_{it}) \Big| \mathcal{D}_{t-1} \Big]
\end{aligned}$$

$$\lesssim \sum_{m\in\mathcal{M}} \mathbb{E}\Big[\mathbb{1}(\beta_{it}\in\mathcal{B}_m)\Big|\int_{\sqrt{H}(P-\beta_{it})}\Big(\nabla_j f_{G_{its}}(z+\sqrt{H}(\beta_{it}-\breve{\beta}_{it}))-\nabla_j f_{G_{its}}(z)\Big)\,\mathrm{d}z\,\Big|\Big|\mathcal{D}_{t-1}\Big]$$
$$\lesssim \frac{\sqrt{H}\mathsf{B}_{\mathsf{FS}}}{J^d}.$$

In the last step we use the following fact: for the first term,

$$\sum_{m\in\mathcal{M}} \mathbb{E}\Big[\mathbb{1}(\beta_{it}\in\mathcal{B}_m)\Big|\int_{\sqrt{H}(P-\beta_{it})}\Big(\nabla_j f_{G_{its}}(z-\sqrt{H}(\breve{\beta}_{it}-\beta_{it}))-\nabla_j f_{G_{its}}(z)\Big)\,\mathrm{d}z\Big|\Big|\mathcal{D}_{t-1}\Big]$$
$$= \sum_{m\in\mathcal{M}} \mathbb{E}\Big[\mathbb{1}(\beta_{it}\in\mathcal{B}_m)\Big|\int_{\sqrt{H}(P-\beta_{it})}[\nabla(\nabla_j f_{G_{its}})(\xi_z)]^\top(\sqrt{H}(\beta_{it}-\breve{\beta}_{it}))\,\mathrm{d}z\Big|\Big|\mathcal{D}_{t-1}\Big]$$
$$\lesssim \sum_{m\in\mathcal{M}} (1+\|m\|)^{-d-1}(\sqrt{H}/J)^d\sqrt{H}\mathsf{B}_{\mathsf{FS}}\mathbb{E}\Big[\mathbb{1}(\beta_{it}\in\mathcal{B}_m)\Big|\mathcal{D}_{t-1}\Big] \lesssim \frac{H^{d/2}}{J^d}\frac{1}{H^{d/2}}\sqrt{H}\mathsf{B}_{\mathsf{FS}}$$

where ξ_z is some point between z and $z-\sqrt{H}(\breve{\beta}_{it}-\beta_{it})$ and we use the tail bound for $\nabla^2 f_{G_{its}}$.

Now, consider the case $J=o(H^{1/2})$. We slightly redefine the partition: divide the whole support $\mathcal{B}$ into rectangles of length proportional to $H^{-1/2}$, labelling these regions by a multi-index

$$m=(m_1,m_1',m_2,m_2',\ldots,m_d,m_d')\in\mathbb{Z}^{2d}.$$

Define $\mathcal{B}_m$ as the set of $b\in\mathcal{B}$ such that $\sqrt{H}(b_k-\ell_k)\in[m_k,m_k+1]$ and $\sqrt{H}(b_k-u_k)\in[m_k',m_k'+1]$ for all $k\in[d]$. That is, when a rectangle is associated with a multi-index with some m_k or m_k' large in absolute value, this region is far away from the facets of the hyper-rectangle $[\ell,u]$. Let $\mathcal{M}$ be the collection of possible indices m. Again, $|\mathcal{M}|\lesssim H^{d/2}$. For each $\mathcal{B}_m$, let $\mathcal{J}_m=\{j\in[d]:\exists b\in\mathcal{B}_m \text{ s.t. } b_j\in[\ell_j,u_j]\}$ and $\mathcal{J}_m^c=[d]\setminus\mathcal{J}_m$. For $\beta_{it}\in\mathcal{B}_m$, we have

$$\int_{\sqrt{H}(P-\beta_{it})}\|\nabla(\nabla_j f_{G_{its}})(\xi_z)\|_\infty\,\mathrm{d}z \lesssim \frac{1}{\prod_{j\in\mathcal{J}_m^c}(1+|\min\{m_j,m_j'\}|)^{1+1/d}}\times\Big(\frac{\sqrt{H}}{J}\Big)^{|\mathcal{J}_m^c|}.$$

Note that for a fixed $j_0\in[d]$, all $\mathcal{B}_m$'s such that $|\mathcal{J}_m|=j_0$ can be classified into $\lesssim(\sqrt{H}/J)^{j_0}$ groups, each containing $\lesssim\sqrt{H}$ subrectangles. Repeating the previous argument, we have

$$\sum_{m\in\mathcal{M}} \mathbb{E}\Big[\mathbb{1}(\beta_{it}\in\mathcal{B}_m)\Big|\int_{\sqrt{H}(P-\beta_{it})}[\nabla(\nabla_j f_{G_{its}})(\xi_z)]^\top(\sqrt{H}(\beta_{it}-\breve{\beta}_{it}))\,\mathrm{d}z\Big|\Big|\mathcal{D}_{t-1}\Big]$$
$$\lesssim \sum_{j_0=1}^{d}\Big(\frac{\sqrt{H}}{J}\Big)^{j_0}\times\Big(\frac{\sqrt{H}}{J}\Big)^{d-j_0}\sqrt{H}\mathsf{B}_{\mathsf{FS}}\times\frac{1}{H^{d/2}} \lesssim \frac{\sqrt{H}\mathsf{B}_{\mathsf{FS}}}{J^d}.$$

Then, we conclude that

$$\frac{1}{H}\sum_{s=1}^{H} I_{1,s} \lesssim_{\mathbb{P}} \frac{1}{\sqrt{H}J^d}+\frac{\sqrt{H}\mathsf{B}_{\mathsf{FS}}}{J^d}.$$

Step 4: Bound $I_{2,s}$. We use a similar argument:

$$\mathbb{E}\bigg[(\nabla\mu_t(\beta_{it}))^\top c_{its}c_{its}^\top\mathbb{E}\bigg[\eta_{i(t-s)}^3\int_{\sqrt{H}(P-\breve{\beta}_{it})}\int_0^1(1-v)\nabla^2 f_{G_{its}}(z-v\delta_{its})c_{its}\,\mathrm{d}v\,\mathrm{d}z\bigg|\mathcal{E}_{t-1}\bigg]\bigg|\mathcal{D}_{t-1}\bigg]$$
$$=c_{ts}^\top\mathbb{E}\bigg[\eta_{i(t-s)}^3\varsigma_{i(t-s)}^3\int_{\sqrt{H}(P-\breve{\beta}_{it})}\int_0^1(1-v)\nabla^2 f_{G_{its}}(z-v\delta_{its})c_{ts}\,\mathrm{d}v\,\mathrm{d}z(\nabla\mu_t(\beta_{it}))^\top\bigg|\mathcal{D}_{t-1}\bigg]c_{ts}.$$

Again, we only need to bound the operator norm of the matrix in the middle. Furthermore,

$$\mathrm{vec}\left(\mathbb{E}\bigg[\eta_{i(t-s)}^3\varsigma_{i(t-s)}^3\int_{\sqrt{H}(P-\breve{\beta}_{it})}\int_0^1(1-v)\nabla^2 f_{G_{its}}(z-v\delta_{its})c_{ts}\,\mathrm{d}v\,\mathrm{d}z(\nabla\mu_t(\beta_{it}))^\top\bigg|\mathcal{D}_{t-1}\bigg]\right)$$
$$=\mathbb{E}\bigg[\nabla\mu_t(\beta_{it})\otimes\eta_{i(t-s)}^3\varsigma_{i(t-s)}^3\int_{\sqrt{H}(P-\breve{\beta}_{it})}\int_0^1(1-v)\nabla^2 f_{G_{its}}(z-v\delta_{its})\,\mathrm{d}v\,\mathrm{d}z\bigg|\mathcal{D}_{t-1}\bigg]c_{ts}.$$

We only need to bound each entry of the expected matrix (a $d^2\times d$ matrix). Consider a generic entry: for some j, j', j'',

$$\mathbb{E}\bigg[\nabla_j\mu_t(\beta_{it})\eta_{i(t-s)}^3\varsigma_{i(t-s)}^3\int_{\sqrt{H}(P-\breve{\beta}_{it})}\int_0^1(1-v)\nabla_{j',j''}^2 f_{G_{its}}(z-v\delta_{its})\,\mathrm{d}v\,\mathrm{d}z\bigg|\mathcal{D}_{t-1}\bigg]$$
$$=\varsigma_{i(t-s)}^3\int_{\mathcal{B}}\nabla_j\mu_t(b)\int_{\sqrt{H}(P-b)}\mathbb{E}\Big[\eta_{i(t-s)}^3\int_0^1(1-v)\nabla_{j',j''}^2 f_{G_{its}}(z-v\delta_{its})\,\mathrm{d}v\Big|\mathcal{E}_{t-1}\Big]\,\mathrm{d}z f_{\beta_{it}}(b|\mathcal{D}_{t-1})\,\mathrm{d}b$$
$$+\mathbb{E}\bigg[\nabla_j\mu_t(\beta_{it})\bigg(\varsigma_{i(t-s)}^3\int_{\sqrt{H}(P-\breve{\beta}_{it})}\eta_{i(t-s)}^3\int_0^1(1-v)\nabla_{j',j''}^2 f_{G_{its}}(z-v\delta_{its})\,\mathrm{d}v\,\mathrm{d}z$$
$$-\varsigma_{i(t-s)}^3\int_{\sqrt{H}(P-\beta_{it})}\eta_{i(t-s)}^3\int_0^1(1-v)\nabla_{j',j''}^2 f_{G_{its}}(z-v\delta_{its})\,\mathrm{d}v\,\mathrm{d}z\bigg)\bigg|\mathcal{D}_{t-1}\bigg].$$

Let g_η denote the Lebesgue density of η_{it} and let $M_0>0$ be fixed. For the first term,

$$\int_{\mathcal{B}}\nabla_j\mu_t(b)\int_{\sqrt{H}(P-b)}\mathbb{E}\Big[\eta_{i(t-s)}^3\int_0^1(1-v)\nabla_{j',j''}^2 f_{G_{its}}(z-v\delta_{its})\,\mathrm{d}v\Big|\mathcal{E}_{t-1}\Big]\,\mathrm{d}z\,f_{\beta_{it}}(b|\mathcal{D}_{t-1})\,\mathrm{d}b$$
$$=\int_{\mathbb{R}}\eta^3 g_\eta(\eta)\int_0^1(1-v)\int_{\mathcal{B}}\int_{\sqrt{H}(P-b)}\nabla_j\mu_t(b)\nabla_{j',j''}^2 f_{G_{its}}(z-v\delta_{its})f_{\beta_{it}}(b|\mathcal{D}_{t-1})\,\mathrm{d}z\,\mathrm{d}b\times$$
$$\Big(\mathbb{1}(\|c_{ts}\||\eta|\le M_0\sqrt{H})+\mathbb{1}(\|c_{ts}\||\eta|>M_0\sqrt{H})\Big)\,\mathrm{d}v\,\mathrm{d}\eta$$
$$\lesssim_{\mathbb{P}} J^{-d}.$$

Note that for $\|c_{ts}\||\eta|\le M_0\sqrt{H}$, $\|v\delta_{its}\|\lesssim\|c_{ts}\||\eta|/\sqrt{H}\lesssim 1$, and thus in the last line we apply the previous partitioning argument again to obtain

$$\Big|\int_{\mathcal{B}}\int_{\sqrt{H}(P-b)}\nabla_j\mu_t(b)\nabla_{j',j''}^2 f_{G_{its}}(z-v\delta_{its})f_{\beta_{it}}(b|\mathcal{D}_{t-1})\,\mathrm{d}z\,\mathrm{d}b\,\mathbb{1}(\|c_{ts}\||\eta|\le M_0\sqrt{H})\Big|$$
$$\lesssim\int_{\mathcal{B}}\int_{\sqrt{H}(P-b)}|\nabla_{j',j''}^2 f_{G_{its}}(z-v\delta_{its})|\,\mathrm{d}b\,\mathrm{d}z\mathbb{1}(\|c_{ts}\||\eta|\le M_0\sqrt{H})$$
$$=\sum_{m\in\mathcal{M}}\int_{\mathcal{B}_m}\int_{\sqrt{H}(P-b)}|\nabla_{j',j''}^2 f_{G_{its}}(z-v\delta_{its})|\mathbb{1}(\|c_{ts}\||\eta|\le M_0\sqrt{H})\,\mathrm{d}b\,\mathrm{d}z$$
$$\lesssim J^{-d}.$$

On the other hand,

$$\begin{aligned}
&\int_{\mathcal{B}} \nabla_j \mu_t(b) \int_{\sqrt{H}(P-b)} \mathbb{E}\Big[\eta_{i(t-s)}^3 \int_0^1 (1-v)\nabla^2_{j',j''} f_{G_{its}}(z - v\delta_{its})\,\mathrm{d}v \\
&\qquad\qquad \times \mathbb{1}(\|c_{ts}\||\eta_{i(t-s)}| > M_0\sqrt{H})\Big|\mathcal{E}_{t-1}\Big]\,\mathrm{d}z\, f_{\beta_{it}}(b|\mathcal{D}_{t-1})\,\mathrm{d}b \\
&\lesssim \mathbb{E}[|\eta_{i(t-s)}|^{3+d}]\|c_{ts}\|^d H^{-d/2} \times \Big(\frac{\sqrt{H}}{J}\Big)^d \\
&\lesssim \|c_{ts}\|^d J^{-d}.
\end{aligned}$$

In addition, repeating the previous argument in Step 3 yields

$$\begin{aligned}
&\mathbb{E}\Bigg[\nabla_j \mu_t(\beta_{it})\Bigg(\varsigma_{i(t-s)}^3 \int_{\sqrt{H}(P-\check{\beta}_{it})} \eta_{i(t-s)}^3 \int_0^1 (1-v)\nabla^2_{j',j''} f_{G_{its}}(z - v\delta_{its})\,\mathrm{d}v\,\mathrm{d}z \\
&\qquad - \varsigma_{i(t-s)}^3 \int_{\sqrt{H}(P-\beta_{it})} \eta_{i(t-s)}^3 \int_0^1 (1-v)\nabla^2_{j',j''} f_{G_{its}}(z - v\delta_{its})\,\mathrm{d}v\,\mathrm{d}z\Bigg)\Bigg|\mathcal{D}_{t-1}\Bigg] \\
&\lesssim \frac{\sqrt{H}\mathsf{B}_{\mathsf{FS}}}{J^d}.
\end{aligned}$$

Therefore, we conclude that

$$\frac{1}{H}\sum_{s=1}^H I_{2,s} \lesssim_{\mathbb{P}} \frac{1}{J^d}.$$

Also, note that in the above calculation we need $\frac{1}{H}\sum_{s=1}^H \|c_{ts}\|^{3+d} \lesssim_{\mathbb{P}} 1$.

Step 5: Density bounds for G_{its}. We prove (SA-II.4.4).

Step 5.1: Factor moments and projection lower bound. The density bounds use two properties, both holding with probability approaching one. First,

$$\max_{H+1\le t\le T} \frac{1}{H}\sum_{s=1}^H |f_{t-s}|^m \lesssim 1, \qquad m = 3+d.$$

This is guaranteed by the imposed moment conditions and the maximal Fuk–Nagaev inequality described in Lemma SA-II.2.2. The process $|f_{t-s}|^m$ is $L_{q/m}$-NED with $q/m > 2$, so

$$\max_{H+1\le t\le T}\left|\frac{1}{H}\sum_{s=1}^H (|f_{t-s}|^m - \mathbb{E}[|f_{t-s}|^m])\right| = o_{\mathbb{P}}(1)$$

by the same argument as in the proof of Lemma SA-II.2.2. Second, for $v \in \mathbb{R}^d$, $m_v = |\{l : |v^\top c_{itl}| \ge a\|v\|\}| \ge \pi_0(H-1)$ for a constant $a > 0$ and a positive sequence π_0 with $\pi_0 H \to \infty$. To see this, note that $\frac{1}{H}\sum_{\substack{l=1\\l\neq s}}^H |v^\top c_{itl}|^2/\|v\|^2 \ge \lambda_{\min}(\frac{1}{H}\sum_{\substack{l=1\\l\neq s}}^H c_{itl}c_{itl}^\top) \ge \underline{\lambda}$ for some $\underline{\lambda}$, w.p.a.1. Let $a = \sqrt{\underline{\lambda}/2}$. We have

$$\frac{1}{H-1}\sum_{\substack{l=1\\l\neq s}}^H |v^\top c_{itl}|^2/\|v\|^2 \le a^2\left(1 - \frac{m_v}{H-1}\right) + \frac{\bar{c}_f^2 m_v}{H-1} = a^2 + \frac{(\bar{c}_f^2 - a^2)m_v}{H-1},$$

which implies that $\frac{m_v}{H-1} \geq \pi_0$ with $\pi_0 = \frac{\lambda/2}{\bar{c}_f^2 - \lambda/2}$. Given the moment condition imposed, $\bar{c}_f \lesssim_{\mathbb{P}} T^{1/q}$. Therefore, under the rate condition $T^{2/q}/H = o(1)$, we can guarantee $\pi_0 H \to \infty$.

Step 5.2: Uniform boundedness of density derivatives. Use ∇^α to denote the partial derivative of order $\alpha \in \mathbb{Z}_+^d$. Note that

$$\sup_{z\in\mathbb{R}^d} |\nabla^\alpha f_{G_{its}}(z)| = \sup_{z\in\mathbb{R}^d} \Big| \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} (-iu)^\alpha e^{-iu^\top z} \varphi_{G_{its}}(u)\, \mathrm{d}u \Big| \lesssim \int_{\mathbb{R}^d} \|u\|^{|\alpha|} |\varphi_{G_{its}}(u)|\, \mathrm{d}u.$$

The assumption $g_\eta^{(1)} \in L^1(\mathbb{R})$ implies that $\varphi_\eta(u) \to 0$ and $|\varphi_\eta(u)| \lesssim |u|^{-1}$ as $|u| \to \infty$. Together with the second-order expansion of φ_η at the origin, there are constants $c_\eta, C_\eta > 0$ and $B_\eta < \infty$ such that, for all $x \in \mathbb{R}$,

$$|\varphi_\eta(x)| \leq \exp\{-c_\eta \min(x^2, 1)\}, \qquad |\varphi_\eta(x)| \leq \frac{C_\eta}{|x|} \quad \text{if } |x| \geq B_\eta.$$

Let $\bar{c}_f = \max_{itl} \|c_{itl}\|$ and $z_l(u) = c_{itl}^\top u/\sqrt{H}$. For $\|u\| \leq B\sqrt{H}$ with fixed $B < \infty$, the first bound gives

$$|\varphi_{G_{its}}(u)| \leq \exp\left\{ -c_\eta \sum_{\substack{l=1 \\ l\neq s}}^{H} \min\left(z_l(u)^2, 1\right) \right\}.$$

Since $\lambda_{\min}\big(H^{-1} \sum_{\substack{l=1 \\ l\neq s}}^{H} c_{itl} c_{itl}^\top\big) \gtrsim 1$ w.p.a.1 and $\max_l z_l(u)^2 \leq \bar{c}_f^2 \|u\|^2/H$,

$$\sum_{\substack{l=1 \\ l\neq s}}^{H} \min\left(z_l(u)^2, 1\right) \gtrsim \frac{\sum_{\substack{l=1 \\ l\neq s}}^{H} z_l(u)^2}{1 + \max_{\substack{l=1 \\ l\neq s}}^{H} z_l(u)^2} \gtrsim \frac{\|u\|^2}{1 + \bar{c}_f^2 \|u\|^2/H} \gtrsim \min\left\{ \|u\|^2, \frac{H}{\bar{c}_f^2} \right\}.$$

Hence

$$\int_{\|u\| \leq B\sqrt{H}} \|u\|^{|\alpha|} |\varphi_{G_{its}}(u)|\, \mathrm{d}u \lesssim 1 + H^{(d+|\alpha|)/2} \exp\{-cH/\bar{c}_f^2\} \lesssim_{\mathbb{P}} 1,$$

using $\bar{c}_f \lesssim_{\mathbb{P}} T^{1/q}$ and the imposed rate condition $T^{2(3+d)/q}/H = o(1)$.

It remains to consider $\|u\| > B\sqrt{H}$ for a fixed B large enough. Write $v = u/\sqrt{H}$ and let $\mathcal{I}(u) = \{l : |v^\top c_{itl}| \geq a\|v\|\}$. By Step 5.1, $|\mathcal{I}(u)| \geq \pi_0 H$ for some $a > 0$ and a positive sequence π_0 satisfying $\pi_0 H \to \infty$ w.p.a.1. For $l \in \mathcal{I}(u)$, $|z_l(u)| \geq a\|v\|$. Choosing B so that $aB \geq B_\eta$, the tail bound on φ_η yields

$$\int_{\|u\| > B\sqrt{H}} \|u\|^{|\alpha|} |\varphi_{G_{its}}(u)|\, \mathrm{d}u \lesssim H^{\frac{d+|\alpha|}{2}} \int_{\|v\| > B} \|v\|^{|\alpha|} \Big(\frac{C_\eta}{a\|v\|} \Big)^{\pi_0 H} \mathrm{d}v \lesssim_{\mathbb{P}} 1,$$

for $\pi_0 H$ sufficiently large. This proves the boundedness of the density and its derivatives.

Step 5.3: Polynomial tail bound for density derivatives. Fix $\|z\| \geq 1$ and choose $r \in [d]$ such that $|z_r| \geq \|z\|/\sqrt{d}$. Let $\gamma = (d+1)e_r$ and define $\Psi(u) = (-iu)^\alpha \varphi_{G_{its}}(u)$. By integration by parts in the r-th coordinate,

$$|z_r|^{d+1} \big| f_{G_{its}}^{(\alpha)}(z) \big| \leq \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left| \frac{\partial^{d+1}}{\partial u_r^{d+1}} \Psi(u) \right| \mathrm{d}u.$$

Thus it suffices to show that the last integral is finite uniformly, because then $|f^{(\alpha)}_{G_{its}}(z)| \lesssim \|z\|^{-d-1}$.

Note that

$$[u^\alpha \varphi_{G_{its}}(u)]^{(\gamma)} = \sum_{\gamma_1 \le \gamma} \binom{\gamma}{\gamma_1} (u^\alpha)^{(\gamma_1)} \left(\varphi_{G_{its}}(u)\right)^{(\gamma-\gamma_1)}$$

where $\gamma_1 \le \gamma$ means $\gamma_{1,j} \le \gamma_j$ for every $j \in [d]$ and $\binom{\gamma}{\gamma_1} = \prod_{j\in[d]} \binom{\gamma_j}{\gamma_{1,j}}$. The derivative of u^α is still polynomial in u. For the derivative of $\varphi_{G_{its}}$, the product rule, the mean-zero expansion of φ_η around the origin, and the moment condition on η_{it} imply the following envelope for every fixed multi-index τ with $|\tau| \le d+1+|\alpha|$:

$$|\varphi^{(\tau)}_{G_{its}}(u)| \lesssim_{\mathbb{P}} (1+\|u\|)^{|\tau|} \max_{\substack{A\subset\{1,\dots,H\} \\ |A|\le|\tau|}} \prod_{l\notin A} \left|\varphi_\eta\left(\frac{u^\top c_{itl}}{\sqrt{H}}\right)\right|.$$

The finite set A only removes a fixed number of factors from the product, so the bounds in Step 5.2 continue to hold after replacing H by $H-|A|$. Hence, for a constant $M_{\alpha d} < \infty$ that depends only on d and $|\alpha|$,

$$|\Psi^{(\gamma)}(u)| \lesssim_{\mathbb{P}} (1+\|u\|)^{M_{\alpha d}} \max_{\substack{A\subset\{1,\dots,H\} \\ |A|\le d+1+|\alpha|}} \prod_{l\notin A} \left|\varphi_\eta\left(\frac{u^\top c_{itl}}{\sqrt{H}}\right)\right|.$$

On $\|u\| \le B\sqrt{H}$, the same Gaussian bound as in Step 5.2 applies, with H reduced by at most a fixed constant; on $\|u\| > B\sqrt{H}$, Step 5.1 gives at least $\pi_0 H$ minus a fixed constant projected arguments larger than $a\|u\|/\sqrt{H}$. Therefore, for some fixed $\bar{C} < \infty$,

$$\int_{\mathbb{R}^d} |\Psi^{(\gamma)}(u)|\,\mathrm{d}u \lesssim_{\mathbb{P}} 1 + H^{(d+M_{\alpha d})/2} \int_{\|v\|>B} (1+\|v\|)^{M_{\alpha d}} \left(\frac{C_\eta}{a\|v\|}\right)^{\pi_0 H - \bar{C}} \mathrm{d}v \lesssim_{\mathbb{P}} 1.$$

This proves the polynomial tail bound in (SA-II.4.4).

Step 6: Localized-basis extension. Fix a scalar component of $p_{\mathfrak{q}}(\cdot, P)$; the vector statement follows because the number of components is fixed. Conditional on $\mathcal{E}_{t-1}$ and $\eta_{i(t-s)}$, write

$$\widehat{\beta}_{it} = \breve{\beta}_{it} + H^{-1/2} G_{its} + H^{-1} c_{its} \eta_{i(t-s)}.$$

The term containing $p_{\mathfrak{q}}(\beta_{it}, P)$ has conditional mean zero after multiplication by $\eta_{i(t-s)}$. For the generated-regressor term, define the truncated moment functional

$$M_P(a; b_0) = \int_P p_{\mathfrak{q}}(v, P) H^{d/2} f_{G_{its}}\left(\sqrt{H}(v - b_0 - a)\right) \mathrm{d}v.$$

Then the conditional expectation is the expectation, over β_{it} and $\eta_{i(t-s)}$, of a bounded predictable multiplier times

$$\eta_{i(t-s)} \left[M_P\left(H^{-1} c_{its}\eta_{i(t-s)}; \breve{\beta}_{it}\right) - M_P(0; \breve{\beta}_{it}) \right].$$

For each fixed component, $M_P(a; b_0)$ is differentiable in a and

$$\nabla_a M_P(a; b_0) = -H^{(d+1)/2} \int_P p_{\mathfrak{q}}(v, P) \nabla f_{G_{its}}\left(\sqrt{H}(v - b_0 - a)\right) \mathrm{d}v.$$

Thus the first-order expansion in a differentiates the density $f_{G_{its}}$, not the normalized polynomial basis. Equivalently,

$$M_P\Big(H^{-1}c_{its}\eta_{i(t-s)};\breve{\beta}_{it}\Big)-M_P(0;\breve{\beta}_{it})=H^{-1}\eta_{i(t-s)}c_{its}^{\top}\int_0^1\nabla_a M_P\Big(rH^{-1}c_{its}\eta_{i(t-s)};\breve{\beta}_{it}\Big)\,\mathrm{d}r.$$

To make explicit why no derivative of the basis appears, let $a_k(b)=[\nabla\mu_t(b)]_k f_{\beta_{it}}(b|\mathcal{D}_{t-1})$ denote a generic bounded smooth multiplier. For any coordinate r,

$$\begin{aligned}&H^{(d+1)/2}\int_{\mathcal{B}}\int_P p_{\mathfrak{q}}(v,P)\partial_r f_{G_{its}}\Big(\sqrt{H}(v-b-a)\Big)\,a_k(b)\,\mathrm{d}v\,\mathrm{d}b\\&\quad=H^{d/2}\int_P p_{\mathfrak{q}}(v,P)\int_{\mathcal{B}}f_{G_{its}}\Big(\sqrt{H}(v-b-a)\Big)\,\partial_r a_k(b)\,\mathrm{d}b\,\mathrm{d}v+\mathcal{R}_{\partial,P}.\end{aligned}$$

The boundary term $\mathcal{R}_{\partial,P}$ is negligible uniformly over the cells in the lemma because P is separated from the boundary of $\mathcal{B}$ by more than $\log(nT)/\sqrt{H}$ and the same innovation truncation used above controls the tails of G_{its}. The main term is $\lesssim_{\mathbb{P}} J^{-d}$ because $p_{\mathfrak{q}}(\cdot,P)$, $\partial_r a_k$, and the conditional convolution density are uniformly bounded and $\mathrm{vol}(P)\asymp J^{-d}$. Multiplying by the increment $H^{-1}c_{its}\eta_{i(t-s)}$ gives $\lesssim_{\mathbb{P}}(HJ^d)^{-1}$, with no additional derivative factor in J. Replacing $\breve{\beta}_{it}$ by β_{it} in the density argument gives the same deterministic-drift term as before, namely $\lesssim_{\mathbb{P}}\mathsf{B}_{\mathtt{FS}}/J^d$. Averaging over s gives the displayed bound for the localized basis. When $\mathfrak{q}=0$, this argument coincides with the switching-bias calculation above. □

SA-II.4.5 Proof of Lemma SA-II.4.1(b)

The notation is as in part (a). We first prove the switching-variance bound, which is the case $\mathfrak{q}=0$, and then add the same-cell polynomial-coordinate movement that arises when $\mathfrak{q}\geq 1$. As in (SA-II.4.3), every $\mathcal{D}_{t-1}$-conditional moment below is split over $\mathcal{A}_{it,\mathtt{FS}}$ and $\mathcal{A}^c_{it,\mathtt{FS}}$. The complement is absorbed by the displayed baseline variance terms using Assumption SA-FSB; on $\mathcal{A}_{it,\mathtt{FS}}$, the signal displacement is bounded by $\mathsf{B}_{\mathtt{FS}}$. Define $\tilde{c}_{its}=[\nabla\mu_t(\beta_{it})]^{\top}c_{its}$. Let $\tilde{P}=\sqrt{H}[\ell-\breve{\beta}_{it},u-\breve{\beta}_{it}]$ and $\delta_{its}=H^{-1/2}c_{its}\eta_{i(t-s)}$.

Step 1: Switching summand variances. Note that

$$\begin{aligned}&\mathbb{V}\mathrm{ar}\left[\Big(\mathbb{1}(\widehat{\beta}_{it}\in P)-\mathbb{1}(\beta_{it}\in P)\Big)\sum_{s=1}^{H}\tilde{c}_{its}\eta_{i(t-s)}\Big|\mathcal{D}_{t-1}\right]\\&=\sum_{s=1}^{H}\mathbb{V}\mathrm{ar}\left[\Big(\mathbb{1}(\widehat{\beta}_{it}\in P)-\mathbb{1}(\beta_{it}\in P)\Big)\tilde{c}_{its}\eta_{i(t-s)}\Big|\mathcal{D}_{t-1}\right]\\&\quad+\sum_{\substack{s,s'=1\\s\neq s'}}^{H}\mathbb{C}\mathrm{ov}\left[\Big(\mathbb{1}(\widehat{\beta}_{it}\in P)-\mathbb{1}(\beta_{it}\in P)\Big)\tilde{c}_{its}\eta_{i(t-s)},\Big(\mathbb{1}(\widehat{\beta}_{it}\in P)-\mathbb{1}(\beta_{it}\in P)\Big)\tilde{c}_{its'}\eta_{i(t-s')}\Big|\mathcal{D}_{t-1}\right]\end{aligned}$$

For the variance of each summand, first note the crude bound

$$\mathbb{V}\mathrm{ar}\left[\Big(\mathbb{1}(\widehat{\beta}_{it}\in P)-\mathbb{1}(\beta_{it}\in P)\Big)\tilde{c}_{its}\eta_{i(t-s)}\Big|\mathcal{D}_{t-1}\right]\leq\mathbb{E}\left[\tilde{c}^2_{its}\eta^2_{i(t-s)}\Big|\mathcal{D}_{t-1}\right]\lesssim\|c_{ts}\|^2,$$

using that the squared indicator difference is bounded by one. The last inequality uses the bounded gradient and scale to obtain $|\tilde{c}_{its}|\lesssim\|c_{ts}\|$, together with the $\mathcal{D}_{t-1}$-measurability of c_{ts} and $\mathbb{E}[\eta^2_{i(t-s)}|\mathcal{D}_{t-1}]=1$. The sharper bound uses the leave-one-out expansion rather than multiplying dependent conditional expectations.

On the other hand, we also have the following bound

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}\left[\Big(\mathbb{1}(\widehat{\beta}_{it}\in P)-\mathbb{1}(\beta_{it}\in P)\Big)\tilde{c}_{its}\eta_{i(t-s)}\Big|\mathcal{D}_{t-1}\right]\\
&\lesssim \|c_{ts}\|^2\mathbb{E}\left[(\mathbb{1}(\widehat{\beta}_{it}\in P)-\mathbb{1}(\beta_{it}\in P))^2\eta_{i(t-s)}^2\Big|\mathcal{D}_{t-1}\right]\\
&=\|c_{ts}\|^2\mathbb{E}\left[\mathbb{E}\left[(\mathbb{1}(\widehat{\beta}_{it}\in P)-\mathbb{1}(\beta_{it}\in P))^2\eta_{i(t-s)}^2\Big|\mathcal{E}_{t-1},\eta_{i(t-s)}\right]\Big|\mathcal{D}_{t-1}\right]\\
&=\|c_{ts}\|^2\mathbb{E}\left[\eta_{i(t-s)}^2\Big(F_{G_{its}}(\tilde{P}-\delta_{its})-2\mathbb{1}(\beta_{it}\in P)F_{G_{its}}(\tilde{P}-\delta_{its})+\mathbb{1}(\beta_{it}\in P)\Big)\Big|\mathcal{D}_{t-1}\right]\\
&=\|c_{ts}\|^2\mathbb{E}\Bigg[\eta_{i(t-s)}^2\Big[\Big(F_{G_{its}}(\tilde{P})\\
&\quad+\int_{\tilde{P}}\Big(-\nabla f_{G_{its}}(z)^\top\delta_{its}+\frac{1}{2}\delta_{its}^\top\nabla^2 f_{G_{its}}(\xi_{its})\delta_{its}\Big)\,\mathrm{d}z\Big)\Big(1-2\mathbb{1}(\beta_{it}\in P)\Big)+\mathbb{1}(\beta_{it}\in P)\Big]\Big|\mathcal{D}_{t-1}\Bigg]\\
&\lesssim \|c_{ts}\|^2\mathbb{E}\left[\eta_{i(t-s)}^2\Big(F_{G_{its}}(\tilde{P})(1-2\mathbb{1}(\beta_{it}\in P))+\mathbb{1}(\beta_{it}\in P)\Big)\Big|\mathcal{D}_{t-1}\right]\\
&\quad-\frac{\|c_{ts}\|^2}{\sqrt{H}}\mathbb{E}\left[\int_{\tilde{P}}\nabla f_{G_{its}}(z)^\top\,\mathrm{d}z\;c_{its}(1-2\mathbb{1}(\beta_{it}\in P))\eta_{i(t-s)}^3\Big|\mathcal{D}_{t-1}\right]\\
&\quad+\frac{\|c_{ts}\|^2}{2H}\mathbb{E}\left[c_{its}^\top\int_{\tilde{P}}\nabla^2 f_{G_{its}}(\xi_{its})c_{its}\,\mathrm{d}z(1-2\mathbb{1}(\beta_{it}\in P))\eta_{i(t-s)}^4\Big|\mathcal{D}_{t-1}\right]\\
&\lesssim \|c_{ts}\|^2\mathbb{E}\left[\Big(F_{G_{its}}(\tilde{P})(1-2\mathbb{1}(\beta_{it}\in P))+\mathbb{1}(\beta_{it}\in P)\Big)\Big|\mathcal{D}_{t-1}\right]\\
&\quad+\Big(\frac{\|c_{ts}\|^3\mathbb{E}[|\eta_{it}|^3]}{\sqrt{H}J^d}+\frac{\|c_{ts}\|^4}{HJ^d}\Big)(1+\sqrt{H}\mathsf{B}_{\mathsf{FS}})
\end{aligned}
$$

with ξ_{its} some point between z and $z-\delta_{its}$, and the last step follows by the same argument in Step 4 of the proof of part (a).

Now, for the leading term, we apply the partitioning argument again. Specifically, we partition the whole support $\mathcal{B}$ of β_{it} into subrectangles of length proportional to $H^{-1/2}$. Recall that we label these subrectangles by a multi-index $m=(m_1,m_1',m_2,m_2',\cdots,m_d,m_d')\in\mathbb{Z}^{2d}$. Define $\mathcal{B}_m$ as the set of $b\in\mathcal{B}$ such that $\sqrt{H}(b_k-\ell_k)\in[m_k,m_k+1]$ and $\sqrt{H}(b_k-u_k)\in[m_k',m_k'+1]$ for all $k\in[d]$. That is, when a rectangle is associated with a multi-index with some m_k or m_k' large in absolute value, this region is far away from the facets of the hyper-rectangle $[\ell,u]$. Let $\mathcal{M}$ be the collection of possible indices m. Note that $|\mathcal{M}|\lesssim H^{d/2}$. Let

$$
\begin{aligned}
\mathcal{M}_0=&\Big\{m\in\mathcal{M}:\forall k,m_km_k'<0,\ \text{and}\ \exists l,|m_l|\le\tau_n\ \text{or}\ |m_l'|\le\tau_n\Big\}\\
&\cup\Big\{m\in\mathcal{M}:\exists k,m_km_k'\ge 0;\ |m_l|\le\tau_n\ \text{or}\ |m_l'|\le\tau_n\ \text{whenever}\ m_lm_l'\ge 0\Big\}
\end{aligned}
$$

for $\tau_n\asymp\log(nT)$. That is, $\cup_{m\in\mathcal{M}_0}\mathcal{B}_m$ is the "neighborhood" of every facet of P. Then,

$$
\begin{aligned}
&\mathbb{E}\left[F_{G_{its}}(\tilde{P})(1-2\mathbb{1}(\beta_{it}\in P))+\mathbb{1}(\beta_{it}\in P)\Big|\mathcal{D}_{t-1}\right]\\
&=\sum_{m\in\mathcal{M}}\int_{\mathcal{B}_m}\Big(F_{G_{its}}(\sqrt{H}[\ell-b,u-b])(1-2\mathbb{1}(b\in P))+\mathbb{1}(b\in P)\Big)f_{\beta_{it}}(b|\mathcal{D}_{t-1})\,\mathrm{d}b\\
&\quad+\mathbb{E}\Big[\Big(F_{G_{its}}(\sqrt{H}[\ell-\breve{\beta}_{it},u-\breve{\beta}_{it}])-F_{G_{its}}(\sqrt{H}[\ell-\beta_{it},u-\beta_{it}])\Big)(1-2\mathbb{1}(\beta_{it}\in P))\Big|\mathcal{D}_{t-1}\Big]\\
&=\sum_{m\in\mathcal{M}_0}\int_{\mathcal{B}_m}\Big(F_{G_{its}}(\sqrt{H}[\ell-b,u-b])(1-2\mathbb{1}(b\in P))+\mathbb{1}(b\in P)\Big)f_{\beta_{it}}(b|\mathcal{D}_{t-1})\,\mathrm{d}b
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{m\in\mathcal{M}_0} \mathbb{E}\Big[\mathbb{1}(\beta_{it}\in\mathcal{B}_m)\Big(F_{G_{its}}(\sqrt{H}[\ell-\breve{\beta}_{it},u-\breve{\beta}_{it}]) - F_{G_{its}}(\sqrt{H}[\ell-\beta_{it},u-\beta_{it}])\Big) \\
&\qquad\qquad \times (1-2\mathbb{1}(\beta_{it}\in P))\Big|\mathcal{D}_{t-1}\Big] \\
&+ \sum_{m\in\mathcal{M}\setminus\mathcal{M}_0} \int_{\mathcal{B}_m} \Big(F_{G_{its}}(\sqrt{H}[\ell-b,u-b])(1-2\mathbb{1}(b\in P)) + \mathbb{1}(b\in P)\Big) f_{\beta_{it}}(b|\mathcal{D}_{t-1})\,\mathrm{d}b \\
&+ \sum_{m\in\mathcal{M}\setminus\mathcal{M}_0} \mathbb{E}\Big[\mathbb{1}(\beta_{it}\in\mathcal{B}_m)\Big(F_{G_{its}}(\sqrt{H}[\ell-\breve{\beta}_{it},u-\breve{\beta}_{it}]) - F_{G_{its}}(\sqrt{H}[\ell-\beta_{it},u-\beta_{it}])\Big) \\
&\qquad\qquad \times (1-2\mathbb{1}(\beta_{it}\in P))\Big|\mathcal{D}_{t-1}\Big] \\
\lesssim_{\mathbb{P}}\ & \frac{\tau_n^d}{J^{d-1}\max\{J,\sqrt{H}\}} + \mathbb{P}(\|G_{its}\|>\tau_n|\mathcal{E}_{t-1}),
\end{aligned}
$$

Here

$$
\sqrt{H}\|\breve{\beta}_{it}-\beta_{it}\| \le \sqrt{H}\mathsf{B}_{\mathtt{FS}} = o(1) \qquad \text{on } \mathcal{A}_{it,\mathtt{FS}},
$$

and the complementary contribution is absorbed by (SA-II.4.3). The neighborhood of the facets of P has conditional mass bounded by the smaller of the boundary-strip order $J^{-(d-1)}\tau_n^d/\sqrt{H}$ and the whole-cell order $\tau_n^d J^{-d}$. Whenever b lies in $\mathcal{B}_m$ with $m\notin\mathcal{M}_0$, either $F_{G_{its}}(\sqrt{H}[\ell-b,u-b]) \le \mathbb{P}(\|G_{its}\|_\infty>\tau_n|\mathcal{E}_{t-1}) \lesssim H^{-d/2}$ ($b\notin P$) or $1-F_{G_{its}}(\sqrt{H}[\ell-b,u-b]) \le \mathbb{P}(\|G_{its}\|_\infty>\tau_n|\mathcal{E}_{t-1}) \lesssim H^{-d/2}$ ($b\in P$) w.p.a.1 (apply the Bernstein inequality as in Step 3 of the proof of part (a), but now we set $M = H^{\frac{d}{2(q-2)}}$). Combining all results together, we conclude that the first term in the variance is $\lesssim_{\mathbb{P}} \tau_n^d/(J^{d-1}\max\{J,\sqrt{H}\})$.

Step 2: Switching summand covariances. Now, consider the covariance between each pair of summands. For $s\neq s'$,

$$
\begin{aligned}
&\mathbb{C}\mathrm{ov}\left[\Big(\mathbb{1}(\widehat{\beta}_{it}\in P)-\mathbb{1}(\beta_{it}\in P)\Big)\tilde{c}_{its}\eta_{i(t-s)},\ \Big(\mathbb{1}(\widehat{\beta}_{it}\in P)-\mathbb{1}(\beta_{it}\in P)\Big)\tilde{c}_{its'}\eta_{i(t-s')}\Big|\mathcal{D}_{t-1}\right] \\
&= \mathbb{E}\left[\Big(\mathbb{1}(\widehat{\beta}_{it}\in P)-\mathbb{1}(\beta_{it}\in P)\Big)^2\tilde{c}_{its}\tilde{c}_{its'}\eta_{i(t-s)}\eta_{i(t-s')}\Big|\mathcal{D}_{t-1}\right] - \mathfrak{m}_{its}\mathfrak{m}_{its'} \\
&= \mathbb{E}\Bigg[\tilde{c}_{its}\tilde{c}_{its'}\eta_{i(t-s)}\eta_{i(t-s')}\Bigg[\Big(F_{G_{itss'}}(\tilde{P}) \\
&\quad + \int_{\tilde{P}}\Big(-\nabla f_{G_{itss'}}(z)^\top\delta_{itss'} + \delta_{itss'}^\top\int_0^1(1-v)\nabla^2 f_{G_{itss'}}(z-v\delta_{itss'})\,\mathrm{d}v\delta_{itss'}\Big)\mathrm{d}z\Big)\Big(1-2\mathbb{1}(\beta_{it}\in P)\Big) \\
&\qquad + \mathbb{1}(\beta_{it}\in P)\Bigg]\Big|\mathcal{D}_{t-1}\Bigg] - \mathfrak{m}_{its}\mathfrak{m}_{its'}
\end{aligned}
$$

where $\mathfrak{m}_{its} = \mathbb{E}[(\mathbb{1}(\widehat{\beta}_{it}\in P)-\mathbb{1}(\beta_{it}\in P))\tilde{c}_{its}\eta_{i(t-s)}|\mathcal{D}_{t-1}]$, $F_{G_{itss'}}$ and $f_{G_{itss'}}$, analogous to $F_{G_{its}}$ and $f_{G_{its}}$, are the distribution and density of $\frac{1}{\sqrt{H}}\sum_{\substack{l=1\\ l\notin\{s,s'\}}}^H c_{itl}\eta_{i(t-l)}$ conditional on $\sigma(\eta_{i(t-s)},\eta_{i(t-s')})\vee\mathcal{E}_{t-1}$, and $\delta_{itss'} = \frac{1}{\sqrt{H}}(c_{its}\eta_{i(t-s)}+c_{its'}\eta_{i(t-s')})$. The density bounds from Step 5 of the proof of part (a) apply to $f_{G_{itss'}}$ as well, because removing a fixed number of summands preserves the rolling-window Gram lower bound and the characteristic-function product bounds. When the diagonal and off-diagonal pieces are recombined, the mean-product contribution to the variance of the normalized sum is $-\big(H^{-1}\sum_{s=1}^H\mathfrak{m}_{its}\big)^2$. This term is

nonpositive. Part (a) of Lemma SA-II.4.1 bounds its absolute value by

$$\left|\frac{1}{H}\sum_{s=1}^{H}\mathfrak{m}_{its}\right|^2 \lesssim_{\mathbb{P}} \left(\frac{1}{HJ^d}+\frac{\mathsf{B}_{\mathsf{FS}}}{J^d}\right)^2,$$

which is the squared-mean term retained in the final display.

Since for any $s \neq s'$,

$$\mathbb{E}[\eta_{i(t-s)}\eta_{i(t-s')}|\mathcal{E}_{t-1}]=0, \qquad \mathbb{E}[\eta^2_{i(t-s)}\eta_{i(t-s')}|\mathcal{E}_{t-1}]=\mathbb{E}[\eta_{i(t-s)}\eta^2_{i(t-s')}|\mathcal{E}_{t-1}]=0,$$

all Taylor terms that are linear in one of the two innovations vanish after multiplication by $\eta_{i(t-s)}\eta_{i(t-s')}$; the second-order Hessian term remains because it contains products such as $\eta^2_{i(t-s)}\eta^2_{i(t-s')}$. Thus the leading term simplifies to

$$\frac{1}{H}\mathbb{E}\Bigg[\left(c_{its}\eta_{i(t-s)}+c_{its'}\eta_{i(t-s')}\right)^\top \tilde{c}_{its}\tilde{c}_{its'}\eta_{i(t-s)}\eta_{i(t-s')}\times \\ \int_{\tilde{P}}\int_0^1 (1-v)\nabla^2 f_{G_{itss'}}(z-v\delta_{itss'})\,\mathrm{d}v\,\mathrm{d}z(1-2\mathbb{1}(\beta_{it}\in P))(c_{its}\eta_{i(t-s)}+c_{its'}\eta_{i(t-s')})\Big|\mathcal{D}_{t-1}\Bigg].$$

Consider a generic Hessian element $f^{(\alpha)}_{G_{itss'}}$ with $|\alpha|=2$. Choose a coordinate $j\in[d]$ such that $\alpha_j\geq 1$. Writing $P=[\ell,u]$ and letting $(x,z_{-j})_j$ denote the vector with j-th coordinate x and remaining coordinates z_{-j}, the fundamental theorem of calculus gives

$$\int_{\sqrt{H}(P-\breve{\beta}_{it})} f^{(\alpha)}_{G_{itss'}}(z-v\delta_{itss'})\,\mathrm{d}z = \int_{\sqrt{H}(P_{-j}-\breve{\beta}_{it,-j})}\Big[f^{(\alpha-e_j)}_{G_{itss'}}\big((\sqrt{H}(u_j-\breve{\beta}_{it,j}),z_{-j})_j-v\delta_{itss'}\big) \\ - f^{(\alpha-e_j)}_{G_{itss'}}\big((\sqrt{H}(\ell_j-\breve{\beta}_{it,j}),z_{-j})_j-v\delta_{itss'}\big)\Big]\,\mathrm{d}z_{-j}.$$

The bracket contains derivatives of order one, evaluated on two facets of the rectangle. Keep these facet terms inside the outer conditional expectation. The density and tail bounds, together with the same innovation truncation, give

$$\mathbb{E}\left[\left(1+|\eta_{i(t-s)}|+|\eta_{i(t-s')}|\right)^4\times\sup_{v\in[0,1]}\left|\int_{\sqrt{H}(P-\breve{\beta}_{it})} f^{(\alpha)}_{G_{itss'}}(z-v\delta_{itss'})\,\mathrm{d}z\right|\,\middle|\,\mathcal{D}_{t-1}\right] \\ \lesssim_{\mathbb{P}} \begin{cases}\tau_n^d/(J^{d-1}\sqrt{H}), & J\leq\sqrt{H},\\ J^{-d}, & J>\sqrt{H},\end{cases} \\ \lesssim_{\mathbb{P}} \frac{\tau_n^d}{J^{d-1}\max\{J,\sqrt{H}\}}.$$

For the first case, integration over β_{it} localizes its coordinate normal to each facet. For the second, Fubini's theorem is applied to the Hessian integral before separating the two facets. Hence the leading term in the previous expansion is $\lesssim_{\mathbb{P}} \tau_n^d/[HJ^{d-1}\max\{J,\sqrt{H}\}]$.

Combining all these results yields the desired bound:

$$\mathbb{V}\mathrm{ar}\left[\frac{1}{H}\sum_{s=1}^{H}\Big(\mathbb{1}(\widehat{\beta}_{it}\in P)-\mathbb{1}(\beta_{it}\in P)\Big)\tilde{c}_{its}\eta_{i(t-s)}\Big|\mathcal{D}_{t-1}\right]$$

$$\lesssim_{\mathbb{P}} H^{-2}\Big[H \times \frac{\log^d(nT)}{J^{d-1}\max\{\sqrt{H}, J\}} + H^2 \times \frac{\log^d(nT)}{HJ^{d-1}\max\{J, \sqrt{H}\}}\Big] + \Big(\frac{1}{HJ^d} + \frac{\mathsf{B}_{\mathtt{FS}}}{J^d}\Big)^2$$
$$\lesssim_{\mathbb{P}} \frac{\log^d(nT)}{HJ^{d-1}\max\{J, \sqrt{H}\}} + \Big(\frac{1}{HJ^d} + \frac{\mathsf{B}_{\mathtt{FS}}}{J^d}\Big)^2.$$

The imposed conditions guarantee that, uniformly over $t = H+1, \ldots, T$, $\frac{1}{H}\sum_{s=1}^{H} \|c_{ts}\|^4 \lesssim 1$ and $\frac{1}{H^2}\sum_{\substack{s,s'=1 \\ s\neq s'}}^{H} \|c_{ts'}\|^2\|c_{ts}\|^2 \lesssim 1$ w.p.a.1.

Step 3: Localized-basis extension. Fix a scalar component of $p_{\mathfrak{q}}(\cdot, P)$. Write

$$p_{\mathfrak{q}}(\widehat{\beta}_{it}, P) - p_{\mathfrak{q}}(\beta_{it}, P) = \Delta^{\mathrm{bd}}_{itP} + \Delta^{\mathrm{in}}_{itP},$$

where $\Delta^{\mathrm{bd}}_{itP}$ is supported on the event that exactly one of $\widehat{\beta}_{it}$ and β_{it} belongs to P, and $\Delta^{\mathrm{in}}_{itP}$ is the remaining same-cell coordinate movement. Because $p_{\mathfrak{q}}(\cdot, P)$ is bounded, the boundary part is controlled by the switching calculation above and the squared mean bound from part (a). This gives $\log^d(nT)/\{HJ^{d-1}\max\{J, \sqrt{H}\}\} + \{1/(HJ^d) + \mathsf{B}_{\mathtt{FS}}/J^d\}^2$. When $\mathfrak{q} = 0$, the basis is constant on P and $\Delta^{\mathrm{in}}_{itP} = 0$. When $\mathfrak{q} \geq 1$, the normalized basis has derivative of order J, so on the same-cell event

$$\|\Delta^{\mathrm{in}}_{itP}\| \lesssim J\|\widehat{\beta}_{it} - \beta_{it}\|\mathbb{1}(\beta_{it} \in P).$$

This derivative bound is used only for the centered variance. From (SA-II.2.2),

$$\widehat{\beta}_{it} - \breve{\beta}_{it} = \frac{1}{H}\sum_{s=1}^{H} S\widehat{A}(t)^{-1}X_s(t)K\Big(\frac{s}{H}\Big)\varepsilon_{i(t-s)}.$$

Conditional Rosenthal bounds for the martingale sum, together with the moment bounds on the local-polynomial weights, give $\mathbb{E}[\|\widehat{\beta}_{it} - \breve{\beta}_{it}\|^2|\mathcal{E}_{t-1}] \lesssim_{\mathbb{P}} H^{-1}$ and $\mathbb{E}[\|\widehat{\beta}_{it} - \breve{\beta}_{it}\|^4|\mathcal{E}_{t-1}] \lesssim_{\mathbb{P}} H^{-2}$, uniformly over i and t. On the intersection of the same-cell event with $\mathcal{A}_{it,\mathtt{FS}}$,

$$\|\Delta^{\mathrm{in}}_{itP}\| \lesssim J(\|\widehat{\beta}_{it} - \breve{\beta}_{it}\| + \mathsf{B}_{\mathtt{FS}})\mathbb{1}(\beta_{it} \in P)$$

and the complementary contribution is absorbed by (SA-II.4.3) as above. Using the conditional independence of the innovations in $\widehat{\beta}_{it} - \breve{\beta}_{it}$ from β_{it} given $\mathcal{E}_{t-1}$, and the regular-event bound above, we obtain

$$\mathbb{E}\Big[\|\Delta^{\mathrm{in}}_{itP}\|^2\|\widehat{\beta}_{it} - \breve{\beta}_{it}\|^2\mathbb{1}(\mathcal{A}_{it,\mathtt{FS}}) \mid \mathcal{D}_{t-1}\Big] \lesssim_{\mathbb{P}} J^2\mathbb{P}(\beta_{it} \in P|\mathcal{D}_{t-1})\big(H^{-2} + \mathsf{B}^2_{\mathtt{FS}}H^{-1}\big) \lesssim_{\mathbb{P}} \frac{J^{2-d}}{H^2}.$$

On $\mathcal{A}^c_{it,\mathtt{FS}}$, boundedness of the localized basis, the conditional second-moment bound for $\widehat{\beta}_{it} - \breve{\beta}_{it}$, and the first cell-probability bound above give

$$\mathbb{E}\Big[\|\Delta^{\mathrm{in}}_{itP}\|^2\|\widehat{\beta}_{it} - \breve{\beta}_{it}\|^2\mathbb{1}(\mathcal{A}^c_{it,\mathtt{FS}})\Big|\mathcal{D}_{t-1}\Big] \lesssim_{\mathbb{P}} \frac{1}{H}\min\big\{\mathbb{P}(\mathcal{A}^c_{it,\mathtt{FS}} \mid \mathcal{D}_{t-1}), J^{-d}\big\} \lesssim_{\mathbb{P}} \frac{1}{HJ^{d-1}\max\{J, \sqrt{H}\}}.$$

Therefore, for the same-cell centered component,

$$\mathbb{V}\mathrm{ar}\left[\Delta^{\mathrm{in}}_{itP}[\nabla\mu_t(\beta_{it})]^\top\frac{1}{H}\sum_{s=1}^{H} S\widehat{A}(t)^{-1}X_s(t)K\Big(\frac{s}{H}\Big)\varepsilon_{i(t-s)}\,\Big|\,\mathcal{D}_{t-1}\right]$$

$$\leq \mathbb{E}\left[\|\Delta^{\text{in}}_{itP}\|^2 \left\| \frac{1}{H}\sum_{s=1}^{H} S\widehat{A}(t)^{-1}X_s(t)K\Big(\frac{s}{H}\Big)\varepsilon_{i(t-s)}\right\|^2 \,\middle|\, \mathcal{D}_{t-1}\right] \lesssim_{\mathbb{P}} \frac{1}{HJ^{d-1}\max\{J,\sqrt{H}\}} + \frac{J^{2-d}}{H^2}.$$

Combining the boundary and interior pieces gives the displayed localized-basis variance bound. □

SA-II.4.6 Proof of Lemma SA-II.4.2

We first record the direct bound for $\mathscr{L}_{\mathtt{B}}(\beta)$, and then focus on the expansion of $\mathscr{L}_{\mathtt{V}}(\beta)$.

By Lemma SA-II.3.3, with probability approaching one, $\|\widehat{Q}_t^{-1}\| \lesssim J^d$ and $\|Q_t^{-1}\| \lesssim J^d$ uniformly over $t = H+1, \cdots, T$. Since the gradients are uniformly bounded under Assumption SA-DGP, and since the local-polynomial weights have bounded absolute sum under the Gram eigenvalue bound,

$$|\mathscr{L}_{\mathtt{B}}(\beta)| \lesssim_{\mathbb{P}} \max_{H+1\leq t\leq T}\max_{i\in[n_t]} \|\breve{\beta}_{it} - \beta_{it}\| \lesssim_{\mathbb{P}} \mathsf{B}_{\mathtt{FS}}.$$

Step 1: Bounds for $\mathscr{A}(\beta)$ and $\mathscr{A}(0)$. The term $\mathscr{A}(\beta)$ is the re-indexed version of the reference-basis expression with the reference Gram matrix Q_t. It can be written as

$$\mathscr{A}(\beta) = \frac{1}{T-H}\sum_{t=H+1}^{T} \widehat{p}_t(\beta)^\top Q_t^{-1}\left[\frac{1}{n_t}\sum_{i=1}^{n_t} p_t(\beta_{it})[\nabla\mu_t(\beta_{it})]^\top \frac{1}{H}\sum_{s=1}^{H} K(s/H)SA(t)^{-1}X_s(t)\varepsilon_{i(t-s)}\right]$$

The summands in the bracketed double sum are centered and independent conditional on $\mathcal{E}_{t-1}$, are supported on reference cells, and have conditional variance of order $J^{-d}/(nH)$ per time period. The evaluation block and inverse Gram matrix only select and weight elements of the same rectangle-polynomial class; the operator norm contributes at most the factor J^d. Since this class is VC-type with bounded envelope, a standard maximal inequality gives

$$\sup_{\beta\in\mathcal{B}^\circ} |\mathscr{A}(\beta)| \lesssim_{\mathbb{P}} \left(\frac{J^d\log(nT)}{nH}\right)^{1/2}.$$

For the special case $\beta = 0$, the partition convention and boundary tie-breaking assign 0 to a unique cell at each date; write its label as $j_0(u)$ at date u. For compactness, write $P_{0,u} = P_{j_0(u),u}$, $p_{0,u} = p_{j_0(u),u}$, and $Q_{0,u} = Q_{j_0(u),u}$. By the split-at-zero construction and the midpoint and side-length normalization in Section SA-II.3, $\widehat{p}_{j_0(u),u}(0) = p_{0,u}(0)$ for every $\mathfrak{q} \geq 0$. Let $\mathcal{S}_t = \{s \in [H] : H+1 \leq t+s \leq T\}$. Thus,

$$\begin{aligned}\mathscr{A}(0) = \frac{1}{T-H}\sum_{t=1}^{T}\sum_{i=1}^{n_t}\frac{1}{H}\sum_{s\in\mathcal{S}_t}\frac{1}{n_{t+s}} & p_{0,t+s}(0)^\top Q_{0,t+s}^{-1} \\ & \times p_{0,t+s}(\beta_{i(t+s)})[\nabla\mu_{t+s}(\beta_{i(t+s)})]^\top SA(t+s)^{-1}K\Big(\frac{s}{H}\Big)X_s(t+s)\mathbb{1}(i\leq n_{t+s})\varsigma_{it}\eta_{it}.\end{aligned}$$

The display is obtained by re-indexing the rolling-window shocks: the primitive innovation η_{it} enters future first-step estimates at dates $t+s$. After the zero-cell identity above, all coefficients in $\mathscr{A}(0)$ are $\mathcal{E}_\infty$-measurable; conditional on $\mathcal{E}_\infty$, the primitive innovations are independent over (i,t) and have mean zero. For the variance display, write $\Xi_s(u) = SA(u)^{-1}K\big(s/H\big)X_s(u)$. Jensen's inequality over the rolling-window

average and the bounds $\|Q_{0,u}^{-1}\| \lesssim_{\mathbb{P}} J^d$ and $\|p_{0,u}(0)\|\|p_{0,u}(\beta_{iu})\| \lesssim \mathbb{1}(\beta_{iu} \in P_{0,u})$ give

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[\mathscr{A}(0)|\mathcal{E}_\infty] &\lesssim \frac{1}{T^2}\sum_{t=1}^{T}\sum_{i=1}^{n_t}\left(\frac{1}{H}\sum_{s\in\mathcal{S}_t}\frac{1}{n_{t+s}}p_{0,t+s}(0)^\top Q_{0,t+s}^{-1}p_{0,t+s}(\beta_{i(t+s)})[\nabla\mu_{t+s}(\beta_{i(t+s)})]^\top\right.\\
&\qquad\left.\times\,\Xi_s(t+s)\mathbb{1}(i\le n_{t+s})\right)^2\\
&\lesssim \frac{J^{2d}}{T^2}\sum_{t=1}^{T}\sum_{i=1}^{n_t}\frac{1}{H}\sum_{s\in\mathcal{S}_t}\frac{1}{n_{t+s}^2}\mathbb{1}(\beta_{i(t+s)}\in P_{0,t+s})\mathbb{1}(i\le n_{t+s})\|\Xi_s(t+s)\|^2\\
&\lesssim \frac{J^{2d}}{T^2}\sum_{u=H+1}^{T}\frac{1}{n_u^2}\sum_{i=1}^{n_u}\mathbb{1}(\beta_{iu}\in P_{0,u})\frac{1}{H}\sum_{s=1}^{H}\|\Xi_s(u)\|^2\\
&\lesssim_{\mathbb{P}} \frac{J^{2d}}{nT}\frac{1}{J^d}.
\end{aligned}$$

The last line uses the true-exposure reference second-moment bound from Lemma SA-II.3.3 and the local-polynomial Gram bound. Then, $\mathscr{A}(0) \lesssim_{\mathbb{P}} \sqrt{J^d/(nT)}$.

Step 2: First-step Gram replacement and inverse-Gram perturbation. Let $\mathscr{A}^\dagger(\beta)$ denote the same re-indexed expression as $\mathscr{A}(\beta)$, but with $A(t+s)^{-1}$ replaced by $\widehat{A}(t+s)^{-1}$. The mean-square version of Lemma SA-II.2.2, the resolvent identity, and the bound for $\mathscr{A}(\beta)$ in Step 1 give

$$\sup_{\beta\in\mathcal{B}^\circ}|\mathscr{A}^\dagger(\beta)-\mathscr{A}(\beta)| \lesssim_{\mathbb{P}} \left(\frac{T^{2/q}}{H}+\frac{1}{\sqrt{H}}\right)\sqrt{\frac{J^d\log(nT)}{nH}} = o_{\mathbb{P}}\left(\frac{1}{H}\right),$$

where the last equality follows from (SA-II.4.1) and $J^d\log(nT)/n = o(1)$. Thus this replacement error is absorbed by the H^{-1} term in the residual bound.

Next, replace the generated inverse Gram $\widehat{Q}_t^{-1}$ in $\mathscr{L}_{\mathsf{V}}(\beta)$ by the reference inverse Gram Q_t^{-1}. By Lemma SA-II.3.3, $\max_{H+1\le t\le T}\|\widehat{Q}_t^{-1}-Q_t^{-1}\| \lesssim_{\mathbb{P}} J^{2d}\mathsf{R}_{\mathtt{Q}}$.

The cost of this replacement is $\min\{J^d\mathsf{R}_{\mathtt{Q}},1\}\sqrt{J^d\log(nT)/(nH)}$. Indeed, if $J^d\mathsf{R}_{\mathtt{Q}} \lesssim 1$, the resolvent identity gives an additional relative factor $J^d\mathsf{R}_{\mathtt{Q}}$ compared to the centered reference-basis bound in Step 1; otherwise, the crude bound by the same centered reference-basis scale gives the truncation at one. In the remaining steps we work after this inverse-Gram replacement and add this perturbation cost at the end.

Step 3: Generated-regressor switch term. For a fixed admissible partition $\Delta = \{P_j : j \in [\bar{J}_t]\}$ equipped with its cell normalizations, let $p(\cdot;\Delta)$ denote the corresponding stacked local-polynomial basis. Consider the following two terms:

$$\begin{aligned}
\mathscr{T}_{\mathtt{G}}(\beta) &= \frac{1}{T-H}\sum_{t=H+1}^{T}\widehat{p}_t(\beta)^\top Q_t^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\left[p(\widehat{\beta}_{it};\widehat{\Delta}_t)-p(\beta_{it};\widehat{\Delta}_t)\right][\nabla\mu_t(\beta_{it})]^\top\left(\widehat{\beta}_{it}-\breve{\beta}_{it}\right),\\
\mathscr{T}_{\mathtt{P}}(\beta) &= \frac{1}{T-H}\sum_{t=H+1}^{T}\widehat{p}_t(\beta)^\top Q_t^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t}\left[p(\beta_{it};\widehat{\Delta}_t)-p(\beta_{it};\Delta_t)\right][\nabla\mu_t(\beta_{it})]^\top\left(\widehat{\beta}_{it}-\breve{\beta}_{it}\right).
\end{aligned}$$

Consider the generated-regressor switch term $\mathscr{T}_{\mathtt{G}}(\beta)$. For any $\beta \in \mathcal{B}^\circ$ (corresponding to an index set $\mathcal{J}_t$ that excludes bins close to the boundary), let $\mathcal{P}^\circ$ denote the collection of quasi-uniform rectangles $P =$

$[\ell, u] \subset \mathcal{B}$ with $\mathrm{diam}(P) \asymp J^{-1}$ and $\min_{k\in[d]} \min\{|\ell_k - \underline{\beta}_k|, |\bar{\beta}_k - u_k|\} \gg \log(nT)/\sqrt{H}$.

$$
\begin{aligned}
|\mathscr{T}_{\mathsf{G}}(\beta)| &\lesssim_{\mathbb{P}} \max_{H+1\le t\le T} \max_{j\in\mathcal{J}_t} \Bigg\| \frac{J^d}{n}\frac{1}{H}\sum_{i=1}^{n_t}\sum_{s=1}^{H} \big[p_{\mathsf{q}}(\widehat{\beta}_{it}; \widehat{P}_{jt}) - p_{\mathsf{q}}(\beta_{it}; \widehat{P}_{jt})\big] \\
&\qquad \times [\nabla\mu_t(\beta_{it})]^\top S\widehat{A}(t)^{-1} X_s(t) K\Big(\frac{s}{H}\Big)\varepsilon_{i(t-s)} \Bigg\| \\
&\lesssim_{\mathbb{P}} \max_{H+1\le t\le T} \sup_{P\in\mathcal{P}^\circ} \Bigg\| \frac{J^d}{nH}\sum_{i=1}^{n_t}\sum_{s=1}^{H} \big[p(\widehat{\beta}_{it}; P) - p(\beta_{it}; P)\big][\nabla\mu_t(\beta_{it})]^\top \\
&\qquad \times S\widehat{A}(t)^{-1} X_s(t) K\Big(\frac{s}{H}\Big)\varepsilon_{i(t-s)} \Bigg\| \\
&\lesssim \sup_{t,P} \Bigg\| \frac{J^d}{nH}\sum_{i=1}^{n_t}\sum_{s=1}^{H} \mathbb{E}[g_{its}|\mathcal{D}_{t-1}] \Bigg\| + \sup_{t,P} \Bigg\| \frac{J^d}{nH}\sum_{i=1}^{n_t}\sum_{s=1}^{H} \Big(g_{its} - \mathbb{E}[g_{its}|\mathcal{D}_{t-1}]\Big) \Bigg\|
\end{aligned}
$$

where

$$
g_{its} = \big[p_{\mathsf{q}}(\widehat{\beta}_{it}; P) - p_{\mathsf{q}}(\beta_{it}; P)\big][\nabla\mu_t(\beta_{it})]^\top S\widehat{A}(t)^{-1} X_s(t) K\Big(\frac{s}{H}\Big)\varepsilon_{i(t-s)},
$$

and the shorthand $\sup_{t,P}$ in the preceding display ranges over $H+1 \le t \le T$ and $P \in \mathcal{P}^\circ$.

By Lemma SA-II.4.1(a),

$$
\max_{H+1\le t\le T} \sup_{P\in\mathcal{P}^\circ} \left| \frac{J^d}{nH}\sum_{i=1}^{n_t}\sum_{s=1}^{H} \mathbb{E}[g_{its}|\mathcal{D}_{t-1}] \right| \lesssim_{\mathbb{P}} \frac{1}{H} + \mathsf{B}_{\mathsf{FS}}.
$$

On the other hand, conditional on $\mathcal{D}_{t-1}$, the second term is a sum of terms independent over i with zero means. By Lemma SA-II.4.1(b),

$$
\mathbb{V}\mathrm{ar}\left[\frac{J^d}{nH}\sum_{i=1}^{n_t}\sum_{s=1}^{H} g_{its} \,\middle|\, \mathcal{D}_{t-1} \right] \lesssim_{\mathbb{P}} \frac{J^{2d}}{n}\left[\frac{\log^d(nT)}{HJ^{d-1}\max\{J, \sqrt{H}\}} + \left(\frac{1}{HJ^d} + \frac{\mathsf{B}_{\mathsf{FS}}}{J^d}\right)^2 + \mathbb{1}(\mathsf{q}\ge 1)\frac{J^{2-d}}{H^2} \right].
$$

Define

$$
\begin{aligned}
g_i(t, P) &= \frac{1}{H}\sum_{s=1}^{H} g_{its}, \\
\mathcal{S} &= \{g(\cdot; t, P) : P \in \mathcal{P}^\circ, H+1 \le t \le T\}.
\end{aligned}
$$

Here $g(\cdot; t, P)$ is the function that evaluates to $g_i(t, P)$ on the full input history for asset i, including the common quantities and all observations entering the rolling window. The envelope satisfies

$$
\begin{aligned}
\sup_{t,P} \|g_i(t, P)\| &\lesssim \sup_{H+1\le t\le T} \left\| \frac{1}{H}\sum_{s=1}^{H} X_s(t) K\Big(\frac{s}{H}\Big)\varepsilon_{i(t-s)} \right\|, \\
\max_{i\in[n]} \sup_{t,P} |g_i(t, P)| &\lesssim_{\mathbb{P}} \frac{\sqrt{\log(nT)}}{\sqrt{H}} + \frac{(nT^2)^{1/q}\log(nT)}{H}.
\end{aligned}
$$

The class indexed by P is VC-type with a finite exponent because it is generated by rectangular cell indicators

and bounded polynomial basis functions on normalized cells. To transfer this covering bound to the rolling-window summands, fix t and a scalar basis coordinate $r \in [K_{\mathfrak{q}}]$. Let

$$b_{P,r}(x,z) = p_{\mathfrak{q},r}(x,P) - p_{\mathfrak{q},r}(z,P),$$
$$M_{it} = [\nabla \mu_t(\beta_{it})]^\top \frac{1}{H} \sum_{s=1}^{H} S\widehat{A}(t)^{-1} X_s(t) K(s/H) \varepsilon_{i(t-s)},$$

where $p_{\mathfrak{q},r}$ is the rth coordinate of $p_{\mathfrak{q}}$. The corresponding coordinate of $g_i(t,P)$ is exactly $M_{it} b_{P,r}(\widehat{\beta}_{it}, \beta_{it})$, and M_{it} does not depend on P. On the full input-history space, denote the functions corresponding to M_{it} and $b_{P,r}(\widehat{\beta}_{it}, \beta_{it})$ by M and $B_{P,r}$. For any probability measure $\mathbb{Q}$ with $0 < \mathbb{E}_{\mathbb{Q}}[M^2] < \infty$, define the probability measure

$$\mathrm{d}\mathbb{Q}_M = \frac{M^2}{\mathbb{E}_{\mathbb{Q}}[M^2]} \, \mathrm{d}\mathbb{Q}.$$

Then, for any $P, \widetilde{P} \in \mathcal{P}^\circ$,

$$\|M(B_{P,r} - B_{\widetilde{P},r})\|_{L_2(\mathbb{Q})} = \|M\|_{L_2(\mathbb{Q})} \|B_{P,r} - B_{\widetilde{P},r}\|_{L_2(\mathbb{Q}_M)}.$$

The bounded basis-difference class $\{B_{P,r} : P \in \mathcal{P}^\circ\}$ has a polynomial covering bound uniformly over probability measures, with an exponent depending only on the fixed d and $\mathfrak{q}$. The identity transfers this bound to $\{MB_{P,r} : P \in \mathcal{P}^\circ\}$ with envelope $C|M|$, where C bounds the basis differences. If $\mathbb{E}_{\mathbb{Q}}[M^2] = 0$, the multiplied class is zero $\mathbb{Q}$-almost surely. Applying the argument to the fixed number of basis coordinates establishes the asserted VC-type bound.

The other index t can only take a finite number $(T - H)$ values. Taking the union over these dates multiplies the covering bound by at most $T - H$. The concentration argument is applied for each fixed t conditional on $\mathcal{D}_{t-1}$, using the conditional independence across i stated above. Given these facts, by the maximal inequality for i.n.i.d. empirical processes in Cattaneo et al. (2024, Lemma SA24),

$$|\mathscr{T}_{\mathtt{G}}(\beta)| \lesssim_{\mathbb{P}} \left(\frac{1}{H} + \mathsf{B}_{\mathtt{FS}}\right) + \sqrt{\frac{J^{2d} \log^{d+1}(nT)}{nHJ^{d-1} \max\{J, \sqrt{H}\}}} + \mathbb{1}(\mathfrak{q} \geq 1) \sqrt{\frac{J^{d+2} \log(nT)}{nH^2}},$$

under the restriction

$$\frac{J^{d-1} \max\{J, \sqrt{H}\} \log^2(nT)}{\min\{n^{1-2/q} T^{1-4/q} h, n\}} \lesssim 1.$$

This restriction ensures that

$$\frac{\log^{3/2}(nT)}{\sqrt{n}} + \frac{(nT^2)^{1/q} \log^2(nT)}{\sqrt{nH}} \lesssim \left(J^{d-1} \max\{J, \sqrt{H}\}\right)^{-1/2} \log^{(d+1)/2}(nT).$$

Step 4: Partition switch term. For $\mathscr{T}_{\mathtt{P}}(\beta)$, by Lemma SA-II.3.2, for all $\beta \in \mathcal{B}^\circ$,

$$|\mathscr{T}_{\mathtt{P}}(\beta)| \lesssim_{\mathbb{P}} \max_{H+1 \leq t \leq T} \max_{j \in \mathcal{J}_t} \left\| \frac{J^d}{n} \frac{1}{H} \sum_{i=1}^{n_t} \sum_{s=1}^{H} \left[p_{\mathfrak{q}}(\beta_{it}; \widehat{P}_{jt}) - p_{\mathfrak{q}}(\beta_{it}; P_{jt}) \right] \right.$$

$$\times [\nabla \mu_t(\beta_{it})]^\top S\widehat{A}(t)^{-1} X_s(t) K\Big(\frac{s}{H}\Big)\varepsilon_{i(t-s)}\Big\|$$

$$\lesssim \max_{H+1\le t\le T}\max_{j\in\mathcal{J}_t}\sup_{\widetilde{P}\in\mathcal{P}_{jt}}\Big\|\frac{J^d}{n}\frac{1}{H}\sum_{i=1}^{n_t}\sum_{s=1}^{H}\big[p_{\mathfrak{q}}(\beta_{it};\widetilde{P}_{jt})-p_{\mathfrak{q}}(\beta_{it};P_{jt})\big][\nabla\mu_t(\beta_{it})]^\top$$
$$\times S\widehat{A}(t)^{-1}X_s(t)K\Big(\frac{s}{H}\Big)\varepsilon_{i(t-s)}\Big\|$$

where, by Lemma SA-II.3.2, we take the supremum over $\mathcal{P}_{jt}=\{[a,b]:\max\{\|a-\ell\|,\|b-u\|\}\le\sqrt{\log T/n}\}$ for a generic cell $P_{jt}=[\ell,u]$. Again, conditional on $\mathcal{E}_{t-1}$, the summands are centered, independent over i, and have zero cross terms across different lags by the innovation orthogonality. The symmetric difference between $\widetilde{P}$ and P_{jt} has $\mathcal{D}_{t-1}$-conditional probability at most of order $J^{-(d-1)}\sqrt{\log T/n}$, uniformly over $\widetilde{P}\in\mathcal{P}_{jt}$. A VC count bound conditional on $\mathcal{D}_{t-1}$ and a union over dates give

$$\max_{H+1\le t\le T}\max_{j\in\mathcal{J}_t}\sup_{\widetilde{P}\in\mathcal{P}_{jt}}\sum_{i=1}^{n_t}\mathbb{1}\big(\beta_{it}\in\widetilde{P}\triangle P_{jt}\big)\lesssim_{\mathbb{P}} nJ^{-(d-1)}\sqrt{\frac{\log T}{n}}+\log(nT).$$

The same argument and $J^d\log(nT)/n=o(1)$ bound the true-beta count in each reference cell by $O_{\mathbb{P}}(n/J^d)$, uniformly over dates and cells. Conditional on $\mathcal{E}_{t-1}$, the realized counts are fixed; the strip and reference-cell count bounds control the boundary and coordinate-normalization contributions, respectively. Since $H^{-1}\sum_{s=1}^{H}\|S\widehat{A}(t)^{-1}X_s(t)K\big(\frac{s}{H}\big)\|^2\lesssim_{\mathbb{P}}1$, the comparison of boundaries and the coordinate-normalization component between two cells gives the following variance bound:

$$\frac{J^{2d}}{n^2H^2}\sum_{i=1}^{n_t}\sum_{s=1}^{H}\|\big[p_{\mathfrak{q}}(\beta_{it};\widetilde{P}_{jt})-p_{\mathfrak{q}}(\beta_{it};P_{jt})\big]\|^2\Big\|S\widehat{A}(t)^{-1}X_s(t)K\Big(\frac{s}{H}\Big)\Big\|^2$$
$$\lesssim_{\mathbb{P}}\frac{J^{d+1}\sqrt{\log T/n}}{nH}+\frac{J^{2d}\log(nT)}{n^2H}+\mathbb{1}(\mathfrak{q}\ge1)\frac{J^{d+2}\log T}{n^2H}.$$

In addition, the class of local polynomial basis functions is of VC-type with a finite exponent, and the multiplier argument from Step 3 applies because the rolling-score multiplier does not depend on the perturbed cell. Using the kernel-weighted rolling-score envelope, datewise truncation, and bounded empirical-process concentration together with Cattaneo et al. (2024, Lemma SA24), then taking a union over dates and retaining the envelope terms, we conclude that

$$|\mathscr{T}_{\mathfrak{p}}(\beta)|\lesssim_{\mathbb{P}}\sqrt{\frac{J^{d+1}\sqrt{\log T/n}\log(nT)}{nH}}+\mathbb{1}(\mathfrak{q}\ge1)\sqrt{\frac{J^{d+2}\log T\log(nT)}{n^2H}}$$
$$+\frac{J^d}{n}\left[\frac{\log^{3/2}(nT)}{\sqrt{H}}+\frac{(nT^2)^{1/q}\log^2(nT)}{H}\right],$$

and this bound holds uniformly over $\beta\in\mathcal{B}^\circ$. The additional boundary-count variance term contributes at most $J^d\log(nT)/(n\sqrt{H})$ after concentration and is covered by the displayed envelope terms.

Step 5: Combine the bounds. Multiplying the final inequality in Step 3 by $J^d/\sqrt{nH}$ gives

$$\frac{J^d}{n}\left[\frac{\log^{3/2}(nT)}{\sqrt{H}}+\frac{(nT^2)^{1/q}\log^2(nT)}{H}\right]\lesssim\sqrt{\frac{J^{2d}\log^{d+1}(nT)}{nHJ^{d-1}\max\{J,\sqrt{H}\}}}.$$

Thus the extra terms retained in Step 4 are absorbed by the existing generated-regressor switch component of $\mathsf{R}_{\mathtt{LA}}$. Combining the first-step Gram replacement and inverse-Gram perturbation from Step 2, the generated-regressor switch bound from Step 3, and the partition-switch bound from Step 4 proves

$$\sup_{\beta\in\mathcal{B}^\circ}|\mathscr{L}_{\mathtt{V}}(\beta)-\mathscr{A}(\beta)|\lesssim_{\mathbb{P}}\mathsf{R}_{\mathtt{LA}}.$$

□

SA-II.5 Convergence Rate

We now combine the first-step and partitioning results from Sections SA-II.2 and SA-II.3 with the general results from Part I. The first result is the direct specialization of the general beta-sorted portfolio convergence theorem to the classical construction. The second uses the stochastic linearization and is sharper at each fixed evaluation point.

SA-II.5.1 Results

Corollary SA-II.5.1 (General Beta-Sorted Portfolios)**.**
Suppose that Assumptions SA-DGP, SA-FS, SA-PA, SA-SEP, and SA-FSB hold and that $s_\mu\geq 1$. Assume further that

$$\frac{T^{(3+d)/q}}{H}+\sqrt{\frac{\log T}{H}}=o(1),\qquad \mathsf{R}_{\mathtt{FS}}=o(1),$$
$$\frac{J^d\log(nT)}{n}=o(1),\qquad \frac{(nT)^{2/q}J^d\log(nT)}{n}\lesssim 1.$$

Then, for each fixed $\beta\in\mathcal{B}^\circ$ and any $\ell_n\to\infty$,

$$\left|\widehat{\mu}(\beta)-\bar{\mu}_T(\beta;H)\right|\lesssim_{\mathbb{P}}\ell_n\left[\mathsf{R}_{\mathtt{FS}}+J^{-\min\{s_\mu,\mathtt{q}+1\}}+\sqrt{\frac{J^d\log(nT)}{n}}+\frac{\|\beta\|}{\sqrt{T}}\right].$$

The next corollary uses the stochastic linearization to recover the time-series averaging in the leading sampling terms. The generated-regressor correction $\mathscr{A}(\beta)$ is the only additional term extracted from $\mathscr{L}_{\mathtt{V}}(\beta)$; the nonconstant-basis movement that appears when $\mathtt{q}\geq 1$ is controlled uniformly and absorbed by $\mathsf{R}_{\mathtt{LA}}$ in Lemma SA-II.4.2. To state the leading first-step correction compactly, define

$$\mathsf{R}_{\mathtt{A}}(\beta)=\begin{cases}\sqrt{\dfrac{J^d\log(nT)}{nH}} & \text{if } \beta\neq 0\\ \sqrt{\dfrac{J^d}{nT}} & \text{if } \beta=0\end{cases}.$$

This rate collects the pointwise bound for the first-step correction $\mathscr{A}(\beta)$ obtained in Lemma SA-II.4.2. The Gram-perturbation and basis-switching pieces enter through the explicit $\mathsf{R}_{\mathtt{Q}}$ and $\mathsf{R}_{\mathtt{P}}$ terms displayed below. Under the stronger idiosyncratic-scale localization condition $J\mathsf{R}_{\mathtt{FS}}=o(1)$, these terms are absorbed by $\sqrt{J^d/(nT)}$; fixed nonzero-β inference only needs them to be negligible at the factor scale.

Corollary SA-II.5.2 (Classical Pointwise Convergence Rate)**.** *Suppose that Assumptions SA-DGP, SA-FS, SA-PA, SA-SEP, and SA-FSB hold and that $s_\mu \geq 2$. Assume further that*

$$\frac{T^{2/q}}{H} + \sqrt{\frac{\log T}{H}} = o(1), \qquad \mathsf{R}_{\mathtt{FS}} = o(1),$$
$$\frac{J^d \log(nT)}{n} = o(1), \qquad \frac{J^2 \log T}{n} = o(1),$$
$$\frac{J^{d-1} \max\{J, \sqrt{H}\} \log^2(nT)}{\min\{n^{1-2/q} T^{1-4/q} h, n\}} \lesssim 1, \qquad \frac{\max\{J^{2/(q-2)} T^{2/q}, T^{2(3+d)/q}\}}{H} = o(1).$$

Then, for each fixed $\beta \in \mathcal{B}^\circ$,

$$\begin{aligned}\left|\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H)\right| \lesssim_{\mathbb{P}} \frac{\|\beta\|}{\sqrt{T}} + \sqrt{\frac{J^d}{nT}} + \sqrt{\frac{J^{3d} \mathsf{R}_{\mathtt{Q}}^2}{nT}} \\ + \sqrt{\frac{J^{2d} \mathsf{R}_{\mathtt{P}}}{nT}} + \mathsf{R}_{\mathtt{A}}(\beta) + \mathsf{R}_{\mathtt{LA}} + J^{-\min\{s_\mu, \mathtt{q}+1\}} + \frac{\mathsf{R}_{\mathtt{FS}}}{\sqrt{T}} + \mathsf{R}_{\mathtt{FS}}^2 \\ + \mathbb{1}(\mathtt{q} = 0) \frac{1}{J\sqrt{T}}.\end{aligned}$$

SA-II.5.2 Proof of Corollary SA-II.5.1

By the first displayed rate condition, Lemma SA-II.2.3 gives

$$\widehat{\mathsf{R}}_{\mathtt{FS}} \lesssim_{\mathbb{P}} \mathsf{R}_{\mathtt{FS}}.$$

Because $\mathsf{R}_{\mathtt{FS}} = o(1)$ and $\mathcal{B}^+$ is a fixed neighborhood of $\mathcal{B}$, the first-step estimates lie in $\mathcal{B}^+$ uniformly over i and t, w.p.a.1.

Next consider the local partition conditions along the cells in $\mathcal{J}_t$. The first-step bias condition, the Gram condition for the rolling-window design, and $J^d \log(nT)/n = o(1)$ imply the assumptions of Lemmas SA-II.3.1, SA-II.3.2, and SA-II.3.3. These results imply that the estimated partition covers $\mathcal{B}^\circ$ and all raw generated regressors, and that the reference and estimated cells indexed by $\mathcal{J}_t$ have side lengths of order J^{-1}. Consequently,

$$\widehat{\mathsf{R}}_{\mathtt{D}}^\circ \lesssim_{\mathbb{P}} J^{-1}.$$

The same lemma gives the required counts only over the relevant interior cells:

$$\frac{n}{J^d} \lesssim_{\mathbb{P}} \widehat{n}_{\min}^\circ \leq \widehat{n}_{\max}^\circ \lesssim_{\mathbb{P}} \frac{n}{J^d}.$$

Consequently,

$$\frac{(nT)^{2/q} \log(nT)}{\widehat{n}_{\min}^\circ} \lesssim_{\mathbb{P}} \frac{(nT)^{2/q} J^d \log(nT)}{n} \lesssim 1.$$

The scaled Gram conclusion in Lemma SA-II.3.3 gives the required eigenvalue bounds uniformly over $j \in \mathcal{J}_t$. The pointwise proof of Theorem SA-I.2.1 is therefore applicable cellwise to the block containing each fixed $\beta \in \mathcal{B}^\circ$; its argument never uses an unselected tail block. Applying that local argument and substituting $\widehat{\mathsf{R}}_{\mathtt{FS}} \lesssim_{\mathbb{P}} \mathsf{R}_{\mathtt{FS}}$, $\widehat{\mathsf{R}}_{\mathtt{D}}^\circ \lesssim_{\mathbb{P}} J^{-1}$, and the displayed cell-count chain gives the displayed pointwise bound. □

SA-II.5.3 Proof of Corollary SA-II.5.2

By the first displayed rate condition, Lemma SA-II.2.3 gives

$$\widehat{\mathsf{R}}_{\mathtt{FS}} \lesssim_{\mathbb{P}} \mathsf{R}_{\mathtt{FS}}.$$

Since $J^d \log(nT)/n = o(1)$ implies $J \log(nT)/n = o(1)$, the displayed rate conditions imply the conditions of Lemma SA-II.3.3. Thus,

$$\widehat{\mathsf{R}}_{\mathtt{Q}} \lesssim_{\mathbb{P}} \mathsf{R}_{\mathtt{Q}}, \qquad \widehat{\mathsf{R}}_{\mathtt{P}} \lesssim_{\mathbb{P}} \mathsf{R}_{\mathtt{P}}, \qquad \widehat{\mathsf{R}}^{\circ}_{\mathtt{D}} \lesssim_{\mathbb{P}} J^{-1},$$

and the generated Gram bounds hold. The parts of the proof of Lemma SA-I.3.1 that keep the generated design fixed give

$$|\mathscr{G}(\beta)| \lesssim_{\mathbb{P}} \frac{\|\beta\|}{\sqrt{T}}, \qquad |\mathscr{B}(\beta)| \lesssim_{\mathbb{P}} J^{-\min\{s_\mu, \mathfrak{q}+1\}}, \qquad |\mathscr{R}_3(\beta)| \lesssim_{\mathbb{P}} \mathbb{1}(\mathfrak{q}=0)\frac{1}{J\sqrt{T}},$$
$$|\mathscr{R}_4(\beta)| \lesssim_{\mathbb{P}} \frac{\mathsf{R}_{\mathtt{FS}}}{\sqrt{T}}, \qquad |\mathscr{R}_5(\beta)| \lesssim_{\mathbb{P}} \mathsf{R}_{\mathtt{FS}}^2.$$

Lemma SA-II.4.2 gives

$$|\mathscr{L}_{\mathtt{B}}(\beta)| \lesssim_{\mathbb{P}} \mathsf{B}_{\mathtt{FS}}, \qquad |\mathscr{A}(\beta)| \lesssim_{\mathbb{P}} \mathsf{R}_{\mathtt{A}}(\beta),$$

and the $\mathscr{L}_{\mathtt{B}}(\beta)$ bound is absorbed by $\mathsf{R}_{\mathtt{LA}}$ because $\mathsf{R}_{\mathtt{LA}}$ contains $\mathsf{B}_{\mathtt{FS}}$. The design-comparison argument in Lemma SA-I.3.1, using the generated and reference Gram bounds verified above, yields

$$|\mathscr{E}(\beta)| \lesssim_{\mathbb{P}} \sqrt{\frac{J^d}{nT}}, \qquad |\mathscr{R}_1(\beta)| \lesssim_{\mathbb{P}} \sqrt{\frac{J^{3d}\mathsf{R}_{\mathtt{Q}}^2}{nT}}, \qquad |\mathscr{R}_2(\beta)| \lesssim_{\mathbb{P}} \sqrt{\frac{J^{2d}\mathsf{R}_{\mathtt{P}}}{nT}}.$$

The remaining generated-regressor residual is exactly the linear-approximation error controlled in Lemma SA-II.4.2: $|\mathscr{L}_{\mathtt{V}}(\beta) - \mathscr{A}(\beta)| \lesssim_{\mathbb{P}} \mathsf{R}_{\mathtt{LA}}$. Combining these displays with the algebraic decomposition

$$\begin{aligned}\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H) = \mathscr{G}(\beta) + \mathscr{E}(\beta) - \mathscr{L}(\beta) + \mathscr{B}(\beta) \\ + \mathscr{R}_1(\beta) + \mathscr{R}_2(\beta) + \mathscr{R}_3(\beta) + \mathscr{R}_4(\beta) - \mathscr{R}_5(\beta),\end{aligned}$$

and $\mathscr{L} = \mathscr{L}_{\mathtt{B}} + \mathscr{L}_{\mathtt{V}}$ gives the stated rate. □

SA-II.6 Pointwise-in-β Distributional Approximation

We now isolate the numerator for the distributional approximation:

$$\bar{\mathscr{S}}(\beta) = \mathscr{G}(\beta) + \mathscr{E}(\beta) - \mathscr{A}(\beta) = \frac{1}{T-H}\sum_{t=H+1}^{T} \beta^\top \widetilde{f}_t + \frac{1}{T-H}\sum_{t=1}^{T}\sum_{i=1}^{n_t} \varsigma_{it}\eta_{it}W_{it}(\beta), \tag{SA-II.6.1}$$

where the direct second-step and first-step primitive-shock weights are

$$W_{it}(\beta) = \frac{1}{n_t}\widehat{p}_t(\beta)^\top Q_t^{-1} p_t(\beta_{it})\mathbb{1}(t \geq H+1) - w_{it}(\beta).$$

Therefore, we obtain

$$\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H) = \bar{\mathscr{S}}(\beta) + \mathscr{R}(\beta),$$

where

$$\begin{aligned}\mathscr{R}(\beta) = \mathscr{B}(\beta) + \mathscr{R}_1(\beta) + \mathscr{R}_2(\beta) + \mathscr{R}_3(\beta) + \mathscr{R}_4(\beta) - \mathscr{R}_5(\beta) \\ - \mathscr{L}_{\mathtt{B}}(\beta) - \Big[\mathscr{L}_{\mathtt{V}}(\beta) - \mathscr{A}(\beta)\Big].\end{aligned} \tag{SA-II.6.2}$$

Thus (SA-II.6.2) lists all residual terms explicitly: the smoothing bias, the design-comparison and Taylor remainders, the first-step bias contribution, and the residual generated-regressor approximation error from Lemma SA-II.4.2. The centered nonconstant-basis term that appears when $\mathfrak{q} \geq 1$ is included in that last residual and controlled uniformly by $\mathsf{R}_{\mathtt{LA}}$.

For every evaluation point $\beta \in \mathcal{B}^\circ$, define the baseline normalization as the sum of the factor variance and the reference idiosyncratic scale:

$$\begin{aligned}\Sigma(\beta) &= \Sigma_f(\beta) + \Sigma_\varepsilon(\beta), \\ \Sigma_f(\beta) &= \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \beta^\top \mathbb{V}\mathrm{ar}(f_t \mid \mathcal{G}_{t-1})\beta, \\ \Sigma_\varepsilon(\beta) &= \frac{1}{(T-H)^2} \sum_{t=1}^{T} \sum_{i=1}^{n_t} \varsigma_{it}^2 \big(W_{it}(\beta)\big)^2,\end{aligned}$$

where $W_{it}(\beta)$ is the generated-regressor-adjusted reference weight. At an anchored evaluation point satisfying $\widehat{p}_t(\beta) = p_t(\beta)$, the weights $W_{it}(\beta)$ are measurable with respect to the innovation-free environment, and $\Sigma_\varepsilon(\beta)$ is the predictable conditional variance of the idiosyncratic leading term. At a generic nonanchored evaluation point, $\Sigma_\varepsilon(\beta)$ is used only as a reference quadratic scale. The proofs below control the corresponding stochastic term through the separate bounds for $\mathscr{E}(\beta)$ and $\mathscr{A}(\beta)$.

SA-II.6.1 Results

The first result isolates the Gaussian approximation for the leading term. The proof treats the origin and nonzero evaluation points separately, but the conclusion is stated as a single Kolmogorov-distance approximation. Away from the origin, the factor variance is nondegenerate under Assumption SA-DGP(iv); the only additional variance condition is needed at the origin, where the factor component vanishes.

Lemma SA-II.6.1 (Gaussian Approximation for the Leading Terms)**.** *Suppose that Assumptions SA-DGP, SA-FS, SA-PA, SA-SEP, and SA-FSB hold. Assume further that*

$$\begin{aligned}&\frac{T^{2/q}}{H} + \sqrt{\frac{\log T}{H}} = o(1), \\ &\frac{J^d \log(nT)}{n} = o(1), \\ &\frac{J^{d-1} \max\{J, \sqrt{H}\} \log^2(nT)}{\min\{n^{1-2/q} T^{1-4/q} h, n\}} \lesssim 1, \qquad \frac{\max\{J^{2/(q-2)} T^{2/q}, T^{2(3+d)/q}\}}{H} = o(1).\end{aligned}$$

If the fixed evaluation point is $\beta \neq 0$, assume additionally that

$$\frac{J^d \log(nT)}{nh} = o(1).$$

If the fixed evaluation point is $\beta = 0$, assume additionally that

$$\frac{nT}{J^d}\Sigma_\varepsilon(0) \gtrsim_{\mathbb{P}} 1.$$

Then, for each fixed $\beta \in \mathcal{B}^\circ$,

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\bar{\mathscr{I}}(\beta)}{\sqrt{\Sigma(\beta)}} \leq x\right) - \Phi(x)\right| \to 0.$$

The next theorem combines the Gaussian approximation for the leading term with the pointwise stochastic linearization error. We state the conclusion directly in Kolmogorov distance, so the approximation can be read as the infeasible size control for a pointwise normal test. Let

$$\mathsf{R}_{\mathtt{CLT}} = J^{-\min\{s_\mu,\mathfrak{q}+1\}} + \sqrt{\frac{J^{3d}\mathsf{R}_{\mathfrak{Q}}^2}{nT}} + \sqrt{\frac{J^{2d}\mathsf{R}_{\mathfrak{P}}}{nT}} + \mathsf{R}_{\mathtt{LA}} + \frac{\mathsf{R}_{\mathtt{FS}}}{\sqrt{T}} + \mathsf{R}_{\mathtt{FS}}^2 + \mathbb{1}(\mathfrak{q}=0)\frac{1}{J\sqrt{T}}.$$

This is the non-leading pointwise remainder after removing the leading components $\mathscr{G}(\beta)$, $\mathscr{E}(\beta)$, and $\mathscr{A}(\beta)$.

Theorem SA-II.6.2 (Distributional Approximation for the Infeasible t-Statistic)**.** *Suppose that the conditions of Lemmas SA-II.4.2 and SA-II.6.1 hold. For fixed $\beta \in \mathcal{B}^\circ$, suppose further that*

$$\mathsf{R}_{\mathtt{FS}} = o(1), \qquad \frac{J^2 \log T}{n} = o(1), \qquad \frac{\mathsf{R}_{\mathtt{CLT}}}{\sqrt{\Sigma(\beta)}} = o_{\mathbb{P}}(1).$$

Then

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\widehat{\mu}(\beta) - \bar{\mu}_T(\beta;H)}{\sqrt{\Sigma(\beta)}} \leq x\right) - \Phi(x)\right| \to 0.$$

SA-II.6.2 Discussion

(i) *Normalization.* The theorem imposes the high-level normalization condition $\mathsf{R}_{\mathtt{CLT}}/\sqrt{\Sigma(\beta)} = o_{\mathbb{P}}(1)$ directly. For each fixed $\beta \neq 0$, Assumption SA-DGP(iv) gives $\Sigma(\beta) \asymp_{\mathbb{P}} T^{-1}$, so it is enough that

$$\sqrt{T}\mathsf{R}_{\mathtt{CLT}} \to 0.$$

At the origin, Lemma SA-II.6.1 imposes $(nT/J^d)\Sigma_\varepsilon(0) \gtrsim_{\mathbb{P}} 1$, and the leading-weight upper bound used below gives $\Sigma(0) \asymp_{\mathbb{P}} J^d/(nT)$. Thus it is enough that

$$\sqrt{\frac{nT}{J^d}}\mathsf{R}_{\mathtt{CLT}} \to 0.$$

(ii) *Polynomial order.* Higher-order second-step polynomials reduce the smoothing-bias component in the CLT remainder: $J^{-\min\{s_\mu,\mathfrak{q}+1\}}$. Under the factor nondegeneracy condition in Assumption SA-

DGP(iv), the relevant bias restrictions are

$$\frac{T}{J^{2\min\{s_\mu,\mathfrak{q}+1\}}} \to 0 \qquad \text{when } \beta \neq 0,$$
$$\frac{nT}{J^{d+2\min\{s_\mu,\mathfrak{q}+1\}}} \to 0 \qquad \text{when } \beta = 0.$$

The stronger restriction at the origin appears because the factor component is absent and $\Sigma(0) \asymp_{\mathbb{P}} J^d/(nT)$. At the origin, the sufficient rate conditions stated here are incompatible with $\mathfrak{q} = 0$.

SA-II.6.3 Proof of Lemma SA-II.6.1

Decompose the leading term into its factor and idiosyncratic components:

$$\bar{\mathscr{S}}(\beta) = \bar{\mathscr{S}}_f(\beta) + \bar{\mathscr{S}}_\varepsilon(\beta),$$
$$\bar{\mathscr{S}}_f(\beta) = \frac{1}{T-H}\sum_{t=H+1}^{T} \beta^\top \widetilde{f}_t,$$
$$\bar{\mathscr{S}}_\varepsilon(\beta) = \frac{1}{T-H}\sum_{t=1}^{T}\sum_{i=1}^{n_t} \varsigma_{it}\eta_{it}W_{it}(\beta).$$

The displayed rate conditions imply the conditions of Lemma SA-II.3.3. Work on the high-probability event from that lemma on which the relevant cell counts, Gram matrices, and local-polynomial basis norms are uniformly controlled. By local support of the partition basis and the quasi-uniform cell-count bounds, define, only for this proof,

$$C_{it}(\beta) = \frac{1}{n_t}\widehat{p}_t(\beta)^\top Q_t^{-1} p_t(\beta_{it})\mathbb{1}(t \geq H+1).$$

Then

$$\sum_{i=1}^{n_t} |C_{it}(\beta)|^m \lesssim_{\mathbb{P}} \frac{J^{(m-1)d}}{n^{m-1}}, \qquad m = 2,3,$$

uniformly in t. For the rolling-window weights, Jensen's inequality and the same local-support argument give, for each t,

$$\sum_{i=1}^{n_t} |w_{it}(\beta)|^m \lesssim_{\mathbb{P}} \frac{J^{(m-1)d}}{n^{m-1}}(1 + \|f_t\|^m), \qquad m = 2,3.$$

It follows from $T^{-1}\sum_{t=1}^{T}(1+\|f_t\|^m) = O_{\mathbb{P}}(1)$ that

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n_t} |w_{it}(\beta)|^m \lesssim_{\mathbb{P}} \frac{J^{(m-1)d}}{n^{m-1}}, \qquad m = 2,3.$$

Since ς_{it} is uniformly bounded, it follows that, for each fixed $\beta \in \mathcal{B}^\circ$,

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n_t} |\varsigma_{it}W_{it}(\beta)|^3 \lesssim_{\mathbb{P}} \frac{J^{2d}}{n^2} \qquad \text{and} \qquad \Sigma_\varepsilon(\beta) = \frac{1}{(T-H)^2}\sum_{t=1}^{T}\sum_{i=1}^{n_t} \varsigma_{it}^2\big(W_{it}(\beta)\big)^2 \lesssim_{\mathbb{P}} \frac{J^d}{nT}.$$

At the origin, the split partition fixes the relevant cell containing zero, and the midpoint and side-length normalization gives $\widehat{p}_t(0) = p_t(0)$. Thus the leading weights at the origin are already expressed in reference zero-cell coordinates and are $\mathcal{E}_\infty$-measurable. Conditional on $\mathcal{E}_\infty$, the shocks $\{\eta_{it}\}$ are independent, mean zero, and have uniformly bounded third moments. Therefore the conditional Berry–Esseen inequality gives

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\bar{\mathscr{S}}_\varepsilon(0)}{\sqrt{\Sigma_\varepsilon(0)}}\le x\,\middle|\,\mathcal{E}_\infty\right)-\Phi(x)\right|\lesssim_{\mathbb{P}}\frac{TJ^{2d}/n^2}{\left((T-H)^2\Sigma_\varepsilon(0)\right)^{3/2}}.$$

At $\beta=0$, the variance nondegeneracy assumption gives $(T-H)^2\Sigma_\varepsilon(0)\gtrsim_{\mathbb{P}} TJ^d/n$, and the displayed bound is $\lesssim_{\mathbb{P}}\sqrt{J^d/(nT)}=o_{\mathbb{P}}(1)$. Hence

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\bar{\mathscr{S}}_\varepsilon(0)}{\sqrt{\Sigma_\varepsilon(0)}}\le x\right)-\Phi(x)\right|\to 0.$$

Since $\bar{\mathscr{S}}_f(0)=0$ and $\Sigma(0)=\Sigma_\varepsilon(0)$, the claim follows for $\beta=0$.

Now take $\beta\neq 0$. The process $\{\beta^\top\widetilde{f}_t\}$ is a martingale-difference sequence with respect to $\{\mathcal{G}_t\}$. Define the deterministic reference variance

$$\Sigma_{f,0}(\beta):=\frac{\beta^\top\Omega_f\beta}{T-H}.$$

By (SA-II.1.1), the sum of the conditional variances of the normalized martingale differences satisfies

$$\frac{1}{(T-H)\beta^\top\Omega_f\beta}\sum_{t=H+1}^{T}\mathbb{V}\mathrm{ar}(\beta^\top\widetilde{f}_t\mid\mathcal{G}_{t-1})\to_{\mathbb{P}}1.$$

For every $\epsilon>0$, the conditional fourth-moment bound verifies the Lindeberg condition using the deterministic threshold $\epsilon\sqrt{(T-H)\beta^\top\Omega_f\beta}$:

$$\begin{aligned}&\frac{1}{(T-H)\beta^\top\Omega_f\beta}\sum_{t=H+1}^{T}\mathbb{E}\left[|\beta^\top\widetilde{f}_t|^2\mathbb{1}\left(|\beta^\top\widetilde{f}_t|>\epsilon\sqrt{(T-H)\beta^\top\Omega_f\beta}\right)\,\middle|\,\mathcal{G}_{t-1}\right]\\&\quad\lesssim_{\mathbb{P}}\frac{\sum_{t=H+1}^{T}\mathbb{E}[|\beta^\top\widetilde{f}_t|^4\mid\mathcal{G}_{t-1}]}{(T-H)^2(\beta^\top\Omega_f\beta)^2}\lesssim_{\mathbb{P}}(T-H)^{-1}=o_{\mathbb{P}}(1).\end{aligned}$$

The martingale central limit theorem and Polya's theorem therefore give

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\bar{\mathscr{S}}_f(\beta)}{\sqrt{\Sigma_{f,0}(\beta)}}\le x\right)-\Phi(x)\right|\to 0.$$

Finally,

$$\frac{\Sigma_f(\beta)}{\Sigma_{f,0}(\beta)}=\frac{\beta^\top\left[(T-H)^{-1}\sum_{t=H+1}^{T}\mathbb{V}\mathrm{ar}(f_t\mid\mathcal{G}_{t-1})\right]\beta}{\beta^\top\Omega_f\beta}\to_{\mathbb{P}}1.$$

Slutsky's lemma and continuity of the standard normal cdf thus imply

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\bar{\mathscr{S}}_f(\beta)}{\sqrt{\Sigma_f(\beta)}}\le x\right)-\Phi(x)\right|\to 0.$$

For the idiosyncratic component, use the exact decomposition $\bar{\mathscr{S}}_\varepsilon(\beta)=\mathscr{E}(\beta)-\mathscr{A}(\beta)$. The direct idiosyncratic bound from Lemma SA-I.3.1 and the uniform first-step-correction bound from Lemma SA-II.4.2 give

$$|\mathscr{E}(\beta)|\lesssim_{\mathbb{P}}\sqrt{\frac{J^d}{n(T-H)}},\qquad |\mathscr{A}(\beta)|\lesssim_{\mathbb{P}}\sqrt{\frac{J^d\log(nT)}{nH}}.$$

Because $\Sigma_f(\beta)\asymp_{\mathbb{P}}(T-H)^{-1}$, the displayed rate conditions imply

$$\frac{\mathscr{E}(\beta)}{\sqrt{\Sigma_f(\beta)}}=o_{\mathbb{P}}(1),\qquad \frac{\mathscr{A}(\beta)}{\sqrt{\Sigma_f(\beta)}}=o_{\mathbb{P}}(1),$$

and hence $\bar{\mathscr{S}}_\varepsilon(\beta)/\sqrt{\Sigma_f(\beta)}=o_{\mathbb{P}}(1)$. Separately, the reference idiosyncratic scale is negligible relative to the factor variance because

$$\frac{\Sigma_\varepsilon(\beta)}{\Sigma_f(\beta)}\lesssim_{\mathbb{P}}\frac{J^d/(nT)}{1/T}=\frac{J^d}{n}=o(1).$$

Thus $\bar{\mathscr{S}}_\varepsilon(\beta)/\sqrt{\Sigma(\beta)}=o_{\mathbb{P}}(1)$ and $\Sigma_f(\beta)/\Sigma(\beta)=1+o_{\mathbb{P}}(1)$. Hence full joint convergence of the factor and idiosyncratic components is not needed for fixed $\beta\neq 0$; Slutsky's argument only requires the factor approximation and the negligibility of the idiosyncratic component. Since the standard normal distribution is continuous, the preceding Kolmogorov approximation is preserved under this $o_{\mathbb{P}}(1)$ perturbation, and hence

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\bar{\mathscr{S}}(\beta)}{\sqrt{\Sigma(\beta)}}\le x\right)-\Phi(x)\right|\to 0.$$

This gives the stated conclusion for $\beta\neq 0$. □

SA-II.6.4 Proof of Theorem SA-II.6.2

By the exact decomposition at the beginning of this section,

$$\frac{\widehat{\mu}(\beta)-\bar{\mu}_T(\beta;H)}{\sqrt{\Sigma(\beta)}}=\frac{\bar{\mathscr{S}}(\beta)}{\sqrt{\Sigma(\beta)}}+\frac{\mathscr{R}(\beta)}{\sqrt{\Sigma(\beta)}}.$$

The componentwise bounds from Lemma SA-I.3.1, the partition bounds from Lemma SA-II.3.3, and the generated-regressor refinement in Lemma SA-II.4.2 imply

$$|\mathscr{R}(\beta)|\lesssim_{\mathbb{P}}\mathsf{R}_{\mathtt{CLT}}.$$

Here the Gram-perturbation and basis-switching pieces are not reabsorbed into the local-support idiosyncratic scale; they are part of $\mathsf{R}_{\mathtt{CLT}}$ through the explicit $\mathsf{R}_{\mathtt{Q}}$ and $\mathsf{R}_{\mathtt{P}}$ terms in its definition. The normalization condition $\mathsf{R}_{\mathtt{CLT}}/\sqrt{\Sigma(\beta)}=o_{\mathbb{P}}(1)$ therefore removes the full remainder directly. Lemma SA-II.6.1 gives the

Kolmogorov approximation for the leading terms, and the standard perturbation argument for continuous limiting cdfs gives the displayed result. The sufficient conditions for $\beta \neq 0$ and $\beta = 0$ follow from $\Sigma(\beta) \asymp_{\mathbb{P}} T^{-1}$ and $\Sigma(0) \asymp_{\mathbb{P}} J^d/(nT)$, respectively. □

SA-II.7 Pointwise-in-β Variance Estimation

This section constructs feasible estimators of the components of the baseline normalization in Section SA-II.6. Recall that $\Sigma(\beta) = \Sigma_f(\beta) + \Sigma_\varepsilon(\beta)$. The idiosyncratic variance result has two parts. At fixed evaluation points whose basis coordinates are anchored in the estimated and reference partitions, the weights $W_{it}(\beta)$ are $\mathcal{E}_\infty$-measurable, $\Sigma(\beta)$ is the predictable quadratic variation of the leading sum in (SA-II.6.1), and the factor–idiosyncratic cross term vanishes by Assumption SA-DGP(ii)(a). At a generic nonanchored point, $\Sigma_\varepsilon(\beta)$ remains a reference scale. For fixed $\beta \neq 0$, only a local-support upper bound for the jackknife estimator is needed because the factor variance is of order T^{-1}.

Let $m_t = \mathbb{E}[f_t \mid \mathcal{G}_{t-1}]$, and let $\widehat{m}_t$ be a $\mathcal{G}_{t-1}$-measurable predictor of m_t. For instance, $\widehat{m}_t$ can be a recursive past sample mean of $\{f_s : s < t\}$ when $\mathbb{E}[f_t \mid \mathcal{G}_{t-1}] = \mu_f$ for every t and the recursive mean is consistent for μ_f, or a fitted predictive model using only information in $\mathcal{G}_{t-1}$ when richer factor dynamics are imposed. Define the estimated factor innovation $\widehat{U}_t = f_t - \widehat{m}_t$. The factor-variance estimator is

$$\widehat{\Sigma}_f(\beta) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \beta^\top \widehat{U}_t \widehat{U}_t^\top \beta.$$

For a fixed evaluation point, we propose a bin-restricted delete-i jackknife estimator. At an anchored point it estimates $\Sigma_\varepsilon(\beta)$ at the idiosyncratic scale; at a generic nonzero point we use only its order bound. Let $\widehat{M}_t^{(-i)}(\beta)$ denote the second-step estimator with unit i removed at time t, holding the partition $\widehat{\Delta}_t$ fixed. Define the leave-one-out pseudo-value

$$\widehat{d}_{it}(\beta) = \widehat{M}_t^{(-i)}(\beta) - \widehat{M}_t(\beta),$$

with the convention that $\widehat{d}_{it}(\beta) = 0$ when unit i is absent from the cross-section at time t. The bin-restricted jackknife estimator is

$$\widehat{\Sigma}_\varepsilon(\beta) = \frac{1}{(T-H)^2} \sum_{i=1}^{n} \sum_{t=H+1}^{T} \sum_{s=H+1}^{T} \widehat{d}_{it}(\beta) \widehat{d}_{is}(\beta) \mathbb{1}(|t-s| \leq H),$$

where $[n]$ is the common index set from the general setup and absent units have zero pseudo-values. The Sherman–Morrison formula for deleting the i-th observation from the period-t least-squares normal equations gives

$$\widehat{d}_{it}(\beta) = -\widehat{C}_{it}(\beta)\widehat{\varepsilon}_{it}, \qquad \widehat{C}_{it}(\beta) = \frac{1}{n_t} \frac{\widehat{p}_t(\beta)^\top \widehat{Q}_t^{-1} \widehat{p}_t(\widehat{\beta}_{it})}{1 - n_t^{-1} \widehat{p}_t(\widehat{\beta}_{it})^\top \widehat{Q}_t^{-1} \widehat{p}_t(\widehat{\beta}_{it})}, \qquad \widehat{\varepsilon}_{it} = R_{it} - \widehat{M}_t(\widehat{\beta}_{it}).$$

When $\mathfrak{q} = 0$, this reduces to the more familiar bin-count expression

$$\widehat{\Sigma}_\varepsilon(\beta) = \frac{1}{(T-H)^2} \sum_{i=1}^{n} \sum_{t=H+1}^{T} \sum_{s=H+1}^{T} \frac{\mathbb{1}_{it}^{(\widehat{\beta})} \mathbb{1}_{is}^{(\widehat{\beta})}}{(\widehat{n}_{j_0(t)t} - 1)(\widehat{n}_{j_0(s)s} - 1)} \widehat{\varepsilon}_{it} \widehat{\varepsilon}_{is} \mathbb{1}(|t-s| \leq H),$$

where $\mathbb{1}_{it}^{(\widehat{\beta})} = \mathbb{1}(\widehat{\beta}_{it} \in \widehat{P}_{j_0(t)t})$ and $j_0(t)$ denotes the cell containing the evaluation point β at time t.

Combining the two components, the feasible variance estimator is

$$\widehat{\Sigma}(\beta) = \widehat{\Sigma}_f(\beta) + \widehat{\Sigma}_\varepsilon(\beta).$$

Recall that for $\beta = 0$, $\Sigma_f(0) = 0$ and $\widehat{\Sigma}(0) = \widehat{\Sigma}_\varepsilon(0)$ suffices; for each fixed $\beta \neq 0$, $\Sigma_f(\beta) \asymp T^{-1}$.

SA-II.7.1 Results

The first result controls the factor component of the variance estimator and isolates the cost of replacing the predictable factor mean by a fitted predictor.

Lemma SA-II.7.1 (Factor Variance Estimation)**.** *Suppose that Assumption SA-DGP(iv) holds and $H \leq cT$ for some fixed $c \in (0,1)$. If $\widehat{m}_t$ is $\mathcal{G}_{t-1}$-measurable and*

$$\frac{1}{T-H}\sum_{t=H+1}^{T} \|m_t - \widehat{m}_t\|^2 = o_{\mathbb{P}}(T),$$

then, for each fixed $\beta \in \mathcal{B}$,

$$\widehat{\Sigma}_f(\beta) - \Sigma_f(\beta) - \mathscr{B}_{\mathtt{F}}(\beta) = \begin{cases} o_{\mathbb{P}}(\|\beta\|^2 T^{-1}), & \beta \neq 0, \\ 0, & \beta = 0, \end{cases}$$

where

$$\mathscr{B}_{\mathtt{F}}(\beta) = \frac{1}{(T-H)^2}\sum_{t=H+1}^{T} \beta^\top (m_t - \widehat{m}_t)(m_t - \widehat{m}_t)^\top \beta \geq 0$$

is the squared factor-prediction bias.

The next result first gives, at every fixed evaluation point, the upper bound needed to make the entire idiosyncratic component negligible relative to the factor variance. This bound does not require the selected block to be anchored. The split-at-zero construction is anchored at $\beta = 0$, and the generalized construction in Section SA-II.9 is anchored at each prespecified contrast point. At these points the result further gives sharp additive consistency at the idiosyncratic scale.

Lemma SA-II.7.2 (Bin-Restricted Jackknife)**.** *Fix $\beta \in \mathcal{B}^\circ$. Suppose Assumptions SA-DGP, SA-FS, SA-PA, SA-SEP, and SA-FSB hold, and that $s_\mu \geq 2$. Assume further that*

$$\frac{T^{2/q}}{H} + \sqrt{\frac{\log T}{H}} = o(1),$$
$$\frac{J^2 \log T}{n} = o(1),$$
$$\frac{J^{d-1}\max\{J, \sqrt{H}\}\log^2(nT)}{\min\{n^{1-2/q}T^{1-4/q}h, n\}} \lesssim 1, \qquad \frac{\max\{J^{2/(q-2)}T^{2/q}, T^{2(3+d)/q}\}}{H} = o(1).$$

Assume further that

$$H\log(nT)\big(J^{-2\min\{s_\mu,\mathsf{q}+1\}}+\mathsf{R}_{\mathtt{FS}}^4+\mathsf{B}_{\mathtt{FS}}^2\big)+\frac{\log(nT)}{H}+\frac{J^d\log^{d+2}(nT)}{n}=o(1). \tag{SA-II.7.1}$$

Then the following conclusions hold.

(i) For every fixed $\beta\in\mathcal{B}^\circ$,

$$\left|\widehat{\Sigma}_\varepsilon(\beta)\right|=O_{\mathbb{P}}\left(\frac{J^d}{nT}\right).$$

(ii) If, in addition, $J\mathsf{R}_{\mathtt{FS}}=o(1)$ and

$$\widehat{p}_t(\beta)=p_t(\beta),\qquad t=H+1,\dots,T, \tag{SA-II.7.2}$$

then

$$\widehat{\Sigma}_\varepsilon(\beta)=\Sigma_\varepsilon(\beta)+o_{\mathbb{P}}\left(\frac{J^d}{nT}\right).$$

Combining the two pieces gives the feasible variance estimator used in the pointwise Gaussian approximation.

Theorem SA-II.7.3 (Feasible Variance Estimation)**.** *Fix $\beta\in\mathcal{B}^\circ$. Suppose Assumptions SA-DGP, SA-FS, SA-PA, SA-SEP, and SA-FSB hold, and that $s_\mu\geq 2$. Assume that $\widehat{m}_t$ is $\mathcal{G}_{t-1}$-measurable,*

$$\frac{1}{T-H}\sum_{t=H+1}^{T}\|m_t-\widehat{m}_t\|^2=o_{\mathbb{P}}(T),$$

and that the following rate restrictions hold:

$$\begin{aligned}
&\frac{T^{2/q}}{H}+\sqrt{\frac{\log T}{H}}=o(1),\\
&\frac{J^2\log T}{n}=o(1),\\
&\frac{J^{d-1}\max\{J,\sqrt{H}\}\log^2(nT)}{\min\{n^{1-2/q}T^{1-4/q}h,n\}}\lesssim 1,\qquad \frac{\max\{J^{2/(q-2)}T^{2/q},T^{2(3+d)/q}\}}{H}=o(1),\\
&H\log(nT)\big(J^{-2\min\{s_\mu,\mathsf{q}+1\}}+\mathsf{R}_{\mathtt{FS}}^4+\mathsf{B}_{\mathtt{FS}}^2\big)+\frac{\log(nT)}{H}+\frac{J^d\log^{d+2}(nT)}{n}=o(1).
\end{aligned}$$

Then

$$\widehat{\Sigma}(\beta)-\Sigma(\beta)-\mathscr{B}_{\mathtt{F}}(\beta)=o_{\mathbb{P}}(T^{-1}).$$

For $\beta=0$, if, in addition, equation (SA-II.7.2) holds, $J\mathsf{R}_{\mathtt{FS}}=o(1)$, and $(nT/J^d)\Sigma_\varepsilon(0)\gtrsim_{\mathbb{P}}1$, then

$$\widehat{\Sigma}(0)-\Sigma(0)=o_{\mathbb{P}}\Big(\frac{J^d}{nT}\Big).$$

SA-II.7.2 Fama-MacBeth Variance Estimator

The Fama-MacBeth variance estimator is widely used because it treats the time-series dispersion of the per-period second-step estimates as the sampling variance. Its target generally differs from the baseline normalization in Section SA-II.6, which equals the predictable variance of the leading sum at anchored evaluation points. Recall that

$$\mathscr{B}(\beta) = \frac{1}{T-H}\sum_{t=H+1}^{T} \mathscr{B}_t(\beta),$$

where $\mathscr{B}_t(\beta)$ is the period-t smoothing-bias component from Section SA-II.4. The classical Fama-MacBeth estimator and its oracle bias-corrected variant are

$$\widehat{\Sigma}_{\mathtt{FM}}(\beta) = \frac{1}{(T-H)^2}\sum_{t=H+1}^{T} \big(\widehat{M}_t(\beta) - \widehat{\mu}(\beta)\big)^2,$$
$$\widetilde{\Sigma}_{\mathtt{FM}}(\beta) = \frac{1}{(T-H)^2}\sum_{t=H+1}^{T} \Big[\big(\widehat{M}_t(\beta) - \mathscr{B}_t(\beta)\big) - \big(\widehat{\mu}(\beta) - \mathscr{B}(\beta)\big)\Big]^2.$$

The second estimator is an oracle benchmark because $\mathscr{B}_t(\beta)$ depends on the unknown function $\mu_t(\cdot)$.

For the comparison at the origin, define the direct oracle idiosyncratic term

$$\mathscr{E}_{\mathtt{o},t}(\beta) = \sum_{i=1}^{n_t} \varsigma_{it}\eta_{it}\left(\frac{1}{n_t}\widehat{p}_t(\beta)^\top Q_t^{-1} p_t(\beta_{it})\right),$$

and its pre-shock conditional variance

$$\Sigma_{\varepsilon,\mathtt{o}}(\beta) = \frac{1}{(T-H)^2}\sum_{t=H+1}^{T} \mathbb{E}\big[\mathscr{E}_{\mathtt{o},t}(\beta)^2 \mid \mathcal{C}_t\big] = \frac{1}{(T-H)^2}\sum_{t=H+1}^{T}\sum_{i=1}^{n_t} \varsigma_{it}^2\left(\frac{1}{n_t}\widehat{p}_t(\beta)^\top Q_t^{-1} p_t(\beta_{it})\right)^2.$$

Here $\mathcal{C}_t$, defined in Section SA-II.1.3, fixes the current innovation-free scales and the past generated design while excluding the current primitive innovations. This is the variance of the direct cross-sectional idiosyncratic component. It differs from $\Sigma_\varepsilon(\beta)$ in Section SA-II.6, whose weight $W_{it}(\beta)$ also subtracts the first-step generated-regressor correction.

The following theorem makes the comparison precise by characterizing the probability limit of the Fama-MacBeth estimator. Define

$$\mathsf{R}_{\mathtt{FM}} = \mathsf{R}_{\mathtt{LA}} + \frac{\log(nT)}{H} + \sqrt{\frac{J^{2d}\mathsf{R}_{\mathtt{P}}}{n}} + \frac{\mathbb{1}(\mathfrak{q}=0)}{J} + \mathsf{R}_{\mathtt{FS}}^2.$$

Theorem SA-II.7.4 (Fama-MacBeth Variance Estimation)**.**
Suppose Assumptions SA-DGP, SA-FS, SA-PA, SA-SEP, and SA-FSB hold, and that $s_\mu \geq 2$. Assume further that

$$\frac{T^{2/q}}{H} + \sqrt{\frac{\log T}{H}} = o(1), \qquad \mathsf{R}_{\mathtt{FS}} = o(1),$$
$$\frac{J^2 \log T}{n} = o(1),$$

$$\frac{J^{d-1}\max\{J,\sqrt{H}\}\log^2(nT)}{\min\{n^{1-2/q}T^{1-4/q}h,n\}}\lesssim 1, \qquad \frac{\max\{J^{2/(q-2)}T^{2/q},T^{2(3+d)/q}\}}{H}=o(1).$$

The anchored-basis condition is required for the origin statement below and holds automatically under Assumption SA-PA. If, in addition,

$$J^{-\min\{s_\mu,\mathsf{q}+1\}}+\mathsf{R}_{\mathtt{FM}}=o(1),$$

then, for each fixed $\beta\neq 0$ in $\mathcal{B}^\circ$,

$$\widehat{\Sigma}_{\mathtt{FM}}(\beta)=\Sigma_f(\beta)+\Sigma_\mu(\beta)+o_{\mathbb{P}}(T^{-1}),$$

where

$$\Sigma_\mu(\beta)=\frac{1}{(T-H)^2}\sum_{t=H+1}^{T}\big(\mu_t(\beta)-\bar{\mu}_T(\beta;H)\big)^2.$$

For $\beta=0$, if, in addition, $J\mathsf{R}_{\mathtt{FS}}=o(1)$,

$$\Big|\widetilde{\Sigma}_{\mathtt{FM}}(0)-\Sigma_{\varepsilon,\mathsf{o}}(0)-\Sigma_\mu(0)\Big|\lesssim_{\mathbb{P}}\frac{\mathsf{R}_{\mathtt{FM}}^2}{T}+\left(\sqrt{\frac{J^d}{nT}}+\frac{\mathsf{R}_{\mathtt{FM}}}{\sqrt{T}}\right)\sqrt{\Sigma_\mu(0)}+o_{\mathbb{P}}\Big(\frac{J^d}{nT}\Big).$$

SA-II.7.3 Discussion

(i) *Truncated jackknife.* The estimator $\widehat{\Sigma}_\varepsilon(\beta)$ uses the delete-i jackknife perturbations only for time pairs with $|t-s|\le H$. After the rolling-window first-step correction is isolated, pairs farther apart do not share primitive idiosyncratic shocks and are conditionally mean zero. The truncation therefore keeps the serial dependence induced by the first-step window while reducing the number of nonzero pairs from order T^2 to order TH.

(ii) *Scale.* At every fixed evaluation point, the selected-block argument gives $|\widehat{\Sigma}_\varepsilon(\beta)|=O_{\mathbb{P}}(J^d/(nT))$. For fixed $\beta\neq 0$, this is sufficient: $J^d/n=o(1)$ makes both the estimator and the reference idiosyncratic scale negligible relative to the first-order factor variance. At an anchored evaluation point, Lemma SA-II.7.2 further gives additive consistency for the idiosyncratic variance at the scale $J^d/(nT)$. At $\beta=0$, this is the relevant first-order variance scale, so relative consistency additionally requires the nondegeneracy condition $(nT/J^d)\Sigma_\varepsilon(0)\gtrsim_{\mathbb{P}}1$.

(iii) *Factor prediction.* The factor variance estimator separates sampling variation from prediction bias. Lemma SA-II.7.1 controls the centered estimation error, while $\mathscr{B}_{\mathtt{F}}(\beta)$ records the bias from using $\widehat{m}_t$ in place of m_t. Thus relative consistency of $\widehat{\Sigma}(\beta)$ for fixed $\beta\neq 0$ requires this prediction bias to be negligible relative to $\Sigma(\beta)$; otherwise the estimator is conservative.

(iv) *Approximation bias in the jackknife.* The jackknife condition (SA-II.7.1) uses the actual second-step approximation order $J^{-\min\{s_\mu,\mathsf{q}+1\}}$. Up to the displayed $\log(nT)$ factor, it gives the familiar HJ^{-2} restriction for $\mathsf{q}=0$, and HJ^{-4} for a degree-one second step when $s_\mu\ge 2$. This variance-estimation restriction is distinct from the same-cell nonconstant-basis term absorbed by $\mathsf{R}_{\mathtt{LA}}$ in the stochastic linearization.

(v) *Primitive conditions.* The displayed conditions $J^2 \log T/n = o(1)$ and $T^{2/q}/H + \sqrt{\log T/H} = o(1)$ verify the common partition and first-step moment requirements. The additional condition $J\mathsf{R}_{\mathtt{FS}} = o(1)$ in the anchored case verifies the comparison with a single reference block, including $\Delta_{\mathtt{JK}} = o(1)$ and $\mathsf{R}_{\mathtt{JKMD}} = o(1)$. The selected-block upper bound does not make that comparison and does not require this stronger localization.

(vi) *Fama-MacBeth.* Theorem SA-II.7.4 shows that the Fama-MacBeth variance has a different target. For fixed $\beta \neq 0$, it includes the time-series variation $\Sigma_\mu(\beta)$ of the conditional mean path. At $\beta = 0$, even the oracle bias-corrected version targets $\Sigma_{\varepsilon,\mathtt{o}}(0) + \Sigma_\mu(0)$, where $\Sigma_{\varepsilon,\mathtt{o}}(0)$ is based on the direct second-step idiosyncratic weights rather than the generated-regressor-adjusted weights in $\Sigma_\varepsilon(0)$. The origin comparison uses (SA-II.7.35) and conditions on current innovation-free information and earlier primitive shocks. It does not require the sharp jackknife rate conditions. The extra condition $J^{-\min\{s_\mu, \mathtt{q}+1\}} + \mathsf{R}_{\mathtt{FM}} = o(1)$ is only needed for the fixed nonzero-β probability limit.

SA-II.7.4 Proof of Lemma SA-II.7.1

Let $\widetilde{\Sigma}_f(\beta) = \frac{1}{(T-H)^2}\sum_{t=H+1}^{T} \beta^\top \widetilde{f}_t \widetilde{f}_t^\top \beta$, and recall $\widetilde{f}_t = f_t - \mathbb{E}[f_t \mid \mathcal{G}_{t-1}]$. Decompose $\widehat{\Sigma}_f(\beta) = \widetilde{\Sigma}_f(\beta) + \mathscr{B}_{\mathtt{F}}(\beta) + \mathscr{V}_{\mathtt{F}}(\beta)$, where

$$\mathscr{V}_{\mathtt{F}}(\beta) = \frac{2}{(T-H)^2}\sum_{t=H+1}^{T} \beta^\top \widetilde{f}_t\,(m_t - \widehat{m}_t)^\top \beta.$$

Conditional on $\mathcal{F}_{t-1}$, the summands in $\mathscr{V}_{\mathtt{F}}(\beta)$ form a martingale-difference sequence, and their predictable quadratic variation is bounded by

$$\begin{aligned}\frac{4\|\beta\|^4}{(T-H)^4}\sum_{t=H+1}^{T} \|m_t - \widehat{m}_t\|^2 \big\|\mathbb{V}\mathrm{ar}(f_t \mid \mathcal{F}_{t-1})\big\| &\lesssim_{\mathbb{P}} \frac{\|\beta\|^4}{(T-H)^4}\sum_{t=H+1}^{T} \|m_t - \widehat{m}_t\|^2 \\ &= \|\beta\|^4 o_{\mathbb{P}}(T^{-2}).\end{aligned}$$

Assumption SA-DGP(iv) and conditional Cauchy–Schwarz give $\max_{1\le t\le T}\|\mathbb{V}\mathrm{ar}(f_t \mid \mathcal{F}_{t-1})\| \lesssim_{\mathbb{P}} 1$. The Lenglart–Rebolledo inequality, applied after localization, gives $\mathscr{V}_{\mathtt{F}}(\beta) = \|\beta\|^2 o_{\mathbb{P}}(T^{-1})$, while $\mathscr{V}_{\mathtt{F}}(0) = 0$ by construction. It remains to compare $\widetilde{\Sigma}_f(\beta)$ with $\Sigma_f(\beta)$. Write $\xi_t = (\beta^\top \widetilde{f}_t)^2 - \mathbb{E}[(\beta^\top \widetilde{f}_t)^2 \mid \mathcal{G}_{t-1}]$, so that $\{\xi_t, \mathcal{G}_t\}$ is a martingale difference sequence and

$$\widetilde{\Sigma}_f(\beta) - \Sigma_f(\beta) = \frac{1}{(T-H)^2}\sum_{t=H+1}^{T} \xi_t.$$

By the unconditional higher-moment bound in Assumption SA-DGP(iv), $\sup_t \mathbb{E}[\|\widetilde{f}_t\|^4] \lesssim 1$, so $\mathbb{E}[\xi_t^2] \le \mathbb{E}[(\beta^\top \widetilde{f}_t)^4] \lesssim \|\beta\|^4$. The variance of $\widetilde{\Sigma}_f(\beta) - \Sigma_f(\beta)$ satisfies

$$\frac{1}{(T-H)^4}\sum_{t=H+1}^{T} \mathbb{E}[\xi_t^2] \lesssim \frac{\|\beta\|^4}{(T-H)^3}.$$

For $\beta \neq 0$, applying Chebyshev's inequality gives

$$\left|\widetilde{\Sigma}_f(\beta) - \Sigma_f(\beta)\right| \lesssim_{\mathbb{P}} \frac{\|\beta\|^2}{T^{3/2}}.$$

For $\beta = 0$, both $\widetilde{\Sigma}_f(0)$ and $\Sigma_f(0)$ vanish identically, so the result is immediate. □

SA-II.7.5 Preliminary Results for Proof of Lemma SA-II.7.2

We next record the unit-level residual bounds used in the jackknife proof. They are stated separately because the aggregate linearization in Lemmas SA-I.3.1 and SA-II.4.2 does not by itself control the banded products of fitted residuals. Put $T_0 = T - H$ and, for $B = C$ or $B = \widehat{C}$, write

$$\mathscr{H}_B(U,V) = \frac{1}{T_0^2}\sum_{i=1}^{n}\sum_{t=H+1}^{T}\sum_{s=H+1}^{T} B_{it}(\beta)B_{is}(\beta)U_{it}V_{is}\mathbb{1}(|t-s| \leq H),$$

where coefficients and arrays are set to zero for invalid unit–date pairs. The map $\mathscr{H}_B$ is only a convenient signed bilinear form; the flat band matrix $[\mathbb{1}(|t-s| \leq H)]_{t,s}$ is not positive semidefinite, so $\mathscr{H}_B$ is not used as an inner product.

For any fixed β, define

$$v_{it} = \frac{1}{H}\sum_{l=1}^{H} K\left(\frac{l}{H}\right) SA(t)^{-1}X_l(t)\varepsilon_{i,t-l},$$
$$Y_{it} = \varepsilon_{it} - \nabla\mu_t(\beta_{it})^\top v_{it}.$$

When β is anchored, also define

$$C_{it}(\beta) = \frac{1}{n_t}p_t(\beta)^\top Q_t^{-1}p_t(\beta_{it}).$$

For $l \in [H]$, define, whenever $t + l \in \{H+1, \ldots, T\}$ and $i \in [n_{t+l}]$,

$$a_{itl}(\beta) = \nabla\mu_{t+l}(\beta_{i,t+l})^\top K\left(\frac{l}{H}\right) SA(t+l)^{-1}X_l(t+l),$$

and set it equal to zero otherwise. On the high-probability event from Lemma SA-II.3.3, local support gives, for every fixed β,

$$\max_{t\in[T]}\max_{i\in[n]}|\widehat{C}_{it}(\beta)| \lesssim_{\mathbb{P}} \frac{J^d}{n},$$
$$\max_{t\in[T]}\sum_{i=1}^{n}|\widehat{C}_{it}(\beta)|^m \lesssim_{\mathbb{P}} \frac{J^{(m-1)d}}{n^{m-1}}, \qquad m = 2,3, \tag{SA-II.7.3}$$

because $\widehat{p}_t(\beta)$ selects only one well-conditioned interior block and its normalized polynomial coordinates are bounded. The same bounds and scalar Cauchy–Schwarz give, uniformly over $t, s \in [T]$,

$$\sum_{i=1}^{n}\widehat{C}_{it}(\beta)^2\widehat{C}_{is}(\beta)^2 \lesssim_{\mathbb{P}} \frac{J^{3d}}{n^3}, \qquad \sum_{i=1}^{n}|\widehat{C}_{it}(\beta)\widehat{C}_{is}(\beta)| \lesssim_{\mathbb{P}} \frac{J^d}{n}. \tag{SA-II.7.4}$$

Under (SA-II.7.2), all these bounds also hold with $\widehat{C}_{it}(\beta)$ replaced by $C_{it}(\beta)$. For a period-t vector $z = (z_1, \ldots, z_{n_t})^\top$ and a row satisfying $\widehat{C}_{it}(\beta) \neq 0$, let

$$(\widehat{\Pi}_t z)_i = \widehat{p}_t(\widehat{\beta}_{it})^\top \widehat{Q}_t^{-1} \frac{1}{n_t} \sum_{k=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{kt}) z_k,$$

and set $(\widehat{\Pi}_t z)_i = 0$ on the unused rows. Thus only the interior block selected by the jackknife weight is inverted below. Recall $M_t(b) = \mu_t(b) + b^\top \widetilde{f}_t$ and, finally, write

$$\begin{aligned} \delta_{it} &= \widehat{\beta}_{it} - \beta_{it}, \\ \mathscr{T}_{it}^{(\mu)} &= \mu_t(\widehat{\beta}_{it}) - \mu_t(\beta_{it}) - \nabla\mu_t(\beta_{it})^\top \delta_{it}. \end{aligned}$$

On every row with $\widehat{C}_{it}(\beta) \neq 0$, the following is an exact definition of the residual remainder; set $\widetilde{\mathscr{R}}_{it} = 0$ on the unused rows:

$$\begin{aligned} \widehat{\varepsilon}_{it} &= Y_{it} + \widetilde{\mathscr{R}}_{it}, \qquad \widehat{C}_{it}(\beta) \neq 0, \\ \widetilde{\mathscr{R}}_{it} &= \Big[(I - \widehat{\Pi}_t) M_t(\widehat{\beta}_{\cdot t})\Big]_i - (\widehat{\Pi}_t \varepsilon_{\cdot t})_i - \nabla\mu_t(\beta_{it})^\top (\delta_{it} - v_{it}) \\ &\quad - \widetilde{f}_t^\top \delta_{it} - \mathscr{T}_{it}^{(\mu)} + \Big\{\widehat{\Pi}_t \Big[\nabla\mu_t(\beta_{\cdot t})^\top \delta_{\cdot t} + \widetilde{f}_t^\top \delta_{\cdot t} + \mathscr{T}_{\cdot t}^{(\mu)}\Big]\Big\}_i. \end{aligned} \tag{SA-II.7.5}$$

Here $M_t(\widehat{\beta}_{\cdot t})$ and the expressions inside $\widehat{\Pi}_t$ denote the corresponding period-t vectors. Number the six terms in the second equality of (SA-II.7.5) in their displayed order. We use the following exact split. The block $\widetilde{\mathscr{R}}_{it}^{(\mathrm{md})}$ contains term (2); the part of term (3) obtained by replacing $\delta_{it} - v_{it}$ with the summation in (SA-II.7.9); and, after writing $\delta_{it} = v_{it} + (\delta_{it} - v_{it})$, the two parts of term (4) containing v_{it} and that same summation. The block $\widetilde{\mathscr{R}}_{it}^{(\mathrm{pr})}$ contains all the remaining pieces. Thus $\widetilde{\mathscr{R}}_{it} = \widetilde{\mathscr{R}}_{it}^{(\mathrm{md})} + \widetilde{\mathscr{R}}_{it}^{(\mathrm{pr})}$. The current-shock pieces of the first block are handled by ordering dates. Its remaining piece and the second block are handled by direct squared-sum bounds.

Define

$$\begin{aligned} \Delta_{\mathtt{JK}} &= \frac{J^d}{n} + J^{2d}\mathsf{R}_{\mathtt{Q}}^2 + J^d \mathsf{R}_{\mathtt{P}}, \\ \mathsf{R}_{\mathtt{JKMD}} &= \frac{J^d}{n} + J^{2d}\mathsf{R}_{\mathtt{Q}}^2 + J^d \mathsf{R}_{\mathtt{P}} + \frac{1}{H} + \frac{1}{\sqrt{T}} + \left(\frac{T^{2/q}}{H} + \sqrt{\frac{\log T}{H}}\right)^2, \\ \mathsf{R}_{\mathtt{JKPR}} &= H\left(\mathsf{R}_{\mathtt{FS}}^4 + J^{-2\min\{s_\mu, \mathfrak{q}+1\}} + \mathsf{B}_{\mathtt{FS}}^2\right) + \frac{J^d \log^{d+1}(nT)}{n} + \frac{1}{H}. \end{aligned}$$

Lemma SA-II.7.5 (Anchored Residual Arrays)**.** *Suppose that the assumptions and rate conditions of Lemma SA-II.7.2 hold and that $J\mathsf{R}_{\mathtt{FS}} = o(1)$. For every fixed evaluation point satisfying (SA-II.7.2),*

$$\sum_{t=H+1}^{T} \sum_{i=1}^{n_t} \big(\widehat{C}_{it}(\beta) - C_{it}(\beta)\big)^2 \left(\varepsilon_{it}^2 + \frac{1}{H}\sum_{l=1}^{H} a_{i,t-l,l}(\beta)^2 \varepsilon_{i,t-l}^2\right) \lesssim_{\mathbb{P}} T\frac{J^d}{n}\Delta_{\mathtt{JK}}, \tag{SA-II.7.6}$$

and

$$\big|\mathscr{H}_{\widehat{C}}(Y,Y) - \mathscr{H}_C(Y,Y)\big| \lesssim_{\mathbb{P}} \frac{J^d}{nT}\Delta_{\mathtt{JK}}^{1/4}, \tag{SA-II.7.7}$$

$$
\begin{aligned}
\left|\mathscr{H}_{\widehat{C}}(Y,\widetilde{\mathscr{R}}^{(\mathrm{md})})\right| &\lesssim_{\mathbb{P}} \frac{J^d}{nT}\mathsf{R}_{\mathtt{JKMD}}^{1/2},\\
\left|\mathscr{H}_{\widehat{C}}(\widetilde{\mathscr{R}}^{(\mathrm{md})},\widetilde{\mathscr{R}}^{(\mathrm{md})})\right| &\lesssim_{\mathbb{P}} \frac{J^d}{nT}\mathsf{R}_{\mathtt{JKMD}},\\
\left|\mathscr{H}_{\widehat{C}}(Y,\widetilde{\mathscr{R}}^{(\mathrm{pr})})\right| &\lesssim_{\mathbb{P}} \frac{J^d}{nT}\sqrt{\mathsf{R}_{\mathtt{JKPR}}\log(nT)},\\
\left|\mathscr{H}_{\widehat{C}}(\widetilde{\mathscr{R}}^{(\mathrm{pr})},\widetilde{\mathscr{R}}^{(\mathrm{pr})})\right| &\lesssim_{\mathbb{P}} \frac{J^d}{nT}\mathsf{R}_{\mathtt{JKPR}},\\
\left|\mathscr{H}_{\widehat{C}}(\widetilde{\mathscr{R}}^{(\mathrm{md})},\widetilde{\mathscr{R}}^{(\mathrm{pr})})\right| &\lesssim_{\mathbb{P}} \frac{J^d}{nT}\sqrt{\mathsf{R}_{\mathtt{JKMD}}\mathsf{R}_{\mathtt{JKPR}}}.
\end{aligned}
\tag{SA-II.7.8}
$$

Proof. Step 1: exact residual expansion. On every row with $\widehat{C}_{it}(\beta)\neq 0$,

$$
\widehat{\varepsilon}_{it}=\left[(I-\widehat{\Pi}_t)\{M_t(\beta_{\cdot t})+\varepsilon_{\cdot t}\}\right]_i.
$$

The unused rows never enter $\mathscr{H}_{\widehat{C}}$. Write

$$
M_t(\beta_{it})=M_t(\widehat{\beta}_{it})-\nabla\mu_t(\beta_{it})^\top\delta_{it}-\widetilde{f}_t^\top\delta_{it}-\mathscr{T}_{it}^{(\mu)}.
$$

Substitution and the definition of Y_{it} give (SA-II.7.5). Moreover, the exact first-step decomposition gives

$$
\delta_{it}-v_{it}=\breve{\beta}_{it}-\beta_{it}+\frac{1}{H}\sum_{l=1}^{H}K\left(\frac{l}{H}\right)S\{\widehat{A}(t)^{-1}-A(t)^{-1}\}X_l(t)\varepsilon_{i,t-l}.
\tag{SA-II.7.9}
$$

Thus the first-step bias and the error from replacing $A(t)^{-1}$ by $\widehat{A}(t)^{-1}$ enter the remainder separately. If $\mathfrak{q}\geq 1$, polynomial reproduction makes the factor part of $[(I-\widehat{\Pi}_t)M_t(\widehat{\beta}_{\cdot t})]_i$ identically zero. If $\mathfrak{q}=0$, that part is the within-cell deviation of $\widehat{\beta}_{it}^\top\widetilde{f}_t$ and is bounded by $J^{-1}\|\widetilde{f}_t\|$. Together with (SA-II.7.9), this verifies that the two blocks defined after (SA-II.7.5) contain every part of the remainder exactly once.

Step 2: shock-weighted generated-design bounds. Put

$$
\omega_{it}=\varepsilon_{it}^2+\frac{1}{H}\sum_{l=1}^{H}a_{i,t-l,l}(\beta)^2\varepsilon_{i,t-l}^2.
$$

We first make the leave-out construction explicit. For a set $\mathcal{I}$ containing one or two distinct primitive-shock indices, define

$$
\widehat{\beta}_{kr}^{[-\mathcal{I}]}=\widehat{\beta}_{kr}-\sum_{(i,u)\in\mathcal{I}}\frac{\mathbb{1}(k=i,\ 1\leq r-u\leq H)}{H}K\left(\frac{r-u}{H}\right)S\widehat{A}(r)^{-1}X_{r-u}(r)\varepsilon_{iu}.
\tag{SA-II.7.10}
$$

Every partition, basis, Gram block, leverage, and coefficient carrying the superscript $[-\mathcal{I}]$ is rebuilt from the entire trajectory in (SA-II.7.10). Thus a removed shock is deleted from all future first-step estimates whose rolling windows contain it. Conditional on $\mathcal{E}_\infty$ and the remaining innovations, the rebuilt objects do not depend on the removed shock or pair.

Let j_t be the anchored block containing β. Apply the deterministic quasi-uniform rectangle bounds used in Lemma SA-II.3.3 to the leave-out trajectories before substituting the empirical cells. The conditional

convolution and boundary-strip calculations in Lemmas SA-II.3.1 and SA-II.4.1 then give

$$\sum_{t=H+1}^{T}\sum_{i=1}^{n_t}\omega_{it}\left\{\|\widehat{p}_{j_tt}(\widehat{\beta}_{it})\|^2+\|p_{j_tt}(\beta_{it})\|^2\right\}\lesssim_{\mathbb{P}}\frac{Tn}{J^d},$$
$$\sum_{t=H+1}^{T}\sum_{i=1}^{n_t}\omega_{it}\|\widehat{p}_{j_tt}(\widehat{\beta}_{it})-p_{j_tt}(\beta_{it})\|^2\lesssim_{\mathbb{P}}Tn\mathsf{R}_{\mathsf{P}}. \tag{SA-II.7.11}$$

To pass from a leave-out empirical partition to the original one, restore the removed observation at every affected date. Each marginal empirical quantile moves by at most one rank. At each affected date, only a fixed number of observations can cross a boundary of the selected block. The conditional moment bounds used above show that their total weighted contribution over all dates is $O_{\mathbb{P}}(T\log(nT))$. After division by Tn, this is of order $\log(nT)/n$, which is already included in R_{q} and R_{P}. For $\mathsf{q}\geq 1$, the nonswitch coordinate displacement is handled by the order-J derivative of the normalized polynomial basis. The weights containing $a_{i,t-l,l}(\beta)^2$ are controlled by the factor fourth-moment and NED time-average bounds. This proves (SA-II.7.11) without replacing the boundary probability by a fractional power.

For completeness, set

$$\widetilde{C}_{it}(\beta)=\frac{1}{n_t}p_t(\beta)^\top\widehat{Q}_t^{-1}\widehat{p}_t(\widehat{\beta}_{it}),\qquad h_{it}=\frac{1}{n_t}\widehat{p}_t(\widehat{\beta}_{it})^\top\widehat{Q}_t^{-1}\widehat{p}_t(\widehat{\beta}_{it}).$$

On the selected block, the exact decomposition

$$\begin{aligned}\widehat{C}_{it}(\beta)-C_{it}(\beta)=&\frac{\widetilde{C}_{it}(\beta)h_{it}}{1-h_{it}}\\&+\frac{1}{n_t}p_t(\beta)^\top(\widehat{Q}_t^{-1}-Q_t^{-1})\widehat{p}_t(\widehat{\beta}_{it})\\&+\frac{1}{n_t}p_t(\beta)^\top Q_t^{-1}\{\widehat{p}_t(\widehat{\beta}_{it})-p_t(\beta_{it})\}\end{aligned}$$

separates leverage, inverse-Gram, and basis perturbations. The leverage bound, the resolvent identity, and (SA-II.7.11) yield

$$\sum_{t=H+1}^{T}\sum_{i=1}^{n_t}\omega_{it}\big(\widehat{C}_{it}(\beta)-C_{it}(\beta)\big)^2\lesssim_{\mathbb{P}}T\frac{J^d}{n}\left\{\frac{J^d}{n}+J^{2d}\mathsf{R}_{\mathsf{q}}^2+J^d\mathsf{R}_{\mathsf{P}}\right\}.$$

This is (SA-II.7.6).

To bound the terms in $\mathscr{H}_C(Y,Y)$ that contain $\varepsilon_{iu}\varepsilon_{iv}$ with $u\neq v$ in the proof of Lemma SA-II.7.2, define

$$\begin{aligned}\Gamma_{iuv}(C)=\sum_{r=H+1}^{T}\sum_{s=H+1}^{T}&C_{ir}(\beta)C_{is}(\beta)\left\{\mathbb{1}(u=r)-\frac{\mathbb{1}(1\leq r-u\leq H)}{H}a_{iu,r-u}(\beta)\right\}\\&\times\left\{\mathbb{1}(v=s)-\frac{\mathbb{1}(1\leq s-v\leq H)}{H}a_{iv,s-v}(\beta)\right\}\mathbb{1}(|r-s|\leq H).\end{aligned}$$

Then

$$T_0^2\mathscr{H}_C(Y,Y)=\sum_{i=1}^{n}\sum_{u=1}^{T}\sum_{v=1}^{T}\Gamma_{iuv}(C)\varepsilon_{iu}\varepsilon_{iv}.$$

We now prove (SA-II.7.7). It bounds the difference between $\mathscr{H}_{\widehat{C}}(Y,Y)$ and $\mathscr{H}_{C}(Y,Y)$. Write $\ell_{it} = \nabla\mu_t(\beta_{it})^\top v_{it}$, so $Y_{it} = \varepsilon_{it} - \ell_{it}$.

Squared-sum bounds for the rolling term. Use the exact leverage, inverse-Gram, and basis decomposition displayed immediately before (SA-II.7.6), now with unit weights. It gives

$$\sum_{t=H+1}^{T}\sum_{i=1}^{n_t}\big(\widehat{C}_{it}(\beta)-C_{it}(\beta)\big)^2 \lesssim_{\mathbb{P}} T\frac{J^d}{n}\Delta_{\mathtt{JK}}, \qquad \max_{i,t}\big|\widehat{C}_{it}(\beta)-C_{it}(\beta)\big| \lesssim_{\mathbb{P}} \frac{J^d}{n}.$$

It follows that

$$\sum_{t=H+1}^{T}\sum_{i=1}^{n_t}\big(\widehat{C}_{it}(\beta)-C_{it}(\beta)\big)^4 \lesssim_{\mathbb{P}} T\frac{J^{3d}}{n^3}\Delta_{\mathtt{JK}}.$$

We next control ℓ_{it} for observations that receive nonzero weight. Anchoring makes the reference and estimated bases select the same cell index at β. On the high-probability events in Lemmas SA-II.2.3, SA-II.3.2, and SA-II.3.3, the cell-endpoint error and the maximum first-step error are both $o(J^{-1})$. Hence, if $C_{it}(\beta)$ or $\widehat{C}_{it}(\beta)$ is nonzero, β_{it} lies in the reference cell containing β or in one of a fixed number of neighboring reference cells. The union of these cells is $\mathcal{E}_\infty$-measurable and contains $O_{\mathbb{P}}(n/J^d)$ units at each date.

We can therefore sum over this union without conditioning on the random selection induced by $\widehat{C}_{it}(\beta)$. The bounded gradient, the conditional fourth moment of ε_{it}, and independence of the primitive shocks give

$$\mathbb{E}[\ell_{it}^4 \mid \mathcal{E}_\infty] \lesssim_{\mathbb{P}} \frac{1}{H^4}\left[\left\{\sum_{l=1}^{H}(1+\|f_{t-l}\|^2)\right\}^2 + \sum_{l=1}^{H}(1+\|f_{t-l}\|^4)\right].$$

To sum this bound over the union and over time, use $(\sum_l x_l)^2 \le H\sum_l x_l^2$ and the time-average fourth-moment bound for the factors. Conditional Markov's inequality then gives

$$\sum_{t=H+1}^{T}\sum_{i=1}^{n_t}\ell_{it}^4\mathbb{1}\Big(|C_{it}(\beta)|+|\widehat{C}_{it}(\beta)|>0\Big) \lesssim_{\mathbb{P}} \frac{Tn}{J^dH^2}.$$

The weights vanish outside the same union. Their pointwise bounds imply

$$\sum_{t=H+1}^{T}\sum_{i=1}^{n_t}\Big(C_{it}(\beta)^4+\widehat{C}_{it}(\beta)^4\Big) \lesssim_{\mathbb{P}} T\frac{J^{3d}}{n^3}.$$

Scalar Cauchy–Schwarz now gives the first bound below. For the second, use the earlier fourth-power bound for $\widehat{C}_{it}(\beta)-C_{it}(\beta)$ in the same way:

$$\begin{aligned}
&\sum_{t=H+1}^{T}\sum_{i=1}^{n_t}\Big(C_{it}(\beta)^2+\widehat{C}_{it}(\beta)^2\Big)\ell_{it}^2 \lesssim_{\mathbb{P}} T\frac{J^d}{nH},\\
&\sum_{t=H+1}^{T}\sum_{i=1}^{n_t}\big(\widehat{C}_{it}(\beta)-C_{it}(\beta)\big)^2\ell_{it}^2 \lesssim_{\mathbb{P}} T\frac{J^d}{nH}\Delta_{\mathtt{JK}}^{1/2}.
\end{aligned} \tag{SA-II.7.12}$$

Current-shock windows. At date t, the weights $C_{it}(\beta)$ and $\widehat{C}_{it}(\beta)$ depend only on the innovation-free environment and innovations dated before t. Thus their products with ε_{it} are martingale differences when the

dates are ordered. The same is true after replacing either weight by their difference. Successive conditioning, (SA-II.7.3), and the unweighted comparison above give

$$\sum_{i=1}^{n}\sum_{s=H+1}^{T}\left(\sum_{\substack{H+1\le t\le T\\ |t-s|\le H}} B_{it}(\beta)\varepsilon_{it}\right)^2 \lesssim_{\mathbb{P}} TH\frac{J^d}{n}, \qquad B=C,\widehat{C},$$

$$\sum_{i=1}^{n}\sum_{s=H+1}^{T}\left(\sum_{\substack{H+1\le t\le T\\ |t-s|\le H}} \big(\widehat{C}_{it}(\beta)-C_{it}(\beta)\big)\varepsilon_{it}\right)^2 \lesssim_{\mathbb{P}} TH\frac{J^d}{n}\Delta_{\mathtt{JK}}. \tag{SA-II.7.13}$$

For unequal dates, condition at the later date; the cross term then has conditional mean zero. Each diagonal term is counted only $O(H)$ times.

In the remaining displays, the common argument β is suppressed. We first compare the current-shock terms. For $B = C, \widehat{C}$, order the off-diagonal sum by its later date. Its summed conditional variance is at most the squared pointwise weight bound times the corresponding one-sided sum of squares. The same successive-conditioning calculation as (SA-II.7.13) bounds this sum of squares at the same order. Hence

$$\left|\mathscr{H}_B(\varepsilon,\varepsilon)-\frac{1}{T_0^2}\sum_{i,t}B_{it}^2\varepsilon_{it}^2\right| \lesssim_{\mathbb{P}} \sqrt{\frac{HJ^{3d}}{n^3T^3}}.$$

For the diagonal part, use $\widehat{C}_{it}^2 - C_{it}^2 = 2C_{it}(\widehat{C}_{it}-C_{it}) + (\widehat{C}_{it}-C_{it})^2$. Equation (SA-II.7.6) and scalar Cauchy–Schwarz give

$$\left|\mathscr{H}_{\widehat{C}}(\varepsilon,\varepsilon)-\mathscr{H}_C(\varepsilon,\varepsilon)\right| \lesssim_{\mathbb{P}} \frac{J^d}{nT}\left\{\Delta_{\mathtt{JK}}^{1/2}+\Delta_{\mathtt{JK}}+\sqrt{\frac{HJ^d}{nT}}\right\}. \tag{SA-II.7.14}$$

For the current–rolling terms, use

$$\widehat{C}_{it}\widehat{C}_{is}-C_{it}C_{is} = (\widehat{C}_{it}-C_{it})\widehat{C}_{is} + C_{it}(\widehat{C}_{is}-C_{is}).$$

Aggregate each current-shock window before applying scalar Cauchy–Schwarz. Equations (SA-II.7.12) and (SA-II.7.13) give

$$\begin{aligned}
&\left|\mathscr{H}_{\widehat{C}}(\varepsilon,\ell)-\mathscr{H}_C(\varepsilon,\ell)\right| \\
&\quad\le \frac{1}{T_0^2}\left\{\sum_{i,s}\widehat{C}_{is}^2\ell_{is}^2\right\}^{1/2}\left\{\sum_{i,s}\left(\sum_{|t-s|\le H}(\widehat{C}_{it}-C_{it})\varepsilon_{it}\right)^2\right\}^{1/2} \\
&\qquad+\frac{1}{T_0^2}\left\{\sum_{i,s}(\widehat{C}_{is}-C_{is})^2\ell_{is}^2\right\}^{1/2}\left\{\sum_{i,s}\left(\sum_{|t-s|\le H}C_{it}\varepsilon_{it}\right)^2\right\}^{1/2} \\
&\quad\lesssim_{\mathbb{P}} \frac{J^d}{nT}\left(\Delta_{\mathtt{JK}}^{1/2}+\Delta_{\mathtt{JK}}^{1/4}\right).
\end{aligned} \tag{SA-II.7.15}$$

Finally, use the order-H absolute row-sum bound rather than treating the signed flat band as positive

semidefinite. This gives

$$\begin{aligned}&\left|\mathscr{H}_{\widehat{C}}(\ell,\ell)-\mathscr{H}_{C}(\ell,\ell)\right|\\&\quad\lesssim\frac{H}{T_0^2}\left\{\sum_{i,t}(\widehat{C}_{it}-C_{it})^2\ell_{it}^2\right\}^{1/2}\left[\left\{\sum_{i,t}\widehat{C}_{it}^2\ell_{it}^2\right\}^{1/2}+\left\{\sum_{i,t}C_{it}^2\ell_{it}^2\right\}^{1/2}\right]\lesssim_{\mathbb{P}}\frac{J^d}{nT}\Delta_{\mathtt{JK}}^{1/4}.\end{aligned}\tag{SA-II.7.16}$$

Expand $Y=\varepsilon-\ell$ and combine (SA-II.7.14)– (SA-II.7.16). Here $J^d/n\le\Delta_{\mathtt{JK}}$ and $H\le T$. Moreover, $\Delta_{\mathtt{JK}}=o(1)$. Hence all the displayed terms are bounded by $\Delta_{\mathtt{JK}}^{1/4}$. This proves (SA-II.7.7).

Step 3: current-shock and inverse-matrix terms. Here and in Step 4, the term numbers refer to the six terms in the second equality of (SA-II.7.5). This step controls $\widetilde{\mathscr{R}}^{(\mathrm{md})}$: term (2), the part of term (3) obtained from the summation in (SA-II.7.9), and the two parts of term (4) containing v_{it} and that same summation. For this proof only, write $\widetilde{\mathscr{R}}^{(\mathrm{md})}=\widetilde{\mathscr{R}}^{(\mathrm{md,cur})}+\widetilde{\mathscr{R}}^{(\mathrm{md,lag})}$. $\widetilde{\mathscr{R}}^{(\mathrm{md,cur})}$ contains term (2) and the two parts of term (4). Each part of term (4) contains the date-t factor innovation multiplying a coefficient fixed before that innovation. $\widetilde{\mathscr{R}}^{(\mathrm{md,lag})}$ is the stated part of term (3).

We first give the calculation for term (2). Conditional on $\mathcal{C}_t$, $\widehat{C}_{it}(\beta)$ and $\widehat{\Pi}_t$ are fixed, whereas the current idiosyncratic innovations are independent and centered. The selected least-squares projection has fixed dimension and does not increase the squared norm. Together with (SA-II.7.3), this gives

$$\sum_{i=1}^{n_t}\mathbb{E}\Big[\widehat{C}_{it}(\beta)^2\{(\widehat{\Pi}_t\varepsilon_{\cdot t})_i\}^2\,\Big|\,\mathcal{C}_t\Big]\lesssim_{\mathbb{P}}\left(\frac{J^d}{n}\right)^2.\tag{SA-II.7.17}$$

The diagonal part of its banded quadratic form is therefore $O_{\mathbb{P}}(J^{2d}/(n^2T))$. For unequal dates, order each pair by its later date. The earlier projected error and both weights are then fixed, while the later projected error has conditional mean zero. After division by T_0^4, the sum of conditional variances is $O_{\mathbb{P}}(\frac{H}{T^3}(\frac{J^d}{n})^4)$. Thus the off-diagonal part is also $O_{\mathbb{P}}(J^{2d}/(n^2T))$, since $H\le T$.

For the two parts of term (4), the factor innovation is common across units. Order unequal dates as above and condition on the past at the later date. The conditional mean of the later factor innovation is zero. The second bound in (SA-II.7.4) controls the sum over units of the weighted coefficient multiplying that innovation. The conditional second moment of v_{it} is of order H^{-1}. Therefore, the time-diagonal part of the banded product for $-\widetilde{f}_t^{\top}v_{it}$ is bounded by $J^d/(nTH)$. For unequal dates, first sum over the lags shared by the two rolling windows and then take the conditional variance. The factor fourth-moment bound yields $O_{\mathbb{P}}((J^d/n)T^{-3/2})$ for the off-diagonal part. Relative to the common factor $J^d/(nT)$, these are the H^{-1} and $T^{-1/2}$ contributions in $\mathsf{R}_{\mathtt{JKMD}}$. For the coefficient containing $\widehat{A}(t)^{-1}-A(t)^{-1}$, use $\widehat{A}(t)^{-1}-A(t)^{-1}=\widehat{A}(t)^{-1}\{A(t)-\widehat{A}(t)\}A(t)^{-1}$ and Lemma SA-II.2.2. This gives the additional factor $\{T^{2/q}/H+\sqrt{\log T/H}\}^2$. Combining these calculations with (SA-II.7.17) gives

$$\begin{aligned}\sum_{i,s}\left(\sum_{|t-s|\le H}\widehat{C}_{it}(\beta)\widetilde{\mathscr{R}}_{it}^{(\mathrm{md,cur})}\right)^2&\lesssim_{\mathbb{P}}TH\frac{J^d}{n}\mathsf{R}_{\mathtt{JKMD}},\\\left|\mathscr{H}_{\widehat{C}}(\widetilde{\mathscr{R}}^{(\mathrm{md,cur})},\widetilde{\mathscr{R}}^{(\mathrm{md,cur})})\right|&\lesssim_{\mathbb{P}}\frac{J^d}{nT}\mathsf{R}_{\mathtt{JKMD}},\\\left|\mathscr{H}_{\widehat{C}}(\varepsilon,\widetilde{\mathscr{R}}^{(\mathrm{md,cur})})\right|&\lesssim_{\mathbb{P}}\frac{J^d}{nT}\mathsf{R}_{\mathtt{JKMD}}^{1/2}.\end{aligned}\tag{SA-II.7.18}$$

Consider the part of term (3) in (SA-II.7.5) obtained by substituting the summation in (SA-II.7.9). This is precisely $\widetilde{\mathscr{R}}_{it}^{(\mathrm{md,lag})}$. On the enlarged-support event used in Step 2, at most $O_{\mathbb{P}}(n/J^d)$ units enter at each date and $\max_i \widehat{C}_{it}(\beta)^2 \lesssim_{\mathbb{P}} (J^d/n)^2$. The inverse-matrix identity and Lemma SA-II.2.2 bound the squared matrix difference by $\{T^{2/q}/H + \sqrt{\log T/H}\}^2$. Conditional on $\mathcal{E}_\infty$, the rolling sum has second moment of order H^{-1}. Conditional Markov's inequality therefore yields

$$\begin{aligned}\sum_{t=H+1}^{T}\sum_{i=1}^{n_t}\widehat{C}_{it}(\beta)^2\{\widetilde{\mathscr{R}}_{it}^{(\mathrm{md,lag})}\}^2 &\lesssim_{\mathbb{P}} \left(\frac{J^d}{n}\right)^2 T\frac{n}{J^d}\frac{1}{H}\left(\frac{T^{2/q}}{H}+\sqrt{\frac{\log T}{H}}\right)^2 \\ &= T\frac{J^d}{nH}\left(\frac{T^{2/q}}{H}+\sqrt{\frac{\log T}{H}}\right)^2 \lesssim_{\mathbb{P}} T\frac{J^d}{nH}\mathsf{R}_{\mathtt{JKMD}}. \end{aligned} \tag{SA-II.7.19}$$

This calculation sums conditional second moments over dates and therefore does not require different rolling windows to be independent.

Cauchy–Schwarz within each lag window and (SA-II.7.19) give the same squared-window bound as the first line of (SA-II.7.18). Hence

$$\sum_{i,s}\left(\sum_{|t-s|\le H}\widehat{C}_{it}(\beta)\widetilde{\mathscr{R}}_{it}^{(\mathrm{md})}\right)^2 \lesssim_{\mathbb{P}} TH\frac{J^d}{n}\mathsf{R}_{\mathtt{JKMD}}. \tag{SA-II.7.20}$$

The band has at most $2H+1$ entries in each row. Thus (SA-II.7.19) bounds the lag–lag product. Scalar Cauchy–Schwarz, using the current-part window bound and the lagged squared sum, bounds the current–lag product. Together with (SA-II.7.18), this gives the second line below. For ε paired with the current array, use the third line of (SA-II.7.18). For ε paired with the lagged array, use (SA-II.7.13) and Cauchy–Schwarz. For ℓ, use (SA-II.7.12) with the corresponding current or lagged squared-sum bound. Therefore

$$\begin{aligned}\left|\mathscr{H}_{\widehat{C}}(Y,\widetilde{\mathscr{R}}^{(\mathrm{md})})\right| &\lesssim_{\mathbb{P}} \frac{J^d}{nT}\mathsf{R}_{\mathtt{JKMD}}^{1/2}, \\ \left|\mathscr{H}_{\widehat{C}}(\widetilde{\mathscr{R}}^{(\mathrm{md})},\widetilde{\mathscr{R}}^{(\mathrm{md})})\right| &\lesssim_{\mathbb{P}} \frac{J^d}{nT}\mathsf{R}_{\mathtt{JKMD}}. \end{aligned} \tag{SA-II.7.21}$$

Step 4: terms controlled by squared sums. This step controls $\widetilde{\mathscr{R}}^{(\mathrm{pr})}$: term (1), the first-step-bias parts of terms (3) and (4), term (5), and every projected piece in term (6). We first bound these terms before forming their banded products. Polynomial reproduction and local approximation bound term (1) by $J^{-2\min\{s_\mu,\mathtt{q}+1\}}$ after squaring. The first-step-bias parts of terms (3) and (4) contribute $\mathsf{B}_{\mathtt{FS}}^2$, and the second-order Taylor remainder in term (5) contributes $\mathsf{R}_{\mathtt{FS}}^4$.

For the projected rolling terms in (6), we bound the conditional mean induced by sorting on the estimated betas. First consider $z_{it} = \nabla\mu_t(\beta_{it})^\top(\widehat{\beta}_{it} - \breve{\beta}_{it})$ for $i \in [n_t]$ and $t = H+1,\dots,T$. For a deterministic admissible rectangle $P \in \mathcal{P}^\circ$, with $\mathcal{P}^\circ$ as in the proof of Lemma SA-II.4.2, define the vector

$$Z_t(P) = \frac{J^d}{n_t}\sum_{i=1}^{n_t} p_{\mathtt{q}}(\widehat{\beta}_{it};P)z_{it}.$$

For the reference basis, conditioning first on the innovation-free environment makes the rolling term centered.

Applying Lemma SA-II.4.1(a) then gives

$$\max_{H+1\le t\le T}\sup_{P\in\mathcal{P}^\circ}\|\mathbb{E}[Z_t(P)\mid\mathcal{D}_{t-1}]\|\lesssim_{\mathbb{P}} H^{-1}+\mathsf{B}_{\mathtt{FS}}.$$

This mean need not have order H^{-1} alone.

We next bound the centered part before substituting the empirical cell. Conditional on $\mathcal{D}_{t-1}$, the unit histories and primitive innovations entering $Z_t(P)$ are independent over i. The reference-basis term has conditional second moment of order J^{-d}/H. Since $J\mathsf{R}_{\mathtt{FS}}=o(1)$ implies $J/\sqrt{H}=o(1)$, Lemma SA-II.4.1(b), together with $\mathsf{B}_{\mathtt{FS}}=o(H^{-1/2})$, bounds the conditional second moment of each generated-basis summand by $\log^d(nT)/(HJ^d)$. The rectangle and normalized-polynomial class is VC-type with fixed exponent. The conditional L_2 maximal bound, obtained by combining the empirical-process and Hoffmann–Jørgensen inequalities in Cattaneo et al. (2024, Lemmas SA24 and SA41), gives

$$\mathbb{E}\left[\sup_{P\in\mathcal{P}^\circ}\|Z_t(P)-\mathbb{E}[Z_t(P)\mid\mathcal{D}_{t-1}]\|^2\mid\mathcal{D}_{t-1}\right]\lesssim_{\mathbb{P}}\frac{J^d\log^{d+1}(nT)}{nH}+\frac{J^{2d}\log^2(nT)}{n^2}\mathbb{E}\left[\max_{i\in[n_t]}|z_{it}|^2\mid\mathcal{D}_{t-1}\right].$$

Localize the $\mathcal{D}_{t-1}$-measurable constants in these conditional bounds and the first-step Gram bound. The empirical-cell and second-step inverse-Gram events are used only in the pathwise projection bound below. Independent-innovation truncation gives

$$\mathbb{E}\left[\max_{i\in[n_t]}|z_{it}|^2\mid\mathcal{D}_{t-1}\right]\lesssim\frac{\log(nT)}{H}+\frac{n^{2/q}\log^2(nT)}{H^2}\left\{\sum_{l=1}^{H}(1+\|f_{t-l}\|)^q\right\}^{2/q}.$$

To verify this bound, truncate each coordinate's primitive summands at $H^{-1}\{n\sum_{l=1}^{H}(1+\|f_{t-l}\|)^q\}^{1/q}$ and center the truncated and discarded parts separately. Bernstein's inequality bounds the squared maximum of the truncated sums by the displayed right-hand side. Independence over lags and the qth innovation moment bound the squared L_2 norm of the maximum discarded sum by the squared truncation level. The time average of the last factor is $O_{\mathbb{P}}(H^{2/q})$ by the factor moment bound and concavity. Thus the time average of the envelope term is bounded by

$$\frac{J^{2d}}{n^2}\left\{\frac{\log^3(nT)}{H}+\frac{(nH)^{2/q}\log^4(nT)}{H^2}\right\}\lesssim\frac{J^d\log^{d+1}(nT)}{nH}.$$

After division by the right-hand side, the two ratios are bounded by constants times $(J/\sqrt{H})\log^{-d}(nT)$ and $(J/\sqrt{H})(H/T^2)^{2/q}\log^{1-d}(nT)$, respectively. This follows from the common ratio restriction, using each of its two denominator branches. Taking expectations after localization and applying Markov's inequality to the nonnegative sum over dates now give

$$\sum_{t=H+1}^{T}\sup_{P\in\mathcal{P}^\circ}\|Z_t(P)-\mathbb{E}[Z_t(P)\mid\mathcal{D}_{t-1}]\|^2\lesssim_{\mathbb{P}}\frac{TJ^d\log^{d+1}(nT)}{nH}.$$

This is a bound for a sum of squares; it does not require independence between rolling windows.

The estimated selected cell belongs to $\mathcal{P}^\circ$ w.p.a.1. Bounded basis coordinates, the selected inverse-Gram

bound, and (SA-II.7.3) imply, pathwise on the same events,

$$\sum_{i=1}^{n_t} \widehat{C}_{it}(\beta)^2 |(\widehat{\Pi}_t z_{\cdot t})_i|^2 \lesssim \frac{J^d}{n} \sup_{P \in \mathcal{P}^\circ} \|Z_t(P)\|^2.$$

Separate $Z_t(P)$ into its conditional mean and centered part in this inequality. Squaring the mean bound retains $\mathsf{B}_{\mathsf{FS}}^2$ in addition to H^{-2}.

To replace $\widehat{\beta}_{it} - \breve{\beta}_{it}$ by v_{it}, use the inverse-Gram remainder in (SA-II.7.9). Projection contraction bounds the weighted squared norm by a constant times $(J^d/n)^2$ times the sum of squared unprojected terms over the innovation-free enlarged selected support used in Step 2. This support contains $O_{\mathbb{P}}(n/J^d)$ units at each date, giving

$$\sum_{t=H+1}^{T} \sum_{i=1}^{n_t} \widehat{C}_{it}(\beta)^2 \left| \left\{ \widehat{\Pi}_t [\nabla \mu_t(\beta_{\cdot t})^\top (\widehat{\beta}_{\cdot t} - \breve{\beta}_{\cdot t} - v_{\cdot t})] \right\}_i \right|^2 \lesssim_{\mathbb{P}} \frac{T J^d}{n H^2}.$$

For the last bound, conditional on $\mathcal{E}_\infty$ the rolling innovation sum has covariance bounded by a constant times H^{-1} on the uniform Gram event. The resolvent identity and the time-averaged squared Gram bound at the end of the proof of Lemma SA-II.2.2 give

$$\frac{1}{T_0} \sum_{t=H+1}^{T} \|\widehat{A}(t)^{-1} - A(t)^{-1}\|^2 \lesssim_{\mathbb{P}} H^{-1}.$$

Conditional Markov's inequality and the support count then establish the displayed projected inverse-Gram bound. Combining the mean, centered, and Gram-replacement bounds proves

$$\sum_{t=H+1}^{T} \sum_{i=1}^{n_t} \widehat{C}_{it}(\beta)^2 \left| \left\{ \widehat{\Pi}_t [\nabla \mu_t(\beta_{\cdot t})^\top v_{\cdot t}] \right\}_i \right|^2 \lesssim_{\mathbb{P}} T \frac{J^d}{n} \left\{ \frac{J^d \log^{d+1}(nT)}{nH} + \frac{1}{H^2} + \mathsf{B}_{\mathsf{FS}}^2 \right\}.$$

For each coordinate of $\widehat{\beta}_{it} - \breve{\beta}_{it}$, the same argument applies: the basis-perturbation mean and variance calculations use a constant coordinate multiplier in place of $\nabla \mu_t(\beta_{it})$. The resulting projected arrays and weights are $\mathcal{F}_{t-1}$-measurable. Conditioning at date t and using the conditional factor moment bound therefore gives the same displayed order for $\{\widehat{\Pi}_t[\widetilde{f}_t^\top v_{\cdot t}]\}_i$. The projected inverse-Gram terms, with or without the current factor, are covered by the inverse-Gram bound above. Finally, bounded absolute row sums of the selected projection give contributions $T J^d \mathsf{B}_{\mathsf{FS}}^2/n$ and $T J^d \mathsf{R}_{\mathsf{FS}}^4/n$ for the projected first-step-bias and Taylor terms, respectively; use the same conditional factor argument for the factor-times-bias term. The square of their sum is at most a fixed constant times the sum of their squares. Hence

$$\begin{aligned} \sum_{t=H+1}^{T} \sum_{i=1}^{n_t} \widehat{C}_{it}(\beta)^2 \{\widetilde{\mathscr{R}}_{it}^{(\mathrm{pr})}\}^2 &\lesssim_{\mathbb{P}} T \frac{J^d}{n} \left\{ J^{-2\min\{s_\mu, \mathfrak{q}+1\}} + \mathsf{B}_{\mathsf{FS}}^2 + \mathsf{R}_{\mathsf{FS}}^4 + \frac{J^d \log^{d+1}(nT)}{nH} + \frac{1}{H^2} \right\} \\ &= T \frac{J^d}{nH} \mathsf{R}_{\mathsf{JKPR}}. \end{aligned} \tag{SA-II.7.22}$$

Since $2|xy| \le x^2 + y^2$ and each date has at most $2H+1$ partners in the band, this gives

$$\left| \mathscr{H}_{\widehat{C}}(\widetilde{\mathscr{R}}^{(\mathrm{pr})}, \widetilde{\mathscr{R}}^{(\mathrm{pr})}) \right| \lesssim_{\mathbb{P}} \frac{J^d}{nT} \mathsf{R}_{\mathsf{JKPR}}. \tag{SA-II.7.23}$$

Move the factor indexed by s outside the inner sum and apply scalar Cauchy–Schwarz.
Equation (SA-II.7.22) controls the first factor below. Equation (SA-II.7.20) controls the second:

$$\left|\mathscr{H}_{\widehat{C}}(\widetilde{\mathscr{R}}^{(\mathrm{md})},\widetilde{\mathscr{R}}^{(\mathrm{pr})})\right| \leq \frac{1}{T_0^2}\left\{\sum_{i,s}\widehat{C}_{is}(\beta)^2\{\widetilde{\mathscr{R}}_{is}^{(\mathrm{pr})}\}^2\right\}^{1/2}\left\{\sum_{i,s}\left(\sum_{|t-s|\leq H}\widehat{C}_{it}(\beta)\widetilde{\mathscr{R}}_{it}^{(\mathrm{md})}\right)^2\right\}^{1/2} \lesssim_{\mathbb{P}} \frac{J^d}{nT}\sqrt{\mathsf{R}_{\mathtt{JKMD}}\mathsf{R}_{\mathtt{JKPR}}}. \tag{SA-II.7.24}$$

Equation (SA-II.7.13) controls the lag windows of ε. Cauchy–Schwarz within each window and (SA-II.7.12) give the same bound for ℓ. Thus

$$\sum_{i,s}\left(\sum_{|t-s|\leq H}\widehat{C}_{it}(\beta)Y_{it}\right)^2 \lesssim_{\mathbb{P}} TH\frac{J^d}{n}.$$

Applying the same scalar Cauchy–Schwarz bound with Y in place of $\widetilde{\mathscr{R}}^{(\mathrm{md})}$ first gives the right-hand side below without the factor $\sqrt{\log(nT)}$. We retain the weaker displayed bound to match the sufficient condition in (SA-II.7.1):

$$\left|\mathscr{H}_{\widehat{C}}(Y,\widetilde{\mathscr{R}}^{(\mathrm{pr})})\right| \lesssim_{\mathbb{P}} \frac{J^d}{nT}\sqrt{\mathsf{R}_{\mathtt{JKPR}}\log(nT)}. \tag{SA-II.7.25}$$

Equations (SA-II.7.21), (SA-II.7.23), (SA-II.7.24), and (SA-II.7.25) prove (SA-II.7.8). □

Lemma SA-II.7.6 (Selected-Block Residual Bound)**.** *Suppose that the assumptions and common rate conditions stated before parts* (i)–(ii) *of Lemma SA-II.7.2 hold. For every fixed $\beta \in \mathcal{B}^\circ$, without imposing* (SA-II.7.2)*,*

$$\left|\mathscr{H}_{\widehat{C}}(\widehat{\varepsilon},\widehat{\varepsilon})\right| \lesssim_{\mathbb{P}} \frac{J^d}{nT}.$$

Proof. On the rows used by $\widehat{C}_{it}(\beta)$, the exact identity in Step 1 of the preceding proof can be written as

$$\widehat{\varepsilon}_{it} = e_{it} + r_{it}, \qquad e_{it} = \left[(I-\widehat{\Pi}_t)\varepsilon_{\cdot t}\right]_i, \qquad r_{it} = \left[(I-\widehat{\Pi}_t)M_t(\beta_{\cdot t})\right]_i,$$

and all three quantities are set to zero on unused rows. Conditional on $\mathcal{C}_t$, the selected empirical block, $\widehat{C}_{it}(\beta)$, and $\widehat{\Pi}_t$ are fixed, whereas the current innovations entering e_{it} are cross-sectionally independent and centered.

We first record two weighted squared-sum bounds. Projection contraction, the uniform cell-count and inverse-Gram bounds, and conditional second moments give

$$\sum_{t=H+1}^{T}\sum_{i=1}^{n_t}\widehat{C}_{it}(\beta)^2 e_{it}^2 \lesssim_{\mathbb{P}} T\frac{J^d}{n}. \tag{SA-II.7.26}$$

For r_{it}, polynomial reproduction removes the factor part of $M_t(\widehat{\beta}_{it})$ when $\mathfrak{q} \geq 1$, while local polynomial approximation controls its conditional-mean part. When $\mathfrak{q} = 0$, the entire within-cell variation is of order

J^{-1}. Apply the first-step decomposition (SA-II.7.9) inside r_{it}. The local polynomial residual contributes $J^{-2\min\{s_\mu,\mathsf{q}+1\}}$ after squaring. The linear first-step term has selected-cell mean square of order H^{-1}, its bias contributes $\mathsf{B}_{\mathtt{FS}}^2$, and the Taylor remainder contributes $\mathsf{R}_{\mathtt{FS}}^4$. For the projected rolling terms, the coarser H^{-1} order suffices here. On the selected block at each date $t = H+1,\dots,T$, projection contraction and $\max_{i\in[n_t]} \widehat{C}_{it}(\beta)^2 \lesssim_{\mathbb{P}} (J^d/n)^2$ bound the weighted squared norm of a projection by a constant times $(J^d/n)^2$ times the unprojected block's squared norm. The selected-cell rolling moment bound just used therefore also bounds the projected rolling contribution by $O_{\mathbb{P}}(TJ^d/(nH))$. Bounded absolute projection row sums handle the bias and Taylor pieces, and conditioning at date t handles current factor multipliers. This argument does not use the sharp anchored projected bound or impose $J\mathsf{R}_{\mathtt{FS}} = o(1)$. These bounds hold uniformly over the deterministic quasi-uniform rectangle class, so they may be evaluated at the realized selected block. Hence

$$\begin{aligned}\sum_{t=H+1}^{T}\sum_{i=1}^{n_t}\widehat{C}_{it}(\beta)^2 r_{it}^2 &\lesssim_{\mathbb{P}} T\frac{J^d}{n}\left\{\frac{1}{H} + J^{-2\min\{s_\mu,\mathsf{q}+1\}} + \mathsf{R}_{\mathtt{FS}}^4 + \mathsf{B}_{\mathtt{FS}}^2 + \frac{J^d\log^{d+1}(nT)}{nH} + \frac{1}{H^2}\right\}\\ &\lesssim_{\mathbb{P}} T\frac{J^d}{nH}. \end{aligned} \tag{SA-II.7.27}$$

The first term in braces covers both the unprojected and projected rolling first-step terms. The remaining nonnegative terms are bounded by $\mathsf{R}_{\mathtt{JKPR}}/H$, so the last inequality follows from (SA-II.7.1). This calculation does not compare the cell containing β with a reference cell. The underlying weighted moment bounds hold uniformly over the deterministic quasi-uniform rectangle class before the realized empirical block is selected.

The current-innovation array $\{\widehat{C}_{it}(\beta)e_{it}\}$ is a chronological martingale-difference array. Martingale orthogonality and (SA-II.7.26) therefore give

$$\sum_{s=H+1}^{T}\sum_{i=1}^{n}\left(\sum_{\substack{H+1\le t\le T\\ |t-s|\le H}}\widehat{C}_{it}(\beta)e_{it}\right)^2 \lesssim_{\mathbb{P}} TH\frac{J^d}{n}. \tag{SA-II.7.28}$$

In the displays below, $\sum_{s,i}$ uses these same ranges, and invalid unit–date pairs continue to contribute zero. This aggregation is important: bounding the individual time pairs first would lose a factor H. By scalar Cauchy–Schwarz after the lag aggregation, (SA-II.7.27) and (SA-II.7.28) imply

$$\left|\mathscr{H}_{\widehat{C}}(e,r)\right| \le \frac{1}{T_0^2}\left\{\sum_{s,i}\widehat{C}_{is}(\beta)^2 r_{is}^2\right\}^{1/2}\left\{\sum_{s,i}\left(\sum_{|t-s|\le H}\widehat{C}_{it}(\beta)e_{it}\right)^2\right\}^{1/2} \lesssim_{\mathbb{P}} \frac{J^d}{nT}.$$

The absolute band bound and (SA-II.7.27) likewise give

$$\left|\mathscr{H}_{\widehat{C}}(r,r)\right| \lesssim \frac{H}{T_0^2}\sum_{t,i}\widehat{C}_{it}(\beta)^2 r_{it}^2 \lesssim_{\mathbb{P}} \frac{J^d}{nT}.$$

Finally, split $\mathscr{H}_{\widehat{C}}(e,e)$ into its time-diagonal and off-diagonal parts. Conditional second moments and (SA-II.7.26) show that the diagonal is $O_{\mathbb{P}}(J^d/(nT))$. For the off-diagonal part, martingale orthogonality, the

leverage bound, and (SA-II.7.4) give

$$\left|\text{off-diagonal part of } \mathscr{H}_{\widehat{C}}(e,e)\right| = O_{\mathbb{P}}\left(\sqrt{\frac{HJ^{3d}}{n^3T^3}}\right) = o_{\mathbb{P}}\left(\frac{J^d}{nT}\right),$$

where the last equality uses $H \le T$ and $J^d/n = o(1)$. Expanding $\mathscr{H}_{\widehat{C}}(e+r, e+r)$ and combining the preceding bounds proves the result. The proof controls only magnitude and therefore does not require the selected block to be innovation-free. □

SA-II.7.6 Proof of Lemma SA-II.7.2

Step 1: common reduction. By the Sherman–Morrison identity above and the convention for invalid unit–date pairs, $\widehat{\Sigma}_\varepsilon(\beta) = \mathscr{H}_{\widehat{C}}(\widehat{\varepsilon}, \widehat{\varepsilon})$. Expanding (SA-II.7.5) in the symmetric bilinear form gives

$$\begin{aligned}\widehat{\Sigma}_\varepsilon(\beta) = {} & \mathscr{H}_{\widehat{C}}(Y,Y) + 2\mathscr{H}_{\widehat{C}}(Y, \widetilde{\mathscr{R}}^{(\mathrm{md})}) + 2\mathscr{H}_{\widehat{C}}(Y, \widetilde{\mathscr{R}}^{(\mathrm{pr})}) \\ & + \mathscr{H}_{\widehat{C}}(\widetilde{\mathscr{R}}^{(\mathrm{md})}, \widetilde{\mathscr{R}}^{(\mathrm{md})}) + 2\mathscr{H}_{\widehat{C}}(\widetilde{\mathscr{R}}^{(\mathrm{md})}, \widetilde{\mathscr{R}}^{(\mathrm{pr})}) + \mathscr{H}_{\widehat{C}}(\widetilde{\mathscr{R}}^{(\mathrm{pr})}, \widetilde{\mathscr{R}}^{(\mathrm{pr})}).\end{aligned}$$

Step 2: rough bound for every fixed evaluation point. For any fixed $\beta \in \mathcal{B}^\circ$, do not impose either $J\mathsf{R}_{\mathtt{FS}} = o(1)$ or (SA-II.7.2). Apply Lemma SA-II.7.6 directly to the common reduction in Step 1. It gives

$$\left|\widehat{\Sigma}_\varepsilon(\beta)\right| \lesssim_{\mathbb{P}} \frac{J^d}{nT}.$$

The proof of Lemma SA-II.6.1 establishes $\sum_{t,i} \varsigma_{it}^2 W_{it}(\beta)^2 \lesssim_{\mathbb{P}} TJ^d/n$. By the definition of $\Sigma_\varepsilon(\beta)$ following (SA-II.6.1), this gives $\Sigma_\varepsilon(\beta) \lesssim_{\mathbb{P}} J^d/(nT)$. The triangle inequality therefore yields

$$\widehat{\Sigma}_\varepsilon(\beta) - \Sigma_\varepsilon(\beta) = O_{\mathbb{P}}\left(\frac{J^d}{nT}\right).$$

Finally, $J^d/n = o(1)$ under (SA-II.7.1), so the last display is $o_{\mathbb{P}}(T^{-1})$. This argument controls only the magnitude of the realized selected block and therefore requires neither anchoring nor the stronger localization $J\mathsf{R}_{\mathtt{FS}} = o(1)$.

Step 3: anchored evaluation points. Suppose (SA-II.7.2) holds and $J\mathsf{R}_{\mathtt{FS}} = o(1)$. Re-indexing the primitive shocks in Y_{it} gives the reference weight

$$W_{it}(\beta) = C_{it}(\beta) - \frac{1}{H}\sum_{l=1}^{H} C_{i,t+l}(\beta) a_{itl}(\beta). \tag{SA-II.7.29}$$

This is exactly the weight $W_{it}(\beta)$ defined in Section SA-II.6, and it is $\mathcal{E}_\infty$-measurable. Consequently, Lemma SA-II.7.5 yields

$$\begin{aligned}&\widehat{\Sigma}_\varepsilon(\beta) = \mathscr{H}_C(Y,Y) + \mathscr{R}^\star_{\mathtt{JK}}(\beta), \\ &|\mathscr{R}^\star_{\mathtt{JK}}(\beta)| \lesssim_{\mathbb{P}} \frac{J^d}{nT}\rho_{\mathtt{JK}}, \\ &\quad \rho_{\mathtt{JK}} = \Delta_{\mathtt{JK}}^{1/4} + \mathsf{R}_{\mathtt{JKMD}}^{1/2} + \mathsf{R}_{\mathtt{JKMD}} + \mathsf{R}_{\mathtt{JKPR}} + \sqrt{\mathsf{R}_{\mathtt{JKPR}} \log(nT)}.\end{aligned} \tag{SA-II.7.30}$$

The mixed term in (SA-II.7.8) is absorbed here by $\sqrt{ab} \le (a+b)/2$. Lemma SA-II.3.3, $J^2 \log T/n = o(1)$, $J\mathsf{R}_{\mathtt{FS}} = o(1)$, and $T^{2/q}/H + \sqrt{\log T/H} = o(1)$ imply $\Delta_{\mathtt{JK}} = o(1)$ and $\mathsf{R}_{\mathtt{JKMD}} = o(1)$, while (SA-II.7.1) implies $\mathsf{R}_{\mathtt{JKPR}} \log(nT) = o(1)$. Hence $\rho_{\mathtt{JK}} = o(1)$.

It remains to identify the leading part of $\mathscr{H}_C(Y,Y)$. After re-indexing its primitive shocks, write

$$\begin{aligned}
\mathscr{H}_C(Y,Y) &= \mathscr{D}^{\circ}_{\mathtt{JK}}(\beta) + \mathscr{O}^{\circ}_{\mathtt{JK}}(\beta), \\
\mathscr{D}^{\circ}_{\mathtt{JK}}(\beta) &= \frac{1}{(T-H)^2} \sum_{i=1}^{n} \sum_{t=1}^{T} W_{it}(\beta)^2 \varepsilon_{it}^2, \\
\mathscr{O}^{\circ}_{\mathtt{JK}}(\beta) &:= \mathscr{H}_C(Y,Y) - \mathscr{D}^{\circ}_{\mathtt{JK}}(\beta).
\end{aligned}$$

The lag restriction is automatically satisfied between any two appearances of the same primitive shock, so its coefficient is precisely $W_{it}(\beta)^2$. Every term in $\mathscr{O}^{\circ}_{\mathtt{JK}}(\beta)$ is instead a product of two distinct primitive shocks.

Conditional on $\mathcal{E}_\infty$, these shocks are independent and centered, and therefore

$$\mathbb{E}[\mathscr{D}^{\circ}_{\mathtt{JK}}(\beta) \mid \mathcal{E}_\infty] = \Sigma_\varepsilon(\beta), \qquad \mathbb{E}[\mathscr{O}^{\circ}_{\mathtt{JK}}(\beta) \mid \mathcal{E}_\infty] = 0.$$

The local-support bounds in (SA-II.7.3), Jensen's inequality over the rolling window, and $T^{-1}\sum_t (1+\|f_t\|^4) = O_{\mathbb{P}}(1)$ give

$$\sum_{i=1}^{n} \sum_{t=1}^{T} W_{it}(\beta)^2 \lesssim_{\mathbb{P}} T \frac{J^d}{n}, \qquad \sum_{i=1}^{n} \sum_{t=1}^{T} W_{it}(\beta)^4 \lesssim_{\mathbb{P}} T \frac{J^{3d}}{n^3}.$$

The conditional fourth-moment bound thus gives

$$\mathscr{D}^{\circ}_{\mathtt{JK}}(\beta) - \Sigma_\varepsilon(\beta) = O_{\mathbb{P}}\left(\sqrt{\frac{J^{3d}}{n^3 T^3}} \right). \tag{SA-II.7.31}$$

For the off-diagonal term, (SA-II.7.4), Jensen's inequality, and $T^{-1}\sum_t (1+\|f_t\|^4) = O_{\mathbb{P}}(1)$ give the following bound on the sum of the squared coefficients:

$$\sum_{i=1}^{n} \sum_{u \neq v} \Gamma_{iuv}(C)^2 \lesssim_{\mathbb{P}} TH \frac{J^{3d}}{n^3}.$$

Conditional orthogonality of distinct-shock products therefore gives

$$\mathscr{O}^{\circ}_{\mathtt{JK}}(\beta) = O_{\mathbb{P}}\left(\sqrt{\frac{HJ^{3d}}{n^3 T^3}} \right). \tag{SA-II.7.32}$$

Combining the preceding displays preserves the explicit leading expansion

$$\begin{aligned}
\widehat{\Sigma}_\varepsilon(\beta) - \Sigma_\varepsilon(\beta) = \frac{1}{(T-H)^2} \sum_{i=1}^{n} \sum_{t=1}^{T} W_{it}(\beta)^2 \left(\varepsilon_{it}^2 - \varsigma_{it}^2\right) \\
+ \mathscr{O}^{\circ}_{\mathtt{JK}}(\beta) + \mathscr{R}^{\star}_{\mathtt{JK}}(\beta).
\end{aligned} \tag{SA-II.7.33}$$

Since $H \le T$ and $J^d/n = o(1)$, the two leading fluctuations in (SA-II.7.31)–(SA-II.7.32) are $o_{\mathbb{P}}(J^d/(nT))$; the remainder has the same conclusion by (SA-II.7.30). Thus, for every fixed anchored evaluation point—both

$\beta = 0$ and an anchored $\beta \neq 0$—

$$\widehat{\Sigma}_\varepsilon(\beta) = \Sigma_\varepsilon(\beta) + o_{\mathbb{P}}\left(\frac{J^d}{nT}\right).$$

□

SA-II.7.7 Proof of Theorem SA-II.7.3

For $\beta \neq 0$, part (i) of Lemma SA-II.7.2 and $\Sigma_\varepsilon(\beta) \lesssim_{\mathbb{P}} J^d/(nT)$ give

$$\widehat{\Sigma}_\varepsilon(\beta) - \Sigma_\varepsilon(\beta) = O_{\mathbb{P}}\left(\frac{J^d}{nT}\right) = o_{\mathbb{P}}(T^{-1}).$$

Lemma SA-II.7.1 then gives

$$\widehat{\Sigma}(\beta) - \Sigma(\beta) - \mathscr{B}_{\mathtt{F}}(\beta) = o_{\mathbb{P}}(T^{-1}).$$

For $\beta = 0$, $\Sigma_f(0) = \widehat{\Sigma}_f(0) = \mathscr{B}_{\mathtt{F}}(0) = 0$. The additive claim follows directly from part (ii) of Lemma SA-II.7.2. If $(nT/J^d)\Sigma_\varepsilon(0) \gtrsim_{\mathbb{P}} 1$, then $\Sigma(0) = \Sigma_\varepsilon(0) \gtrsim_{\mathbb{P}} J^d/(nT)$, so the additive error is also $o_{\mathbb{P}}(\Sigma(0))$. □

SA-II.7.8 Proof of Theorem SA-II.7.4

Fix an evaluation point β and let u_t be the periodwise estimation error entering the Fama–MacBeth dispersion, with $\bar{u} = (T-H)^{-1}\sum_{t=H+1}^{T} u_t$. For any variance target V_u, adding and subtracting $\mu_t(\beta)$ and $\bar{\mu}_T(\beta; H)$ gives

$$\left| \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \Big[u_t - \bar{u} + \mu_t(\beta) - \bar{\mu}_T(\beta; H) \Big]^2 - V_u - \Sigma_\mu(\beta) \right| \leq \mathscr{V}_{\mathtt{FM},1} + \mathscr{V}_{\mathtt{FM},2} + \mathscr{V}_{\mathtt{FM},3},$$

where

$$\begin{aligned}
\mathscr{V}_{\mathtt{FM},1} &= \left| \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} u_t^2 - V_u \right|, \\
\mathscr{V}_{\mathtt{FM},2} &= \frac{\bar{u}^2}{T-H}, \\
\mathscr{V}_{\mathtt{FM},3} &= \left| \frac{2}{(T-H)^2} \sum_{t=H+1}^{T} u_t \big(\mu_t(\beta) - \bar{\mu}_T(\beta; H) \big) \right|.
\end{aligned}$$

Step 1: The case $\beta \neq 0$. Take $u_t = \widehat{M}_t(\beta) - \mu_t(\beta)$ and $V_u = \Sigma_f(\beta) + \Sigma_\varepsilon(\beta)$. The left-hand side in the common display is $|\widehat{\Sigma}_{\mathtt{FM}}(\beta) - \Sigma_f(\beta) - \Sigma_\varepsilon(\beta) - \Sigma_\mu(\beta)|$.

For $\mathscr{V}_{\mathtt{FM},1}$, the per-period linearization in Lemma SA-II.4.2 gives

$$\widehat{M}_t(\beta) - \mu_t(\beta) = \beta^\top \widetilde{f}_t + \mathscr{I}_{\mathtt{FM},t}(\beta) + \mathscr{R}_{\mathtt{FM},t}(\beta),$$

where $\mathscr{I}_{\mathtt{FM},t}(\beta)$ collects the idiosyncratic leading terms and $\mathscr{R}_{\mathtt{FM},t}(\beta)$ collects the remaining linearization

errors. The same weight calculations as in the proof of Lemma SA-II.7.2 give

$$\frac{1}{(T-H)^2}\sum_{t=H+1}^{T}\mathscr{I}_{\mathtt{FM},t}(\beta)^2 \lesssim_{\mathbb{P}} \frac{J^d}{nT} = o_{\mathbb{P}}(T^{-1}),$$

and the periodwise residual bounds apply. Since

$$J^{-\min\{s_\mu,\mathsf{q}+1\}} + \mathsf{R}_{\mathtt{FM}} = o(1),$$

they give

$$\frac{1}{(T-H)^2}\sum_{t=H+1}^{T}\mathscr{R}_{\mathtt{FM},t}(\beta)^2 = o_{\mathbb{P}}(T^{-1}).$$

Hence all terms involving $\mathscr{I}_{\mathtt{FM},t}(\beta)$ or $\mathscr{R}_{\mathtt{FM},t}(\beta)$ are $o_{\mathbb{P}}(T^{-1})$ by Cauchy–Schwarz. For the factor square,

$$\frac{1}{(T-H)^2}\sum_{t=H+1}^{T}\Big[\big(\beta^\top \widetilde{f}_t\big)^2 - \mathbb{V}\mathrm{ar}(\beta^\top f_t \mid \mathcal{G}_{t-1})\Big] = o_{\mathbb{P}}(T^{-1}),$$

because the summands are martingale differences with conditional second moments bounded by Assumption SA-DGP(iv). Since $\Sigma_\varepsilon(\beta) \lesssim_{\mathbb{P}} J^d/(nT) = o(T^{-1})$, we obtain

$$\mathscr{V}_{\mathtt{FM},1} = o_{\mathbb{P}}\Big(\frac{1}{T}\Big),$$

for $\beta \neq 0$.

Next, the aggregate expansion gives $\widehat{\mu}(\beta) - \bar{\mu}_T(\beta;H) = o_{\mathbb{P}}(1)$: the factor term is $O_{\mathbb{P}}(T^{-1/2})$, the idiosyncratic term is $O_{\mathbb{P}}(\sqrt{J^d/(nT)}) = o_{\mathbb{P}}(T^{-1/2})$, and the residual terms are $o_{\mathbb{P}}(1)$ under the displayed rate condition. Therefore

$$\mathscr{V}_{\mathtt{FM},2} = o_{\mathbb{P}}\Big(\frac{1}{T}\Big).$$

Finally, for $\mathscr{V}_{\mathtt{FM},3}$, write

$$\mathscr{V}_{\mathtt{FM},3} = \Big|\frac{2}{(T-H)^2}\sum_{t=H+1}^{T}\big(\widehat{M}_t(\beta)-\mu_t(\beta)\big)\,\mu_t(\beta) \;-\; \frac{2\,\bar{\mu}_T(\beta;H)}{T-H}\big(\widehat{\mu}(\beta)-\bar{\mu}_T(\beta;H)\big)\Big|.$$

For the first term, by the linearization in Lemma SA-II.4.2, the leading term in $\widehat{M}_t(\beta)-\mu_t(\beta)$ is a martingale difference $\beta^\top \widetilde{f}_t$, and thus the corresponding component in the first term of $\mathscr{V}_{\mathtt{FM},3}$ is

$$\frac{2}{(T-H)^2}\sum_{t=H+1}^{T}\beta^\top \widetilde{f}_t\mu_t(\beta).$$

Since $\mu_t(\beta)$ is uniformly bounded, the predictable quadratic variation of this term is bounded by

$$\frac{\|\beta\|^2}{(T-H)^4}\sum_{t=H+1}^{T}\mu_t(\beta)^2\mathbb{E}[\|\widetilde{f}_t\|^2|\mathcal{G}_{t-1}] \;\lesssim\; \frac{1}{T^3},$$

so that by the Lenglart–Rebolledo inequality,

$$\frac{2}{(T-H)^2}\sum_{t=H+1}^{T}\beta^\top \widetilde{f}_t\mu_t(\beta) \lesssim_{\mathbb{P}} T^{-3/2}.$$

The idiosyncratic leading term and the remaining linearization errors contribute $o_{\mathbb{P}}(T^{-1})$ by the same predictable-variation and Cauchy–Schwarz bounds used for $\mathscr{V}_{\mathtt{FM},1}$, plus the fact that the deterministic periodwise residual scale is $o(1)$.

For the second term in $\mathscr{V}_{\mathtt{FM},3}(\beta)$, since $\bar{\mu}_T(\beta;H) \lesssim_{\mathbb{P}} 1$ and $\widehat{\mu}(\beta) - \bar{\mu}_T(\beta;H) = o_{\mathbb{P}}(1)$ by the preceding aggregate expansion,

$$\frac{\bar{\mu}_T(\beta;H)}{T-H}\big(\widehat{\mu}(\beta) - \bar{\mu}_T(\beta;H)\big) = o_{\mathbb{P}}(T^{-1}).$$

Thus,

$$\mathscr{V}_{\mathtt{FM},3} = o_{\mathbb{P}}\Big(\frac{1}{T}\Big).$$

Combining the bounds for $\mathscr{V}_{\mathtt{FM},1}$, $\mathscr{V}_{\mathtt{FM},2}$, and $\mathscr{V}_{\mathtt{FM},3}$, and using $\Sigma_\varepsilon(\beta) \lesssim_{\mathbb{P}} J^d/(nT) = o(T^{-1})$, yields the desired result for $\beta \neq 0$.

Step 2: The case $\beta = 0$. Consider the bias-corrected estimator $\widetilde{\Sigma}_{\mathtt{FM}}(0)$. The periodwise version of Lemmas SA-I.3.1 and SA-II.4.2, with the smoothing-bias component $\mathscr{B}_t(0)$ removed exactly, gives

$$\widehat{M}_t(0) - \mathscr{B}_t(0) - \mu_t(0) = \mathscr{E}_{\mathtt{o},t}(0) + \mathscr{R}_{\mathtt{FM},t}, \tag{SA-II.7.34}$$

The remainder $\mathscr{R}_{\mathtt{FM},t}$ is detailed below. Equation (SA-II.7.35) will give the time-averaged squared bound used in the proof. The rate $\mathsf{R}_{\mathtt{FM}}$ collects the periodwise versions of the residual terms in the stochastic linearization: the first-step bias and smooth-interior generated-regressor terms in $\mathsf{R}_{\mathtt{LA}}$, the first-step time-series approximation term $\log(nT)/H$, and, when $\mathfrak{q} = 0$, the curvature and endpoint-comparison terms.

In the common decomposition, take

$$u_t = \widehat{M}_t(0) - \mathscr{B}_t(0) - \mu_t(0), \qquad V_u = \Sigma_{\varepsilon,\mathtt{o}}(0).$$

Then $\bar{u} = \widehat{\mu}(0) - \mathscr{B}(0) - \bar{\mu}_T(0;H)$, and the left-hand side is $|\widetilde{\Sigma}_{\mathtt{FM}}(0) - \Sigma_{\varepsilon,\mathtt{o}}(0) - \Sigma_\mu(0)|$.

For $\mathscr{V}_{\mathtt{FM},1}$, insert (SA-II.7.34). Expanding the square gives

$$\begin{aligned}
&\frac{1}{(T-H)^2}\sum_{t=H+1}^{T}\big(\widehat{M}_t(0) - \mathscr{B}_t(0) - \mu_t(0)\big)^2 \\
&= \frac{1}{(T-H)^2}\sum_{t=H+1}^{T}\mathscr{E}_{\mathtt{o},t}(0)^2 + \frac{2}{(T-H)^2}\sum_{t=H+1}^{T}\mathscr{E}_{\mathtt{o},t}(0)\mathscr{R}_{\mathtt{FM},t} + \frac{1}{(T-H)^2}\sum_{t=H+1}^{T}\mathscr{R}_{\mathtt{FM},t}^2.
\end{aligned}$$

At the origin, $\widehat{p}_t(0) = p_t(0)$, so the periodwise leading weights below use the same reference zero-cell coordinates as Lemma SA-II.4.2. The periodwise weights and volatility scales are $\mathcal{E}_\infty$-measurable; the primitive

innovations in $\mathscr{E}_{\mathtt{o},t}(0)$ are conditionally independent over i with mean zero and unit variance. Hence

$$\mathbb{E}\big[\mathscr{E}_{\mathtt{o},t}(0)^2 \mid \mathcal{E}_\infty\big] = \sum_{i=1}^{n_t} \varsigma_{it}^2 \left(\frac{1}{n_t} p_t(0)^\top Q_t^{-1} p_t(\beta_{it})\right)^2.$$

Therefore,

$$\frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \mathscr{E}_{\mathtt{o},t}(0)^2 = \Sigma_{\varepsilon,\mathtt{o}}(0) + o_{\mathbb{P}}\left(\frac{J^d}{nT}\right),$$

where the deviation follows from conditional independence given $\mathcal{E}_\infty$ and the same fourth-moment calculation used for the jackknife variance terms. Also, $\Sigma_{\varepsilon,\mathtt{o}}(0) \lesssim_{\mathbb{P}} J^d/(nT)$. We next control the cross product without applying Cauchy–Schwarz to the full remainder. Write the periodwise summands of the linearization remainders as follows:

$$\mathscr{L}_{\mathtt{B},t}(0) = \widehat{p}_t(0)^\top \widehat{Q}_t^{-1} \frac{1}{n_t} \sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it}) [\nabla \mu_t(\beta_{it})]^\top \big(\breve{\beta}_{it} - \beta_{it}\big),$$

$$\mathscr{L}_{\mathtt{V},t}(0) = \widehat{p}_t(0)^\top \widehat{Q}_t^{-1} \frac{1}{n_t} \sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it}) [\nabla \mu_t(\beta_{it})]^\top \big(\widehat{\beta}_{it} - \breve{\beta}_{it}\big),$$

$$\mathscr{A}_t(0) = p_t(0)^\top Q_t^{-1} \frac{1}{n_t} \sum_{i=1}^{n_t} p_t(\beta_{it}) [\nabla \mu_t(\beta_{it})]^\top \frac{1}{H} \sum_{s=1}^{H} SA(t)^{-1} X_s(t) K\Big(\frac{s}{H}\Big) \varepsilon_{i(t-s)}.$$

The five periodwise remainders from Lemma SA-I.3.1 are

$$\mathscr{R}_{1,t}(0) = \widehat{p}_t(0)^\top \left(\widehat{Q}_t^{-1} - Q_t^{-1}\right) \frac{1}{n_t} \sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it}) \varepsilon_{it},$$

$$\mathscr{R}_{2,t}(0) = \widehat{p}_t(0)^\top Q_t^{-1} \frac{1}{n_t} \sum_{i=1}^{n_t} \big(\widehat{p}_t(\widehat{\beta}_{it}) - p_t(\beta_{it})\big) \varepsilon_{it},$$

$$\mathscr{R}_{3,t}(0) = \widehat{p}_t(0)^\top \widehat{Q}_t^{-1} \left(\frac{1}{n_t} \sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it}) \widehat{\beta}_{it}^\top\right) \widetilde{f}_t,$$

$$\mathscr{R}_{4,t}(0) = \widehat{p}_t(0)^\top \widehat{Q}_t^{-1} \left(\frac{1}{n_t} \sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it}) (\beta_{it} - \widehat{\beta}_{it})^\top\right) \widetilde{f}_t,$$

$$\mathscr{R}_{5,t}(0) = \frac{1}{2} \widehat{p}_t(0)^\top \widehat{Q}_t^{-1} \frac{1}{n_t} \sum_{i=1}^{n_t} \widehat{p}_t(\widehat{\beta}_{it}) \big(\widehat{\beta}_{it} - \beta_{it}\big)^\top \nabla^2 \mu_t(\tilde{\beta}_{it}) \big(\widehat{\beta}_{it} - \beta_{it}\big).$$

Thus the remainder in (SA-II.7.34) is

$$\mathscr{R}_{\mathtt{FM},t} = -\mathscr{A}_t(0) - \mathscr{L}_{\mathtt{B},t}(0) - \big(\mathscr{L}_{\mathtt{V},t}(0) - \mathscr{A}_t(0)\big) + \sum_{j=1}^{4} \mathscr{R}_{j,t}(0) - \mathscr{R}_{5,t}(0).$$

The periodwise bounds in Lemmas SA-I.3.1 and SA-II.4.2, with the sharper origin calculation for $\mathscr{A}_t(0)$,

imply

$$\frac{1}{T-H}\sum_{t=H+1}^{T}\Big[\mathscr{A}_t(0)^2+\mathscr{L}_{\mathtt{B},t}(0)^2+\big(\mathscr{L}_{\mathtt{V},t}(0)-\mathscr{A}_t(0)\big)^2+\sum_{j=1}^{5}\mathscr{R}_{j,t}(0)^2\Big]\lesssim_{\mathbb{P}}\mathsf{R}^2_{\mathtt{FM}}+o_{\mathbb{P}}\left(\frac{J^d}{n}\right). \tag{SA-II.7.35}$$

We spell out the argument for $\mathscr{L}_{\mathtt{B},t}(0)$. This term is $\mathcal{C}_t$-measurable, while $\mathbb{E}[\mathscr{E}_{\mathtt{o},t}(0)\mid\mathcal{C}_t]=0$. Hence the predictable quadratic variation of the cross product satisfies

$$\frac{1}{(T-H)^4}\sum_{t=H+1}^{T}\mathbb{E}\big[\mathscr{E}_{\mathtt{o},t}(0)^2\mathscr{L}_{\mathtt{B},t}(0)^2\mid\mathcal{C}_t\big]\lesssim_{\mathbb{P}}\frac{1}{(T-H)^4}\frac{J^d}{n}\sum_{t=H+1}^{T}\mathscr{L}_{\mathtt{B},t}(0)^2.$$

Therefore,

$$\begin{aligned}\left|\frac{1}{(T-H)^2}\sum_{t=H+1}^{T}\mathscr{E}_{\mathtt{o},t}(0)\mathscr{L}_{\mathtt{B},t}(0)\right|&\lesssim_{\mathbb{P}}\frac{1}{T^{3/2}}\sqrt{\frac{J^d}{n}}\left(\frac{1}{T-H}\sum_{t=H+1}^{T}\mathscr{L}_{\mathtt{B},t}(0)^2\right)^{1/2}\\&\lesssim_{\mathbb{P}}\frac{J^d}{nT^2}+\frac{1}{T(T-H)}\sum_{t=H+1}^{T}\mathscr{L}_{\mathtt{B},t}(0)^2\\&=o_{\mathbb{P}}\left(\frac{J^d}{nT}\right)+O_{\mathbb{P}}\left(\frac{1}{T(T-H)}\sum_{t=H+1}^{T}\mathscr{L}_{\mathtt{B},t}(0)^2\right),\end{aligned}$$

where the last line uses Young's inequality. To bound the components of $\sum_{t=H+1}^{T}\mathscr{E}_{\mathtt{o},t}(0)\mathscr{R}_{\mathtt{FM},t}$ that contain $\mathscr{A}_t(0)$, $\mathscr{L}_{\mathtt{V},t}(0)-\mathscr{A}_t(0)$, $\mathscr{R}_{3,t}(0)$, $\mathscr{R}_{4,t}(0)$, and $\mathscr{R}_{5,t}(0)$, apply the preceding predictable-variation argument used for $\mathscr{L}_{\mathtt{B},t}(0)$ after conditioning on $\mathcal{E}_\infty$ and the primitive shocks dated before t. Use Cauchy–Schwarz for any part without a conditionally mean-zero shock. Summing over these displayed components and using (SA-II.7.35) gives a contribution bounded by $\mathsf{R}^2_{\mathtt{FM}}/T+o_{\mathbb{P}}(J^d/(nT))$.

It remains to handle the two design-comparison terms. For $\mathscr{R}_{1,t}(0)$, Cauchy–Schwarz after the inverse-Gram decomposition and the Gram-convergence bound from Lemma SA-II.3.3 give $o_{\mathbb{P}}(J^d/(nT))$. For $\mathscr{R}_{2,t}(0)$, open the product with $\mathscr{E}_{\mathtt{o},t}(0)$ into diagonal and off-diagonal pieces. The predictable diagonal piece is

$$\frac{2}{(T-H)^2}\sum_{t=H+1}^{T}\sum_{i=1}^{n_t}\varsigma_{it}^2\left(\frac{1}{n_t}p_t(0)^\top Q_t^{-1}p_t(\beta_{it})\right)\left(\frac{1}{n_t}p_t(0)^\top Q_t^{-1}\big(\widehat{p}_t(\widehat{\beta}_{it})-p_t(\beta_{it})\big)\right),$$

and is bounded by the switch-probability and cell-count rates entering $\mathsf{R}^2_{\mathtt{FM}}/T$. The centered diagonal remainder and the off-diagonal piece are conditionally mean-zero sums and are $o_{\mathbb{P}}(J^d/(nT))$ by the same predictable-variation calculation as above. Therefore,

$$\left|\frac{2}{(T-H)^2}\sum_{t=H+1}^{T}\mathscr{E}_{\mathtt{o},t}(0)\mathscr{R}_{\mathtt{FM},t}\right|\lesssim_{\mathbb{P}}\frac{\mathsf{R}^2_{\mathtt{FM}}}{T}+o_{\mathbb{P}}\left(\frac{J^d}{nT}\right),$$

and

$$\frac{1}{(T-H)^2}\sum_{t=H+1}^{T}\mathscr{R}^2_{\mathtt{FM},t}\lesssim_{\mathbb{P}}\frac{\mathsf{R}^2_{\mathtt{FM}}}{T}+o_{\mathbb{P}}\left(\frac{J^d}{nT}\right).$$

Hence, we conclude that

$$\mathscr{V}_{\mathtt{FM},1} \lesssim_{\mathbb{P}} \frac{\mathsf{R}_{\mathtt{FM}}^2}{T} + o_{\mathbb{P}}\left(\frac{J^d}{nT}\right).$$

For $\mathscr{V}_{\mathtt{FM},2}$, the aggregate deviation satisfies $\widehat{\mu}(0) - \mathscr{B}(0) - \bar{\mu}_T(0;H) \lesssim_{\mathbb{P}} \sqrt{J^d/(nT)} + \mathsf{R}_{\mathtt{FM}}$ by (SA-II.7.34), so

$$\mathscr{V}_{\mathtt{FM},2} \lesssim_{\mathbb{P}} \frac{J^d}{nT^2} + \frac{\mathsf{R}_{\mathtt{FM}}^2}{T} = o_{\mathbb{P}}\left(\frac{J^d}{nT}\right) + \frac{\mathsf{R}_{\mathtt{FM}}^2}{T}.$$

For $\mathscr{V}_{\mathtt{FM},3}$, by Cauchy–Schwarz,

$$\mathscr{V}_{\mathtt{FM},3}^2 \lesssim \left(\frac{1}{(T-H)^2}\sum_{t=H+1}^{T}\left(\widehat{M}_t(0) - \mathscr{B}_t(0) - \mu_t(0)\right)^2\right)\left(\frac{1}{(T-H)^2}\sum_{t=H+1}^{T}\left(\mu_t(0) - \bar{\mu}_T(0;H)\right)^2\right),$$

where the first factor is $\lesssim_{\mathbb{P}} J^d/(nT) + \mathsf{R}_{\mathtt{FM}}^2/T$ by (SA-II.7.34), and the second equals $\Sigma_\mu(0)$. Thus,

$$\mathscr{V}_{\mathtt{FM},3} \lesssim_{\mathbb{P}} \left(\sqrt{\frac{J^d}{nT}} + \frac{\mathsf{R}_{\mathtt{FM}}}{\sqrt{T}}\right)\sqrt{\Sigma_\mu(0)}.$$

Combining the bounds for $\mathscr{V}_{\mathtt{FM},1}$, $\mathscr{V}_{\mathtt{FM},2}$, and $\mathscr{V}_{\mathtt{FM},3}$ gives the asserted bound for $\widetilde{\Sigma}_{\mathtt{FM}}(0) - \Sigma_{\varepsilon,\mathsf{o}}(0) - \Sigma_\mu(0)$. □

SA-II.8 Uniform-in-β Inference

The pointwise-in-β Gaussian approximation in Theorem SA-II.6.2 and the feasible variance estimators above give the usual confidence intervals at each fixed nonsingular evaluation point. This section converts those pointwise intervals into a family of confidence intervals whose coverage is controlled uniformly over $\beta \in \mathcal{B}^\circ$, where $\mathcal{B}^\circ$ is the interior evaluation region used in the stochastic-linearization results. The result is not a simultaneous coverage statement for the entire function; it is a uniform statement about the coverage probability of the interval indexed by the evaluation point.

The only nonstandard feature is the transition in the variance as β approaches the origin. Away from zero, the factor variance is first order and $\Sigma(\beta) \asymp_{\mathbb{P}} \|\beta\|^2/T$. At the origin, the factor component vanishes and the idiosyncratic variance is at the local-support scale $J^d/(nT)$. We therefore use the usual feasible standard error outside a shrinking neighborhood of the origin and a conservative variance floor inside that neighborhood. The floor argument deliberately avoids using the pointwise origin Gaussian approximation, so no lower bound on $\Sigma_\varepsilon(0)$ is needed in this section. Throughout this section, take $\mathcal{B}^\circ$ to contain a fixed neighborhood of 0.

SA-II.8.1 Result

We impose the stochastic-linearization and factor-CLT conditions used above, but the variance requirements are weaker than full pointwise jackknife consistency. For the factor component, we use the factor-variance estimator condition in Lemma SA-II.7.1. For the idiosyncratic component, the conservative floor only needs the local-support upper bound verified in the proof of Theorem SA-II.8.1. Thus this section does not import the primitive jackknife localization condition $J\mathsf{R}_{\mathtt{FS}} = o(1)$. We also do not impose the origin-specific lower-

bound condition $(nT/J^d)\Sigma_\varepsilon(0) \gtrsim_{\mathbb{P}} 1$ from Lemma SA-II.6.1; the conservative floor below replaces that requirement. Set

$$\mathsf{R}_{\mathtt{UI}} = \sqrt{T-H}\,\mathsf{R}_{\mathtt{CLT}}.$$

The rate $\mathsf{R}_{\mathtt{UI}}$ is the threshold scale obtained by comparing the pointwise linearization remainder with the factor standard deviation. By definition, it includes the same-cell nonconstant-basis component in $\mathsf{R}_{\mathtt{LA}}$ when $\mathfrak{q} \geq 1$. The first-step correction $\mathscr{A}(\beta)$ is part of the adjusted idiosyncratic leading term, but the uniform bound for this term is the stronger scale from Lemma SA-II.4.2. Define

$$\mathsf{R}_{\mathtt{A}}^{\circ} = \sqrt{T-H}\left(\frac{J^d \log(nT)}{nH}\right)^{1/2}.$$

Thus $\mathsf{R}_{\mathtt{A}}^{\circ}/\sqrt{T-H}$ is the uniform scale for $\mathscr{A}(\beta)$ over $\mathcal{B}^\circ$. The factor-variance estimation error is handled differently: it is a quadratic form in β and is therefore controlled relative to $\Sigma_f(\beta)$ by the factor-prediction condition in Lemma SA-II.7.1.

Let $\delta > 0$ be a possibly data-dependent factor-variance threshold. The estimated factor variance $\widehat{\Sigma}_f(\beta)$ is invariant to nonsingular rescalings of the beta coordinates and factor normalization. The associated normalized factor-risk scale is $\sqrt{(T-H)\widehat{\Sigma}_f(\beta)}$. The condition below compares all error scales that can bind with the corresponding boundary scale $\sqrt{(T-H)\delta}$. We require δ to satisfy

$$(T-H)\delta = o_{\mathbb{P}}(1), \qquad \frac{\mathsf{R}_{\mathtt{UI}} + \mathsf{R}_{\mathtt{A}}^{\circ}}{\sqrt{(T-H)\delta}} = o_{\mathbb{P}}(1) \tag{SA-II.8.1}$$

Let $z_{1-\alpha/2}$ denote the $1-\alpha/2$ quantile of the standard normal distribution. The uniform-in-β $100(1-\alpha)\%$ confidence interval is

$$[\widehat{L}_T(\beta), \widehat{U}_T(\beta)] = \widehat{\mu}(\beta) \pm z_{1-\alpha/2}\sqrt{\widehat{\Sigma}_{\mathtt{UI}}(\beta)},$$

where

$$\widehat{\Sigma}_{\mathtt{UI}}(\beta) = \begin{cases} \widehat{\Sigma}(\beta) & \text{if } \widehat{\Sigma}_f(\beta) > \delta \\ \max\{\widehat{\Sigma}(\beta), \widehat{\Sigma}_\delta\} & \text{if } \widehat{\Sigma}_f(\beta) \leq \delta \end{cases},$$

and, when the factor-variance level set is nonempty,

$$\widehat{\Sigma}_\delta = \sup_{\substack{b\in\mathcal{B}^\circ \\ \widehat{\Sigma}_f(b)=\delta}} \max\{\widehat{\Sigma}(b), \delta\} = \delta + \sup_{\substack{b\in\mathcal{B}^\circ \\ \widehat{\Sigma}_f(b)=\delta}} \max\{\widehat{\Sigma}_\varepsilon(b), 0\}.$$

The floor $\widehat{\Sigma}_\delta$ adds the threshold factor variance δ to the largest nonnegative feasible idiosyncratic component on the boundary of the shrinking factor-variance neighborhood. On the negligible complement event where the level set is empty, set $\widehat{\Sigma}_\delta = \delta$; this convention has no asymptotic effect. The level set is nonempty w.p.a.1 because $\widehat{\Sigma}_f(0) = 0$, $\widehat{\Sigma}_f(\cdot)$ is continuous, and $\mathcal{B}^\circ$ contains a fixed neighborhood of zero. The uniform quadratic-form bound used below, factor nondegeneracy, and $\mathscr{B}_{\mathtt{F}}(b) \geq 0$ give $\widehat{\Sigma}_f(b) \gtrsim_{\mathbb{P}} (T-H)^{-1}$ uniformly on any fixed small sphere around zero. Since $(T-H)\delta = o_{\mathbb{P}}(1)$, continuity then gives a point inside the sphere at which $\widehat{\Sigma}_f(b) = \delta$.

For a family of random variables $\{X_n(\beta) : \beta \in \mathcal{I}_n\}$ and a positive deterministic or data-dependent rate r_n, write $X_n(\beta) = o_{\mathbb{P},\mathcal{I}_n}(r_n)$ if, for every $\epsilon > 0$,

$$\sup_{\beta\in\mathcal{I}_n} \mathbb{P}(|X_n(\beta)| > \epsilon r_n) \to 0,$$

and $X_n(\beta) \lesssim_{\mathbb{P},\mathcal{I}_n} r_n$ if

$$\lim_{M\to\infty} \limsup_{n,T\to\infty} \sup_{\beta\in\mathcal{I}_n} \mathbb{P}(|X_n(\beta)| > M r_n) = 0.$$

These are pointwise-uniform probability statements; they are the right notion for the infimum of pointwise coverage probabilities below and do not assert a realized supremum bound unless one is displayed explicitly.

Theorem SA-II.8.1 (Uniform Coverage of Confidence Intervals)**.** *Assume that the conditions of Corollary SA-II.5.2 hold, that (SA-II.7.1) holds, and that the factor-variance estimator satisfies the conditions of Lemma SA-II.7.1. If δ satisfies (SA-II.8.1), then*

$$\liminf_{n,T\to\infty} \inf_{\beta\in\mathcal{B}^\circ} \mathbb{P}\Big(\bar{\mu}_T(\beta; H) \in [\widehat{L}_T(\beta), \widehat{U}_T(\beta)]\Big) \geq 1 - \alpha.$$

The theorem does not require a pointwise Gaussian approximation at $\beta = 0$, and the origin-specific lower-bound condition $(nT/J^d)\Sigma_\varepsilon(0) \gtrsim_{\mathbb{P}} 1$ is not imposed. Inside the factor-variance neighborhood $\widehat{\Sigma}_f(\beta) \leq \delta$, the interval is calibrated using the boundary variance floor $\widehat{\Sigma}_\delta$. The condition (SA-II.8.1), together with local support of the portfolio weights, ensures that the idiosyncratic leading term, the uniform first-step correction, and the full linearization remainder are negligible relative to that boundary standard-error scale.

SA-II.8.2 Proof of Theorem SA-II.8.1

We first record the uniform consequences of the linearization and variance bounds. The pointwise Gaussian approximation at the origin is not used.

The stochastic linearization in Lemmas SA-I.3.1 and SA-II.4.2, together with Section SA-II.6, gives

$$\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H) = \mathscr{G}(\beta) + \bar{\mathscr{S}}_\varepsilon(\beta) + \mathscr{R}(\beta), \qquad \bar{\mathscr{S}}_\varepsilon(\beta) = \mathscr{E}(\beta) - \mathscr{A}(\beta).$$

Thus the first-step correction $\mathscr{A}(\beta)$ is not part of the non-leading remainder, but it must be controlled at its own uniform scale. The rate $\mathsf{R}_{\mathtt{CLT}}$ controls only $\mathscr{R}(\beta)$, while Lemma SA-II.4.2 controls $\mathscr{A}(\beta)$:

$$|\mathscr{R}(\beta)| \lesssim_{\mathbb{P},\mathcal{B}^\circ} \mathsf{R}_{\mathtt{CLT}},$$
$$|\mathscr{A}(\beta)| \lesssim_{\mathbb{P},\mathcal{B}^\circ} \frac{\mathsf{R}_{\mathtt{A}}^\circ}{\sqrt{T-H}}.$$

The proofs of Lemmas SA-II.6.1 and SA-II.7.6 take their local-support and selected-block bounds uniformly over the deterministic quasi-uniform rectangle class before the block containing β is selected. Their constants therefore do not depend on β, so

$$\Sigma_\varepsilon(\beta) + |\widehat{\Sigma}_\varepsilon(\beta)| \lesssim_{\mathbb{P},\mathcal{B}^\circ} \frac{J^d}{n(T-H)}.$$

This is dominated by $(\mathsf{R}_{\mathtt{A}}^\circ)^2/(T-H)$ under the maintained rolling-window scaling. Hence (SA-II.8.1) and

the definitions of $\mathsf{R}_{\mathtt{UI}}$ and $\mathsf{R}^{\circ}_{\mathtt{A}}$ imply

$$\begin{aligned}\mathscr{R}(\beta) &= o_{\mathbb{P},\mathcal{B}^\circ}(\sqrt{\delta}),\\ \mathscr{A}(\beta) &= o_{\mathbb{P},\mathcal{B}^\circ}(\sqrt{\delta}),\\ \Sigma_{\varepsilon}(\beta) + |\widehat{\Sigma}_{\varepsilon}(\beta)| &= o_{\mathbb{P},\mathcal{B}^\circ}(\delta).\end{aligned}$$

The bound for $\mathscr{A}(\beta)$ is the uniform first-step-correction bound from Lemma SA-II.4.2. For the direct idiosyncratic term $\mathscr{E}(\beta)$, condition period by period on the field $\mathcal{C}_t$ from Section SA-II.1.3. Its coefficients are fixed under this conditioning, and its predictable variance is the direct variance $\Sigma_{\varepsilon,\mathsf{o}}(\beta)$ defined in the Fama–MacBeth variance subsection above, not the adjusted reference scale $\Sigma_{\varepsilon}(\beta)$. The same local-support calculation used for $\mathscr{E}(\beta)$ in Lemma SA-I.3.1 gives

$$\Sigma_{\varepsilon,\mathsf{o}}(\beta) \lesssim_{\mathbb{P},\mathcal{B}^\circ} \frac{J^d}{n(T-H)} = o_{\mathbb{P},\mathcal{B}^\circ}(\delta).$$

Because δ may be data dependent, choose a deterministic sequence $a_n \to \infty$ such that $\mathbb{P}\big(\delta < a_n J^d/[n(T-H)]\big) \to 0$. The preceding conditional-variance bound and conditional Chebyshev's inequality at the deterministic threshold then imply that for every $\epsilon > 0$,

$$\sup_{\beta\in\mathcal{B}^\circ} \mathbb{P}\left(|\mathscr{E}(\beta)| > \epsilon\sqrt{\delta}\right) \le \mathbb{P}\left(\delta < \frac{a_n J^d}{n(T-H)}\right) + \sup_{\beta\in\mathcal{B}^\circ} \mathbb{P}\left(|\mathscr{E}(\beta)| > \epsilon\sqrt{\frac{a_n J^d}{n(T-H)}}\right) \to 0,$$

or equivalently $\mathscr{E}(\beta) = o_{\mathbb{P},\mathcal{B}^\circ}(\sqrt{\delta})$. Here the supremum is taken over the pointwise probabilities after applying Chebyshev's inequality; no realized supremum of $\Sigma_{\varepsilon,\mathsf{o}}(\beta)$ or $\Sigma_{\varepsilon}(\beta)$ is invoked. Consequently, $\bar{\mathscr{S}}_{\varepsilon}(\beta) = \mathscr{E}(\beta) - \mathscr{A}(\beta) = o_{\mathbb{P},\mathcal{B}^\circ}(\sqrt{\delta})$.

Next, record the uniform quadratic-form version of Lemma SA-II.7.1. Write

$$\widehat{\Sigma}_f(\beta) - \Sigma_f(\beta) - \mathscr{B}_{\mathsf{F}}(\beta) = \beta^\top \Delta_{\mathsf{F}} \beta,$$

where

$$\Delta_{\mathsf{F}} = \frac{1}{(T-H)^2}\sum_{t=H+1}^{T}\left\{\widetilde{f}_t\widetilde{f}_t^\top - \mathbb{E}[\widetilde{f}_t\widetilde{f}_t^\top \mid \mathcal{G}_{t-1}]\right\} + \frac{2}{(T-H)^2}\sum_{t=H+1}^{T}\widetilde{f}_t(m_t - \widehat{m}_t)^\top.$$

Because d is fixed and $\mathcal{B}^\circ$ is bounded, the entrywise martingale bounds in the proof of Lemma SA-II.7.1 and the factor-prediction condition in Lemma SA-II.7.1 give

$$\|\Delta_{\mathsf{F}}\| = o_{\mathbb{P}}\big((T-H)^{-1}\big).$$

Then, by the factor nondegeneracy condition in Assumption SA-DGP(iv), we have

$$\omega_n := \sup_{\substack{\beta\in\mathcal{B}^\circ\\ \beta\neq 0}} \frac{\left|\widehat{\Sigma}_f(\beta) - \Sigma_f(\beta) - \mathscr{B}_{\mathsf{F}}(\beta)\right|}{\Sigma_f(\beta)} = o_{\mathbb{P}}(1),$$

while the numerator is zero at $\beta = 0$. Since $\mathscr{B}_{\mathtt{F}}(\beta) \geq 0$, this implies

$$\Big(\Sigma_f(\beta) - \widehat{\Sigma}_f(\beta)\Big)_+ \leq \omega_n \Sigma_f(\beta),$$

where $(a)_+ := \max\{a, 0\}$ for any $a \in \mathbb{R}$. Combining this display with $|\widehat{\Sigma}_\varepsilon(\beta)| = o_{\mathbb{P},\mathcal{B}^\circ}(\delta)$ and $\widehat{\Sigma}(\beta) = \widehat{\Sigma}_f(\beta) + \widehat{\Sigma}_\varepsilon(\beta)$ gives the pointwise-uniform lower bound

$$\left(1 - \frac{\widehat{\Sigma}_{\mathtt{UI}}(\beta)}{\max\{\Sigma_f(\beta), \delta\}}\right)_+ = o_{\mathbb{P},\mathcal{B}^\circ}(1). \tag{SA-II.8.2}$$

To see this, consider the two regions used in the construction. If $\widehat{\Sigma}_f(\beta) > \delta$, then $\widehat{\Sigma}_{\mathtt{UI}}(\beta) = \widehat{\Sigma}(\beta)$ and the centered factor-variance bound gives

$$\widehat{\Sigma}_{\mathtt{UI}}(\beta) \geq (1 - \omega_n)\Sigma_f(\beta) - o_{\mathbb{P},\mathcal{B}^\circ}(\delta), \qquad \widehat{\Sigma}_{\mathtt{UI}}(\beta) \geq \delta - o_{\mathbb{P},\mathcal{B}^\circ}(\delta).$$

Thus (SA-II.8.2) holds on the outer region. If $\widehat{\Sigma}_f(\beta) \leq \delta$, since $\widehat{\Sigma}_f(\beta) \geq (1-\omega_n)\Sigma_f(\beta)$, $\Sigma_f(\beta) \leq (1+o_{\mathbb{P}}(1))\delta$ uniformly on this region. Also, $\widehat{\Sigma}_{\mathtt{UI}}(\beta) \geq \widehat{\Sigma}_\delta \geq \delta$, where the last inequality follows from the definition of the variance floor. This gives the same ratio lower bound on the inner region. In particular,

$$\left(1 - \frac{\widehat{\Sigma}_{\mathtt{UI}}(\beta)}{\delta}\right)_+ = o_{\mathbb{P},\mathcal{B}^\circ}(1). \tag{SA-II.8.3}$$

Next, write

$$\mathscr{G}(\beta) = \beta^\top \frac{1}{T-H} \sum_{t=H+1}^{T} \widetilde{f}_t.$$

Let

$$S_T := \sum_{t=H+1}^{T} \widetilde{f}_t, \qquad V_T := \sum_{t=H+1}^{T} \mathbb{V}\mathrm{ar}(\widetilde{f}_t \mid \mathcal{G}_{t-1}).$$

Then

$$\Sigma_f(\beta) = \frac{\beta^\top V_T \beta}{(T-H)^2}, \qquad \left\| \frac{V_T}{T-H} - \Omega_f \right\| = o_{\mathbb{P}}(1)$$

by (SA-II.1.1).

For $u \in \mathbb{S}^{d-1}$, define the deterministically normalized process

$$Z_T^0(u) := \frac{u^\top S_T}{\sqrt{(T-H)u^\top \Omega_f u}}.$$

For each fixed u, the sum of the conditional variances of its martingale increments converges in probability to one:

$$\frac{u^\top V_T u}{(T-H)u^\top \Omega_f u} \to_{\mathbb{P}} 1.$$

The conditional Lyapunov bound is uniform over directions because $\lambda_{\min}(\Omega_f) > 0$:

$$\sup_{u\in\mathbb{S}^{d-1}} \frac{\sum_{t=H+1}^{T} \mathbb{E}[|u^\top \widetilde{f}_t|^4 \mid \mathcal{G}_{t-1}]}{(T-H)^2(u^\top \Omega_f u)^2} \lesssim_{\mathbb{P}} (T-H)^{-1} = o_{\mathbb{P}}(1).$$

The scalar martingale CLT therefore gives

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}\left(Z_T^0(u) \le x\right) - \Phi(x)\right| \to 0$$

for each fixed u.

We now make this approximation uniform over directions. Fix $\epsilon > 0$ and let $\mathcal{N}_\epsilon$ be a finite ϵ-net of $\mathbb{S}^{d-1}$. The preceding display holds uniformly over the fixed finite set $\mathcal{N}_\epsilon$. Positive definiteness of Ω_f and martingale orthogonality give

$$\sup_{\|u-v\|\le\epsilon} |Z_T^0(u) - Z_T^0(v)| \lesssim \epsilon \frac{\|S_T\|}{\sqrt{T-H}} = O_{\mathbb{P}}(\epsilon), \qquad \frac{\|S_T\|}{\sqrt{T-H}} = O_{\mathbb{P}}(1).$$

The finite-net argument and normal anticoncentration thus imply

$$\sup_{u\in\mathbb{S}^{d-1}} \sup_{x\in\mathbb{R}} \left|\mathbb{P}\left(Z_T^0(u) \le x\right) - \Phi(x)\right| \to 0. \tag{SA-II.8.4}$$

Finally, stabilization gives

$$\sup_{u\in\mathbb{S}^{d-1}} \left| \frac{u^\top V_T u}{(T-H)u^\top \Omega_f u} - 1 \right| \le \frac{\|V_T/(T-H) - \Omega_f\|}{\lambda_{\min}(\Omega_f)} = o_{\mathbb{P}}(1).$$

Since $\sup_{u\in\mathbb{S}^{d-1}} |Z_T^0(u)| \lesssim \|S_T\|/\sqrt{T-H} = O_{\mathbb{P}}(1)$,

$$\sup_{u\in\mathbb{S}^{d-1}} \left| \frac{u^\top S_T}{\sqrt{u^\top V_T u}} - Z_T^0(u) \right| = o_{\mathbb{P}}(1).$$

Combining this uniform replacement with (SA-II.8.4) yields

$$\sup_{u\in\mathbb{S}^{d-1}} \sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\frac{u^\top S_T}{\sqrt{u^\top V_T u}} \le x \right) - \Phi(x) \right| \to 0.$$

Thus the factor Gaussian approximation is uniform after writing $\beta = \|\beta\| u$; at $\beta = 0$, $\mathscr{G}(0) = 0$. Since $\max\{\Sigma_f(\beta), \delta\} \ge \Sigma_f(\beta)$, the event with the conservative normalizer contains the event with the true factor normalizer for every realization of δ. Therefore, for any fixed $c \ge 0$,

$$\inf_{\beta\in\mathcal{B}^\circ} \mathbb{P}\left(|\mathscr{G}(\beta)| \le c\sqrt{\max\{\Sigma_f(\beta), \delta\}} \right) \ge 2\Phi(c) - 1 + o(1).$$

This statement does not condition on the data-dependent inner or outer region; the conservative scale $\max\{\Sigma_f(\beta), \delta\}$ dominates the true factor variance for every β and every realization.

Finally, (SA-II.8.3) and $\bar{\mathscr{S}}_\varepsilon(\beta) + \mathscr{R}(\beta) = o_{\mathbb{P},\mathcal{B}^\circ}(\sqrt{\delta})$ give

$$\frac{|\bar{\mathscr{S}}_\varepsilon(\beta) + \mathscr{R}(\beta)|}{\sqrt{\widehat{\Sigma}_{\mathtt{UI}}(\beta)}} = o_{\mathbb{P},\mathcal{B}^\circ}(1).$$

To assemble the coverage event, fix $\eta > 0$ with $\eta < z_{1-\alpha/2}$. The probability of the following event tends to one, uniformly over $\beta \in \mathcal{B}^\circ$ in the pointwise sense:

$$|\bar{\mathscr{S}}_\varepsilon(\beta) + \mathscr{R}(\beta)| \le \eta\sqrt{\widehat{\Sigma}_{\mathtt{UI}}(\beta)} \quad \text{and} \quad \widehat{\Sigma}_{\mathtt{UI}}(\beta) \ge (1-\eta)\max\{\Sigma_f(\beta), \delta\}.$$

On this event,

$$\left\{|\mathscr{G}(\beta)| \le (z_{1-\alpha/2} - \eta)\sqrt{1-\eta}\,\sqrt{\max\{\Sigma_f(\beta), \delta\}}\right\} \subseteq \left\{|\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H)| \le z_{1-\alpha/2}\sqrt{\widehat{\Sigma}_{\mathtt{UI}}(\beta)}\right\}.$$

The factor bound above with $c = (z_{1-\alpha/2} - \eta)\sqrt{1-\eta}$ gives

$$\inf_{\beta\in\mathcal{B}^\circ} \mathbb{P}\Big(|\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H)| \le z_{1-\alpha/2}\sqrt{\widehat{\Sigma}_{\mathtt{UI}}(\beta)}\Big) \ge 2\Phi\Big((z_{1-\alpha/2} - \eta)\sqrt{1-\eta}\Big) - 1 + o(1).$$

Letting $\eta \downarrow 0$ yields the claimed lower bound $1-\alpha$. The boundary $\widehat{\Sigma}_f(\beta) = \delta$ is covered by the inner formula in the definition of $\widehat{\Sigma}_{\mathtt{UI}}(\beta)$, so no separate boundary argument is needed. The displayed lower bound is the claimed uniform coverage. □

SA-II.9 Factor-Cancelling Linear Contrasts

This section develops inference for finite linear contrasts that eliminate the factor component in the pointwise stochastic expansion. Let K_Ψ be fixed, let $\mathsf{b}_1, \ldots, \mathsf{b}_{K_\Psi} \in \mathcal{B}^\circ$, and let $a_1, \ldots, a_{K_\Psi}$ be fixed real weights. The contrast of interest is

$$\widehat{\Psi} = \sum_{k=1}^{K_\Psi} a_k \widehat{\mu}(\mathsf{b}_k), \qquad \bar{\Psi}_T = \sum_{k=1}^{K_\Psi} a_k \bar{\mu}_T(\mathsf{b}_k; H).$$

The key restriction is the factor-cancelling condition

$$\sum_{k=1}^{K_\Psi} a_k \mathsf{b}_k = 0. \tag{SA-II.9.1}$$

For example, the butterfly spread corresponds to $K_\Psi = 3$, $(a_1, a_2, a_3) = (1, -2, 1)$, and $\mathsf{b}_2 = (\mathsf{b}_1 + \mathsf{b}_3)/2$.

SA-II.9.1 Setup and Linearization

For any scalar term $A(\cdot)$, write its contrast as

$$A_\Psi = \sum_{k=1}^{K_\Psi} a_k A(\mathsf{b}_k).$$

Applying this operator to the pointwise linearization in Section SA-II.6 gives

$$\widehat{\Psi} - \bar{\Psi}_T = \bar{\mathscr{S}}_{\varepsilon\Psi} + \mathscr{R}_\Psi, \tag{SA-II.9.2}$$

because

$$\bar{\mathscr{S}}_{f\Psi} = \frac{1}{T-H}\sum_{t=H+1}^{T}\left(\sum_{k=1}^{K_\Psi} a_k \mathsf{b}_k\right)^\top \widetilde{f}_t = 0.$$

The idiosyncratic leading term can be written as

$$\bar{\mathscr{S}}_{\varepsilon\Psi} = \frac{1}{T-H}\sum_{t=1}^{T}\sum_{i=1}^{n_t} \varsigma_{it}\eta_{it} W_{it,\Psi}, \qquad W_{it,\Psi} = \sum_{k=1}^{K_\Psi} a_k W_{it}(\mathsf{b}_k).$$

Since K_Ψ is fixed, the leading-weight bounds used in Lemma SA-II.6.1 imply

$$\left|\bar{\mathscr{S}}_{\varepsilon\Psi}\right| \lesssim_{\mathbb{P}} \sqrt{\frac{J^d}{nT}}.$$

Because K_Ψ and the contrast weights are fixed, define the deterministic rate

$$\mathsf{R}_\Psi := \left(\sum_{k=1}^{K_\Psi} |a_k|\right) \mathsf{R}_{\mathtt{CLT}}.$$

The triangle inequality and the common pointwise remainder bound give

$$|\mathscr{R}_\Psi| \lesssim_{\mathbb{P}} \mathsf{R}_\Psi.$$

The condition needed for idiosyncratic-rate inference is

$$\sqrt{\frac{nT}{J^d}}\mathsf{R}_\Psi \to 0. \tag{SA-II.9.3}$$

In particular, the second-step smoothing-bias component requires

$$\frac{nT}{J^{d+2\min\{s_\mu,\mathfrak{q}+1\}}} \to 0,$$

the same smoothing-bias restriction as in the pointwise $\beta = 0$ theory. Thus, $\mathfrak{q} = 0$ cannot satisfy the sufficient rate conditions for factor-cancelling contrast inference.

SA-II.9.2 Evaluation-Point-Anchored Partitioning

For factor-cancelling contrasts, the second-step partition is constructed so that the evaluation points themselves are fixed marginal breakpoints. This is the analogue of the split-at-zero construction in Section SA-II.3: the cell containing each evaluation point has a nonrandom index, because each coordinate of the evaluation point is assigned to the subsample immediately to its right. For notational simplicity, we state the construction under the convention that the evaluation points have distinct coordinates in every dimension. If several evaluation points share a coordinate in some dimension, the same construction uses the sorted

unique coordinates in that dimension; the number of fixed breakpoints, and hence the number of marginal subsamples, is then smaller than K_Ψ and $K_\Psi + 1$, respectively.

We describe the construction first. For each coordinate $l \in [d]$, order the fixed evaluation-point coordinates as

$$\underline{\beta}_l < \mathsf{b}_{\pi_l(1),l} < \cdots < \mathsf{b}_{\pi_l(K_\Psi),l} < \bar{\beta}_l,$$

where π_l is a permutation of $[K_\Psi]$. Set $\gamma_{0,l} = \underline{\beta}_l$, $\gamma_{r,l} = \mathsf{b}_{\pi_l(r),l}$ for $r = 1, \ldots, K_\Psi$, and $\gamma_{K_\Psi+1,l} = \bar{\beta}_l$. These fixed breakpoints split coordinate l into the $K_\Psi + 1$ subsamples

$$\mathcal{I}_{r,l} = [\gamma_{r,l}, \gamma_{r+1,l}), \qquad r = 0, \ldots, K_\Psi - 1, \qquad \mathcal{I}_{K_\Psi,l} = [\gamma_{K_\Psi,l}, \gamma_{K_\Psi+1,l}],$$

with ties at a fixed breakpoint assigned to the interval on the right. Thus, if $\mathsf{b}_{k,l} = \gamma_{r,l}$, then $\mathsf{b}_{k,l} \in \mathcal{I}_{r,l}$.

Within each subsample, form marginal empirical quantile-spaced intervals of the generated regressor. Specifically, for $t = H + 1, \ldots, T$, $l \in [d]$, and $r = 0, \ldots, K_\Psi$, let

$$\begin{aligned}
\widehat{N}_{t,l,r} &= \sum_{i=1}^{n_t} \mathbb{1}(\widehat{\beta}_{it,l} \in \mathcal{I}_{r,l}), \\
\widehat{F}_{\widehat{\beta},t,l,r}(z) &= \frac{1}{\widehat{N}_{t,l,r}} \sum_{i=1}^{n_t} \mathbb{1}(\widehat{\beta}_{it,l} \in \mathcal{I}_{r,l}) \mathbb{1}(\widehat{\beta}_{it,l} \le z), \qquad z \in \mathcal{I}_{r,l},
\end{aligned}$$

on the event $\widehat{N}_{t,l,r} > 0$, with generalized inverse on $\mathcal{I}_{r,l}$. Given integers $J_{t,l,r}$, define the empirical breakpoints by

$$\hat{\tau}_{0t,l,r} = \gamma_{r,l}, \qquad \hat{\tau}_{J_{t,l,r}t,l,r} = \gamma_{r+1,l}, \qquad \hat{\tau}_{jt,l,r} = \widehat{F}^{-1}_{\widehat{\beta},t,l,r}\left(\frac{j}{J_{t,l,r}}\right), \qquad j = 1, \ldots, J_{t,l,r} - 1.$$

The marginal partition $\widehat{\Delta}_{t,l}$ is the ordered collection of intervals generated by these breakpoints over $r = 0, \ldots, K_\Psi$, and the full partition is the tensor product $\widehat{\Delta}_t = \otimes_{l \in [d]} \widehat{\Delta}_{t,l}$.

Assumption SA-PA-G (Evaluation-point-anchored partition). *$\mathsf{b}_1, \ldots, \mathsf{b}_{K_\Psi}$ are fixed evaluation points with distinct coordinates in each dimension and lie strictly inside $\mathcal{B}^\circ$. For every $t = H + 1, \ldots, T$, the estimated partition $\widehat{\Delta}_t$ is formed by the evaluation-point-based marginal-quantile construction described above, with $J_{t,l,r} \asymp J$ uniformly over t, l, r, $J \to \infty$, and $J^d/n = o(1)$.*

In this scenario, the reference partition is the conditional counterpart of the same construction. For each t, l, r, define

$$\begin{aligned}
\bar{\pi}_{t,l,r} &= \frac{1}{n_t} \sum_{i=1}^{n_t} \mathbb{P}(\widehat{\beta}_{it,l} \in \mathcal{I}_{r,l} \mid \mathcal{D}_{t-1}), \\
\varrho_{it,l,r} &= \frac{\mathbb{P}(\widehat{\beta}_{it,l} \in \mathcal{I}_{r,l} \mid \mathcal{D}_{t-1})}{\bar{\pi}_{t,l,r}}, \\
\bar{F}_{\widehat{\beta},t,l,r}(z) &= \frac{1}{n_t} \sum_{i=1}^{n_t} \varrho_{it,l,r} \mathbb{P}(\widehat{\beta}_{it,l} \le z \mid \widehat{\beta}_{it,l} \in \mathcal{I}_{r,l}, \mathcal{D}_{t-1}), \qquad z \in \mathcal{I}_{r,l}.
\end{aligned}$$

Let

$$\tau_{0t,l,r} = \gamma_{r,l}, \qquad \tau_{J_{t,l,r}t,l,r} = \gamma_{r+1,l}, \qquad \tau_{jt,l,r} = \bar{F}^{-1}_{\widehat{\beta},t,l,r}\left(\frac{j}{J_{t,l,r}}\right), \qquad j = 1, \ldots, J_{t,l,r} - 1.$$

The reference marginal partition $\Delta_{t,l}$ and tensor-product partition Δ_t are formed from these conditional breakpoints. Cell normalizations, raw-regressor membership conventions, and reference-basis definitions are the same as in Section SA-II.3.

Under Assumption SA-PA-G, each evaluation point b_k belongs to a cell with a deterministic tensor-product label. On that cell, every coordinate of b_k is a fixed lower endpoint of both the estimated and reference cell. Therefore the midpoint and side-length normalization gives

$$\widehat{p}_t(\mathsf{b}_k) = p_t(\mathsf{b}_k), \qquad k = 1, \ldots, K_\Psi,$$

for every polynomial degree $\mathsf{q} \geq 0$, exactly as the split-at-zero construction gives $\widehat{p}_t(0) = p_t(0)$ in Section SA-II.3.

SA-II.9.3 Result

The relevant variance is the variance of the factor-cancelling idiosyncratic leading term,

$$\Sigma_{\varepsilon\Psi} = \frac{1}{(T-H)^2}\sum_{t=1}^{T}\sum_{i=1}^{n_t}\varsigma_{it}^2 W_{it,\Psi}^2.$$

The bin-restricted jackknife from Section SA-II.7 extends by applying the same contrast to the delete-i pseudo-values:

$$\widehat{\Sigma}_{\varepsilon\Psi} = \frac{1}{(T-H)^2}\sum_{i=1}^{n}\sum_{t=H+1}^{T}\sum_{s=H+1}^{T}\Delta_{it\Psi}\Delta_{is\Psi}\mathbb{1}(|t-s| \leq H),$$

where

$$\Delta_{it\Psi} = \sum_{k=1}^{K_\Psi} a_k \Delta_{it}(\mathsf{b}_k), \qquad \Delta_{it}(\beta) = \widehat{M}_t^{(-i)}(\beta) - \widehat{M}_t(\beta).$$

Theorem SA-II.9.1 (Factor-Cancelling Contrast Inference)**.**
Let K_Ψ be fixed, $\mathsf{b}_1, \ldots, \mathsf{b}_{K_\Psi} \in \mathcal{B}^\circ$ be fixed evaluation points, and $a_1, \ldots, a_{K_\Psi}$ be fixed contrast weights satisfying (SA-II.9.1)*. Suppose Assumptions SA-DGP, SA-FS, SA-PA-G, SA-SEP, and SA-FSB hold, and that $s_\mu \geq 2$. Assume that the rate conditions* (SA-II.4.1)*,* (SA-II.7.1)*, and* (SA-II.9.3) *hold, that*

$$\frac{T^{2/q}}{H} + \sqrt{\frac{\log T}{H}} = o(1),$$

and assume further that $J\mathsf{R}_{\mathtt{FS}} = o(1)$ and $J^2 \log T/n = o(1)$. If

$$\frac{nT}{J^d}\Sigma_{\varepsilon\Psi} \gtrsim_{\mathbb{P}} 1,$$

then

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\widehat{\Psi}-\bar{\Psi}_T}{\sqrt{\Sigma_{\varepsilon\Psi}}}\le x\right)-\Phi(x)\right|\to 0,$$

and

$$\widehat{\Sigma}_{\varepsilon\Psi}=\Sigma_{\varepsilon\Psi}+o_{\mathbb{P}}\left(\frac{J^d}{nT}\right).$$

SA-II.9.4 Discussion

(i) *Confidence interval.* The theorem gives both the Gaussian approximation and relative consistency of $\widehat{\Sigma}_{\varepsilon\Psi}$ under the nondegeneracy condition. Therefore Slutsky's lemma implies that

$$\widehat{\Psi}\pm z_{1-\alpha/2}\sqrt{\widehat{\Sigma}_{\varepsilon\Psi}}$$

has asymptotic coverage $1-\alpha$.

(ii) *Butterfly spreads.* The butterfly contrast is the special case $\widehat{\Psi}=\widehat{\mu}(\mathsf{b}_1)-2\widehat{\mu}(\mathsf{b}_2)+\widehat{\mu}(\mathsf{b}_3)$ with $\mathsf{b}_2=(\mathsf{b}_1+\mathsf{b}_3)/2$. The factor component vanishes because $\mathsf{b}_1-2\mathsf{b}_2+\mathsf{b}_3=0$, so no factor variance estimator is needed.

(iii) *Pairwise differences.* The pairwise difference $\widehat{D}=\widehat{\mu}(\mathsf{b}_2)-\widehat{\mu}(\mathsf{b}_1)$ does not cancel the factor component:

$$\mathscr{S}_{fD}=\frac{1}{T-H}\sum_{t=H+1}^{T}(\mathsf{b}_2-\mathsf{b}_1)^\top\widetilde{f}_t.$$

Its factor variance contribution is

$$\frac{1}{(T-H)^2}(\mathsf{b}_2-\mathsf{b}_1)^\top\sum_{t=H+1}^{T}\mathbb{V}\mathrm{ar}(\widetilde{f}_t\mid\mathcal{G}_{t-1})(\mathsf{b}_2-\mathsf{b}_1).$$

A two-point difference with weights $(-1,1)$ cancels the factor term only in the degenerate case $\mathsf{b}_2=\mathsf{b}_1$, so a nontrivial two-point difference is outside Theorem SA-II.9.1.

(iv) *Other difference-in-differences.* A quadruple contrast with weights $(1,-1,-1,1)$ also removes the factor component whenever $\mathsf{b}_1+\mathsf{b}_4=\mathsf{b}_2+\mathsf{b}_3$. It targets the difference-in-differences restriction

$$\bar{\mu}_T(\mathsf{b}_1;H)-\bar{\mu}_T(\mathsf{b}_2;H)=\bar{\mu}_T(\mathsf{b}_3;H)-\bar{\mu}_T(\mathsf{b}_4;H).$$

The theorem applies by taking $K_\Psi=4$ and the corresponding weights.

SA-II.9.5 Proof of Theorem SA-II.9.1

By (SA-II.9.1), the factor component in the pointwise linearization cancels exactly, giving (SA-II.9.2). Under (SA-II.9.3), the residual satisfies

$$|\mathscr{R}_\Psi|=o_{\mathbb{P}}\left(\sqrt{\frac{J^d}{nT}}\right).$$

It remains to verify the Gaussian approximation and the jackknife variance consistency for the combined idiosyncratic weights.

The partition, Gram, and basis-switching bounds from Section SA-II.3 carry over under Assumption SA-PA-G: the number of fixed subsamples is finite, and the interior fixed breakpoints and density bounds give each subsample positive conditional mass, so the marginal quantile-spacing and cell-count arguments apply with constants depending only on K_Ψ.

Under Assumption SA-PA-G, the anchored construction gives $\widehat{p}_t(\mathsf{b}_k) = p_t(\mathsf{b}_k)$ for every contrast point. Thus the combined reference weights $W_{it,\Psi}$ are $\mathcal{E}_\infty$-measurable by the same argument as in Lemma SA-II.7.2. Conditional on $\mathcal{E}_\infty$, $\bar{\mathscr{I}}_{\varepsilon\Psi}$ is a sum of independent mean-zero primitive shocks with weights $W_{it,\Psi}/(T-H)$. Since the number of contrast points is fixed, the single-point local-support bounds, Jensen's inequality over the rolling windows, and $T^{-1}\sum_{t=1}^{T}(1+\|f_t\|^3) = O_{\mathbb{P}}(1)$ imply

$$\sum_{t=1}^{T}\sum_{i=1}^{n_t} W_{it,\Psi}^2 \lesssim_{\mathbb{P}} \frac{TJ^d}{n}, \qquad \sum_{t=1}^{T}\sum_{i=1}^{n_t} |W_{it,\Psi}|^3 \lesssim_{\mathbb{P}} \frac{TJ^{2d}}{n^2}.$$

The conditional Berry–Esseen bound therefore gives

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\bar{\mathscr{I}}_{\varepsilon\Psi}}{\sqrt{\Sigma_{\varepsilon\Psi}}} \le x \,\middle|\, \mathcal{E}_\infty\right) - \Phi(x)\right| \lesssim_{\mathbb{P}} \frac{TJ^{2d}/n^2}{\left((T-H)^2\Sigma_{\varepsilon\Psi}\right)^{3/2}} = o_{\mathbb{P}}(1),$$

where the last equality uses $(nT/J^d)\Sigma_{\varepsilon\Psi} \gtrsim_{\mathbb{P}} 1$ and $J^d/n = o(1)$. Removing the conditioning gives the infeasible Gaussian approximation for $\bar{\mathscr{I}}_{\varepsilon\Psi}$, and the residual bound above transfers it to $\widehat{\Psi} - \bar{\Psi}_T$.

For the variance estimator, apply the proof of Lemma SA-II.7.2 to the combined delete-i perturbation $\Delta_{it\Psi}$. Each diagonal, off-diagonal, and remainder term in that proof is quadratic in the relevant weights. Replacing $W_{it}(\beta)$ by the fixed finite sum $W_{it,\Psi} = \sum_{k=1}^{K_\Psi} a_k W_{it}(\mathsf{b}_k)$ only changes the constants in the second- and fourth-power sums and the squared-coefficient bound in that proof. The same argument therefore yields

$$\widehat{\Sigma}_{\varepsilon\Psi} = \Sigma_{\varepsilon\Psi} + o_{\mathbb{P}}\left(\frac{J^d}{nT}\right).$$

This proves the theorem. □

SA-II.10 Fixed-Grid Test of Linearity

SA-II.10.1 Testing Problem

We test whether the conditional mean $\mathbb{E}[R_{it}|\mathcal{F}_{t-1}]$ is linear in β_{it} at every date. Specifically, the null hypothesis is

$$\mathsf{H}_0^{\mathtt{LIN}}: \qquad \mu_t(b) = a_t + c_t^\top b, \qquad b \in \mathcal{B}, \quad t = 1,\dots,T, \tag{SA-II.10.1}$$

for some a_t and c_t that may vary over time and are $\mathcal{G}_{t-1}$-measurable. Under the null, realized returns satisfy

$$R_{it} = a_t + \varpi_t^\top \beta_{it} + \varepsilon_{it}, \qquad \varpi_t = c_t + \widetilde{f}_t.$$

Thus, the cross-sectional return slope is ϖ_t, not c_t: the linear fit must absorb the realized factor innovation.

The test uses two adjacent first-step windows. The partition at date t is formed from $\widehat{\beta}_{it}$, whose window ends at $t-1$. The test outcome is dated $t+H$, and $\widehat{\beta}_{i(t+H)}$ uses dates $t,\dots,t+H-1$. The two windows do not overlap. The linear fit uses a generic instrument available before the future window; it need not be $\widehat{\beta}_{it}$.

For $t = H+1,\dots,T-H$, define the future-window conditioning field

$$\mathcal{K}_t = \mathcal{E}_\infty \vee \sigma(\eta_{jr} : j \in [n],\ r \le t-1)\,.$$

To accommodate an unbalanced panel, for each base date $t = H+1,\dots,T-H$, let $\mathcal{I}_t^{\texttt{LIN}} \subseteq [n]$ collect the units observed throughout both rolling windows and at the outcome date $t+H$, and write $n_t^{\texttt{LIN}} = |\mathcal{I}_t^{\texttt{LIN}}|$. Thus both $\widehat{\beta}_{it}$ and $\widehat{\beta}_{i(t+H)}$, as well as $R_{i(t+H)}$, are available for every $i \in \mathcal{I}_t^{\texttt{LIN}}$. All date-$t$ cross-sectional objects in this section, including the partition and Gram matrix, are constructed on $\mathcal{I}_t^{\texttt{LIN}}$. We require $\mathcal{I}_t^{\texttt{LIN}}$ to be measurable with respect to the conditioning field $\mathcal{K}_t$, so sample availability does not select on the future primitive innovations, and assume that, for some fixed $c > 0$,

$$\min_{H+1\le t\le T-H} n_t^{\texttt{LIN}} \ge cn$$

with probability approaching one. The DGP, partition, and Gram conditions invoked below are transferred to these matched cross-sections through the following random-missingness condition. For this purpose, $\{(\alpha_{ir},\beta_{ir},\varsigma_{ir},\eta_{ir}) : i \in [n],\ r \in [T]\}$ is understood as a latent full-data array satisfying Assumption SA-DGP before selection. The auxiliary variables entering $\mathcal{E}_\infty$ are also understood to include the innovation-free availability process.

Assumption SA-LIN-M (Persistent random matched-sample availability). *Let $D_{it}^{\texttt{LIN}} = \mathbb{1}(i \in \mathcal{I}_t^{\texttt{LIN}})$ and let $\mathcal{Z}_t^{\texttt{LIN}}$ be the sigma-field generated by the full-data variables*

$$\{f_r : t-H \le r \le t+H\} \quad \textit{and} \quad \{(\alpha_{ir},\beta_{ir},\varsigma_{ir},\eta_{ir}) : i \in [n],\ t-H \le r \le t+H\}.$$

Uniformly over $t = H+1,\dots,T-H$:

(i) The availability vector $(D_{1t}^{\texttt{LIN}},\dots,D_{nt}^{\texttt{LIN}})$ is $\mathcal{K}_t$-measurable. Given $\mathcal{D}_{t-1}$, this vector has mutually independent coordinates and is independent of $\mathcal{Z}_t^{\texttt{LIN}}$.

(ii) For some fixed $\pi_0 > 0$, $\min_{i\in[n]} \mathbb{P}(D_{it}^{\texttt{LIN}} = 1 \mid \mathcal{D}_{t-1}) \ge \pi_0$ almost surely.

Define the matched design field

$$\mathcal{D}_{t-1}^{\texttt{LIN}} = \mathcal{D}_{t-1} \vee \sigma(D_{1t}^{\texttt{LIN}},\dots,D_{nt}^{\texttt{LIN}}).$$

Conditional on this field, the retained old-window predictable histories and old first-step estimates remain independent over i and retain their original conditional laws. Future-window primitive shocks remain independent conditional on $\mathcal{K}_t$. Although $\mathcal{D}_{t-1}^{\texttt{LIN}}$ need not be contained in $\mathcal{F}_{t-1}$, it is contained in $\mathcal{K}_t$. The matched versions of Lemmas SA-II.3.1, SA-II.3.2, and SA-II.3.3 therefore follow by the same conditional arguments, with n_t replaced by $n_t^{\texttt{LIN}}$ and with the reference partitions conditioning on $\mathcal{D}_{t-1}^{\texttt{LIN}}$. Whenever a sum is written over the common index set $[n]$, the contribution of unit i is set to zero if $i \notin \mathcal{I}_t^{\texttt{LIN}}$.

SA-II.10.2 IV Fit and Test Statistic

For $t = H+1, \dots, T-H$, define the future regressor

$$\widehat{x}_{i(t+H)} = (1, \widehat{\beta}_{i(t+H)}^\top)^\top.$$

Let $\zeta_{it} \in \mathbb{R}^{d+1}$ be a generic instrument, whose first coordinate is one to accommodate the intercept. Set

$$\widehat{\mathsf{M}}_t = \frac{1}{n_t^{\mathtt{LIN}}} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \zeta_{it} \widehat{x}_{i(t+H)}^\top, \quad \mathsf{M}_t^\circ = \mathbb{E}[\widehat{\mathsf{M}}_t \mid \mathcal{K}_t]. \tag{SA-II.10.2}$$

We impose the following high-level conditions on the instrument.

Assumption SA-LIN-IV (High-level instrument). *The following conditions hold uniformly over $t = H+1, \dots, T-H$:*

(i) ***Predetermination.*** *ζ_{it} is $\mathcal{F}_{t-1}$-measurable for every $i \in \mathcal{I}_t^{\mathtt{LIN}}$, and hence is $\mathcal{K}_t$-measurable.*

(ii) ***Moment and leverage regularity.***

$$\max_{H+1 \le t \le T-H} \frac{1}{n_t^{\mathtt{LIN}}} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \|\zeta_{it}\|^4 \lesssim_{\mathbb{P}} 1, \qquad \max_{H+1 \le t \le T-H} \max_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \|\zeta_{it}\| \lesssim_{\mathbb{P}} 1.$$

(iii) ***Relevance and concentration.***

$$\max_{H+1 \le t \le T-H} \left\{ \|\mathsf{M}_t^\circ\| + \|(\mathsf{M}_t^\circ)^{-1}\| \right\} \lesssim_{\mathbb{P}} 1, \qquad \max_{H+1 \le t \le T-H} \|\widehat{\mathsf{M}}_t - \mathsf{M}_t^\circ\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(nT)}{n}}.$$

Predetermination is the exclusion condition. It ensures that the instrument is fixed before the innovations entering $\widehat{\beta}_{i(t+H)}$ and $R_{i(t+H)}$ are observed. Moment and leverage regularity control both aggregate moments and cell-specific instrument leverage in the covariance proof. The first bound in part (iii) is the relevance condition for the intercept and realized cross-sectional slope; the second is a high-level uniform cross-sectional concentration condition on the sample instrument–regressor matrix.

One natural choice is $\zeta_{it} = (1, \widehat{\beta}_{it}^\top)^\top$. Lemma SA-II.10.4 below verifies Assumption SA-LIN-IV for this choice under an oracle old–future beta relevance condition and the primitive first-step conditions. Other predetermined characteristics or generated instruments may be used whenever Assumption SA-LIN-IV holds.

Given this generic instrument ζ_{it}, define

$$\widehat{\theta}_{t+H}^{\mathtt{IV}} = \widehat{\mathsf{M}}_t^{-1} \frac{1}{n_t^{\mathtt{LIN}}} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \zeta_{it} R_{i(t+H)}. \tag{SA-II.10.3}$$

Write $\widehat{\theta}_{t+H}^{\mathtt{IV}} = (\widehat{a}_{t+H}, \widehat{\varpi}_{t+H}^\top)^\top$ and

$$\widehat{e}_{i(t+H)} = R_{i(t+H)} - \widehat{x}_{i(t+H)}^\top \widehat{\theta}_{t+H}^{\mathtt{IV}}, \qquad i \in \mathcal{I}_t^{\mathtt{LIN}}, \quad t = H+1, \dots, T-H.$$

Set $\widehat{e}_{i(t+H)} = 0$ when $i \notin \mathcal{I}_t^{\mathtt{LIN}}$.

For any evaluation point $b \in \mathcal{B}^\circ$, define the projection weight

$$\widehat{w}_{it}(b) = \frac{1}{n_t^{\mathtt{LIN}}} \widehat{p}_t(b)^\top \widehat{Q}_t^{-1} \widehat{p}_t(\widehat{\beta}_{it}).$$

Set $\widehat{w}_{it}(b) = 0$ when $i \notin \mathcal{I}_t^{\mathtt{LIN}}$. When $\mathsf{q} = 0$, this is the usual equal weight in the cell containing b. The test statistic is

$$\widehat{\nu}(b) = \frac{1}{T-2H} \sum_{t=H+1}^{T-H} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \widehat{w}_{it}(b) \widehat{e}_{i(t+H)}. \tag{SA-II.10.4}$$

More generally, for distinct evaluation points $\mathsf{b}_1, \ldots, \mathsf{b}_{K_{\mathtt{LIN}}} \in \mathcal{B}^\circ$ with $K_{\mathtt{LIN}}$ fixed, define

$$\widehat{\boldsymbol{\nu}}_{\mathtt{LIN}} = \big(\widehat{\nu}(\mathsf{b}_1), \ldots, \widehat{\nu}(\mathsf{b}_{K_{\mathtt{LIN}}})\big)^\top.$$

Let $\Sigma_{\mathtt{LIN}}$ denote the predictable covariance matrix of the fixed-grid leading term under the null. The variance-stabilization condition below makes it asymptotically equivalent to an $\mathcal{E}_\infty$-measurable reference covariance $\overline{\Sigma}_{\mathtt{LIN}}$. Let $\widehat{\Sigma}_{\mathtt{LIN}}$ denote the feasible estimator of the predictable covariance; these objects are defined explicitly below. The Wald test statistic is

$$\mathsf{W}_{\mathtt{LIN}} := \widehat{\boldsymbol{\nu}}_{\mathtt{LIN}}^\top \widehat{\Sigma}_{\mathtt{LIN}}^{-1} \widehat{\boldsymbol{\nu}}_{\mathtt{LIN}}. \tag{SA-II.10.5}$$

Its null validity is established in Theorem SA-II.10.3.

SA-II.10.3 Decomposition

From (SA-II.2.2), write the centered first-step error as

$$v_{it} = \widehat{\beta}_{it} - \breve{\beta}_{it} = \frac{1}{H} \sum_{s=1}^{H} w_s(t) \varepsilon_{i(t-s)}.$$

Under the null, let $\theta_t = (a_t, \varpi_t^\top)^\top$. The future residual has the exact decomposition

$$\begin{aligned} \widehat{e}_{i(t+H)} = \varepsilon_{i(t+H)} &- \varpi_{t+H}^\top v_{i(t+H)} - \varpi_{t+H}^\top (\breve{\beta}_{i(t+H)} - \beta_{i(t+H)}) \\ &- \widehat{x}_{i(t+H)}^\top (\widehat{\theta}_{t+H}^{\mathtt{IV}} - \theta_{t+H}). \end{aligned} \tag{SA-II.10.6}$$

Substitution into (SA-II.10.4) gives

$$\widehat{\nu}(b) = \mathscr{S}_{\mathtt{LIN}}(b) + \mathscr{B}_{\mathtt{LIN}}(b) + \mathscr{R}_{\mathtt{LIN}}(b), \tag{SA-II.10.7}$$

where

$$\mathscr{S}_{\mathtt{LIN}}(b) = \frac{1}{T-2H} \sum_{t=H+1}^{T-H} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \widehat{w}_{it}(b) \left[\varepsilon_{i(t+H)} - \varpi_{t+H}^\top v_{i(t+H)} \right],$$

$$\mathscr{B}_{\mathtt{LIN}}(b) = -\frac{1}{T-2H} \sum_{t=H+1}^{T-H} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \widehat{w}_{it}(b) \varpi_{t+H}^\top (\breve{\beta}_{i(t+H)} - \beta_{i(t+H)}),$$

$$\mathscr{R}_{\mathtt{LIN}}(b) = -\frac{1}{T-2H}\sum_{t=H+1}^{T-H}\left(\sum_{i\in\mathcal{I}_t^{\mathtt{LIN}}}\widehat{w}_{it}(b)\widehat{x}_{i(t+H)}^{\top}\right)(\widehat{\theta}_{t+H}^{\mathtt{IV}} - \theta_{t+H}).$$

The first term is centered and determines the limiting distribution. The second is the smooth first-step bias. The third is the error from the IV fit.

Note that the forward split removes the leave-in bias from the first term. By Assumption SA-LIN-M(i), the matched set is also fixed conditional on $\mathcal{K}_t$. Thus $\widehat{w}_{it}(b)$ and ζ_{it} are fixed, while $v_{i(t+H)}$ uses innovations dated $t,\dots,t+H-1$ and $\varepsilon_{i(t+H)}$ is dated $t+H$. Hence

$$\mathbb{E}\big[\widehat{w}_{it}(b)v_{i(t+H)} \mid \mathcal{K}_t\big] = 0, \tag{SA-II.10.8}$$
$$\mathbb{E}\big[\widehat{w}_{it}(b)\varepsilon_{i(t+H)} \mid \mathcal{K}_t\big] = 0. \tag{SA-II.10.9}$$

By Assumption SA-LIN-IV(i), the same argument gives

$$\mathbb{E}\Big[\zeta_{it}v_{i(t+H)}^{\top} \mid \mathcal{K}_t\Big] = 0, \qquad \mathbb{E}\big[\zeta_{it}\varepsilon_{i(t+H)} \mid \mathcal{K}_t\big] = 0. \tag{SA-II.10.10}$$

Therefore the generic IV score is centered at the true realized slope, up to the smooth first-step bias.

For the limiting distribution, set $\widehat{w}_{it}(b) = 0$ when $t \notin \{H+1,\dots,T-H\}$ or $i \notin \mathcal{I}_t^{\mathtt{LIN}}$. Re-indexing the shocks in $\mathscr{S}_{\mathtt{LIN}}(b)$ gives

$$\mathscr{S}_{\mathtt{LIN}}(b) = \frac{1}{T-2H}\sum_{r=1}^{T}\sum_{i=1}^{n} W_{ir}(b)\varepsilon_{ir}, \tag{SA-II.10.11}$$

where

$$W_{ir}(b) = \widehat{w}_{i,r-H}(b) - \frac{1}{H}\sum_{s=1}^{H}\widehat{w}_{i,r+s-H}(b)\varpi_{r+s}^{\top}w_s(r+s). \tag{SA-II.10.12}$$

Invalid terms are set to zero. Every coefficient on ε_{ir} uses a partition whose first-step window ends before r. Hence $W_{ir}(b)$ is predictable. For $b, b' \in \mathcal{B}^{\circ}$, define

$$\Sigma_{\mathtt{LIN}}(b,b') = \frac{1}{(T-2H)^2}\sum_{r=1}^{T}\sum_{i=1}^{n}\varsigma_{ir}^2 W_{ir}(b)W_{ir}(b'). \tag{SA-II.10.13}$$

Let

$$\Sigma_{\mathtt{LIN}} = \big(\Sigma_{\mathtt{LIN}}(\mathsf{b}_k,\mathsf{b}_l)\big)_{k,l\in[K_{\mathtt{LIN}}]},$$

which is the terminal predictable covariance of the fixed-grid leading term under the null.

SA-II.10.4 Results

The first lemma gives the infeasible fixed-grid Gaussian limit for a generic instrument satisfying Assumption SA-LIN-IV.

Lemma SA-II.10.1 (Forward-Split Asymptotic Normality)**.** *Suppose Assumptions SA-DGP, SA-FS, SA-PA, SA-LIN-M, SA-LIN-IV, SA-SEP, and SA-FSB hold. Assume further that:*

(i) $T^{(3+d)/q}/H + \sqrt{\log T/H} = o(1)$.

(ii) $J^d \log(nT)/n = o(1)$.

(iii) The following rate restrictions hold:

$$\frac{T-2H}{nH} = o(1), \tag{SA-II.10.14}$$

$$\sqrt{\frac{n(T-2H)}{J^d}}\left(\mathsf{B}_{\mathtt{FS}} + \frac{\log(nT)}{n}\right) = o(1). \tag{SA-II.10.15}$$

(iv) Variance stabilization. *There exists an* $\mathcal{E}_\infty$*-measurable positive-definite matrix* $\overline{\Sigma}_{\mathtt{LIN}}$ *such that*

$$\lambda_{\min}\left(\frac{n(T-2H)}{J^d}\overline{\Sigma}_{\mathtt{LIN}}\right) \gtrsim_{\mathbb{P}} 1, \qquad \lambda_{\max}\left(\frac{n(T-2H)}{J^d}\overline{\Sigma}_{\mathtt{LIN}}\right) \lesssim_{\mathbb{P}} 1, \tag{SA-II.10.16}$$

$$\left\|\overline{\Sigma}_{\mathtt{LIN}}^{-1/2}\Sigma_{\mathtt{LIN}}\overline{\Sigma}_{\mathtt{LIN}}^{-1/2} - I_{K_{\mathtt{LIN}}}\right\| = o_{\mathbb{P}}(1). \tag{SA-II.10.17}$$

Then under the null hypothesis (SA-II.10.1)*,*

$$\Sigma_{\mathtt{LIN}}^{-1/2}\widehat{\boldsymbol{\nu}}_{\mathtt{LIN}} \rightsquigarrow \mathcal{N}(0, I_{K_{\mathtt{LIN}}}).$$

For feasible inference, define

$$\widehat{\Sigma}_{\mathtt{LIN}}(b,b') = \frac{1}{(T-2H)^2}\sum_{i=1}^{n}\sum_{t=H+1}^{T-H}\sum_{s=H+1}^{T-H}\widehat{w}_{it}(b)\widehat{w}_{is}(b')\widehat{e}_{i(t+H)}\widehat{e}_{i(s+H)}\mathbb{1}(|t-s|\le H), \tag{SA-II.10.18}$$

and set

$$\widehat{\Sigma}_{\mathtt{LIN}} = \left(\widehat{\Sigma}_{\mathtt{LIN}}(\mathsf{b}_k,\mathsf{b}_l)\right)_{k,l\in[K_{\mathtt{LIN}}]}.$$

The outer sum in (SA-II.10.18) uses the common index set $[n]$; the zero-weight and zero-residual conventions remove a unit whenever it is unavailable at either base date. The truncation at H is exact for the leading term because two forward residuals share a primitive idiosyncratic shock only when their base dates are at most H periods apart.

The next lemma shows that this feasible variance estimator is consistent.

Lemma SA-II.10.2 (Feasible Variance Consistency)**.** *Suppose Assumptions SA-DGP, SA-FS, SA-PA, SA-LIN-M, SA-LIN-IV, SA-SEP, SA-FSB, and conditions* (i)–(ii) *of Lemma SA-II.10.1 hold. Also, assume*

$$\frac{H\log(nT)}{n} = o(1). \tag{SA-II.10.19}$$

Then under the null hypothesis (SA-II.10.1)*,*

$$\|\widehat{\Sigma}_{\mathtt{LIN}} - \Sigma_{\mathtt{LIN}}\| = o_{\mathbb{P}}\left(\frac{J^d}{n(T-2H)}\right). \tag{SA-II.10.20}$$

The next theorem establishes the validity of the Wald test under the null.

Theorem SA-II.10.3 (Validity of the Forward-Split Wald Test)**.** *Suppose the conditions of Lemma SA-II.10.1 hold and* (SA-II.10.19) *holds. Then* $\widehat{\Sigma}_{\mathtt{LIN}}$ *is nonsingular with probability approaching one and the statistic in* (SA-II.10.5) *satisfies* $\mathsf{W}_{\mathtt{LIN}} \rightsquigarrow \chi^2_{K_{\mathtt{LIN}}}$ *under the null hypothesis* (SA-II.10.1)*. Consequently, the test that rejects when* $\mathsf{W}_{\mathtt{LIN}} > \chi^2_{K_{\mathtt{LIN}},1-\alpha}$ *has asymptotic size* α*.*

Finally, the following lemma verifies Assumption SA-LIN-IV for the natural choice of instrument $\zeta_{it} = (1, \widehat{\beta}_{it}^\top)^\top$ under the primitive first-step conditions and an oracle relevance condition.

Lemma SA-II.10.4 (Old-Beta Instrument Verification)**.** *Let*

$$x_{it} = (1, \beta_{it}^\top)^\top, \qquad \mathsf{M}^0_{t,\beta} = \frac{1}{n_t^{\mathtt{LIN}}} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} x_{it} x_{i(t+H)}^\top.$$

Suppose Assumptions SA-DGP, SA-FS, SA-FSB, and SA-LIN-M hold,

$$\frac{T^{2/q}}{H} + \sqrt{\frac{\log T}{H}} = o(1), \qquad \mathsf{R}_{\mathtt{FS}} + \frac{\log(nT)}{n} = o(1),$$

and the oracle old–future beta cross-moment is uniformly relevant:

$$\max_{H+1 \le t \le T-H} \left\{ \|\mathsf{M}^0_{t,\beta}\| + \|(\mathsf{M}^0_{t,\beta})^{-1}\| \right\} \lesssim_{\mathbb{P}} 1. \tag{SA-II.10.21}$$

Then the choice

$$\zeta_{it} = (1, \widehat{\beta}_{it}^\top)^\top$$

satisfies Assumption SA-LIN-IV.

Condition (SA-II.10.21) isolates the only substantive extra requirement for the old-beta choice: the predetermined beta must retain enough cross-sectional information about the future beta. Predetermination alone does not imply this cross-time relevance.

Remark SA-II.10.5 (Linear specification for excess returns)**.** *When* R_{it} *denotes an excess return, the absence of arbitrage imposes the stronger zero-intercept restriction*

$$\mu_t(b) = c_t^\top b, \qquad b \in \mathcal{B}, \quad t = 1, \dots, T.$$

The corresponding test is obtained by omitting the intercept from the period-by-period cross-sectional IV fit. Specifically, set $\widehat{x}_{i(t+H)} = \widehat{\beta}_{i(t+H)}$ *and use a* d*-dimensional predetermined instrument* ζ_{it} *without a constant, with* $\zeta_{it} = \widehat{\beta}_{it}$ *as the natural choice. Define*

$$\widehat{\mathsf{M}}_t = \frac{1}{n_t^{\mathtt{LIN}}} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \zeta_{it} \widehat{\beta}_{i(t+H)}^\top,$$
$$\widehat{\varpi}_{t+H}^{\mathtt{IV}} = \widehat{\mathsf{M}}_t^{-1} \frac{1}{n_t^{\mathtt{LIN}}} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \zeta_{it} R_{i(t+H)},$$

and construct the residuals as

$$\widehat{e}_{i(t+H)} = R_{i(t+H)} - \widehat{\beta}_{i(t+H)}^\top \widehat{\varpi}_{t+H}^{\mathtt{IV}}.$$

The fixed-grid moments, feasible covariance estimator, and Wald statistic are then constructed as above using these residuals. Under the corresponding d-dimensional version of Assumption SA-LIN-IV, including nonsingularity of the $d \times d$ instrument–regressor matrix, and the remaining conditions of Theorem SA-II.10.3, the same argument yields

$$\mathsf{W}_{\mathtt{LIN}} \rightsquigarrow \chi^2_{K_{\mathtt{LIN}}}.$$

SA-II.10.5 Discussion

(i) *Realized versus conditional slope.* The fitted slope estimates $\varpi_{t+H} = c_{t+H} + \widetilde{f}_{t+H}$. Estimating only c_{t+H} would leave the realized factor component in the residual and would not give an idiosyncratic-scale test.

(ii) *Why IV is necessary.* The forward split makes $v_{i(t+H)}$ independent of the old basis, but this fact alone does not make OLS valid. Specifically, for the OLS estimator of θ_{t+H} from regressing $R_{i(t+H)}$ on $\widehat{x}_{i(t+H)}$, its population score at the true coefficient contains a component

$$\left\| \mathbb{E}\left[v_{i(t+H)} v_{i(t+H)}^\top \mid \mathcal{E}_\infty \right] \varpi_{t+H} \right\| = O_{\mathbb{P}}(H^{-1}).$$

Thus OLS would add an $O_{\mathbb{P}}(H^{-1})$ attenuation bias term to $\mathscr{R}_{\mathtt{LIN}}(b)$ even though $v_{i(t+H)}$ is independent of the old partition. Any instrument satisfying Assumption SA-LIN-IV removes this term. The old beta estimate is one example because its first-step window does not overlap the future window.

(iii) *A primitive verification of old–future relevance.* Condition (SA-II.10.21) can be verified under genuine sample-path smoothness of the beta process and conditionally random missingness. Suppose, for example, that there are possibly random paths $B_{iT} : [0,1] \to \mathcal{B}$ such that $\beta_{it} = B_{iT}(t/T)$ and, for some $s_\beta > 0$, the paths are Hölder-s_β on $[0,1]$ uniformly over $i \in [n]$, with their uniform bounds and Hölder constants bounded in probability. Then, since $H/T = o(1)$,

$$\delta_{\beta,H} := \max_{H+1 \le t \le T-H} \max_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \|\beta_{i(t+H)} - \beta_{it}\| \lesssim_{\mathbb{P}} \left(\frac{H}{T}\right)^{s_\beta \wedge 1} = o_{\mathbb{P}}(1).$$

Define the old-beta Gram matrix on the matched cross-section by

$$Q^0_{t,\beta} = \frac{1}{n_t^{\mathtt{LIN}}} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} x_{it} x_{it}^\top.$$

Compactness of $\mathcal{B}$ gives

$$\max_{H+1 \le t \le T-H} \|\mathsf{M}^0_{t,\beta} - Q^0_{t,\beta}\| \lesssim_{\mathbb{P}} \delta_{\beta,H} = o_{\mathbb{P}}(1).$$

Under Assumption SA-LIN-M, conditioning on the realized matched set preserves the original conditional beta-density bounds and cross-sectional independence. The density lower bound on the full-dimensional rectangle $\mathcal{B}$, $n_t^{\mathtt{LIN}} \asymp n$, and uniform cross-sectional concentration under $\log(nT)/n = o(1)$ therefore imply

$$\min_{H+1 \le t \le T-H} \lambda_{\min}(Q^0_{t,\beta}) \gtrsim_{\mathbb{P}} 1, \qquad \max_{H+1 \le t \le T-H} \|Q^0_{t,\beta}\| \lesssim_{\mathbb{P}} 1.$$

Singular-value perturbation therefore yields

$$\max_{H+1\le t\le T-H}\left\{\|\mathsf{M}^0_{t,\beta}\|+\|(\mathsf{M}^0_{t,\beta})^{-1}\|\right\}\lesssim_{\mathbb{P}} 1,$$

which is (SA-II.10.21). We retain the high-level relevance condition in Lemma SA-II.10.4 because it also accommodates other persistence and missingness mechanisms.

(iv) *Generated partition.* The proof conditions on the old generated partition. It does not compare that partition with a true-beta partition. The basis-switching restriction $J\mathsf{R}_{\mathtt{FS}} = o(1)$ from the contemporaneous theory is therefore not needed for null size.

(v) *Rate conditions.* Consider the proposed MSE-optimal tuning choices in Part **??**. For $(\mathfrak{p},\mathfrak{q}) = (0,0)$, they give

$$h \asymp T^{-1/3}, \qquad H \asymp T^{2/3}, \qquad J \asymp n^{1/(d+2)}.$$

The bias component of (SA-II.10.15) then requires

$$J\sqrt{T}\,\mathsf{B}_{\mathtt{FS}} = o(1), \qquad \text{or equivalently} \qquad \mathsf{B}_{\mathtt{FS}} = o\Big(n^{-1/(d+2)}T^{-1/2}\Big). \tag{SA-II.10.22}$$

The implications depend on the order of the first-step local polynomial.

Local-constant first step. With a nonzero leading bias, if $\mathsf{B}_{\mathtt{FS}} \asymp h \asymp T^{-1/3}$, then $J\sqrt{T}\,\mathsf{B}_{\mathtt{FS}} \asymp JT^{1/6}$ does not vanish, so the proposed MSE-optimal bandwidth does not satisfy (SA-II.10.22).

Local-linear first step. If $s \ge 2$ and $\mathsf{B}_{\mathtt{FS}} \lesssim h^2 \asymp T^{-2/3}$, write $T = n^{\eta}$, suppress logarithmic factors, and maintain the high-level instrument and covariance nondegeneracy conditions. The bias restriction gives the lower bound $\eta > 6/(d+2)$, while the other Gaussian approximation rates give the upper bound $\eta < 2(d+1)/(d+2)$. Thus

$$\frac{6}{d+2} < \eta < \frac{2(d+1)}{d+2}.$$

For the feasible Wald statistic, (SA-II.10.19) further requires $\eta < 3/2$. The resulting window is

$$\frac{6}{d+2} < \eta < \min\left\{\frac{2(d+1)}{d+2}, \frac{3}{2}\right\}. \tag{SA-II.10.23}$$

This window is nonempty when $d > 2$. It contains a subrange with $\eta < 1$, and hence permits $T \ll n$, only when $d > 4$.

Local-quadratic first step. If $s \ge 3$ and $\mathsf{B}_{\mathtt{FS}} \lesssim h^3 \asymp T^{-1}$, the bias restriction now gives $\eta > 2/(d+2)$, so the feasible window becomes

$$\frac{2}{d+2} < \eta < \min\left\{\frac{2(d+1)}{d+2}, \frac{3}{2}\right\}.$$

This window is nonempty for every $d \ge 1$. In particular, when $d = 1$, it contains the subrange $2/3 < \eta < 1$, which permits $T \ll n$. Thus, under the proposed MSE-optimal bandwidth, a higher-order first step is needed to accommodate $T \ll n$ in the lower-dimensional cases.

(vi) *Variance stabilization.* Condition (SA-II.10.17) concerns the full rolling covariance functional, not pointwise convergence of the generated weights. One verification separates dates at which each grid point is away from reference-cell boundaries from boundary-near dates: the former have a stable cell label, while the latter are negligible under a time-averaged margin condition or are covered by local equivalence of the adjacent-cell rolling covariance kernels. Alternatively, the evaluation-point-anchored construction in Section SA-II.9, particularly Assumption SA-PA-G, fixes each grid point's cell label and basis coordinate; conditional quantile and Gram concentration can then control the remaining empirical-to-reference covariance difference. We retain the high-level condition here to accommodate either route.

(vii) *Power against fixed persistent alternatives.* Let $\boldsymbol{\nu}_{0,T,\mathtt{LIN}}$ denote the time-averaged population forward moment, after the population cross-time IV linear fit, evaluated on the fixed grid. If $\boldsymbol{\nu}_{0,T,\mathtt{LIN}} \to_{\mathbb{P}} \boldsymbol{\nu}_{0,\mathtt{LIN}} \neq 0$ and $\widehat{\boldsymbol{\nu}}_{\mathtt{LIN}} - \boldsymbol{\nu}_{0,T,\mathtt{LIN}} = o_{\mathbb{P}}(1)$, while $\widehat{\Sigma}_{\mathtt{LIN}}$ is positive definite with probability approaching one and $\lambda_{\max}(\widehat{\Sigma}_{\mathtt{LIN}}) = o_{\mathbb{P}}(1)$, then $\mathsf{W}_{\mathtt{LIN}} \to_{\mathbb{P}} \infty$ and the test is consistent. The nonzero-moment condition requires persistence and rules out cancellation over dates and grid points; without it, a nonlinear fixed alternative need not be detected.

(viii) *Scope of the test.* Taking $K_{\mathtt{LIN}} = 1$ gives a pointwise specification test; fixed $K_{\mathtt{LIN}}$ gives the Wald test above. A diverging-grid omnibus test requires a separate high-dimensional Gaussian approximation. The test uses an exactly identified $(d+1)$-vector of instruments. An overidentified specification can be accommodated after replacing (SA-II.10.3) by the corresponding GMM fit and imposing the analogous Jacobian and weight-matrix conditions.

(ix) *A fixed-period OLS alternative.* At one prespecified date $t > H$, one could instead compare the degree-one beta-sorted fit with the same-date full-cross-sectional OLS fit. Degree-one reproduction removes the within-cell discrepancy under the linear null. The nonparametric component has standard deviation $\sqrt{J^d/n}$, whereas the OLS sampling error is of smaller order $n^{-1/2}$. Unlike the forward-split test, this approach does not remove the generated-regressor and OLS attenuation distortions; it instead requires the sufficient centering condition $\mathsf{B}_{\mathtt{FS}} + H^{-1} = o\Big(\sqrt{J^d/n}\Big)$. Use the paper's proposed local-constant point-estimation MSE tuning,

$$h \asymp T^{-1/3}, \qquad H \asymp T^{2/3}, \qquad J \asymp n^{1/(d+2)},$$

but construct both steps with degree-one fits. Then if $\mathsf{B}_{\mathtt{FS}} \lesssim T^{-2/3}$, the main restriction reduces, up to logarithmic factors from uniform remainder bounds, to $T \gg n^{3/[2(d+2)]}$. Thus $T = n^{\eta}$ with $3/[2(d+2)] < \eta < 1$ is admissible for every $d \geq 1$, permitting $T \ll n$.

SA-II.10.6 Proof of Lemma SA-II.10.1

Step 1: Remainders. Put

$$u^{\circ}_{it} = \varepsilon_{i(t+H)} - \varpi_{t+H}^{\top} v_{i(t+H)}, \qquad g_t = \frac{1}{n_t^{\mathtt{LIN}}} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \zeta_{it} u^{\circ}_{it},$$

$$A_t(b) = \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \widehat{w}_{it}(b) \widehat{x}_{i(t+H)}^{\top}, \qquad A_t^{\circ}(b) = \mathbb{E}[A_t(b) \mid \mathcal{K}_t].$$

For notational convenience, set $u_{it}^{\circ} = 0$ when $i \notin \mathcal{I}_t^{\mathtt{LIN}}$. The IV normal equations give

$$\widehat{\theta}_{t+H}^{\mathtt{IV}} - \theta_{t+H} = \widehat{\mathsf{M}}_t^{-1} g_t + \mathscr{B}_{t+H}^{\mathtt{IV}}, \tag{SA-II.10.24}$$

$$\mathscr{B}_{t+H}^{\mathtt{IV}} := -\widehat{\mathsf{M}}_t^{-1} \frac{1}{n_t^{\mathtt{LIN}}} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \zeta_{it} \varpi_{t+H}^{\top} (\breve{\beta}_{i(t+H)} - \beta_{i(t+H)}). \tag{SA-II.10.25}$$

Thus $\mathscr{B}_{t+H}^{\mathtt{IV}}$ is the component of the IV estimation error induced by the smooth first-step bias. In particular,

$$\mathscr{R}_{\mathtt{LIN}}(b) = -\frac{1}{T-2H} \sum_{t=H+1}^{T-H} A_t(b) \widehat{\mathsf{M}}_t^{-1} g_t - \frac{1}{T-2H} \sum_{t=H+1}^{T-H} A_t(b) \mathscr{B}_{t+H}^{\mathtt{IV}}.$$

The partition, Gram, first-step moment, and instrument moment bounds imply

$$\max_{\substack{H+1 \le t \le T-H \\ k \in [K_{\mathtt{LIN}}]}} \{ \|A_t(\mathsf{b}_k)\| + \|A_t^{\circ}(\mathsf{b}_k)\| \} \lesssim_{\mathbb{P}} 1.$$

Hence the bias terms in (SA-II.10.7) and (SA-II.10.24) satisfy

$$\max_{k \in [K_{\mathtt{LIN}}]} \left\{ |\mathscr{B}_{\mathtt{LIN}}(\mathsf{b}_k)| + \left| \frac{1}{T-2H} \sum_{t=H+1}^{T-H} A_t(\mathsf{b}_k) \mathscr{B}_{t+H}^{\mathtt{IV}} \right| \right\} \lesssim_{\mathbb{P}} \mathsf{B}_{\mathtt{FS}}. \tag{SA-II.10.26}$$

Conditional on $\mathcal{K}_t$, g_t is centered by (SA-II.10.10). Cross-sectional independence, which is preserved after selection by Assumption SA-LIN-M, Assumption SA-LIN-IV(ii), the first-step moment bounds, and the weight bounds give

$$\mathbb{E}[\|g_t\|^2 \mid \mathcal{K}_t] \lesssim_{\mathbb{P}} \frac{1 + \|\varpi_{t+H}\|^2/H}{n_t^{\mathtt{LIN}}} \lesssim_{\mathbb{P}} \frac{1 + \|\varpi_{t+H}\|^2/H}{n}, \tag{SA-II.10.27}$$

$$\max_{\substack{H+1 \le t \le T-H \\ k \in [K_{\mathtt{LIN}}]}} \mathbb{E}[\|A_t(\mathsf{b}_k) - A_t^{\circ}(\mathsf{b}_k)\|^2 \mid \mathcal{K}_t] \lesssim_{\mathbb{P}} \frac{J^d}{nH}.$$

Boundedness of μ_t on the fixed rectangle $\mathcal{B}$ bounds c_t under the null. Assumption SA-DGP(iv) and Markov's inequality therefore give $(T-2H)^{-1} \sum_{t=H+1}^{T-H} (1 + \|\varpi_{t+H}\|^4) \lesssim_{\mathbb{P}} 1$. The multiplier $A_t^{\circ}(\mathsf{b}_k)(\mathsf{M}_t^{\circ})^{-1} \zeta_{it}$ attached to each summand of g_t is fixed before the future-window shocks begin. Re-indexing the direct outcome shocks therefore gives a martingale sum. The rolling-window component is also centered; it has periodwise variance $O_{\mathbb{P}}((nH)^{-1})$ and overlaps for at most H adjacent base dates. Thus, for fixed $K_{\mathtt{LIN}}$,

$$\max_{k \in [K_{\mathtt{LIN}}]} \left| \frac{1}{T-2H} \sum_{t=H+1}^{T-H} A_t^{\circ}(\mathsf{b}_k)(\mathsf{M}_t^{\circ})^{-1} g_t \right| \lesssim_{\mathbb{P}} \frac{1}{\sqrt{n(T-2H)}}. \tag{SA-II.10.28}$$

It remains to replace the conditional multipliers by their sample counterparts. Assumption SA-LIN-IV(iii) and a matrix perturbation bound imply that $\widehat{\mathsf{M}}_t$ and $\widehat{\mathsf{M}}_t^{-1}$ are uniformly bounded in probability and

$$r_{\mathtt{M}} := \max_{H+1 \le t \le T-H} \|\widehat{\mathsf{M}}_t - \mathsf{M}_t^{\circ}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(nT)}{n}},$$

$$\max_{H+1 \le t \le T-H} \|\widehat{\mathsf{M}}_t^{-1} - (\mathsf{M}_t^{\circ})^{-1}\| \lesssim_{\mathbb{P}} r_{\mathtt{M}}.$$

Use the exact decomposition

$$A_t\widehat{\mathsf{M}}_t^{-1} - A_t^\circ(\mathsf{M}_t^\circ)^{-1} = (A_t - A_t^\circ)(\mathsf{M}_t^\circ)^{-1} + A_t\{\widehat{\mathsf{M}}_t^{-1} - (\mathsf{M}_t^\circ)^{-1}\}.$$

For the first term on the right-hand side, (SA-II.10.27), the uniform bound on $(\mathsf{M}_t^\circ)^{-1}$, the time-averaged factor-moment bound, and Markov's inequality give, uniformly over the fixed grid,

$$\frac{1}{T-2H}\sum_{t=H+1}^{T-H}\|A_t(\mathsf{b}_k) - A_t^\circ(\mathsf{b}_k)\|^2 \lesssim_{\mathbb{P}} \frac{J^d}{nH}, \qquad \frac{1}{T-2H}\sum_{t=H+1}^{T-H}\|g_t\|^2 \lesssim_{\mathbb{P}} \frac{1}{n}.$$

Cauchy–Schwarz over the base dates therefore yields

$$\max_{k\in[K_{\mathtt{LIN}}]}\left|\frac{1}{T-2H}\sum_{t=H+1}^{T-H}\{A_t(\mathsf{b}_k) - A_t^\circ(\mathsf{b}_k)\}(\mathsf{M}_t^\circ)^{-1}g_t\right| \lesssim_{\mathbb{P}} \frac{1}{n}\sqrt{\frac{J^d}{H}}. \tag{SA-II.10.29}$$

Relative to the local standard-deviation scale, this bound satisfies

$$\sqrt{\frac{n(T-2H)}{J^d}}\,\frac{1}{n}\sqrt{\frac{J^d}{H}} = \sqrt{\frac{T-2H}{nH}} = o(1)$$

by (SA-II.10.14).

For the inverse-matrix term, (SA-II.10.27), the time-averaged factor-moment bound, and Markov's inequality give

$$\frac{1}{T-2H}\sum_{t=H+1}^{T-H}\|g_t\| \le \left\{\frac{1}{T-2H}\sum_{t=H+1}^{T-H}\|g_t\|^2\right\}^{1/2} \lesssim_{\mathbb{P}} \frac{1}{\sqrt{n}}.$$

Together with the uniform bound on A_t, this gives

$$\max_{k\in[K_{\mathtt{LIN}}]}\left|\frac{1}{T-2H}\sum_{t=H+1}^{T-H}A_t(\mathsf{b}_k)\{\widehat{\mathsf{M}}_t^{-1} - (\mathsf{M}_t^\circ)^{-1}\}g_t\right| \lesssim_{\mathbb{P}} \frac{r_{\mathtt{M}}}{\sqrt{n}} \lesssim_{\mathbb{P}} \frac{\sqrt{\log(nT)}}{n}.$$

The last bound is absorbed by $O_{\mathbb{P}}\{\log(nT)/n\}$. Combining the two parts of the exact multiplier decomposition gives

$$\max_{k\in[K_{\mathtt{LIN}}]}\left|\frac{1}{T-2H}\sum_{t=H+1}^{T-H}\left\{A_t(\mathsf{b}_k)\widehat{\mathsf{M}}_t^{-1} - A_t^\circ(\mathsf{b}_k)(\mathsf{M}_t^\circ)^{-1}\right\}g_t\right| \lesssim_{\mathbb{P}} \frac{1}{n}\sqrt{\frac{J^d}{H}} + \frac{\log(nT)}{n}. \tag{SA-II.10.30}$$

Combining (SA-II.10.26)– (SA-II.10.30) gives

$$\max_{k\in[K_{\mathtt{LIN}}]}\{|\mathscr{R}_{\mathtt{LIN}}(\mathsf{b}_k)| + |\mathscr{B}_{\mathtt{LIN}}(\mathsf{b}_k)|\} \lesssim_{\mathbb{P}} \mathsf{B}_{\mathtt{FS}} + \frac{1}{\sqrt{n(T-2H)}} + \frac{1}{n}\sqrt{\frac{J^d}{H}} + \frac{\log(nT)}{n}. \tag{SA-II.10.31}$$

The restriction $J^{-d/2} = o(1)$ controls the normalized $\{n(T-2H)\}^{-1/2}$ term. Equations (SA-II.10.14) and (SA-II.10.15) control the remaining terms. Thus both non-leading terms in (SA-II.10.7), uniformly over the fixed grid, are $o_{\mathbb{P}}(\sqrt{J^d/(n(T-2H))})$.

Step 2: Leading term. The matched version of Lemma SA-II.3.3 described after Assumption SA-LIN-M,

Jensen's inequality over the H rolling lags, the first-step Gram bound, and $cn \le n_t^{\mathtt{LIN}} \le n$ imply, under the absent-unit convention and uniformly over $k \in [K_{\mathtt{LIN}}]$,

$$\sum_{r=1}^{T}\sum_{i=1}^{n} W_{ir}(\mathsf{b}_k)^2 \lesssim_{\mathbb{P}} \frac{(T-2H)J^d}{n}, \qquad \sum_{r=1}^{T}\sum_{i=1}^{n} |W_{ir}(\mathsf{b}_k)|^3 \lesssim_{\mathbb{P}} \frac{(T-2H)J^{2d}}{n^2}. \tag{SA-II.10.32}$$

Consequently, $\lambda_{\max}(\Sigma_{\mathtt{LIN}}) \lesssim_{\mathbb{P}} J^d/\{n(T-2H)\}$. Condition (iv) implies that all eigenvalues of both $\overline{\Sigma}_{\mathtt{LIN}}$ and $\Sigma_{\mathtt{LIN}}$ have this order.

Let $\boldsymbol{W}_{ir} = \big(W_{ir}(\mathsf{b}_1), \dots, W_{ir}(\mathsf{b}_{K_{\mathtt{LIN}}})\big)^\top$. For fixed $a \in \mathbb{R}^{K_{\mathtt{LIN}}}$, put

$$\widetilde{W}_{ir}(a) := a^\top \overline{\Sigma}_{\mathtt{LIN}}^{-1/2} \boldsymbol{W}_{ir}.$$

The reference covariance is $\mathcal{E}_\infty$-measurable. Hence, after ordering the primitive innovations first by date and then by unit, $\widetilde{W}_{ir}(a)$ is measurable before η_{ir} is observed. The sum of the conditional variances of the corresponding normalized martingale differences (terminal predictable quadratic variation of their sum) is

$$\frac{1}{(T-2H)^2}\sum_{r=1}^{T}\sum_{i=1}^{n} \varsigma_{ir}^2 \widetilde{W}_{ir}(a)^2 = a^\top \overline{\Sigma}_{\mathtt{LIN}}^{-1/2} \Sigma_{\mathtt{LIN}} \overline{\Sigma}_{\mathtt{LIN}}^{-1/2} a \to_{\mathbb{P}} \|a\|^2$$

by (SA-II.10.17). Let

$$\mathcal{J}_{ir-}^{\mathtt{LIN}} := \mathcal{E}_\infty \vee \sigma(\eta_{js} : s < r, \text{ or } s = r \text{ and } j < i).$$

Since $K_{\mathtt{LIN}}$ is fixed, the third-moment bound in (SA-II.10.32) and (SA-II.10.16) also give the conditional Lyapunov bound

$$\frac{1}{(T-2H)^3}\sum_{r=1}^{T}\sum_{i=1}^{n} \mathbb{E}\Big[|\widetilde{W}_{ir}(a)\varepsilon_{ir}|^3 \mid \mathcal{J}_{ir-}^{\mathtt{LIN}}\Big] \lesssim_{\mathbb{P}} \sqrt{\frac{J^d}{n(T-2H)}} = o_{\mathbb{P}}(1).$$

The martingale central limit theorem and the Cramér–Wold device therefore give

$$\overline{\Sigma}_{\mathtt{LIN}}^{-1/2} \mathscr{S}_{\mathtt{LIN}} \rightsquigarrow \mathcal{N}(0, I_{K_{\mathtt{LIN}}}),$$

where $\mathscr{S}_{\mathtt{LIN}}$ stacks the leading terms over the fixed grid. Condition (SA-II.10.16) and the Step 1 remainder bound then imply

$$\overline{\Sigma}_{\mathtt{LIN}}^{-1/2} \widehat{\boldsymbol{\nu}}_{\mathtt{LIN}} \rightsquigarrow \mathcal{N}(0, I_{K_{\mathtt{LIN}}}). \tag{SA-II.10.33}$$

Finally, (SA-II.10.16) and (SA-II.10.17), together with a matrix square-root perturbation bound, give

$$\Big\|\Sigma_{\mathtt{LIN}}^{-1/2}\overline{\Sigma}_{\mathtt{LIN}}^{1/2} - I_{K_{\mathtt{LIN}}}\Big\| = o_{\mathbb{P}}(1).$$

Therefore

$$\Sigma_{\mathtt{LIN}}^{-1/2}\widehat{\boldsymbol{\nu}}_{\mathtt{LIN}} = \Sigma_{\mathtt{LIN}}^{-1/2}\overline{\Sigma}_{\mathtt{LIN}}^{1/2}\left(\overline{\Sigma}_{\mathtt{LIN}}^{-1/2}\widehat{\boldsymbol{\nu}}_{\mathtt{LIN}}\right) \rightsquigarrow \mathcal{N}(0, I_{K_{\mathtt{LIN}}})$$

by (SA-II.10.33) and Slutsky's theorem. □

SA-II.10.7 Proof of Lemma SA-II.10.2

For this proof, put

$$\mathcal{T}_H = \{H+1,\dots,T-H\}, \qquad L_T = |\mathcal{T}_H| = T-2H, \qquad \ell_{nT} = \frac{\log(nT)}{n},$$
$$e_{nT}^2 = \mathsf{B}_{\mathsf{FS}}^2 + \ell_{nT}, \qquad \rho_{nT}^2 = \frac{1}{H} + \ell_{nT}, \qquad u_{it}^\circ = \varepsilon_{i(t+H)} - \varpi_{t+H}^\top v_{i(t+H)},$$
$$g_t = \frac{1}{n_t^{\mathtt{LIN}}} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \zeta_{it} u_{it}^\circ.$$

Unless stated otherwise, sums over t and s below are over $\mathcal{T}_H$, and shifted dates outside $\mathcal{T}_H$ are omitted. The generated weight $\widehat{w}_{it}$ and the fitted residual-error components introduced below are set to zero outside the corresponding matched set. The latent block u_{it}° and its primitive shocks remain defined for every $i \in [n]$; zero weights remove unavailable units from all sample sums.

The assumptions used in this proof imply

$$\frac{H}{L_T} = o(1), \qquad \frac{J^d}{n} = o(1), \qquad He_{nT}^2 = o_{\mathbb{P}}(1), \qquad H\rho_{nT}e_{nT} = o_{\mathbb{P}}(1),$$
$$\frac{H}{n} = o(1), \qquad \sqrt{\frac{HJ^d}{nL_T}} = o(1), \qquad H\sqrt{\frac{J^d}{n^2 L_T}} = o(1). \tag{SA-II.10.34}$$

For example, $H^2\rho_{nT}^2 e_{nT}^2 = (1+H\ell_{nT})(H\mathsf{B}_{\mathsf{FS}}^2 + H\ell_{nT}) = o_{\mathbb{P}}(1)$. The matched-sample local-support and Gram bounds also give, uniformly over the fixed grid,

$$\max_{t,k} \sum_i \widehat{w}_{it}(\mathsf{b}_k)^2 \lesssim_{\mathbb{P}} \frac{J^d}{n}, \qquad \max_{i,t,k} |\widehat{w}_{it}(\mathsf{b}_k)| \lesssim_{\mathbb{P}} \frac{J^d}{n},$$
$$\max_{t,s,k,\ell} \sum_i \widehat{w}_{it}(\mathsf{b}_k)^2 \widehat{w}_{is}(\mathsf{b}_\ell)^2 \lesssim_{\mathbb{P}} \frac{J^{3d}}{n^3}. \tag{SA-II.10.35}$$

We use two primitive-shock calculations repeatedly. First, define

$$\kappa_{tr} := \mathbb{1}(r = t+H) - \frac{1}{H}\mathbb{1}(t \le r \le t+H-1)\varpi_{t+H}^\top w_{t+H-r}(t+H).$$

The uniform first-step Gram bound gives

$$\max_{s \in \mathcal{T}_H} \frac{1}{H} \sum_{a=1}^{H} \|w_a(s+H)\|^2 \lesssim_{\mathbb{P}} 1.$$

Therefore $u_{it}^\circ = \sum_{r=1}^T \kappa_{tr}\varepsilon_{ir}$ and

$$\sum_{r=1}^T \kappa_{tr}^2 \lesssim_{\mathbb{P}} 1 + \frac{\|\varpi_{t+H}\|^2}{H}. \tag{SA-II.10.36}$$

More generally, if Z_{is} is $\mathcal{K}_s$-measurable, then

$$\sum_s Z_{is} u^\circ_{is} = \sum_{r=1}^{T} \Gamma_{ir}(Z)\varepsilon_{ir},$$

$$\Gamma_{ir}(Z) := Z_{i,r-H} - \frac{1}{H}\sum_{a=1}^{H} Z_{i,r+a-H}\varpi_{r+a}^\top w_a(r+a),$$

$$\sum_{i,r} |\Gamma_{ir}(Z)|^2 \lesssim_{\mathbb{P}} \sum_{i,s}(1+\|\varpi_{s+H}\|^2)|Z_{is}|^2, \tag{SA-II.10.37}$$

where invalid subscripts are zero. Each $\Gamma_{ir}(Z)$ is measurable before ε_{ir} is observed.

Second, index primitive shocks by $a=(i,r)$ and $b=(j,r')$, write $\varepsilon_a=\varepsilon_{ir}$ and $\varepsilon_b=\varepsilon_{jr'}$, and order these compound indices lexicographically by date and unit. Let

$$\mathcal{J}^{\mathtt{LIN}}_{ir-} := \mathcal{E}_\infty \vee \sigma(\eta_{js} : s<r, \text{ or } s=r \text{ and } j<i).$$

Suppose that $a<b$, A_{ab} is measurable before ε_a, and, for a deterministic sequence r_n, $\sum_{a<b}|A_{ab}|^2 = O_{\mathbb{P}}(r_n^2)$. Then bounded conditional fourth moments and successive conditioning give

$$\sum_{a<b} A_{ab}\varepsilon_a\varepsilon_b = O_{\mathbb{P}}(r_n). \tag{SA-II.10.38}$$

For a quadratic sum over $a\neq b$, we first combine the coefficients on (a,b) and (b,a), which multiply the same shock product. We also use the analogous bound for a predictable linear shock sum: $\sum_{i,r}\Gamma_{ir}\varepsilon_{ir} = O_{\mathbb{P}}(r_n)$ whenever the coefficients are predictable and $\sum_{i,r}|\Gamma_{ir}|^2 = O_{\mathbb{P}}(r_n^2)$.

Step 1: Oracle covariance. For $k,\ell\in[K_{\mathtt{LIN}}]$, let $\widetilde{\Sigma}_{k\ell}$ denote (SA-II.10.18) with $(b,b')=(\mathsf{b}_k,\mathsf{b}_\ell)$ and $\widehat{e}_{i(t+H)}$ replaced by u°_{it}. Because u°_{it} is a linear combination of primitive shocks dated in $[t,t+H]$, the truncation at H is exact. For example, if $t+H<s$, then $\widehat{w}_{it}(\mathsf{b}_k)u^\circ_{it}\widehat{w}_{is}(\mathsf{b}_\ell)$ is $\mathcal{K}_s$-measurable, whereas $\mathbb{E}[u^\circ_{is}\mid\mathcal{K}_s]=0$; the reverse ordering is symmetric. Re-indexing the retained products by primitive shock gives

$$\widetilde{\Sigma}_{k\ell} = \mathscr{D}_{k\ell} + \mathscr{O}_{k\ell}, \tag{SA-II.10.39}$$

$$\mathscr{D}_{k\ell} = \frac{1}{L_T^2}\sum_{r=1}^{T}\sum_{i=1}^{n} W_{ir}(\mathsf{b}_k)W_{ir}(\mathsf{b}_\ell)\varepsilon_{ir}^2,$$

$$\mathscr{O}_{k\ell} = \frac{1}{L_T^2}\sum_{i=1}^{n}\sum_{r\neq r'} M^{k\ell}_{i,rr'}\varepsilon_{ir}\varepsilon_{ir'}.$$

Here $M^{k\ell}_{i,rr'}$ collects the coefficients on products of distinct primitive shocks. The diagonal coefficient is exactly the product of the primitive weights in (SA-II.10.12): every pair of residual terms containing the same primitive shock is retained because their base dates differ by at most H.

Diagonal component. The first two bounds in (SA-II.10.35), $|x+y|^4\le 8(|x|^4+|y|^4)$, Jensen's inequality over the rolling average, and $L_T^{-1}\sum_t(1+\|\varpi_{t+H}\|^4)\lesssim_{\mathbb{P}} 1$ give

$$\sum_{r=1}^{T}\sum_{i=1}^{n}|W_{ir}(\mathsf{b}_k)|^4 \lesssim_{\mathbb{P}} \frac{L_T J^{3d}}{n^3}. \tag{SA-II.10.40}$$

Since $W_{ir}(\mathsf{b}_k)W_{ir}(\mathsf{b}_\ell)$ is $\mathcal{J}^{\mathtt{LIN}}_{ir-}$-measurable, $W_{ir}(\mathsf{b}_k)W_{ir}(\mathsf{b}_\ell)(\varepsilon_{ir}^2-\varsigma_{ir}^2)$ is a martingale difference and

$$\mathscr{D}_{k\ell}-\Sigma_{\mathtt{LIN}}(\mathsf{b}_k,\mathsf{b}_\ell)=\frac{1}{L_T^2}\sum_{r=1}^{T}\sum_{i=1}^{n}W_{ir}(\mathsf{b}_k)W_{ir}(\mathsf{b}_\ell)\left(\varepsilon_{ir}^2-\varsigma_{ir}^2\right).$$

Its conditional variance is bounded by (SA-II.10.40) and the conditional fourth-moment bound for ε_{ir}. Therefore

$$\mathscr{D}_{k\ell}-\Sigma_{\mathtt{LIN}}(\mathsf{b}_k,\mathsf{b}_\ell)=O_{\mathbb{P}}\left(\sqrt{\frac{J^{3d}}{n^3L_T^3}}\right)=o_{\mathbb{P}}\left(\frac{J^d}{nL_T}\right). \tag{SA-II.10.41}$$

Off-diagonal component. For $t<s\le t+H$, define the pre-s and s-onward components by

$$u_{it}^{<s}:=-\frac{1}{H}\sum_{r=t}^{s-1}\varpi_{t+H}^{\top}w_{t+H-r}(t+H)\varepsilon_{ir},\qquad\qquad u_{it}^{\ge s}:=u_{it}^{\circ}-u_{it}^{<s}.$$

Set $u_{it}^{<s}=0$ and $u_{it}^{\ge s}=u_{it}^{\circ}$ when $s\le t$; the reverse ordering is handled by interchanging t and s. Put $\Lambda_{t,h}=1+\|\varpi_{t+h+H}\|^2$. The first-step Gram and factor-moment bounds, $2|xy|\le x^2+y^2$, and $L_T^{-1}\sum_t(1+\|\varpi_{t+H}\|^4)\lesssim_{\mathbb{P}}1$ give

$$\begin{aligned}\max_{k\in[K_{\mathtt{LIN}}]}\sum_{h=1}^{H}\frac{1}{L_T}\sum_t\sum_i\Lambda_{t,h}\widehat{w}_{it}(\mathsf{b}_k)^2\left(u_{it}^{<t+h}\right)^2&\lesssim_{\mathbb{P}}\frac{J^d}{n},\\ \sum_{h=1}^{H}\frac{1}{L_T}\sum_t\frac{\Lambda_{t,h}}{n}\sum_i\left(u_{it}^{<t+h}\right)^2&\lesssim_{\mathbb{P}}1,\end{aligned} \tag{SA-II.10.42}$$

where boundary terms are omitted.

For the products involving the s-onward component $u_{it}^{\ge s}$, all primitive shocks are dated at or after the later base date, so each coefficient is measurable before the earlier of the two shocks. Let $\mathscr{O}^{\mathrm{F}}_{k\ell}$ denote their off-diagonal contribution and define $\{M^{k\ell,\mathrm{F}}_{i,rr'}\}$ as the collected coefficients in

$$\mathscr{O}^{\mathrm{F}}_{k\ell}:=\frac{1}{L_T^2}\sum_{i=1}^{n}\sum_{r\ne r'}M^{k\ell,\mathrm{F}}_{i,rr'}\varepsilon_{ir}\varepsilon_{ir'}.$$

The coefficients include the matched-set indicators and hence vanish when the relevant unit is unavailable. Equations (SA-II.10.35) and (SA-II.10.36), together with Jensen's inequality over the rolling averages, give

$$\sum_{r\ne r'}\sum_{i=1}^{n}|M^{k\ell,\mathrm{F}}_{i,rr'}|^2\lesssim_{\mathbb{P}}\frac{L_THJ^{3d}}{n^3}. \tag{SA-II.10.43}$$

Combining the coefficients on (r,r') and (r',r), which multiply the same product $\varepsilon_{ir}\varepsilon_{ir'}$, (SA-II.10.38) bounds this predictable quadratic shock sum by $O_{\mathbb{P}}\{\sqrt{L_THJ^{3d}/n^3}/L_T^2\}$.

For the pre-s component when s is the later base date, put

$$P^{k\ell}_{is}:=\widehat{w}_{is}(\mathsf{b}_\ell)\sum_{h=1}^{H}\widehat{w}_{i,s-h}(\mathsf{b}_k)u_{i,s-h}^{<s}.$$

This coefficient is $\mathcal{K}_s$-measurable. The maximum-weight bound in (SA-II.10.35), Cauchy–Schwarz over h, and (SA-II.10.42) imply

$$\max_{k,\ell} \sum_i \sum_s (1 + \|\varpi_{s+H}\|^2)|P_{is}^{k\ell}|^2 \lesssim_{\mathbb{P}} \frac{L_T H J^{3d}}{n^3}.$$

Equation (SA-II.10.37), applied with $Z_{is} = P_{is}^{k\ell}$, therefore rewrites $L_T^{-2} \sum_{i,s} P_{is}^{k\ell} u_{is}^{\circ}$ as a predictable linear shock sum whose coefficient-square sum is $O_{\mathbb{P}}(L_T H J^{3d}/n^3)$. The same argument applies after interchanging the base dates. Combining the two orderings of the base dates with the preceding off-diagonal contribution yields

$$\mathscr{O}_{k\ell} = O_{\mathbb{P}}\left(\sqrt{\frac{H J^{3d}}{n^3 L_T^3}}\right) = o_{\mathbb{P}}\left(\frac{J^d}{n L_T}\right). \tag{SA-II.10.44}$$

Step 2: Fitted-residual error. We next replace u_{it}° by the fitted residual. From (SA-II.10.24), the IV matrix bounds, and (SA-II.10.27),

$$\frac{1}{L_T} \sum_{t=H+1}^{T-H} \|\widehat{\theta}_{t+H}^{\mathtt{IV}} - \theta_{t+H}\|^2 \lesssim_{\mathbb{P}} e_{nT}^2. \tag{SA-II.10.45}$$

Moreover,

$$\widehat{e}_{i(t+H)} - u_{it}^{\circ} = -\varpi_{t+H}^{\top}(\breve{\beta}_{i(t+H)} - \beta_{i(t+H)}) - \widehat{x}_{i(t+H)}^{\top}(\widehat{\theta}_{t+H}^{\mathtt{IV}} - \theta_{t+H}).$$

For $i \in \mathcal{I}_t^{\mathtt{LIN}}$, let the left-hand side be δ_{it}, and set $\delta_{it} = 0$ otherwise. The first bound in (SA-II.10.35), the time-averaged factor moments, and (SA-II.10.45) imply

$$\max_{k \in [K_{\mathtt{LIN}}]} \frac{1}{L_T} \sum_{t=H+1}^{T-H} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \widehat{w}_{it}(\mathsf{b}_k)^2 \delta_{it}^2 \lesssim_{\mathbb{P}} \frac{J^d}{n} e_{nT}^2. \tag{SA-II.10.46}$$

To make the residual replacement explicit, for arrays $a = \{a_{it}\}$ and $b = \{b_{it}\}$ define the banded bilinear form

$$\mathcal{H}_{k\ell}(a,b) := \frac{1}{L_T^2} \sum_{i=1}^{n} \sum_{t=H+1}^{T-H} \sum_{s=H+1}^{T-H} \widehat{w}_{it}(\mathsf{b}_k) \widehat{w}_{is}(\mathsf{b}_\ell) a_{it} b_{is} \mathbb{1}(|t-s| \le H),$$

using the zero-weight convention for unavailable units. Since $\|a\|_k^2 := L_T^{-1} \sum_{t=H+1}^{T-H} \sum_{i=1}^{n} \widehat{w}_{it}(\mathsf{b}_k)^2 a_{it}^2$, banded Cauchy–Schwarz gives

$$|\mathcal{H}_{k\ell}(a,b)| \le \frac{2H+1}{L_T} \|a\|_k \|b\|_\ell. \tag{SA-II.10.47}$$

Since $\widehat{e}_{i(t+H)} = u_{it}^{\circ} + \delta_{it}$,

$$\widehat{\Sigma}_{\mathtt{LIN}}(\mathsf{b}_k, \mathsf{b}_\ell) - \widetilde{\Sigma}_{k\ell} = \mathcal{H}_{k\ell}(u^{\circ}, \delta) + \mathcal{H}_{k\ell}(\delta, u^{\circ}) + \mathcal{H}_{k\ell}(\delta, \delta). \tag{SA-II.10.48}$$

The inequality $2|xy| \le x^2 + y^2$, the fact that each base date has at most $2H+1$ retained partners, and

(SA-II.10.46) give

$$\max_{k,\ell} |\mathcal{H}_{k\ell}(\delta,\delta)| \lesssim_{\mathbb{P}} \frac{J^d}{nL_T} H e_{nT}^2. \tag{SA-II.10.49}$$

To decompose the two remaining cross terms, put

$$\begin{aligned}
&d_{it} = \breve{\beta}_{i(t+H)} - \beta_{i(t+H)}, && \breve{x}_{i(t+H)} = (1, \breve{\beta}_{i(t+H)}^\top)^\top, && C_t^\circ = (\mathsf{M}_t^\circ)^{-1},\\
&h_t := \frac{1}{n_t^{\mathtt{LIN}}} \sum_{j \in \mathcal{I}_t^{\mathtt{LIN}}} \zeta_{jt} \varpi_{t+H}^\top d_{jt}, && \mathscr{B}_{t+H}^{\circ,\mathtt{IV}} := -C_t^\circ h_t,\\
&b_{it}^\circ := -\varpi_{t+H}^\top d_{it} - \breve{x}_{i(t+H)}^\top \mathscr{B}_{t+H}^{\circ,\mathtt{IV}}, && q_{it}^\circ := -\breve{x}_{i(t+H)}^\top C_t^\circ g_t.
\end{aligned}$$

For $i \in \mathcal{I}_t^{\mathtt{LIN}}$, also define

$$r_{it} := -\left\{ \widehat{x}_{i(t+H)}^\top \widehat{\mathsf{M}}_t^{-1} - \breve{x}_{i(t+H)}^\top C_t^\circ \right\} (g_t - h_t).$$

Set $b_{it}^\circ = q_{it}^\circ = r_{it} = 0$ outside the matched set. Direct substitution gives the exact identity $\delta_{it} = b_{it}^\circ + q_{it}^\circ + r_{it}$. Consequently, the two cross terms in (SA-II.10.48) satisfy

$$\begin{aligned}
&\mathcal{H}_{k\ell}(u^\circ,\delta) + \mathcal{H}_{k\ell}(\delta,u^\circ)\\
&\quad = \mathcal{H}_{k\ell}(u^\circ,b^\circ) + \mathcal{H}_{k\ell}(b^\circ,u^\circ) + \mathcal{H}_{k\ell}(u^\circ,q^\circ) + \mathcal{H}_{k\ell}(q^\circ,u^\circ)\\
&\qquad + \mathcal{H}_{k\ell}(u^\circ,r) + \mathcal{H}_{k\ell}(r,u^\circ).
\end{aligned} \tag{SA-II.10.50}$$

Step 3: Cross terms. Multiplier remainder. Assumption SA-LIN-IV(iii) gives

$$\max_{H+1 \le t \le T-H} \|\widehat{\mathsf{M}}_t^{-1} - (\mathsf{M}_t^\circ)^{-1}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(nT)}{n}}.$$

For $i \in \mathcal{I}_t^{\mathtt{LIN}}$,

$$\widehat{x}_{i(t+H)}^\top \widehat{\mathsf{M}}_t^{-1} - \breve{x}_{i(t+H)}^\top C_t^\circ = (0, v_{i(t+H)}^\top) \widehat{\mathsf{M}}_t^{-1} + \breve{x}_{i(t+H)}^\top (\widehat{\mathsf{M}}_t^{-1} - C_t^\circ).$$

The conditional score bound (SA-II.10.27), $\max_{i,t} \|d_{it}\| \le \mathsf{B}_{\mathtt{FS}}$ w.p.a.1, and the time-averaged factor-moment bound imply

$$\frac{1}{L_T} \sum_t \|g_t - h_t\|^2 \lesssim_{\mathbb{P}} e_{nT}^2, \qquad \frac{1}{L_T} \sum_t \mathbb{E}[\|g_t - h_t\|^4 \mid \mathcal{K}_t] \lesssim_{\mathbb{P}} e_{nT}^4.$$

Together with $\mathbb{E}[\|v_{i(t+H)}\|^2 \mid \mathcal{K}_t] \lesssim_{\mathbb{P}} H^{-1}$, the corresponding conditional fourth-moment bounds, and the maximum-instrument bound in Assumption SA-LIN-IV(ii), the multiplier identity above gives

$$\max_{k \in [K_{\mathtt{LIN}}]} \frac{1}{L_T} \sum_{t=H+1}^{T-H} \sum_{i \in \mathcal{I}_t^{\mathtt{LIN}}} \widehat{w}_{it}(\mathsf{b}_k)^2 r_{it}^2 \lesssim_{\mathbb{P}} \frac{J^d}{n} \rho_{nT}^2 e_{nT}^2. \tag{SA-II.10.51}$$

The corresponding weighted mean square for u° is $O_{\mathbb{P}}(J^d/n)$. Hence (SA-II.10.47) gives

$$\max_{k,\ell}\{|\mathcal{H}_{k\ell}(u^\circ,r)|+|\mathcal{H}_{k\ell}(r,u^\circ)|\}\lesssim_{\mathbb{P}}\frac{J^d}{nL_T}H\rho_{nT}e_{nT}=o_{\mathbb{P}}\left(\frac{J^d}{nL_T}\right), \tag{SA-II.10.52}$$

by (SA-II.10.34).

First-step bias. Use the pre-s/s-onward split introduced in Step 1. By the definition of κ_{tr} above,

$$u_{it}^{\geq s}=\sum_{r=1}^{T}\kappa_{tr}\varepsilon_{ir}\mathbb{1}(r\geq\max(t,s)).$$

Define the s-onward and pre-s pieces explicitly by

$$\begin{aligned}\mathscr{F}_{k\ell}^{u,b}&:=\frac{1}{L_T^2}\sum_{i=1}^{n}\sum_{t}\sum_{s}\widehat{w}_{it}(\mathsf{b}_k)\widehat{w}_{is}(\mathsf{b}_\ell)u_{it}^{\geq s}b_{is}^\circ\mathbb{1}(|t-s|\leq H),\\ \mathscr{P}_{k\ell}^{u,b}&:=\frac{1}{L_T^2}\sum_{i=1}^{n}\sum_{t}\sum_{s}\widehat{w}_{it}(\mathsf{b}_k)\widehat{w}_{is}(\mathsf{b}_\ell)u_{it}^{<s}b_{is}^\circ\mathbb{1}(|t-s|\leq H),\\ \mathcal{H}_{k\ell}(u^\circ,b^\circ)&=\mathscr{F}_{k\ell}^{u,b}+\mathscr{P}_{k\ell}^{u,b}.\end{aligned}$$

The two weights and b_{is}° use only innovations dated before their base dates, and κ_{tr} is $\mathcal{E}_\infty$-measurable. Define

$$F_{ir}^{k\ell}:=\sum_{t}\sum_{s}\widehat{w}_{it}(\mathsf{b}_k)\widehat{w}_{is}(\mathsf{b}_\ell)b_{is}^\circ\kappa_{tr}\mathbb{1}(r\geq\max(t,s))\mathbb{1}(|t-s|\leq H).$$

Then $F_{ir}^{k\ell}$ is measurable before ε_{ir}, and

$$\mathscr{F}_{k\ell}^{u,b}=\frac{1}{L_T^2}\sum_{i=1}^{n}\sum_{r=1}^{T}F_{ir}^{k\ell}\varepsilon_{ir}.$$

Equation (SA-II.10.35), the bound $\max_{i,t}\|d_{it}\|\leq\mathsf{B}_{\mathsf{FS}}$ w.p.a.1, and the instrument moment, inverse-matrix, and time-averaged factor-moment bounds give

$$\max_{k,\ell}\sum_{i=1}^{n}\sum_{r=1}^{T}|F_{ir}^{k\ell}|^2\lesssim_{\mathbb{P}}\frac{L_TH^2J^{3d}}{n^3}\mathsf{B}_{\mathsf{FS}}^2.$$

The predictable linear-shock-sum bound therefore yields

$$\max_{k,\ell}|\mathscr{F}_{k\ell}^{u,b}|\lesssim_{\mathbb{P}}\sqrt{\frac{H^2J^{3d}}{n^3L_T^3}}\,\mathsf{B}_{\mathsf{FS}}.$$

For the pre-$(t+h)$ component, we use a direct bound rather than martingale cancellation. For $h=1,\ldots,H$, let

$$\begin{aligned}E_{k,h}^{u}&:=\frac{1}{L_T}\sum_{t}\sum_{i}\widehat{w}_{it}(\mathsf{b}_k)^2(u_{it}^{<t+h})^2,\\ E_{\ell,h}^{b}&:=\frac{1}{L_T}\sum_{t}\sum_{i}\widehat{w}_{i,t+h}(\mathsf{b}_\ell)^2(b_{i,t+h}^\circ)^2,\end{aligned}$$

where boundary terms are omitted. Equation (SA-II.10.42), the maximum-instrument bound, $n_t^{\mathtt{LIN}} \asymp n$, and $\sup_t \|C_t^\circ\| \lesssim_{\mathbb{P}} 1$ give

$$\sum_{h=1}^{H} E_{k,h}^{u} \lesssim_{\mathbb{P}} \frac{J^d}{n}, \qquad \sum_{h=1}^{H} E_{\ell,h}^{b} \lesssim_{\mathbb{P}} \frac{HJ^d}{n}\mathsf{B}_{\mathtt{FS}}^2.$$

Cauchy–Schwarz first over (i,t) and then over h therefore bounds the pre-s piece explicitly as

$$\max_{k,\ell} |\mathscr{P}_{k\ell}^{u,b}| \le \frac{1}{L_T}\sum_{h=1}^{H}\sqrt{E_{k,h}^{u}E_{\ell,h}^{b}} \lesssim_{\mathbb{P}} \frac{J^d}{nL_T}\sqrt{H}\,\mathsf{B}_{\mathtt{FS}}.$$

By interchanging t and s,

$$\mathcal{H}_{k\ell}(b^\circ,u^\circ) = \mathcal{H}_{\ell k}(u^\circ,b^\circ) = \mathscr{F}_{\ell k}^{u,b} + \mathscr{P}_{\ell k}^{u,b}.$$

Consequently,

$$\max_{k,\ell}\{|\mathcal{H}_{k\ell}(u^\circ,b^\circ)| + |\mathcal{H}_{k\ell}(b^\circ,u^\circ)|\} \lesssim_{\mathbb{P}} \frac{J^d}{nL_T}\left\{\sqrt{H}\,\mathsf{B}_{\mathtt{FS}} + H\mathsf{B}_{\mathtt{FS}}\sqrt{\frac{J^d}{nL_T}}\right\} = o_{\mathbb{P}}\left(\frac{J^d}{nL_T}\right). \tag{SA-II.10.53}$$

The last equality follows from $\sqrt{H}\,\mathsf{B}_{\mathtt{FS}} = o(1)$ and (SA-II.10.34).

IV score. The maximum-instrument bound in Assumption SA-LIN-IV(ii) and (SA-II.10.35) imply

$$\max_{t,k}\sum_{i\in\mathcal{I}_t^{\mathtt{LIN}}} \widehat{w}_{it}(\mathsf{b}_k)^2\|\zeta_{it}\|^2 \lesssim_{\mathbb{P}} \frac{J^d}{n}. \tag{SA-II.10.54}$$

Substituting $g_s = (n_s^{\mathtt{LIN}})^{-1}\sum_{j\in\mathcal{I}_s^{\mathtt{LIN}}}\zeta_{js}u_{js}^\circ$ into q_{is}°, for $i\in\mathcal{I}_s^{\mathtt{LIN}}$, gives

$$\begin{aligned} q_{is}^{\circ,\mathrm{same}} &:= -\frac{\breve{x}_{i(s+H)}^\top C_s^\circ\zeta_{is}}{n_s^{\mathtt{LIN}}}u_{is}^\circ, \\ q_{is}^{\circ,\mathrm{cross}} &:= -\frac{1}{n_s^{\mathtt{LIN}}}\sum_{\substack{j\in\mathcal{I}_s^{\mathtt{LIN}}\\ j\neq i}}\breve{x}_{i(s+H)}^\top C_s^\circ\zeta_{js}u_{js}^\circ, \qquad q_{is}^\circ = q_{is}^{\circ,\mathrm{same}} + q_{is}^{\circ,\mathrm{cross}}. \end{aligned}$$

Set both components to zero for $i\notin\mathcal{I}_s^{\mathtt{LIN}}$. Thus the displayed decomposition is exact for every $i\in[n]$. Accordingly, define the same-unit and cross-unit pieces of the two score cross-products by

$$\begin{aligned} \mathscr{Q}_{k\ell}^{\mathrm{same}} &:= \mathcal{H}_{k\ell}(u^\circ,q^{\circ,\mathrm{same}}) + \mathcal{H}_{k\ell}(q^{\circ,\mathrm{same}},u^\circ), \\ \mathscr{Q}_{k\ell}^{\mathrm{cross}} &:= \mathcal{H}_{k\ell}(u^\circ,q^{\circ,\mathrm{cross}}) + \mathcal{H}_{k\ell}(q^{\circ,\mathrm{cross}},u^\circ), \\ \mathcal{H}_{k\ell}(u^\circ,q^\circ) + \mathcal{H}_{k\ell}(q^\circ,u^\circ) &= \mathscr{Q}_{k\ell}^{\mathrm{same}} + \mathscr{Q}_{k\ell}^{\mathrm{cross}}. \end{aligned}$$

The same-unit piece need not be centered. The multiplier $\breve{x}_{i(s+H)}^\top C_s^\circ\zeta_{is}$ is uniformly bounded in probability and $(n_s^{\mathtt{LIN}})^{-1}\lesssim n^{-1}$. Applying $2|xy|\le x^2+y^2$ to the banded sum, then using (SA-II.10.35) and the weighted

second-moment bound for u° used above, gives

$$\max_{k,\ell} |\mathscr{Q}_{k\ell}^{\text{same}}| \lesssim_{\mathbb{P}} \frac{J^d}{nL_T} \frac{H}{n}. \tag{SA-II.10.55}$$

For $j \neq i$, Assumption SA-LIN-M preserves conditional independence of future primitive shocks across units after selection. For each pair t, s, split the residual associated with the earlier base date into its component before the later base date and its component from the later base date onward. In the latter component, all weights, instruments, and IV multipliers are fixed before both primitive shocks. Let $\mathscr{Q}_{k\ell}^{\text{cross,F}}$ denote this contribution and define $\{N_{ij,rr'}^{k\ell,\text{F}}\}$ as the collected coefficients in

$$\mathscr{Q}_{k\ell}^{\text{cross,F}} := \frac{1}{L_T^2} \sum_{\substack{i,j\in[n]\\ i\neq j}} \sum_{r=1}^{T} \sum_{r'=1}^{T} N_{ij,rr'}^{k\ell,\text{F}} \varepsilon_{ir} \varepsilon_{jr'}.$$

These coefficients include all matched-set indicators and are zero if a required unit is unavailable. Assumption SA-LIN-IV(ii), (SA-II.10.35), (SA-II.10.36), and Jensen's inequality give

$$\sum_{\substack{i,j\in[n]\\ i\neq j}} \sum_{r=1}^{T} \sum_{r'=1}^{T} |N_{ij,rr'}^{k\ell,\text{F}}|^2 \lesssim_{\mathbb{P}} \frac{L_T H^2 J^{3d}}{n^4}. \tag{SA-II.10.56}$$

After ordering the compound shock indices, combine the coefficients on each pair and its reversal, which multiply the same shock product. The predictable quadratic-shock-sum bound in (SA-II.10.38) then gives

$$\max_{k,\ell} |\mathscr{Q}_{k\ell}^{\text{cross,F}}| \lesssim_{\mathbb{P}} \sqrt{\frac{H^2 J^{3d}}{n^4 L_T^3}}.$$

Consider first $\mathcal{H}_{k\ell}(u^\circ, q^{\circ,\text{cross}})$ in the case $t < s$, so the IV-score term indexed by s has the later base date. Its pre-s contribution is linear in u_{js}°. For $j \in \mathcal{I}_s^{\texttt{LIN}}$, its $\mathcal{K}_s$-measurable coefficient is

$$P_{js}^{k\ell,t<s} := -\frac{1}{n_s^{\texttt{LIN}}} \sum_{h=1}^{H} \sum_{\substack{i\in\mathcal{I}_s^{\texttt{LIN}}\\ i\neq j}} \widehat{w}_{i,s-h}(\mathsf{b}_k) \widehat{w}_{is}(\mathsf{b}_\ell) \{\breve{x}_{i(s+H)}^\top C_s^\circ \zeta_{js}\} u_{i,s-h}^{<s}.$$

As stipulated above, shifted base dates outside $\mathcal{T}_H$ are omitted. Set this coefficient to zero for $j \notin \mathcal{I}_s^{\texttt{LIN}}$. Since $n_s^{\texttt{LIN}} \asymp n$, Cauchy–Schwarz, (SA-II.10.42), (SA-II.10.35), and the uniform bounds on ζ_{js}, $\breve{x}_{i(s+H)}$, and C_s° give

$$\max_{k,\ell} \sum_j \sum_s (1 + \|\varpi_{s+H}\|^2) |P_{js}^{k\ell,t<s}|^2 \lesssim_{\mathbb{P}} \frac{L_T H J^{3d}}{n^3}.$$

Equation (SA-II.10.37) therefore gives $O_{\mathbb{P}}\Big(\sqrt{HJ^{3d}/(n^3 L_T^3)}\Big)$ for the corresponding predictable linear shock sum.

Next consider the same bilinear form in the case $s < t$, so the IV-score term indexed by s has the earlier

base date. For $s,t \in \mathcal{T}_H$ with $s<t$ and $i \in \mathcal{I}_s^{\mathtt{LIN}}$, define

$$G_{i,s,t}^{<t} := \frac{1}{n_s^{\mathtt{LIN}}} \sum_{\substack{j \in \mathcal{I}_s^{\mathtt{LIN}} \\ j \neq i}} \breve{x}_{i(s+H)}^{\top} C_s^{\circ} \zeta_{js} u_{js}^{<t}.$$

Set $G_{i,s,t}^{<t}=0$ outside this domain. For $t \in \mathcal{T}_H$ and $i \in \mathcal{I}_t^{\mathtt{LIN}}$, define

$$P_{it}^{k\ell,s<t} := -\widehat{w}_{it}(\mathsf{b}_k) \sum_{h=1}^{H} \widehat{w}_{i,t-h}(\mathsf{b}_\ell) G_{i,t-h,t}^{<t}.$$

Set $P_{it}^{k\ell,s<t}=0$ outside this domain. Then $P_{it}^{k\ell,s<t}$ is $\mathcal{K}_t$-measurable and is the coefficient multiplying the residual u_{it}° at the later base date t. Assumption SA-LIN-IV(ii) gives $|G_{i,s,t}^{<t}|^2 \lesssim_{\mathbb{P}} n^{-1} \sum_{j \in \mathcal{I}_s^{\mathtt{LIN}}} (u_{js}^{<t})^2$. Cauchy–Schwarz over h, (SA-II.10.35), and the second bound in (SA-II.10.42) therefore give

$$\max_{k,\ell} \sum_i \sum_t (1+\|\varpi_{t+H}\|^2) |P_{it}^{k\ell,s<t}|^2 \lesssim_{\mathbb{P}} \frac{L_T H J^{3d}}{n^3}.$$

Equation (SA-II.10.37) gives the same rate in the case $s<t$. The identity $\mathcal{H}_{k\ell}(q^{\circ,\mathrm{cross}}, u^{\circ}) = \mathcal{H}_{\ell k}(u^{\circ}, q^{\circ,\mathrm{cross}})$ covers the other cross term. Consequently,

$$\max_{k,\ell} |\mathscr{Q}_{k\ell}^{\mathrm{cross}}| \lesssim_{\mathbb{P}} \sqrt{\frac{H^2 J^{3d}}{n^4 L_T^3}} + \sqrt{\frac{H J^{3d}}{n^3 L_T^3}} = o_{\mathbb{P}}\left(\frac{J^d}{nL_T}\right). \tag{SA-II.10.57}$$

Relative to $J^d/(nL_T)$, the same-unit, cross-unit s-onward, and cross-unit pre-s bounds are

$$\frac{H}{n}, \qquad H\sqrt{\frac{J^d}{n^2 L_T}}, \qquad \text{and} \qquad \sqrt{\frac{HJ^d}{nL_T}}.$$

All three are $o(1)$ by (SA-II.10.34).

Conclusion. The quadratic and random-multiplier terms are controlled by (SA-II.10.49) and (SA-II.10.52). The bias and score terms are controlled by (SA-II.10.53), (SA-II.10.55), and (SA-II.10.57). Substitution into (SA-II.10.48) and (SA-II.10.50) gives, uniformly over the fixed grid,

$$\begin{aligned} \max_{k,\ell \in [K_{\mathtt{LIN}}]} \left| \widehat{\Sigma}_{\mathtt{LIN}}(\mathsf{b}_k, \mathsf{b}_\ell) - \widetilde{\Sigma}_{k\ell} \right| &\lesssim_{\mathbb{P}} \frac{J^d}{nL_T} H e_{nT}^2 + o_{\mathbb{P}}\left(\frac{J^d}{nL_T}\right) \\ &= o_{\mathbb{P}}\left(\frac{J^d}{nL_T}\right). \end{aligned} \tag{SA-II.10.58}$$

Step 1, specifically (SA-II.10.41) and (SA-II.10.44), also gives

$$\max_{k,\ell \in [K_{\mathtt{LIN}}]} \left| \widetilde{\Sigma}_{k\ell} - \Sigma_{\mathtt{LIN}}(\mathsf{b}_k, \mathsf{b}_\ell) \right| = o_{\mathbb{P}}\left(\frac{J^d}{nL_T}\right).$$

The triangle inequality and fixed $K_{\mathtt{LIN}}$ therefore give

$$\left\| \widehat{\Sigma}_{\mathtt{LIN}} - \Sigma_{\mathtt{LIN}} \right\| \leq K_{\mathtt{LIN}} \max_{k,\ell \in [K_{\mathtt{LIN}}]} \left| \widehat{\Sigma}_{\mathtt{LIN}}(\mathsf{b}_k, \mathsf{b}_\ell) - \Sigma_{\mathtt{LIN}}(\mathsf{b}_k, \mathsf{b}_\ell) \right|$$

$$= o_{\mathbb{P}}\left(\frac{J^d}{nL_T}\right),$$

which is (SA-II.10.20). □

SA-II.10.8 Proof of Theorem SA-II.10.3

By Lemma SA-II.10.2, (SA-II.10.16), and (SA-II.10.17), $\lambda_{\min}(\Sigma_{\texttt{LIN}}) \gtrsim_{\mathbb{P}} J^d/\{n(T-2H)\}$ and hence

$$\left\|\Sigma_{\texttt{LIN}}^{-1/2}\widehat{\Sigma}_{\texttt{LIN}}\Sigma_{\texttt{LIN}}^{-1/2} - I_{K_{\texttt{LIN}}}\right\| = o_{\mathbb{P}}(1).$$

It follows that $\widehat{\Sigma}_{\texttt{LIN}}$ is nonsingular with probability approaching one. Put $Z_{\texttt{LIN}} = \Sigma_{\texttt{LIN}}^{-1/2}\widehat{\boldsymbol{\nu}}_{\texttt{LIN}}$ and $A_{\texttt{LIN}} = \Sigma_{\texttt{LIN}}^{-1/2}\widehat{\Sigma}_{\texttt{LIN}}\Sigma_{\texttt{LIN}}^{-1/2}$. Then

$$\widehat{\boldsymbol{\nu}}_{\texttt{LIN}}^{\top}\widehat{\Sigma}_{\texttt{LIN}}^{-1}\widehat{\boldsymbol{\nu}}_{\texttt{LIN}} = Z_{\texttt{LIN}}^{\top}A_{\texttt{LIN}}^{-1}Z_{\texttt{LIN}},$$
$$Z_{\texttt{LIN}}^{\top}A_{\texttt{LIN}}^{-1}Z_{\texttt{LIN}} \rightsquigarrow \chi^2_{K_{\texttt{LIN}}}.$$

The convergence follows from Lemma SA-II.10.1, Slutsky's theorem, and the continuous mapping theorem. Continuity of the chi-square distribution at its $1-\alpha$ quantile gives the asserted asymptotic size. □

SA-II.10.9 Proof of Lemma SA-II.10.4

Write

$$\widehat{x}_{it}^{\texttt{old}} = (1, \widehat{\beta}_{it}^{\top})^{\top}, \qquad \breve{x}_{it} = (1, \breve{\beta}_{it}^{\top})^{\top}.$$

Because the rolling window defining $\widehat{\beta}_{it}$ ends at $t-1$, $\widehat{x}_{it}^{\texttt{old}}$ is $\mathcal{F}_{t-1}$-measurable and hence $\mathcal{K}_t$-measurable. This proves Assumption SA-LIN-IV(i).

Lemma SA-II.2.3, the boundedness of β_{it} in Assumption SA-DGP(iii), and $\mathsf{R}_{\texttt{FS}} = o(1)$ give

$$\max_{H+1\le t\le T-H}\max_{i\in\mathcal{I}_t^{\texttt{LIN}}}\|\widehat{\beta}_{it}\| \lesssim_{\mathbb{P}} 1.$$

Consequently,

$$\max_{H+1\le t\le T-H}\frac{1}{n_t^{\texttt{LIN}}}\sum_{i\in\mathcal{I}_t^{\texttt{LIN}}}\|\widehat{x}_{it}^{\texttt{old}}\|^4 \lesssim_{\mathbb{P}} 1.$$

The last two displays verify Assumption SA-LIN-IV(ii).

The future-window decomposition and conditional mean-zero property imply

$$\mathsf{M}_t^{\circ} = \frac{1}{n_t^{\texttt{LIN}}}\sum_{i\in\mathcal{I}_t^{\texttt{LIN}}}\widehat{x}_{it}^{\texttt{old}}\breve{x}_{i(t+H)}^{\top}.$$

The first-step rate and boundedness of the true betas therefore give

$$\max_{H+1\le t\le T-H}\|\mathsf{M}_t^{\circ} - \mathsf{M}_{t,\beta}^{0}\| \lesssim_{\mathbb{P}} \mathsf{R}_{\texttt{FS}} = o_{\mathbb{P}}(1).$$

Equation (SA-II.10.21) and a matrix perturbation bound then yield

$$\max_{H+1\le t\le T-H}\left\{\|\mathsf{M}_t^\circ\|+\|(\mathsf{M}_t^\circ)^{-1}\|\right\}\lesssim_{\mathbb{P}} 1.$$

Finally, with $(0, v_{i(t+H)}^\top)$ interpreted as a row vector,

$$\widehat{\mathsf{M}}_t-\mathsf{M}_t^\circ=\frac{1}{n_t^{\mathtt{LIN}}}\sum_{i\in\mathcal{I}_t^{\mathtt{LIN}}}\widehat{x}_{it}^{\mathtt{old}}(0,v_{i(t+H)}^\top).$$

Conditional on $\mathcal{K}_t$, the summands are centered and independent across units; Assumption SA-LIN-M preserves this property after restriction to the matched set. The first-step Gram bound and the conditional moment bounds in Assumption SA-DGP(ii) give

$$\mathbb{E}[\|v_{i(t+H)}\|^2\mid\mathcal{K}_t]\lesssim_{\mathbb{P}} H^{-1}.$$

Applying the conditional Fuk–Nagaev inequality and a union bound over the base dates, using $T^{2/q}/H=o(1)$ and $cn\le n_t^{\mathtt{LIN}}\le n$, yields

$$\max_{H+1\le t\le T-H}\|\widehat{\mathsf{M}}_t-\mathsf{M}_t^\circ\|\lesssim_{\mathbb{P}}\sqrt{\frac{\log(nT)}{nH}}+\frac{(nT)^{1/q}\log(nT)}{nH}\lesssim_{\mathbb{P}}\sqrt{\frac{\log(nT)}{n}}.$$

This proves the concentration part of Assumption SA-LIN-IV(iii) and completes the verification. □

References

Cattaneo, M. D., Feng, Y., Underwood, W. G., 2024. Uniform inference for kernel density estimators with dyadic data. Journal of the American Statistical Association 119, 2695–2708.

Freedman, D. A., 1975. On tail probabilities for martingales. Annals of Probability 3, 100–118.

Rio, E., 2017. Asymptotic Theory of Weakly Dependent Random Processes, vol. 80 of *Probability Theory and Stochastic Modelling*. Springer, Berlin, Heidelberg.

Beta-Sorted Portfolios
Supplemental Appendix: Applied

Matias D. Cattaneo* Richard K. Crump† Yingjie Feng‡ Weining Wang§

September 22, 2026

Abstract

This supplemental appendix provides practical guidance and numerical evidence for beta-sorted portfolio estimation and inference. Part III presents data-driven choices of the first-step bandwidth and the second-step number of portfolios and discusses their use in pointwise-in-β and uniform-in-β inference. Part IV reports Monte Carlo evidence on confidence interval coverage and average length, additional empirical results for the business-credit factor, and an application to capital share risk. The companion Supplemental Appendix: Theory provides the theoretical foundations for these estimation and inference procedures.

*Department of Economics, Princeton University.
†Capital Markets, Federal Reserve Bank of New York.
‡School of Economics and Management, Tsinghua University.
§School of Economics, University of Bristol.

Contents

Part III: Tuning Parameter Selection 1

SA-III.1 Pointwise-in-β 1
- SA-III.1.1 First-Step Bandwidth Selection . . . 2
- SA-III.1.2 Second-Step Selection of the Number of Portfolios . . . 2
- SA-III.1.3 Estimation and Robust Bias-Corrected Inference . . . 4
- SA-III.1.4 Validity of Tuning Parameter Selection . . . 4
- SA-III.1.5 Proof of Lemma SA-III.1.1 . . . 6

SA-III.2 Uniform-in-β 8
- SA-III.2.1 Near-Origin Truncation Parameter . . . 8
- SA-III.2.2 Estimation and Robust Bias-Corrected Inference . . . 9

SA-III.3 Factor-Cancelling Linear Contrast 10

SA-III.4 Default Implementation Choices 10

Part IV: Numerical Evidence 12

SA-IV.1 Simulations 12
- SA-IV.1.1 Data-Generating Process . . . 12
- SA-IV.1.2 Monte Carlo Implementation . . . 14
- SA-IV.1.3 Results . . . 14

SA-IV.2 Empirical Applications 21
- SA-IV.2.1 Additional Results for Business Credit Cycle Factor . . . 21
  - SA-IV.2.1.1 Out-of-Sample Predictive Regressions . . . 21
  - SA-IV.2.1.2 Implementation and Tuning . . . 21
- SA-IV.2.2 Capital Share Risk Factor . . . 22

Part III: Tuning Parameter Selection

This part turns the rate restrictions in Part II into implementable tuning choices for the classical beta-sorted portfolio estimator. Section SA-III.1 gives the pointwise first-step plug-in rule for the bandwidth h and the integrated-MSE plug-in rule for the second-step number of portfolios J. Section SA-III.2 keeps those h and J choices, and proposes a data-driven selector for the near-origin truncation parameter δ. Section SA-III.3 records the separate implication for factor-cancelling linear contrasts, whose idiosyncratic normalization requires dedicated undersmoothing rather than the default plug-in choices.

SA-III.1 Pointwise-in-β

This section summarizes a practical data-driven tuning rule for the classical beta-sorted portfolio estimator and an associated robust bias-corrected confidence interval. Throughout this section, braces on tuning parameters denote the relevant collection over time and units; for example, $\{J_t\}$ denotes $\{J_t : t = H + 1, \ldots, T\}$.

- We write the first-step estimator as $\widehat{\beta}_{it}(\mathfrak{p}, h_{it})$, where $\mathfrak{p}$ is the local-polynomial order and h_{it} is the bandwidth.
- We write the period-t second-step fit as $\widehat{M}_t(\beta; \mathfrak{q}, J_t)$, where $\mathfrak{q}$ is the portfolio polynomial order and J_t is the number of portfolios per beta dimension at time t.

The full second-step estimator averages these period-specific fits:

$$\widehat{\mu}(\beta; \mathfrak{q}, \{J_t\}) = \frac{1}{T-H} \sum_{t=H+1}^{T} \widehat{M}_t(\beta; \mathfrak{q}, J_t).$$

Given choices of bandwidths h_{it} and portfolio counts J_t, the classical beta-sorted portfolio estimator is $\widehat{\mu}(\beta; 0, \{J_t\})$ constructed using $\widehat{\beta}_{it}(0, h_{it})$. The bias-corrected estimator used for inference is $\widehat{\mu}(\beta; 1, \{J_t\})$ constructed using $\widehat{\beta}_{it}(1, h_{it})$.

The bandwidths h_{it} and portfolio counts J_t are chosen once for the classical point estimation objective and then reused for inference. Thus, the empirical rule is simple: choose h_{it} and J_t as if the objective were the classical estimator with $(\mathfrak{p}, \mathfrak{q}) = (0, 0)$, but form confidence intervals with the bias-corrected estimator with $(\mathfrak{p}, \mathfrak{q}) = (1, 1)$. This is the usual robust bias-correction logic: the lower-order fit determines the MSE-optimal tuning, and the higher-order fit removes the leading smoothing bias at that same tuning; see Calonico et al. (2014) and Calonico et al. (2018, 2022).

The plug-in logic is the same in both steps:

1. run a pilot regression at the point estimation order and an auxiliary pilot regression at the higher order;
2. estimate the variance constant from the point estimation pilot;
3. estimate the leading bias constant from the difference between the two pilot fits; and
4. plug the two constants into the MSE-optimal tuning formula.

SA-III.1.1 First-Step Bandwidth Selection

For first-step point estimation of order $\mathfrak{p} \geq 0$, we use squared Euclidean loss for the vector $\widehat{\beta}_{it}(\mathfrak{p}, h_{it})$. Accordingly, let $\mathcal{V}_{\mathtt{FS},it}(\mathfrak{p})$ denote the trace of its leading conditional covariance matrix and let $\mathcal{B}^2_{\mathtt{FS},it}(\mathfrak{p})$ denote the squared Euclidean norm of its leading bias vector. In the smooth-trend case considered in Lemma SA-II.2.4(i) with $s \geq \mathfrak{p}+1$, the point estimation first-step (conditional) MSE has the expansion

$$\mathrm{MSE}\big(\widehat{\beta}_{it}(\mathfrak{p}, h_{it})\big) = \mathcal{V}_{\mathtt{FS},it}(\mathfrak{p})(Th_{it})^{-1} + \mathcal{B}^2_{\mathtt{FS},it}(\mathfrak{p})h_{it}^{2(\mathfrak{p}+1)} + o_{\mathbb{P}}\big((Th_{it})^{-1} + h_{it}^{2(\mathfrak{p}+1)}\big).$$

The MSE-optimal bandwidth is therefore

$$h^{\star}_{it}(\mathfrak{p}) = \left(\frac{\mathcal{V}_{\mathtt{FS},it}(\mathfrak{p})}{2(\mathfrak{p}+1)\mathcal{B}^2_{\mathtt{FS},it}(\mathfrak{p})}\right)^{1/(2\mathfrak{p}+3)} T^{-1/(2\mathfrak{p}+3)},$$

provided that $\mathcal{B}^2_{\mathtt{FS},it}(\mathfrak{p}) > 0$. Algorithm SA-III.1.1 implements this rule for the classical β-estimation underlying the beta-sorted portfolio estimator.

Algorithm SA-III.1.1: First-step MSE-optimal bandwidth ($\mathfrak{p} = 0$)

Input: Pilot bandwidth $h_{it,\mathrm{pilot}}$

Pilot regressions at $h_{it,\mathrm{pilot}}$:

order $\mathfrak{p} = 0$ fit $\widehat{\beta}_{it}(0, h_{it,\mathrm{pilot}})$ of $R_{i,t-s}$ on $X_s(t)$ over $s \in [1, \lceil Th_{it,\mathrm{pilot}} \rceil]$

order $\mathfrak{p} = 1$ fit $\widehat{\beta}_{it}(1, h_{it,\mathrm{pilot}})$ over the same window and with the same kernel

Variance estimate: $\widehat{\mathcal{V}}_{\mathtt{FS},it}(0) \leftarrow$ trace of the sandwich covariance constant from the $\mathfrak{p} = 0$ pilot

Squared-bias estimate: $\widehat{\mathcal{B}}^2_{\mathtt{FS},it}(0) \leftarrow \left\| \big(\widehat{\beta}_{it}(1, h_{it,\mathrm{pilot}}) - \widehat{\beta}_{it}(0, h_{it,\mathrm{pilot}})\big)/h_{it,\mathrm{pilot}} \right\|^2$

Plug-in bandwidth:

$$\widehat{h}^{\star}_{it} \leftarrow \left(\frac{\widehat{\mathcal{V}}_{\mathtt{FS},it}(0)}{2\widehat{\mathcal{B}}^2_{\mathtt{FS},it}(0)}\right)^{1/3} T^{-1/3}$$

Estimates: compute $\widehat{\beta}_{it} = \widehat{\beta}_{it}(0, \widehat{h}^{\star}_{it})$ and $\widehat{\beta}^{\mathrm{BC}}_{it} = \widehat{\beta}_{it}(1, \widehat{h}^{\star}_{it})$

Output: $\widehat{h}^{\star}_{it}$, $\widehat{\beta}_{it}$, and $\widehat{\beta}^{\mathrm{BC}}_{it}$.

SA-III.1.2 Second-Step Selection of the Number of Portfolios

For second-step point estimation using within-portfolio polynomials of order $\mathfrak{q} \geq 0$, we use the same marginal portfolio count J_t in every beta coordinate. Thus J_t is a single scalar selected for period t, and the tensor-product partition contains at most J_t^d cells. Let $\mathcal{V}_{\mathtt{SS},t}(\beta; \mathfrak{q})$ denote the leading variance constant of $\widehat{M}_t(\beta; \mathfrak{q}, J_t)$, and let $\mathcal{B}_{\mathtt{SS},t}(\beta; \mathfrak{q})$ denote the leading approximation-bias constant multiplying $J_t^{-(\mathfrak{q}+1)}$. Up to additive terms that do not depend on the portfolio count, including the factor time-series component present when $\beta \neq 0$, we use the period-t point estimation approximate MSE criterion

$$\mathrm{MSE}_t\big(\widehat{M}_t(\beta; \mathfrak{q}, J_t)\big) = \mathcal{V}_{\mathtt{SS},t}(\beta; \mathfrak{q})\frac{J_t^d}{n_t} + \mathcal{B}_{\mathtt{SS},t}(\beta; \mathfrak{q})^2 J_t^{-2(\mathfrak{q}+1)} + o_{\mathbb{P}}(J_t^d/n_t + J_t^{-2(\mathfrak{q}+1)}).$$

To obtain one portfolio count for the whole beta domain, fix a nonnegative weight function $w(\beta)$ on $\mathcal{B}^{\circ}$

satisfying $0 < \int_{\mathcal{B}^\circ} w(\beta)d\beta < \infty$, and define the integrated variance and squared-bias constants

$$\mathcal{V}^{w}_{\mathtt{SS},t}(\mathfrak{q}) := \int_{\mathcal{B}^\circ} \mathcal{V}_{\mathtt{SS},t}(\beta;\mathfrak{q})w(\beta)d\beta,$$
$$\mathcal{B}^{2,w}_{\mathtt{SS},t}(\mathfrak{q}) := \int_{\mathcal{B}^\circ} \mathcal{B}_{\mathtt{SS},t}(\beta;\mathfrak{q})^2 w(\beta)d\beta.$$

The corresponding approximate integrated MSE is

$$\begin{aligned}\mathrm{IMSE}_t(\mathfrak{q}, J_t; w) = \mathcal{V}^{w}_{\mathtt{SS},t}(\mathfrak{q})\frac{J_t^d}{n_t} + \mathcal{B}^{2,w}_{\mathtt{SS},t}(\mathfrak{q})J_t^{-2(\mathfrak{q}+1)} \\ + o_{\mathbb{P}}(J_t^d/n_t + J_t^{-2(\mathfrak{q}+1)}).\end{aligned}$$

Thus the IMSE-optimal common number of portfolios per beta dimension is

$$J_t^\star(\mathfrak{q}; w) = \left(\frac{2(\mathfrak{q}+1)\mathcal{B}^{2,w}_{\mathtt{SS},t}(\mathfrak{q})}{d\,\mathcal{V}^{w}_{\mathtt{SS},t}(\mathfrak{q})}\right)^{1/(d+2\mathfrak{q}+2)} n_t^{1/(d+2\mathfrak{q}+2)},$$

provided that $\mathcal{B}^{2,w}_{\mathtt{SS},t}(\mathfrak{q}) > 0$. Algorithm SA-III.1.2 implements this rule for the classical beta-sorted portfolio estimator.

Algorithm SA-III.1.2: Second-step IMSE-optimal portfolios ($\mathfrak{q} = 0$)

Input: First-step point estimates $\{\widehat{\beta}_{it}\}$ from Algorithm SA-III.1.1; pilot portfolio counts $\{J_{t,\mathrm{pilot}}\}$ with common scale J_{pilot}; grid $\mathcal{G} = \{\beta_\ell : \ell \in [L]\} \subset \mathcal{B}^\circ$; nonnegative grid weights $\{w_\ell : \ell \in [L]\}$

Pilot portfolio regressions at $\{J_{t,\mathrm{pilot}}\}$:

second-step order $\mathfrak{q} = 0$ portfolio regression $\widehat{M}_t(\cdot; 0, J_{t,\mathrm{pilot}})$

second-step order $\mathfrak{q} = 1$ portfolio regression $\widehat{M}_t(\cdot; 1, J_{t,\mathrm{pilot}})$

Pointwise variance estimates: $\widehat{\mathcal{V}}_{\mathtt{SS},t}(\beta_\ell; 0) \leftarrow$ period-t sandwich variance constant from the $\mathfrak{q} = 0$ pilot for each $\ell \in [L]$

Integrated variance estimate: $\widehat{\mathcal{V}}^{w}_{\mathtt{SS},t}(0) \leftarrow \sum_{\ell=1}^{L} w_\ell \widehat{\mathcal{V}}_{\mathtt{SS},t}(\beta_\ell; 0)$

Integrated squared-bias estimate:

$$\widehat{\mathcal{B}}^{2,w}_{\mathtt{SS},t}(0) \leftarrow \sum_{\ell=1}^{L} w_\ell \left[\widehat{M}_t(\beta_\ell; 1, J_{t,\mathrm{pilot}}) - \widehat{M}_t(\beta_\ell; 0, J_{t,\mathrm{pilot}})\right]^2 J_{t,\mathrm{pilot}}^2$$

Plug-in number of portfolios:

$$\widehat{J}_t^\star \leftarrow \left(\frac{2\widehat{\mathcal{B}}^{2,w}_{\mathtt{SS},t}(0)}{d\,\widehat{\mathcal{V}}^{w}_{\mathtt{SS},t}(0)}\right)^{1/(d+2)} n_t^{1/(d+2)}$$

Output: The common marginal portfolio count $\widehat{J}_t^\star$.

Our default implementation takes $w_\ell = 1$ for every grid point. Hence the estimated squared-bias constant is the sum of the squared pointwise pilot biases and the estimated variance constant is the sum of the pointwise

pilot variance constants. Normalizing the grid weights to sum to one would leave $\widehat{J}_t^\star$ unchanged because the same normalization multiplies the numerator and denominator. On an irregular grid, the w_ℓ may instead combine the desired $w(\beta_\ell)$ with quadrature weights; this changes the discrete approximation to the integral but not the form of the selector.

SA-III.1.3 Estimation and Robust Bias-Corrected Inference

The classical beta-sorted portfolio estimator constructed using the first- and second-step plug-in rules is

$$\widehat{\mu}(\beta) = \widehat{\mu}(\beta; 0, \{\widehat{J}_t^\star\}),$$

which is computed using the $\mathfrak{p} = 0$ first-step estimates $\{\widehat{\beta}_{it}\}$ from Algorithm SA-III.1.1.

A robust bias-corrected (RBC) confidence interval uses the $\mathfrak{p} = \mathfrak{q} = 1$ fits with the selected tuning parameters $\{\widehat{h}_{it}^\star\}$ and $\{\widehat{J}_t^\star\}$, and the bias-corrected first-step point estimates $\{\widehat{\beta}_{it}^{\mathrm{BC}}\}$ from Algorithm SA-III.1.1. Specifically, a $100(1-\alpha)\%$ nominal level RBC confidence interval is

$$\widehat{\mu}^{\mathrm{BC}}(\beta) \pm z_{1-\alpha/2}\sqrt{\widehat{\Sigma}^{\mathrm{BC}}(\beta)},$$

where $\widehat{\mu}^{\mathrm{BC}}(\beta) = \widehat{\mu}(\beta; 1, \{\widehat{J}_t^\star\})$ using $\{\widehat{\beta}_{it}^{\mathrm{BC}}\}$, $\{\widehat{h}_{it}^\star\}$, and $\{\widehat{J}_t^\star\}$, and $\widehat{\Sigma}^{\mathrm{BC}}(\beta) = \widehat{\Sigma}(\beta)$ as in Section SA-II.7 for $\widehat{\mu}^{\mathrm{BC}}(\beta)$. For each fixed $\beta \neq 0$, under the pointwise Gaussian-approximation conditions and the conditions of Theorem SA-II.7.3, this interval has exact pointwise coverage when $T\mathscr{B}_{\mathtt{F}}(\beta) = o_{\mathbb{P}}(1)$; without this additional factor-prediction rate it is conservative because $\mathscr{B}_{\mathtt{F}}(\beta) \geq 0$, and no anchoring condition is needed because Lemma SA-II.7.2 gives $|\widehat{\Sigma}_\varepsilon(\beta)| = O_{\mathbb{P}}(J^d/(nT)) = o_{\mathbb{P}}(T^{-1})$.

SA-III.1.4 Validity of Tuning Parameter Selection

The next lemma checks the polynomial-rate restrictions for the plug-in scaling. The key distinction is the normalization. Pointwise inference at a fixed nonzero β is governed by the factor scale. Pointwise inference at $\beta = 0$ and factor-cancelling contrasts are governed by the idiosyncratic scale, which magnifies the components of $\mathsf{R}_{\mathtt{LA}}$ that are harmless at the factor scale.

Lemma SA-III.1.1 (Validity of Pointwise-in-β Tuning)**.** *Suppose $T = n^\eta$ for some $\eta > 0$, and suppose that Assumptions SA-DGP, SA-FS, SA-PA, and SA-FSB hold and that $s_\mu \geq 2$. Suppose the plug-in choices satisfy, uniformly over their indices,*

$$\widehat{h}_{it}^\star \asymp h \asymp T^{-1/3}, \qquad \widehat{J}_t^\star \asymp J \asymp n^{1/(d+2)},$$

and take the moment exponent in Assumption SA-DGP sufficiently large. Then the following rate implications hold.

(i) Pointwise inference for $\beta \neq 0$. *The plug-in scaling verifies the factor-scale restrictions for fixed-nonzero-β inference if*

$$0 < \eta < \frac{4}{d+2}. \tag{SA-III.1.1}$$

In addition, the deterministic threshold in Assumption SA-FSB must satisfy

$$\sqrt{T}\,\mathsf{B}_{\mathtt{FS}} = o(1).$$

The upper bound is the factor-scale second-step smoothing-bias condition for the $\mathfrak{q} = 1$ *inference fit,*

$$\sqrt{T}J^{-2} \to 0,$$

which is equivalent to $\eta < 4/(d+2)$ *under* $J \asymp n^{1/(d+2)}$ *and* $T = n^{\eta}$.

(ii) Pointwise inference for $\beta = 0$ and factor-cancelling contrasts. *Pointwise inference at* $\beta = 0$ *in Theorem SA-II.6.2 and factor-cancelling contrast inference in Theorem SA-II.9.1 both require the linearization error to be negligible after normalization by* $\sqrt{nT/J^d}$. *Ignoring logarithmic factors, the third component of* $\mathsf{R}_{\mathtt{LA}}$ *in Lemma SA-II.4.2 requires*

$$\sqrt{\frac{nT}{J^d}}\sqrt{\frac{J^{2d}}{nHJ^{d-1}\max\{J,\sqrt{H}\}}} = \sqrt{\frac{TJ}{H\max\{J,\sqrt{H}\}}} \to 0. \tag{SA-III.1.2}$$

With $H \asymp T^{2/3}$ *and* $J \asymp n^{1/(d+2)}$, *the right-hand side is equal to*

$$\begin{cases} T^{1/6}, & J \geq \sqrt{H}, \\ J^{1/2}, & J < \sqrt{H}, \end{cases}$$

and therefore diverges. Thus the plug-in scaling cannot verify the natural origin-scale RBC rate restrictions. In addition, the second-step smoothing-bias condition alone would require

$$\frac{nT}{J^{d+4}} \to 0 \qquad \Longleftrightarrow \qquad \eta < \frac{2}{d+2},$$

while the deterministic H^{-1} *component of* $\mathsf{R}_{\mathtt{LA}}$ *would require, up to logarithmic factors,*

$$\eta > \frac{6}{d+2}.$$

The first-step bias component would separately require

$$\sqrt{\frac{nT}{J^d}}\,\mathsf{B}_{\mathtt{FS}} = o(1).$$

The stronger primitive condition $J\mathsf{R}_{\mathtt{FS}} = o(1)$ is not needed for this fixed-nonzero-β factor-scale calculation; it is an idiosyncratic-scale sufficient condition for absorbing the basis-switching term into the local-support scale. Thus, the recommended choices of h and J, together with $\sqrt{T}\,\mathsf{B}_{\mathtt{FS}} = o(1)$, verify fixed-nonzero-$\beta$ pointwise inference within the growth window in (SA-III.1.1), but do not verify idiosyncratic-scale RBC inference at the origin or for factor-cancelling contrasts.

SA-III.1.5 Proof of Lemma SA-III.1.1

Throughout the proof, exponent comparisons suppress logarithmic factors. The additional first-step bias restriction in part (i) gives $\sqrt{T}\,\mathsf{B}_{\mathtt{FS}} = o(1)$. The first-step variance terms satisfy

$$\sqrt{T}\frac{1}{Th} \asymp T^{-1/6} \to 0,$$

and the remaining first-step terms involving the moment exponent are negligible by taking that exponent sufficiently large. For the second-step portfolio fit with $\mathfrak{q} = 1$ used for inference, the smoothing bias is of order J^{-2}. Since $J \asymp n^{1/(d+2)}$ and $T = n^{\eta}$,

$$\sqrt{T}J^{-2} \to 0 \quad \Longleftrightarrow \quad n^{\eta/2-2/(d+2)} \to 0 \quad \Longleftrightarrow \quad \eta < \frac{4}{d+2},$$

which gives the upper bound in (SA-III.1.1). The main portfolio variance restriction is $J^d/(nh) = o(1)$. With the same substitutions,

$$\frac{J^d}{nh} \asymp n^{d/(d+2)-1+\eta/3},$$

so this term is $o(1)$ whenever $\eta < 6/(d+2)$, which is implied by (SA-III.1.1). This condition also implies $J^d/n = o(1)$, which is the effective nonempty-portfolio requirement in the consistency and variance-estimation arguments. The sharpened jackknife restriction for the $\mathfrak{q} = 1$ inference fit requires $HJ^{-4} = o(1)$ up to logarithmic factors. Since

$$HJ^{-4} \asymp n^{2\eta/3-4/(d+2)},$$

this condition follows from (SA-III.1.1). The remaining terms in (SA-II.7.1) are smaller under the same condition and the first-step rates above. The endpoint condition also holds, since

$$\frac{J^2}{n} \asymp n^{2/(d+2)-1} = o(1).$$

The full linear-approximation remainder is negligible at the factor scale. The deterministic pieces satisfy

$$\sqrt{T}\left(\frac{1}{H} + \mathsf{B}_{\mathtt{FS}}\right) \lesssim T^{-1/6} + \sqrt{T}\,\mathsf{B}_{\mathtt{FS}} = o(1).$$

The empirical-process component in $\mathsf{R}_{\mathtt{LA}}$ satisfies

$$\sqrt{T}\sqrt{\frac{J^{2d}}{nHJ^{d-1}\max\{J,\sqrt{H}\}}} = \begin{cases} \sqrt{TJ^d/(nH)}, & J \geq \sqrt{H}, \\ \sqrt{TJ^{d+1}/(nH^{3/2})}, & J < \sqrt{H}, \end{cases} = o(1),$$

where the first case uses $\eta < 6/(d+2)$ and the second case is $n^{-1/(2(d+2))}$. The Gram-perturbation term is also negligible. Indeed, Lemma SA-II.3.3 gives $J^d\mathsf{R}_{\mathtt{Q}} = o(1)$ and

$$\sqrt{T}\,J^d\mathsf{R}_{\mathtt{Q}}\sqrt{\frac{J^d}{nH}} \leq J^d\mathsf{R}_{\mathtt{Q}}\sqrt{\frac{J^d}{nh}} = o(1),$$

where logarithmic factors are suppressed. The basis-switching term is also negligible at the factor scale without imposing the stronger idiosyncratic-scale condition $J^d \mathsf{R}_{\mathtt{P}} = o(1)$. Since the minimum operators in $\mathsf{R}_{\mathtt{P}}$ are bounded by one,

$$\mathsf{R}_{\mathtt{P}} \lesssim J^{-d} + \sqrt{\frac{J^{-d}\log(nT)}{n}} + \frac{\log(nT)}{n}.$$

Therefore Lemma SA-I.3.1 gives

$$\sqrt{T}\,|\mathscr{R}_2(\beta)| \lesssim_{\mathbb{P}} \sqrt{\frac{J^{2d}\mathsf{R}_{\mathtt{P}}}{n}} = o_{\mathbb{P}}(1),$$

because $J \asymp n^{1/(d+2)}$ implies $J^d/n = o(1)$. The same-cell nonconstant-basis component in $\mathsf{R}_{\mathtt{LA}}$ is negligible away from the origin. At the plug-in scaling, for fixed $\beta \neq 0$,

$$\sqrt{T}\sqrt{\frac{J^{d+2}}{nH^2}} \asymp T^{-1/6} = o(1),$$

because $J^{d+2} \asymp n$ and $H \asymp T^{2/3}$. The endpoint-interior term is smaller:

$$\sqrt{T}\,\sqrt{\frac{J^{d+2}}{n^2 H}} \asymp \sqrt{\frac{TJ^{d+2}}{n^2 H}} = \sqrt{\frac{T}{nH}} = o(1).$$

Thus the infeasible Gaussian approximation holds for each fixed $\beta \neq 0$ at the plug-in scaling.

At the origin, the relevant normalization is $\sqrt{nT/J^d}$. The third component of $\mathsf{R}_{\mathtt{LA}}$ gives the necessary condition in (SA-III.1.2). At the plug-in scaling,

$$\sqrt{\frac{nT}{J^d}}\sqrt{\frac{J^{2d}}{nHJ^{d-1}\max\{J,\sqrt{H}\}}} = \begin{cases} T^{1/6}, & J \geq \sqrt{H}, \\ J^{1/2}, & J < \sqrt{H}, \end{cases}$$

which diverges. Hence $\mathsf{R}_{\mathtt{CLT}}/\sqrt{\Sigma(0)}$ cannot vanish under the plug-in scaling. Even ignoring this obstruction, the smoothing-bias component in $\mathsf{R}_{\mathtt{CLT}}$ would require

$$\sqrt{\frac{nT}{J^d}}J^{-2} \to 0 \quad \Longleftrightarrow \quad \frac{nT}{J^{d+4}} \to 0 \quad \Longleftrightarrow \quad \eta < \frac{2}{d+2}.$$

The deterministic and smooth-interior components of $\mathsf{R}_{\mathtt{LA}}$ also give

$$\sqrt{\frac{nT}{J^d}}\left(\frac{1}{H} + \mathsf{B}_{\mathtt{FS}} + \sqrt{\frac{J^{d+2}}{nH^2}}\right) \gtrsim \frac{J\sqrt{T}}{H},$$

up to logarithmic factors. Under the plug-in scaling,

$$\frac{J\sqrt{T}}{H} \asymp n^{1/(d+2)-\eta/6},$$

so these components would require $\eta > 6/(d+2)$, which is incompatible with the smoothing-bias requirement

above. The first-step bias component would additionally require

$$\sqrt{\frac{nT}{J^d}}\,\mathsf{B}_{\mathtt{FS}} = o(1).$$

The recommended tuning therefore does not verify natural origin-scale RBC inference. The same obstruction applies to factor-cancelling contrasts. In Theorem SA-II.9.1, the factor term cancels, so the condition (SA-II.9.3) normalizes the residual at the idiosyncratic scale $\sqrt{nT/J^d}$. By definition, $\mathsf{R}_{\Psi} = (\sum_{k=1}^{K_\Psi} |a_k|)\mathsf{R}_{\mathtt{CLT}}$, and therefore it contains the same $\mathsf{R}_{\mathtt{LA}}$ component. Thus (SA-II.9.3) imposes the same necessary condition (SA-III.1.2), which diverges under the plug-in scaling. □

SA-III.2 Uniform-in-β

This section gives the tuning and implementation steps for the uniform-in-β confidence intervals of Theorem SA-II.8.1. Throughout, "uniform-in-β" means that the coverage probability is controlled uniformly over evaluation points $\beta \in \mathcal{B}^\circ$, in the pointwise-uniform sense defined in Section SA-II.8; it is not a simultaneous confidence band for the full function $\mu(\cdot)$. The region $\mathcal{B}^\circ$ is the same fixed interior region used in Theorem SA-II.8.1, and it contains a fixed neighborhood of the origin.

We keep the first-step bandwidths and second-step portfolio counts selected in Section SA-III.1. The rate calculation uses the same uniform orders as in Lemma SA-III.1.1, namely $\widehat{h}^\star_{it} \asymp T^{-1/3}$ and $\widehat{J}^\star_t \asymp n_t^{1/(d+2)}$. Algorithm SA-III.1.2 implements a grid approximation to the period-specific weighted IMSE criterion.

We separate the final time-average burn-in from the selected local first-step windows. Let T_0 denote a pre-specified implementation burn-in, chosen as a deterministic cap for the selected first-step windows so that, with probability approaching one,

$$T_0 \geq \max_{1\leq s\leq T}\max_{i\in[n_s]} \widehat{H}^\star_{is}, \qquad \widehat{H}^\star_{is} = \lceil T\widehat{h}^\star_{is}\rceil,$$

and $T_0 \asymp T^{2/3}$. The final time-series averages below use $t = T_0+1,\dots,T$, whereas each first-step local-polynomial fit still uses its own selected window $\widehat{H}^\star_{it}$. The remaining tuning parameter is the near-origin truncation parameter δ, which determines where the usual feasible standard error is replaced by the conservative variance floor from Theorem SA-II.8.1.

SA-III.2.1 Near-Origin Truncation Parameter

Theorem SA-II.8.1 uses a scalar threshold δ in factor-variance units. Its role in the implementation is only to identify a small neighborhood of the origin in which the usual factor-based standard error is replaced by a conservative floor. We therefore use a deliberately simple rule of thumb rather than attempting to estimate every remainder constant.

For the degree-one second step used in the robust bias-corrected implementation, we choose a threshold of the form

$$\widehat{\delta} := c_\delta \widehat{s}_R^2 (\log T)^{d+1}(\log\log T)\max\{\widehat{r}_J, \widehat{r}_H, \widehat{r}_N\}. \tag{SA-III.2.1}$$

Here $\widehat{r}_J$, $\widehat{r}_H$, and $\widehat{r}_N$ estimate the rate components associated with the squared degree-one second-step approximation bias, the squared local-window remainder, and the uniform first-step stochastic effect, re-

spectively; $\widehat{s}_R^2$ puts the threshold in return- variance units. Section SA-III.4 gives the exact finite-sample definitions of these quantities and the default choice of c_δ. The adjustment factor $(\log T)^{d+1}(\log\log T)$ ensures that the threshold rate restrictions required by Theorem SA-II.8.1 are satisfied under our recommended tuning parameter choices, provided that $\mathsf{B}_{\mathtt{FS}} = O(H^{-1})$.

Under independent factors and homoskedastic first-step errors, converting coefficient noise back to factor-variance units cancels the factor variance; local portfolio averaging then gives the rate $J^d/(nH)$ underlying $\widehat{r}_N$.

The rule is scale equivariant in the intended sense. Multiplying returns by a multiplies $\widehat{s}_R^2$ and $\widehat{\delta}$ by a^2. A diagonal rescaling of a beta coordinate paired with the inverse rescaling of its factor leaves the switching comparison unchanged. Coordinate permutations leave the rate candidates unchanged. The rule is a finite-sample calibration device, not a claim that the unknown optimal threshold constant has been consistently estimated. Formal use of Theorem SA-II.8.1 still requires the resulting scalar threshold to satisfy (SA-II.8.1).

SA-III.2.2 Estimation and Robust Bias-Corrected Inference

Given $\widehat{h}_{it}^\star$, the common portfolio-count scale $\widehat{J}_t^\star$, and the threshold $\widehat{\delta}$, compute the same robust bias-corrected estimator as in Subsection SA-III.1.3, but with the common uniform-in-β portfolio-count sequence and the final averaging range $t = T_0+1,\dots,T$:

$$\widehat{\mu}^{\mathrm{BC}}(\beta) = \frac{1}{T-T_0}\sum_{t=T_0+1}^{T} \widehat{M}_t(\beta; 1, \widehat{J}_t^\star),$$

using the bias-corrected first-step estimates $\{\widehat{\beta}_{it}^{\mathrm{BC}}\}$ and the selected bandwidths $\{\widehat{h}_{it}^\star\}$. For this estimator, compute the feasible factor component $\widehat{\Sigma}_f(\beta)$ and the idiosyncratic component $\widehat{\Sigma}_\varepsilon^{\mathrm{BC}}(\beta)$ as in Section SA-II.7, but over the same final averaging range $t = T_0+1,\dots,T$, with the idiosyncratic component using the residuals and weights from the same robust bias-corrected fit. Set

$$\widehat{\Sigma}^{\mathrm{BC}}(\beta) = \widehat{\Sigma}_f(\beta) + \widehat{\Sigma}_\varepsilon^{\mathrm{BC}}(\beta).$$

The conservative variance floor is

$$\widehat{\Sigma}_{\widehat{\delta}}^{\mathrm{BC}} = \sup_{\substack{b\in\mathcal{B}^\circ \\ \widehat{\Sigma}_f(b)=\widehat{\delta}}} \max\{\widehat{\Sigma}^{\mathrm{BC}}(b), \widehat{\delta}\} = \widehat{\delta} + \sup_{\substack{b\in\mathcal{B}^\circ \\ \widehat{\Sigma}_f(b)=\widehat{\delta}}} \max\{\widehat{\Sigma}_\varepsilon^{\mathrm{BC}}(b), 0\}.$$

If the level set is empty in a finite sample, set $\widehat{\Sigma}_{\widehat{\delta}}^{\mathrm{BC}} = \widehat{\delta}$. For numerical work, the boundary supremum is evaluated on the grid $\mathcal{G}$, using interpolation or the nearest grid points to the level set $\widehat{\Sigma}_f(b) = \widehat{\delta}$. This grid approximation is purely computational; the theoretical statement above is the continuum version.

The variance used in the uniform-in-β robust bias-corrected interval is

$$\widehat{\Sigma}_{\mathtt{UI}}^{\mathrm{BC}}(\beta) = \begin{cases} \widehat{\Sigma}^{\mathrm{BC}}(\beta) & \text{if } \widehat{\Sigma}_f(\beta) > \widehat{\delta} \\ \max\{\widehat{\Sigma}^{\mathrm{BC}}(\beta), \widehat{\Sigma}_{\widehat{\delta}}^{\mathrm{BC}}\} & \text{if } \widehat{\Sigma}_f(\beta) \le \widehat{\delta} \end{cases},$$

The resulting $100(1-\alpha)\%$ interval is

$$\widehat{\mu}^{\mathrm{BC}}(\beta) \pm z_{1-\alpha/2}\sqrt{\widehat{\Sigma}_{\mathtt{UI}}^{\mathrm{BC}}(\beta)}.$$

SA-III.3 Factor-Cancelling Linear Contrast

Factor-cancelling contrasts do not introduce a new threshold parameter. Their distinctive feature is the normalization: after the factor component cancels, Theorem SA-II.9.1 works at the idiosyncratic scale $\sqrt{nT/J^d}$. Thus a factor-cancelling contrast is covered by the theory only when the selected first-step bandwidths and second-step portfolio counts verify

$$\sqrt{\frac{nT}{J^d}}\mathsf{R}_{\Psi} \to 0,$$

together with the remaining conditions in Theorem SA-II.9.1. For the fixed finite contrast, the definition above yields

$$\mathsf{R}_{\Psi} = \left(\sum_{k=1}^{K_{\Psi}} |a_k|\right) \mathsf{R}_{\mathtt{CLT}},$$

so this condition is equivalent to imposing the same normalized rate on $\mathsf{R}_{\mathtt{CLT}}$.

The pointwise first-step and IMSE second-step plug-in choices from Algorithm SA-III.1.1 and Algorithm SA-III.1.2 are therefore not a validated default for factor-cancelling inference. Lemma SA-III.1.1 shows that, under those choices, the idiosyncratic-scale normalization magnifies the linear-approximation remainder in $\mathsf{R}_{\mathtt{LA}}$, and the required rate diverges. Consequently, a factor-cancelling confidence interval should use the jackknife variance estimator in Theorem SA-II.9.1 only with a dedicated undersmoothing choice of h and J that satisfies the display above. Operationally, this means choosing h and J to make the estimated non-leading remainder small relative to $\sqrt{J^d/(nT)}$, rather than choosing J to minimize the IMSE of the lower-order estimator.

SA-III.4 Default Implementation Choices

In this section, we describe the implementation conventions used in the two empirical applications and the Monte Carlo simulations.

Tuning parameters We follow the approach described in Section SA-III.1. The pilot choices are $H_{\text{pilot}} =$ 10 years (120 months or 40 quarters) and $J_{\text{pilot}} = 10$. Thus, $h_{\text{pilot}} = 120/T$ for monthly data and $h_{\text{pilot}} = 40/T$ for quarterly data. Here $h = H/T$, so the denominator is the full time-series sample size T, rather than the $T - H$ dates entering the second-step average. We make a number of finite-sample regularizations to ensure that the first-step and second-step estimators are well behaved. For the first step we restrict the selected window to $36 \leq H_{it} \leq 360$ for monthly data and $36 \leq H_{it} \leq 120$ for quarterly data. For an unbalanced panel, we require at least 36 usable observations within the selected window to estimate a beta. In addition, for inference, we use $\mathfrak{p} = 1$ so long as $\widehat{\mathrm{Var}}_{\mathrm{HC1}}\big(\widehat{\beta}_{it}^{(1)}\big) \leq T^{1/10}\widehat{\mathrm{Var}}_{\mathrm{HC1}}\big(\widehat{\beta}_{it}^{(0)}\big)$, and use $\mathfrak{p} = 0$ otherwise. Because $T \to \infty$, this regularization becomes asymptotically inactive under the conditions in SA-Theory. Similarly, if the $\mathfrak{p} = 1$ design matrix is singular, we use the $\mathfrak{p} = 0$ estimate instead. We constrain the number of portfolios to $2 \leq J_t \leq 40$. The lower bound ensures the data are split into separate portfolios whereas the upper bound helps prevent the formation of portfolios containing too few assets. The grids used to select J_t and to report estimates and inference results are chosen separately. In each eligible post-burn-in period, the implementation finds the largest origin-centered interval contained in the cross-sectional range of

the relevant estimated betas. It then uses the largest such symmetric interval that is contained in the realized range in at least 90% of the eligible periods, weighting periods equally. The J_t-selection grid consists of 25 equally spaced points on this interval and is constructed from the point-estimation betas. The reporting and inference grid consists of 11 equally spaced points and is constructed from the first-step beta estimates used for inference. Hence the reporting grid contains zero exactly but does not affect the choice of J_t. For graphical clarity, the business-credit application uses a manually specified reporting grid from -1.25 to 1.25 in increments of 0.25, together with a separate grid of 25 equally spaced points on $[-1, 1]$ for selecting J_t. For simplicity, we use the round endpoints -1 and 1, which are close to those selected by the default support rule. This interval lies within the realized range of the point-estimation betas in 85.7% of months, compared with the default target of 90%.

Variance estimation For the common-factor component of the variance estimator, both empirical applications forecast the risk factor using an AR(1) with an intercept, estimated recursively by OLS using only prior observations. Since the first-step window H_{it} generally varies across assets and dates, the implementation uses a pair-specific lag cutoff in the JN variance estimator. For observations on asset i at dates t and s, it retains the jackknife cross-product whenever $|t-s| \leq \max\{H_{it}, H_{is}\}$. This rule accommodates the dependence induced by overlapping first-step windows of different lengths without applying the largest selected window to every asset-date pair. If $H_{it} = H$ for all i and t, it reduces to the common lag-H truncation in equation (4.11) of the main text.

Uniform-in-β inference For each post-burn-in date t, let $\mathcal{I}_t^\delta$ denote the set of assets included in the second-step estimator used for inference. Let $n_t^\delta = |\mathcal{I}_t^\delta|$ and $\mathcal{I}^\delta = \{(i,t) : i \in \mathcal{I}_t^\delta\}$. Because returns can contain extreme observations, we use the square of the pooled median absolute deviation as a robust estimate of the return variance scale, rescaled to be consistent for the variance under Gaussianity:

$$\widehat{s}^2_{R,\mathrm{MAD}} = \left\{ 1.4826 \operatorname*{median}_{(i,t)\in\mathcal{I}^\delta} \left| R_{it} - \operatorname*{median}_{(j,s)\in\mathcal{I}^\delta} R_{js} \right| \right\}^2 .$$

We compute the three rate components as observation-level averages:

$$\widehat{r}_J = \frac{1}{|\mathcal{I}^\delta|} \sum_{(i,t)\in\mathcal{I}^\delta} J_t^{-4}, \qquad \widehat{r}_H = \frac{1}{|\mathcal{I}^\delta|} \sum_{(i,t)\in\mathcal{I}^\delta} H_{it}^{-2}, \qquad \widehat{r}_N = \frac{1}{|\mathcal{I}^\delta|} \sum_{(i,t)\in\mathcal{I}^\delta} \frac{J_t^d}{n_t^\delta H_{it}}.$$

Finally, we choose $c_\delta = 1/160$ so that

$$\widehat{\delta} = \frac{1}{160} \widehat{s}^2_{R,\mathrm{MAD}} (\log T)^{d+1} (\log\log T) \max\left\{\widehat{r}_J, \widehat{r}_H, \widehat{r}_N\right\}.$$

Each included observation receives equal weight, so periods with larger effective cross sections receive proportionally more weight in these pooled averages. The calibration sample is determined before evaluating the reporting grid; consequently, changing only the reporting grid does not alter $\widehat{\delta}$.

Affine Specification Test For the business-credit application, we implement the affine specification test described in Section 5.2 of the main text, as follows. We estimate both betas using local quadratic regressions, $\mathfrak{p} = 2$, over two adjacent, nonoverlapping windows of $H = 63$ months each. We choose $H = 63$ as it is the time-series median of the cross-sectional median series shown in Figure SA-IV.2.1. We continue to require

at least 36 nonmissing observations in each window. At each date, we regress returns on an intercept and the most-recent beta by using the earlier beta as the instrument. We then form equal-weighted portfolios of the resulting residuals by sorting on the earlier beta, with $\mathfrak{q} = 0$. We select J_t on the matched sample using the MSE criterion with earlier local-constant beta estimates and bounds of 2 and 40 portfolios. The testing grid consists of the seven points from -0.75 to 0.75 in increments of 0.25. The covariance estimator retains cross-products of the same firm's weighted residual contributions for outcome dates at most 63 calendar months apart, accounting for overlap between recent estimation windows and outcome dates. The test sample comprises 534 months, from July 1975 through December 2019. The resulting Wald statistic is 7.871; comparison with a chi-squared distribution with seven degrees of freedom gives a p-value of 0.344.

Part IV: Numerical Evidence

SA-IV.1 Simulations

SA-IV.1.1 Data-Generating Process

We consider two specifications for the risk factor and one volatility state. Let $k \in \{1, 2\}$ index the risk factor specification, with $\rho_1 = 0.5$ and $\rho_2 = 0.9$. For each k, the risk factor follows the stationary Gaussian AR(1)

$$f_{kt} = \rho_k f_{k,t-1} + u_{kt}, \qquad u_{kt} \overset{\text{iid}}{\sim} N(0, \sigma_{u,k}^2). \tag{SA-IV.1.1}$$

We fix the stationary factor standard deviation at $\sigma_f = 0.04$ in both specifications and therefore set $\sigma_{u,k}^2 = (1 - \rho_k^2)\sigma_f^2$. For each specification, we draw the initial condition from $N(0, \sigma_f^2)$.

The volatility state follows

$$z_t = \rho_z z_{t-1} + \sqrt{1 - \rho_z^2}\, v_t, \qquad v_t \overset{\text{iid}}{\sim} N(0, 1), \tag{SA-IV.1.2}$$

where $\rho_z = 0.975$. We initialize z_t from its stationary $N(0, 1)$ distribution. The z_t process is independent of the risk-factor process.

The return equation under factor specification k is

$$R_{it}^{(k)} = \mu_{\mathrm{F}}(\beta_{it}) + \beta_{it} f_{kt} + \varsigma_{it}^{(k)} \eta_{it}. \tag{SA-IV.1.3}$$

We specify the function $\mu_{\mathrm{F}}(\cdot)$, the paths of β_{it}, and the volatility scale $\varsigma_{it}^{(k)}$ below.

Functional form of $\mu_{\mathrm{F}}(\cdot)$ For $\beta \in [-1, 2]$, define the normalized exposure $\widetilde{\beta} = (\beta + 1)/3 \in [0, 1]$ and let

$$\mu_{\mathrm{F}}(\beta) = -\frac{7}{200} + \frac{1}{20}\left\{6\widetilde{\beta}^5 - 15\widetilde{\beta}^4 + 10\widetilde{\beta}^3\right\}.$$

Thus, for all eight designs, the conditional expected return is

$$\mathbb{E}\left[R_{it}^{(k)} \mid \mathcal{F}_{t-1}, \beta_{it} = \beta\right] = \mu_{\mathrm{F}}(\beta) + \beta\rho_k f_{k,t-1}. \tag{SA-IV.1.4}$$

Conditional expected returns therefore vary with the lagged value of the risk factor even though the structural expected-return curve $\mu_{\mathrm{F}}(\cdot)$ has no additional time-varying component.

Paths of β_{it} We consider two specifications for the beta paths. In the time-varying specification, we generate smooth paths using the low-frequency components of a persistent time-series process (see Crump and Gospodinov, 2022):

1. Constant betas: $\beta_{it} = \beta_i$ for $t = 1, \ldots, T$.

2. Time-varying betas:

$$\beta_{it} = \beta_i + \sum_{j=1}^{K_\beta} a_{ij}\varphi_{j,t}, \qquad \varphi_{j,t} = \sqrt{\frac{2}{T}} \cos\left(\frac{t(j-\frac{1}{2})\pi}{T}\right),$$

$$a_{ij} = \sqrt{\lambda_j}\,\xi_{ij}, \qquad \lambda_j = \frac{v_\beta(1-\rho_\beta^2)}{1+\rho_\beta^2 - 2\rho_\beta\cos(\omega_j)}, \qquad \omega_j = \frac{(j-\frac{1}{2})\pi}{T}.$$

   Here a_{ij} and ξ_{ij} are scalar.

We set $\mathcal{B} = [-1, 2]$ and draw $\beta_i \overset{\text{iid}}{\sim} \text{Unif}[-1, 2]$. For the time-varying beta specification, we set $K_\beta = 5$, $\rho_\beta = 0.95$, $v_\beta = 0.04$, and $\xi_{ij} \overset{\text{iid}}{\sim} N(0, 1)$. After drawing the low-frequency coefficients, for each unit we scale the time-varying component by the smaller of one and 99.9% of the largest factor that keeps the entire resulting beta path in $\mathcal{B}$.

Functional form of $\varsigma_{it}^{(k)}$ and η_{it} We assume that $\eta_{it} \overset{\text{iid}}{\sim} N(0, 1)$ for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. We consider two volatility specifications:

1. Constant volatility: $\left(\varsigma_{it}^{(k)}\right)^2 = (1/50)^2$.

2. State-dependent volatility:

$$\left(\varsigma_{it}^{(k)}\right)^2 = \left(\frac{1}{50}\right)^2 \exp\left\{\frac{1}{2}z_{t-1} - \frac{1}{8}\right\}.$$

Because $z_t \sim N(0, 1)$ in the stationary distribution, $\mathbb{E}[(\varsigma_{it}^{(k)})^2] = (1/50)^2$ under both specifications.

Within each design, the risk-factor process, volatility-state process, beta paths, and idiosyncratic return shocks are mutually independent. The complete set of experiments crosses the two persistence choices, two types of beta paths, and two volatility specifications, for a total of eight data-generating processes (DGPs) summarized in Table SA-IV.1.1.

Design	ρ_k	Beta path	Idiosyncratic volatility
S1	0.5	Constant	Constant
S2	0.5	Constant	State-dependent
S3	0.5	Time-varying	Constant
S4	0.5	Time-varying	State-dependent
S5	0.9	Constant	Constant
S6	0.9	Constant	State-dependent
S7	0.9	Time-varying	Constant
S8	0.9	Time-varying	State-dependent

Table SA-IV.1.1: Simulation designs. All other calibration choices are common across designs; the factor-innovation variance varies with ρ_k as described above to hold the unconditional risk-factor variance fixed.

SA-IV.1.2 Monte Carlo Implementation

We generate $S = 1,000$ Monte Carlo replications for each of the eight DGPs. Each replication is a panel dataset with $n = 5{,}000$ and $T = 800$. The first 240 months are excluded from second-step estimation and inference as a burn-in period. In each replication, we select the first-step windows H_{it} and the second-step number of portfolios J_t using the procedure described in Section SA-III.1. Apart from the longer 240-month burn-in and the evaluation grid specified below, we use the default implementation for monthly data described in Section SA-III.4.

For the factor component of the proposed covariance estimator, the implementation recursively fits via OLS the univariate AR(1)

$$f_{k\tau} = a_{k,t} + \rho_{k,t} f_{k,\tau-1} + e_{k\tau}, \qquad \tau < t,$$

using only the history of the selected factor f_k, and uses the resulting one-step residual at date t. The implementation estimates the intercept and persistence recursively and does not use the true value of ρ_k.

We evaluate the SACER at

$$\mathcal{B}_{\text{eval}} = \left\{ -1 + \frac{\ell}{7} : \ell = 1, \ldots, 20 \right\}.$$

For replication s and evaluation point β, the corresponding inferential target is

$$\overline{\mu}_s^{(k)}(\beta) = \frac{1}{T - 240} \sum_{t=241}^{T} \left\{ \mu_{\text{F}}(\beta) + \beta \rho_k f_{k,t-1}^{(s)} \right\}.$$

Since our focus is on inference, our primary estimators are the infeasible oracle with known β_{it} and $\mathfrak{q} = 1$, and the feasible estimator with $(\mathfrak{p}, \mathfrak{q}) = (1, 1)$. For pointwise inference, we report the following procedures:

1. the proposed JN variance estimator;
2. the FM variance estimator formed from the $(\mathfrak{p}, \mathfrak{q}) = (0, 0)$ estimates; and
3. the FM variance estimator formed from the $(\mathfrak{p}, \mathfrak{q}) = (1, 1)$ estimates.

For the proposed JN variance estimator, we report the oracle and feasible estimators separately. For every reported procedure, we record pointwise coverage and average confidence-interval length.

All reported confidence intervals have nominal coverage of 95%. For the feasible estimator paired with the proposed JN variance estimator, we also report pointwise confidence intervals that are uniformly valid over β, as developed in Section SA-III.2. We choose $\widehat{\delta}$ using the default implementation described in Section SA-III.4. We record coverage and average interval length at every evaluation point.

SA-IV.1.3 Results

Figure SA-IV.1.1 reports empirical coverage for the proposed pointwise and uniform-in-β confidence intervals, together with the infeasible oracle procedure. The oracle procedure—which treats the β_{it} as observed—performs well across all eight designs, with average pointwise coverage across the evaluation grid ranging from 93.9% to 96.0% (blue solid line).[1] The feasible JN pointwise interval also performs well away from

[1] Recall that, in our simulation design, $\widehat{\Sigma}_f$ uses the correctly specified dynamic model for f_t.

the origin but can undercover near zero, particularly when the factor is highly persistent and betas vary over time (black dashed line). The uniform-in-β procedure eliminates this sharp deterioration (orange solid line). Its minimum coverage across the interior grid ranges from 93.2% to 95.6%, compared with 72.1% to 90.5% for pointwise inference. This contrast illustrates the distinction between pointwise validity at each fixed evaluation point in Corollary 4.6.1 and the uniform-in-β coverage guarantee discussed in Remark 4.7 of the main text. The uniform-in-β intervals are conservative at the origin, where empirical coverage is 100% in every design. Comparing the left and right columns shows that allowing for state-dependent idiosyncratic volatility has only a modest effect on coverage. Figure SA-IV.1.3 reports the corresponding average confidence interval lengths for all three procedures. The feasible and oracle JN pointwise intervals have nearly identical average lengths throughout the evaluation grid. The uniform-in-β adjustment increases interval length in a neighborhood of the origin, where the pointwise procedure can undercover. Outside that neighborhood, the uniform-in-β and pointwise intervals coincide exactly. Finally, note that $\widehat{\Sigma}_\varepsilon(\beta;1)$ was never negative at any evaluation point in the simulations across all eight designs.

Figure SA-IV.1.2 reports pointwise coverage for the two Fama-MacBeth procedures we consider. The procedure based on $(\mathfrak{p},\mathfrak{q}) = (0,0)$ for both point and variance estimation corresponds to the conventional implementation in empirical work. In contrast, the modified procedure is based on $(\mathfrak{p},\mathfrak{q}) = (1,1)$. Both procedures exhibit a sharp deterioration at $\beta = 0$. Coverage at the origin ranges from 26.0% to 39.2% for the procedure based on $(\mathfrak{p},\mathfrak{q}) = (0,0)$ and from 60.7% to 85.7% for the modified procedure based on $(\mathfrak{p},\mathfrak{q}) = (1,1)$. Thus, the modified procedure performs substantially better near the origin, but it remains below the nominal level. This undercoverage is consistent with the discussion following Theorem 4.6 of the main text, which explains why the FM variance estimator is not generally valid at the origin. Figure SA-IV.1.4 reports the corresponding average confidence interval lengths. The curves for the two Fama-MacBeth procedures are nearly indistinguishable in every design. This similarity is consistent with Theorems 4.1 and 4.6 of the main text, which show that, at each fixed nonzero evaluation point, $\mathfrak{q} = 0$ and $\mathfrak{q} = 1$ have the same leading asymptotic variance and the same first-order Fama-MacBeth variance target. Consequently, the improvement from the modified procedure near the origin reflects the use of the higher-order point estimator rather than wider confidence intervals.

Finally, Figure SA-IV.1.5 directly compares the average lengths of the JN uniform-in-β intervals and the modified Fama-MacBeth pointwise intervals. Away from the origin, the Fama-MacBeth intervals are wider, especially when the factor is highly persistent. Their greater width is reflected in the overcoverage shown in Figure SA-IV.1.2. As shown in Theorem 4.6 of the main text, this difference arises because the FM variance estimator includes period-by-period variation in conditional expected returns that is not part of the estimation uncertainty around the SACER. Near the origin, the Fama-MacBeth intervals can instead be shorter, but Figure SA-IV.1.2 shows that this ostensible gain in precision is accompanied by substantial undercoverage. Taken together, the results support the uniform-in-β JN procedure as a reliable approach to inference across both zero and nonzero evaluation points.

Figure SA-IV.1.1: **Coverage probability.** Each panel reports empirical coverage at the 20 interior evaluation points for one of the eight designs in Table SA-IV.1.1. The black dashed line is the feasible JN pointwise interval, the orange solid line is the feasible JN uniform-in-β interval using the default calibration of $\widehat{\delta}$, and the blue solid line is the infeasible oracle JN pointwise interval. The horizontal grey dotted line denotes the nominal coverage probability of 0.95. These results are based on 1,000 Monte Carlo replications per design.

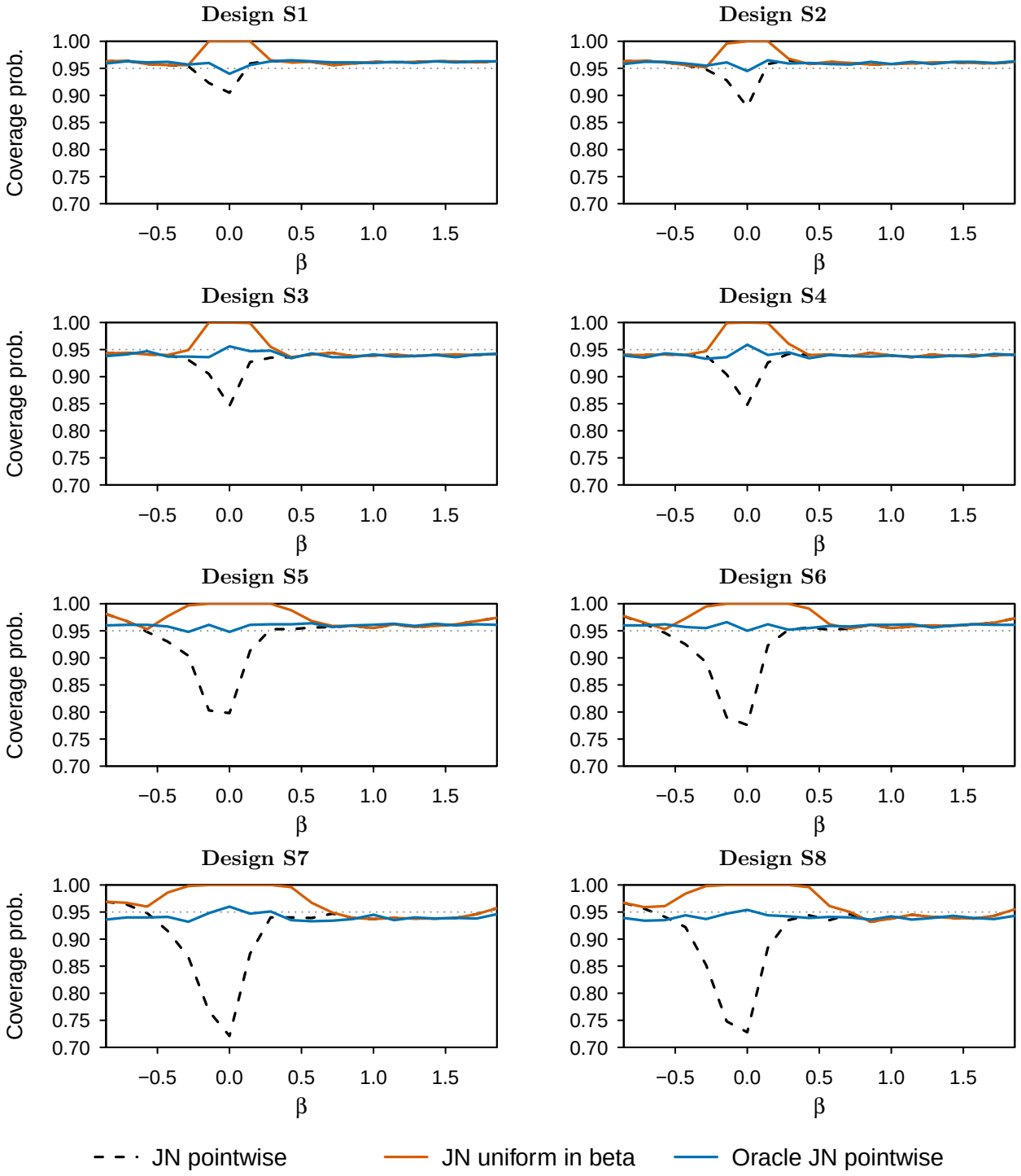


Figure SA-IV.1.2: **Pointwise coverage using Fama-MacBeth variance estimators.** Each panel reports empirical coverage at the 20 interior evaluation points for one of the eight designs in Table SA-IV.1.1. The blue solid line uses the $(\mathfrak{p}, \mathfrak{q}) = (0, 0)$ estimates and the corresponding Fama-MacBeth variance estimator. The orange dashed line is the modified Fama-MacBeth procedure based on $(\mathfrak{p}, \mathfrak{q}) = (1, 1)$. The horizontal grey dotted line denotes the nominal coverage probability of 0.95. These results are based on 1,000 Monte Carlo replications per design.

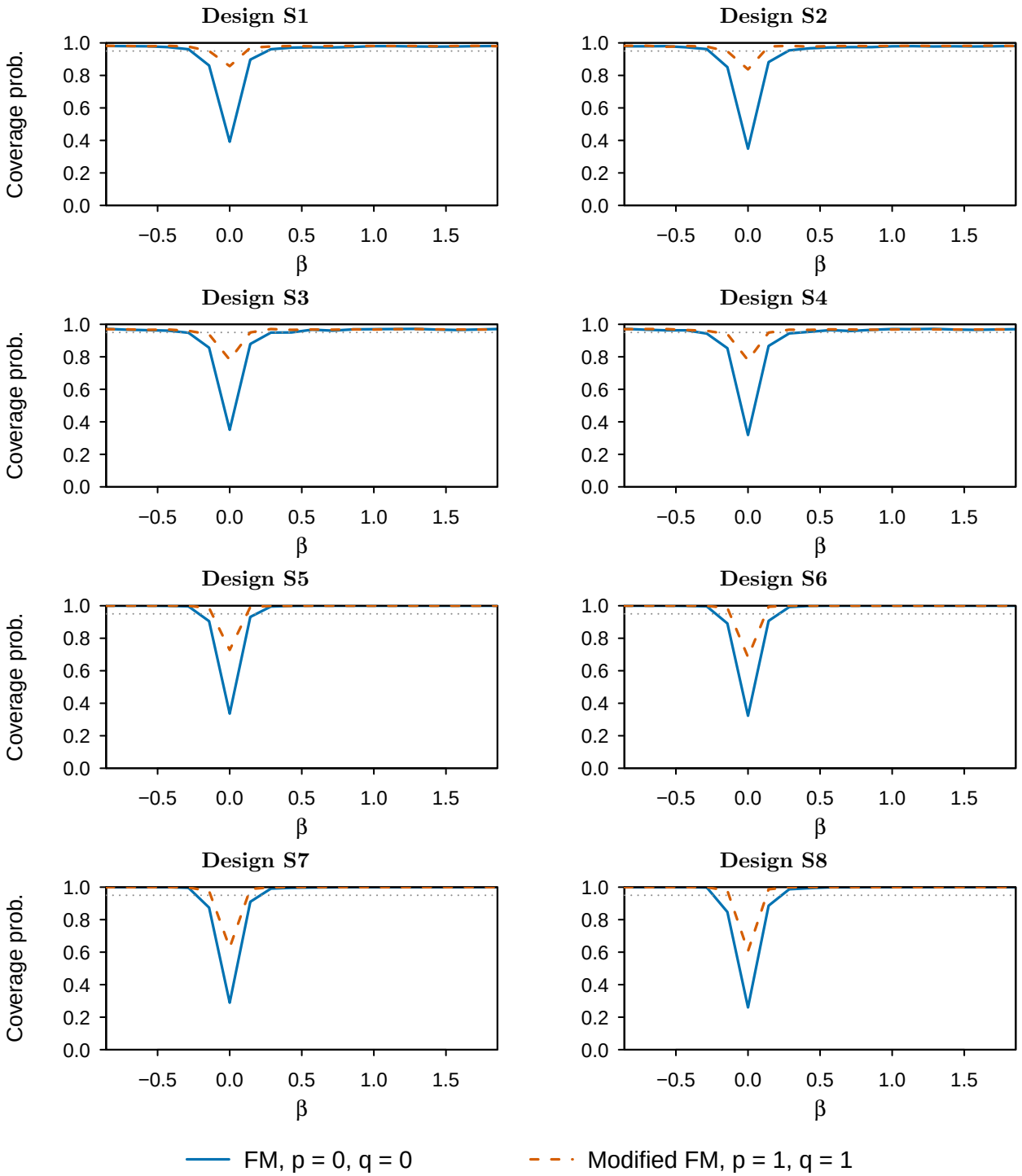


Figure SA-IV.1.3: **Average confidence-interval length.** Each panel reports the average length of the 95% confidence interval at the 20 interior evaluation points for one of the eight designs in Table SA-IV.1.1. The black dashed line is the feasible JN pointwise interval, the orange solid line is the feasible JN uniform-in-β interval using the default calibration of $\widehat{\delta}$, and the blue solid line is the infeasible oracle JN pointwise interval. These results are based on 1,000 Monte Carlo replications per design.

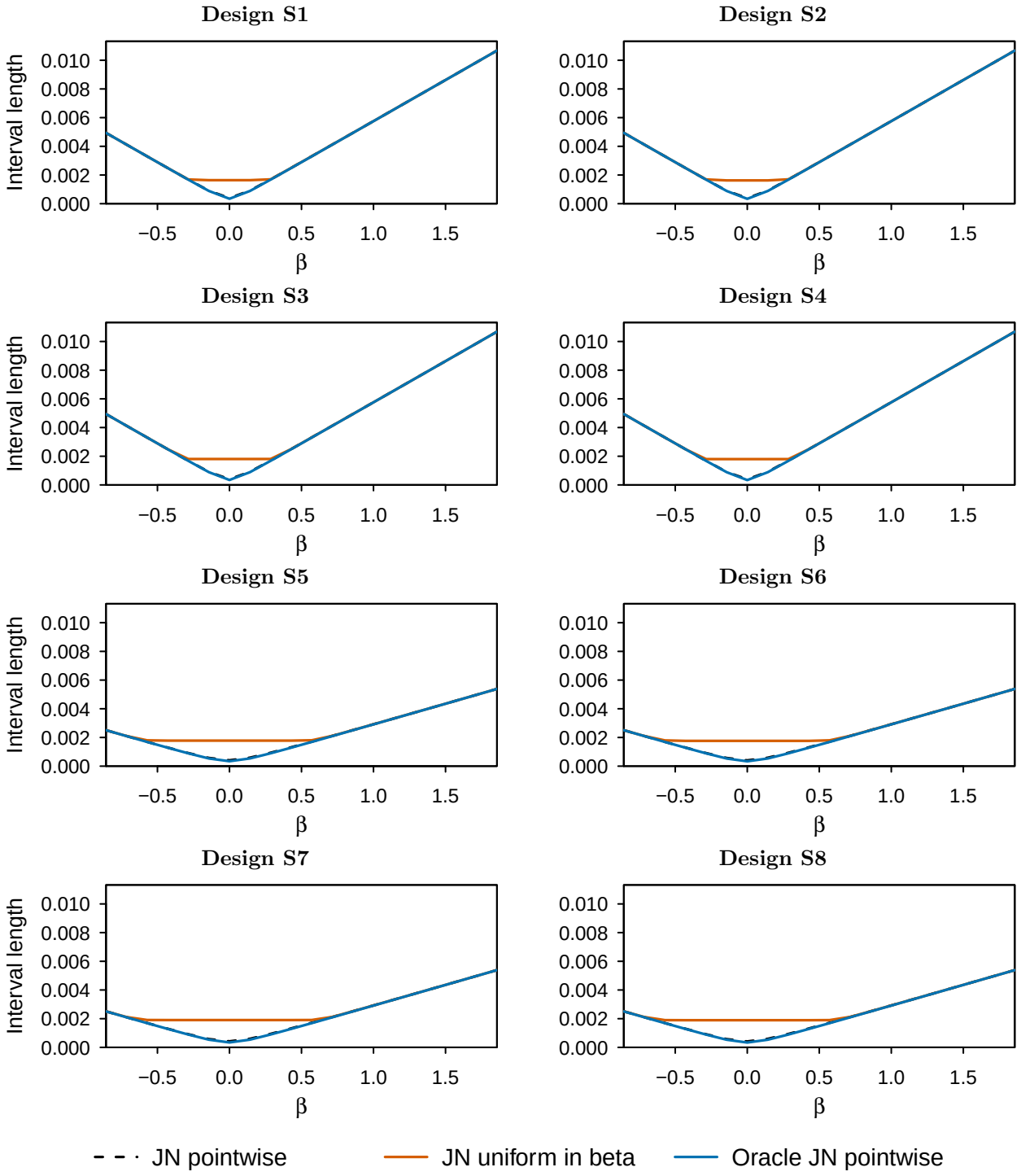


Figure SA-IV.1.4: **Average confidence-interval length using Fama-MacBeth variance estimators.** Each panel reports the average length of the 95% confidence interval at the 20 interior evaluation points for one of the eight designs in Table SA-IV.1.1. The blue solid line uses the $(\mathfrak{p}, \mathfrak{q}) = (0, 0)$ estimates and the corresponding Fama-MacBeth variance estimator. The orange dashed line is the modified Fama-MacBeth procedure based on $(\mathfrak{p}, \mathfrak{q}) = (1, 1)$. These results are based on 1,000 Monte Carlo replications per design.

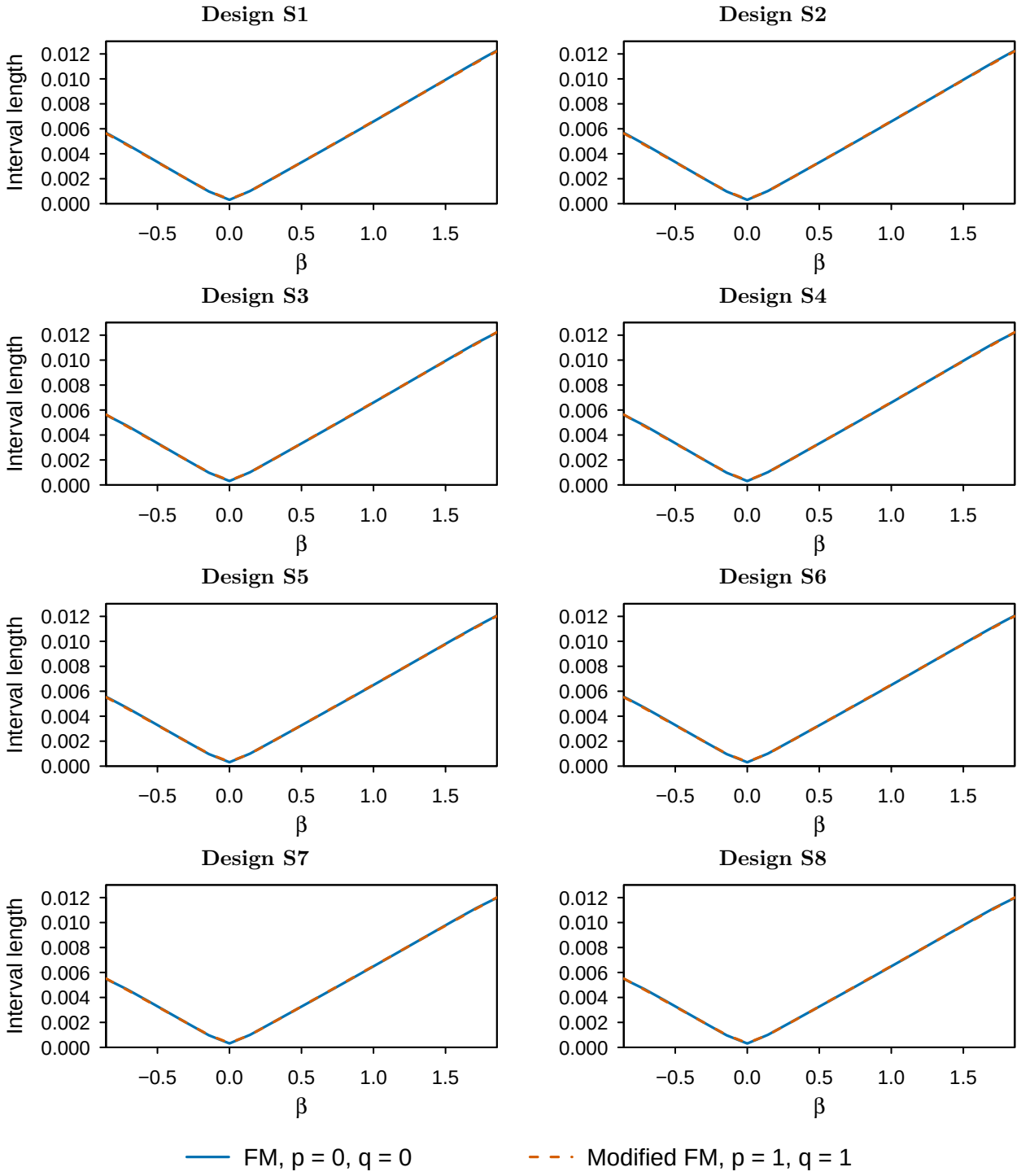
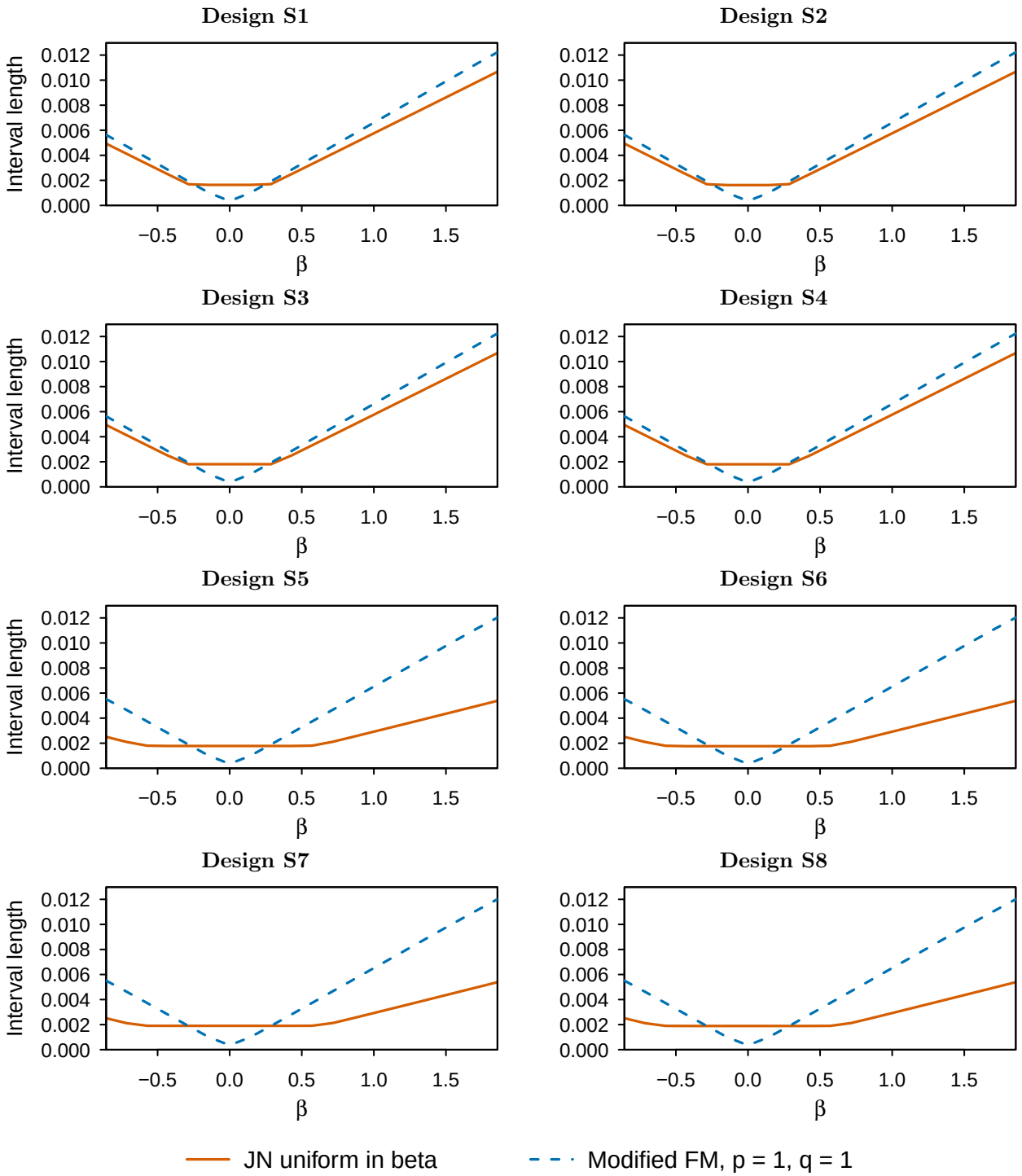


Figure SA-IV.1.5: **Average confidence-interval length: JN uniform-in-β and modified Fama-MacBeth.** Each panel reports the average length of the 95% confidence interval at the 20 interior evaluation points for one of the eight designs in Table SA-IV.1.1. The orange solid line is the JN uniform-in-β interval using the default calibration of $\widehat{\delta}$. The blue dashed line is the modified Fama-MacBeth pointwise interval. Both procedures use the same feasible estimates. Results are based on 1,000 Monte Carlo replications per design.



SA-IV.2 Empirical Applications

In this section, we provide additional results illustrating our new procedures. First, we provide additional evidence and implementation details for the business-credit factor introduced in the main text. Second, we revisit the capital share risk factor of Lettau et al. (2019), providing another application of our new estimation and inference procedures.

SA-IV.2.1 Additional Results for Business Credit Cycle Factor

SA-IV.2.1.1 Out-of-Sample Predictive Regressions

As an additional check on the predictive performance of the business-credit factor, Table SA-IV.2.1 reports out-of-sample R^2 from expanding-window predictive regressions. The out-of-sample R^2 compares the mean squared forecast error from each predictive regression with that from the recursively estimated historical mean forecast. We use data through 1989 as the initial training sample and evaluate the forecasts from January 1990 through December 2019. The business-credit factor produces positive out-of-sample R^2 at all four horizons and outperforms each of the benchmark predictors.

Table SA-IV.2.1: **Out-of-Sample Predictive Regressions** This table reports out-of-sample R^2 (in percent) from expanding window predictive regressions of cumulative market excess returns on an individual predictor from Welch and Goyal (2008) for horizons of 1, 3, 6, and 12 months ahead. The row labeled "CGP" reports results for the SLOOS only portion of our risk factor as studied in Chava et al. (2015). The row labeled "BCF" reports results for our proposed risk factor. Positive values have been bolded. The training period is 1965m1–1989m12 and the evaluation sample is 1990m1–2019m12.

	$\ell = 1$	$\ell = 3$	$\ell = 6$	$\ell = 12$
(log) Dividend Price Ratio	-1.79	-5.51	-12.00	-27.91
(log) Dividend Yield	-1.94	-5.35	-12.12	-26.96
(log) Earnings Price Ratio	-0.90	-3.70	-8.55	-14.87
(log) Dividend Payout Ratio	-1.50	-4.51	-6.51	-5.08
Stock Variance	**0.79**	-4.15	-1.54	-0.35
Book-to-Market Ratio	-0.60	-2.01	-4.68	-11.79
Net Equity Expansion	-2.35	-7.20	-16.86	-24.49
Treasury Bill Yield	**0.02**	-0.08	-0.38	-0.55
Long Term Treasury Yield	-0.03	-0.77	-2.51	-8.44
Long Term Treasury Return	-0.14	-1.10	-0.84	-0.98
Term Spread	-0.98	-2.12	-1.68	**4.26**
Default Yield Spread	-2.43	-5.89	-6.84	-5.46
Default Return Spread	-0.57	-1.09	-0.43	-0.95
(lagged) Inflation	-0.61	-0.08	**2.56**	**2.58**
CGP	**1.67**	**3.98**	**4.41**	**3.88**
BCF	**3.00**	**9.60**	**13.90**	**11.88**

SA-IV.2.1.2 Implementation and Tuning

Here we provide further details on the implementation of our main empirical application. The values of H_{it} and J_t are obtained following the procedure described in Section SA-III.1. The grid used to select J_t is described in Section SA-III.4. Figure SA-IV.2.1 summarizes the resulting data-driven tuning choices and the

effective cross-sectional sample size over time. The selected first-step windows vary substantially across firms and months: H_{it} equals its lower bound for 15.5% and its upper bound for 2.2% of firm-month observations assigned to portfolios. The selected number of portfolios ranges from 5 to 40 and reaches its upper bound in 4.8% of the sample months.

The effective cross-sectional sample size ranges from 2,153 to 5,122 firms, with an average of 3,949 firms per month. Even when the upper bound on J_t is attained, the smallest realized portfolio contains 54 firms. The sharp increase in the effective cross-sectional sample size in January 1976 reflects the beginning of CRSP's Nasdaq coverage in December 1972. Because many Nasdaq firms have their first usable return in January 1973, they first satisfy the requirement of 36 prior monthly observations in January 1976 and then enter the portfolio sorts.

Finally, the uniform-in-β calibration described in Section SA-III.4 gives $\widehat{\delta} = 0.0151$. With this threshold, the uniform variance floor applies at the seven reporting grid points from $\beta = -0.75$ through $\beta = 0.75$.

SA-IV.2.2 Capital Share Risk Factor

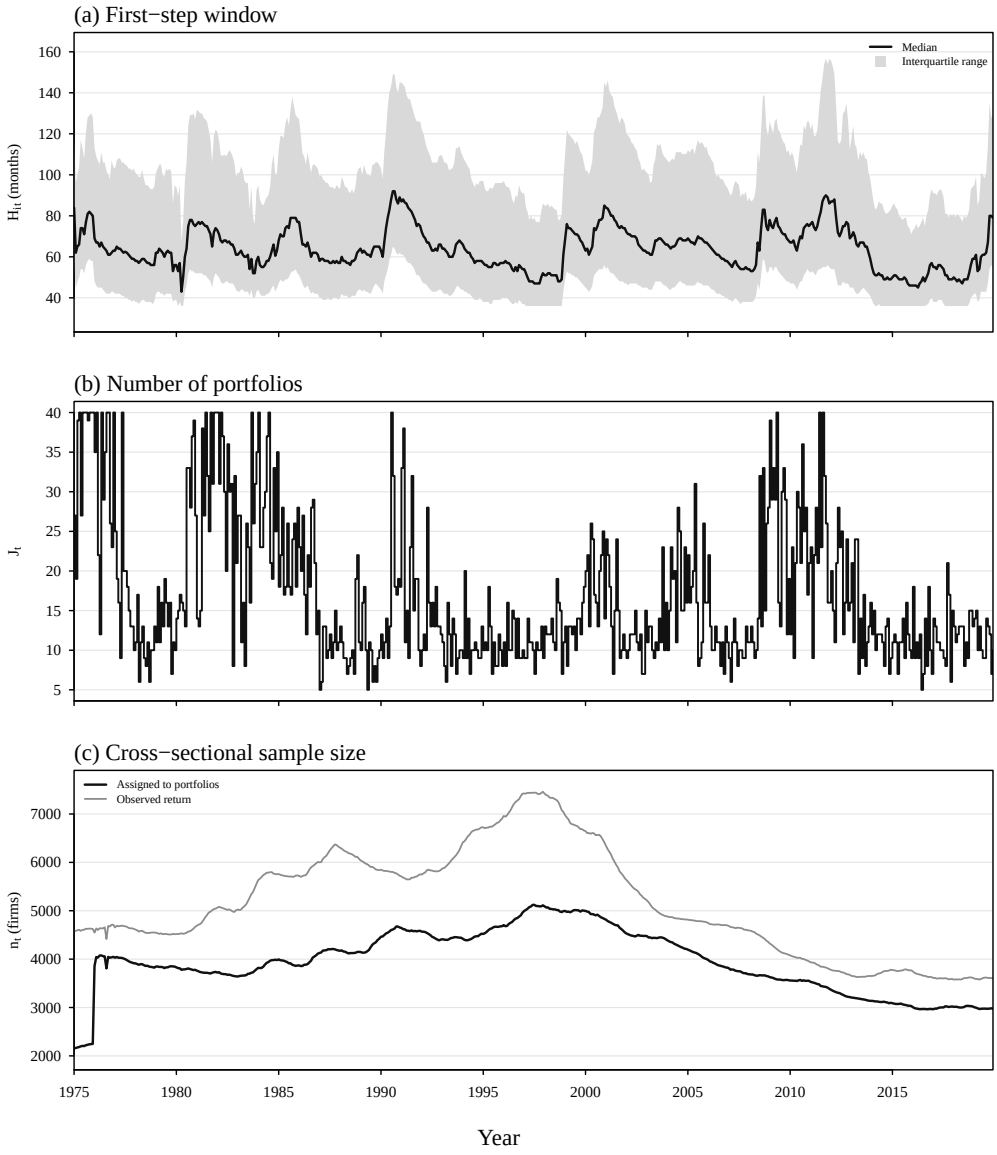
In this section, we revisit the capital share risk factor proposed by Lettau et al. (2019) using our new estimation and inference procedures. Lettau et al. (2019) argue that an empirical model in which growth in the capital share of aggregate income is the single source of macroeconomic risk explains a substantial fraction of the cross-sectional variation in expected returns across a range of equity characteristic portfolios and non-equity asset classes, with positive exposure associated with higher expected returns. Their benchmark capital share measure is constructed from quarterly data on the labor share of the nonfarm business sector from the Bureau of Labor Statistics' Productivity and Costs release.[2] Specifically, they define the capital share as $KS_t = 1 - LS_t$ and measure its growth over eight quarters. In our implementation, we multiply this log growth rate by 100 and construct the risk factor as $f_t = 100[\log(KS_t) - \log(KS_{t-8})]$, where LS_t denotes the labor share at time t.[3]

We use the same CRSP stock universe and treatment of delisting returns as in Section 6 of the main text. We form quarterly raw returns by compounding the three monthly returns within each quarter and use the one-factor quarterly implementation choices described in Section SA-III.4. The values of H_{it} and J_t are obtained following the procedure described in Section SA-III.1. Applying the default support rule to the point-estimation betas yields a J_t-selection grid of 25 equally spaced points with endpoints -1.92 and 1.92, while the corresponding rule applied to the inferential betas yields a reporting and inference grid of 11 equally spaced points with endpoints -2.37 and 2.37, where all endpoints are rounded. The full reporting interval lies within the realized cross-sectional range of the $\widehat{\beta}_{it}$ values used for inference in 90.3% of the estimation quarters. Figure SA-IV.2.2 summarizes the resulting data-driven tuning choices and effective cross-sectional sample size over time. The selected first-step window H_{it} equals its lower bound for 73.3% and its upper bound for 2.5% of firm-quarter observations assigned to portfolios. The selected number of portfolios ranges from 6 to 40 and reaches its upper bound in 1.3% of the sample quarters. The effective cross-sectional sample size ranges from 1,269 to 2,905 firms, with an average of 2,285 firms per quarter; the smallest realized portfolio contains 40 firms. Finally, the uniform-in-β calibration gives $\widehat{\delta} = 0.0548$. The

[2] We obtain these data from the paper's replication package available at https://onlinelibrary.wiley.com/doi/full/10.1111/jofi.12772.

[3] Gospodinov and Robotti (2021) reexamine these findings and argue that the original evidence of statistical significance may reflect the persistence induced by using overlapping long-horizon returns and capital-share growth. These authors recommend using nonoverlapping data or alternative empirical methods. This critique motivates our alternative implementation: we retain the eight-quarter capital-share growth factor of Lettau et al. (2019) but use nonoverlapping one-quarter returns and our beta-sorted portfolio procedure rather than their two-pass risk-premium procedure based on overlapping returns.

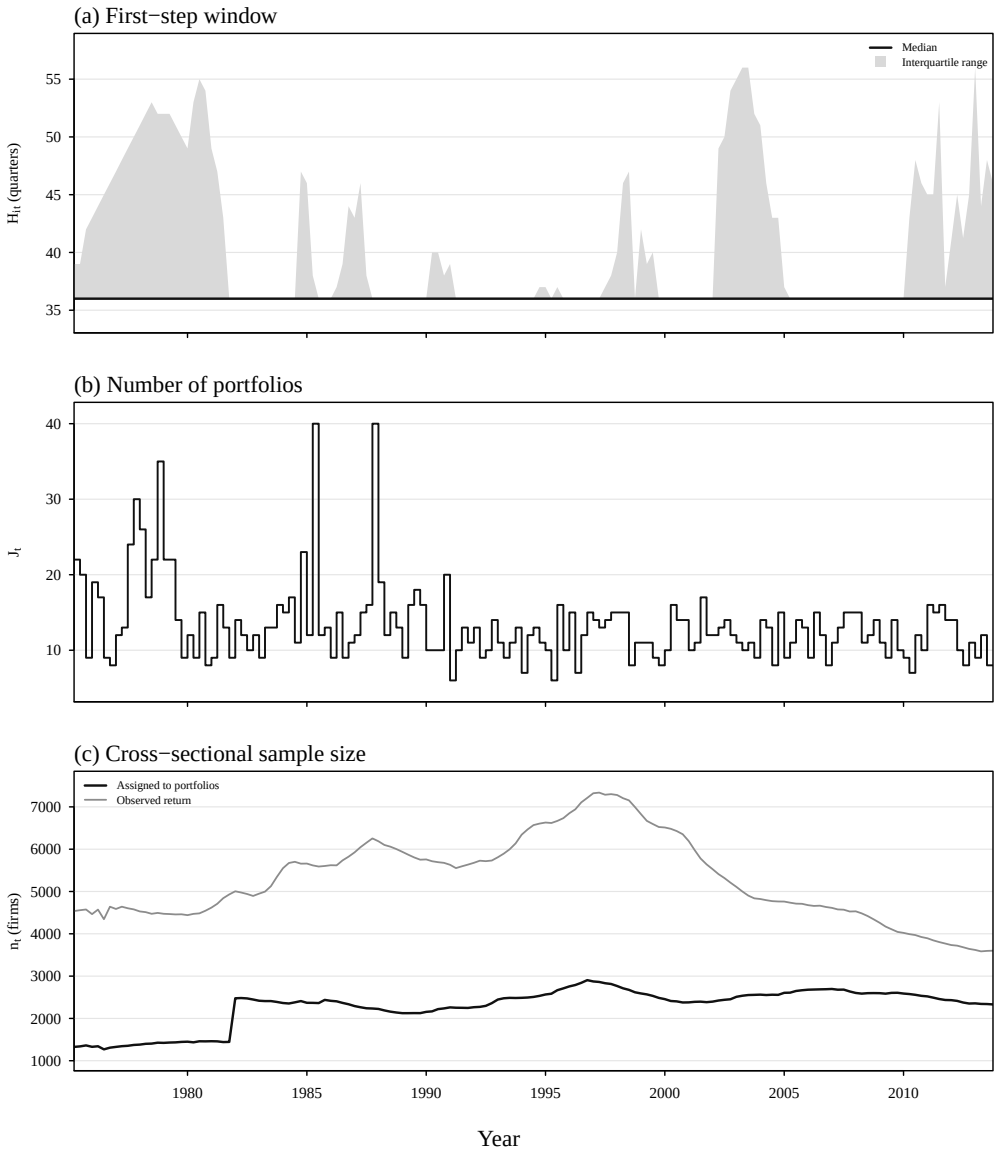
Figure SA-IV.2.1: **Tuning parameters and effective sample size over time.** Panel A reports the monthly median of the selected first-step window H_{it} across firms, with the shaded region denoting the interquartile range. Panel B reports the selected number of portfolios, J_t. Panel C reports the effective cross-sectional sample size, n_t, defined as the number of firms with a finite beta estimate that are assigned to a portfolio; the grey line reports the number of firms with an observed return before imposing the first-step estimation requirement. The sample period is January 1975 through December 2019.



estimated risk-factor variance component reaches this threshold at approximately $\beta = -1.27$ and $\beta = 1.27$. Accordingly, the uniform variance floor applies at the five reporting grid points from $\beta = -0.95$ through $\beta = 0.95$.

Figure SA-IV.2.3 reports the estimated SACER and associated inference using the capital share risk factor. The point estimate based on $(\mathfrak{p}, \mathfrak{q}) = (0, 0)$ is given by the black line. The estimated relation generally slopes upward; specifically, the estimated SACER increases from 4.17% per quarter at $\beta = -2.37$ to 5.02% at $\beta = 2.37$. This pattern is qualitatively consistent with the positive price of capital share risk

Figure SA-IV.2.2: **Tuning parameters and effective sample size over time for the capital share application.** Panel A reports the quarterly median of the selected first-step window H_{it} across firms, with the shaded region denoting the interquartile range. Panel B reports the selected number of portfolios, J_t. Panel C reports the effective cross-sectional sample size, n_t, defined as the number of firms with a finite beta estimate that are assigned to a portfolio; the grey line reports the number of firms with an observed complete quarterly return before imposing the first-step estimation requirement. The estimation period is 1975Q2 through 2013Q4.



documented by Lettau et al. (2019). The corresponding high-minus-low contrast point estimate based on $(\mathfrak{p}, \mathfrak{q}) = (0, 0)$ is 0.85 percentage points. Table SA-IV.2.2 reports inference for the same contrast using the $(\mathfrak{p}, \mathfrak{q}) = (1, 1)$ estimate; consequently, the confidence intervals need not be centered on the black-line estimate. The JN standard error is 1.030, giving a two-sided p-value of 0.605, while the modified FM standard error is 0.634, giving a two-sided p-value of 0.401. The $\widehat{\Sigma}_f$-only benchmark also fails to reject, so none of the three procedures rejects a zero contrast.

Figure SA-IV.2.3: **Inference on the SACER.** The black line reports the $(\mathfrak{p}, \mathfrak{q}) = (0, 0)$ estimate of the SACER. The grey region is the 95% JN pointwise interval. The red dashed lines are the 95% JN uniform-in-beta intervals, and the green dotted lines are the 95% modified FM pointwise intervals. All three intervals use $(\mathfrak{p}, \mathfrak{q}) = (1, 1)$. Results are reported for quarterly raw returns (in percent), formed by compounding the three monthly returns within each quarter. The sample period is 1965Q2–2013Q4 with a 40-quarter burn-in period, so the estimation period is 1975Q2–2013Q4.

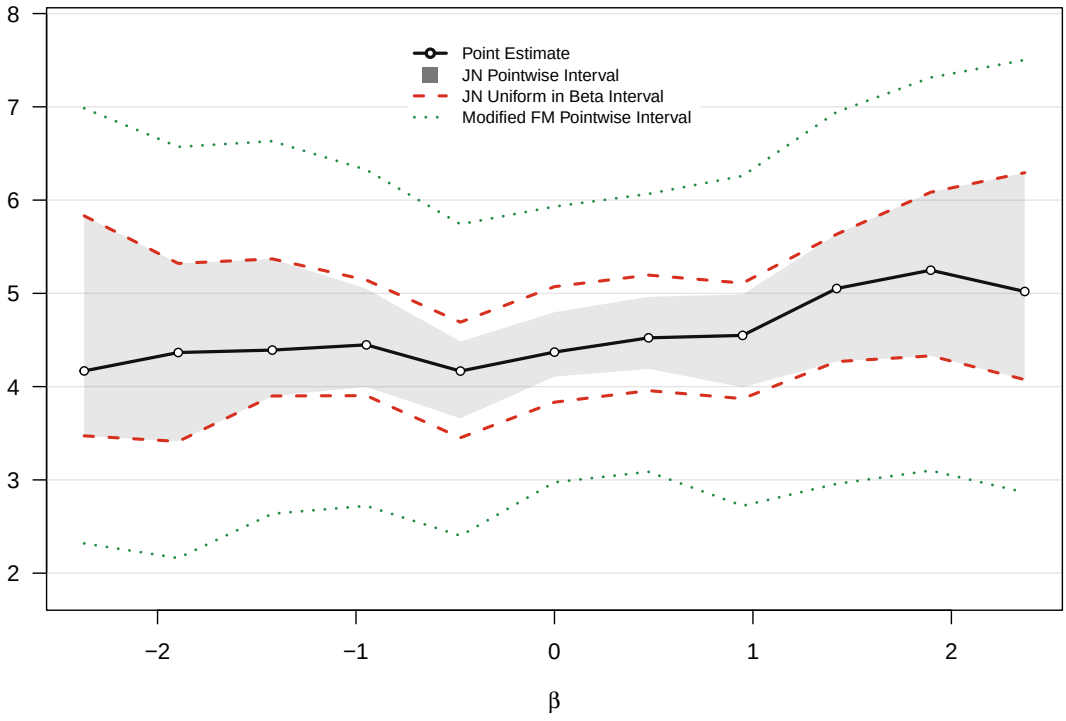


Table SA-IV.2.2: **High Minus Low Contrast.** The table reports inference for $\widehat{\mu}(2.37) - \widehat{\mu}(-2.37)$ in quarterly percentage points. Each row reports the standard error, 90% confidence interval, and two-sided normal-approximation p-value under the JN, modified FM, or $\widehat{\Sigma}_f$-only variance estimator. The confidence intervals are centered at the $(\mathfrak{p}, \mathfrak{q}) = (1, 1)$ estimate.

Variance estimator	SE	90% CI	p-value
JN	1.030	$[-1.162, 2.227]$	0.605
Modified FM	0.634	$[-0.511, 1.576]$	0.401
$\widehat{\Sigma}_f$ only	0.874	$[-0.904, 1.969]$	0.542

References

Calonico, S., Cattaneo, M. D., Farrell, M. H., 2018. On the effect of bias estimation on coverage accuracy in nonparametric inference. Journal of the American Statistical Association 113, 767–779.

Calonico, S., Cattaneo, M. D., Farrell, M. H., 2022. Coverage error optimal confidence intervals for local polynomial regression. Bernoulli 28, 2998–3022.

Calonico, S., Cattaneo, M. D., Titiunik, R., 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica 82, 2295–2326.

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

Crump, R. K., Gospodinov, N., 2022. On the factor structure of bond returns. Econometrica 90, 295–314.

Gospodinov, N., Robotti, C., 2021. Capital share risk in U.S. asset pricing: A reappraisal. Journal of Finance, Replications and Comments (web-only).

Lettau, M., Ludvigson, S. C., Ma, S., 2019. Capital share risk in U.S. asset pricing. Journal of Finance 74, 1753–1792.

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