Beta-Sorted Portfolios*

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

November 8, 2024

Abstract

Beta-sorted portfolios—portfolios comprised of assets with similar covariation to selected risk factors—are a popular tool in empirical finance to analyze models of (conditional) expected returns. Despite their widespread use, little is known of their 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 portfolios construction. Our framework rationalizes the well-known estimation algorithm with precise economic and statistical assumptions on the general data generating process. We provide conditions 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 shortcomings of the widely-used Fama-MacBeth variance estimator. To address these limitations, we propose a new variance estimator. In an empirical application, we introduce a novel risk factor—a measure of the business credit cycle—and show that it is strongly predictive of both the cross-section and time-series behavior of U.S. stock returns.

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

1 Introduction

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

The implementation of beta-sorted portfolios entails the following two-step procedure, which incorporates a beta-adaptive portfolios construction (see, e.g., Bali et al. 2016). In a first step, time-varying risk factor exposures are estimated through (backward-looking) rolling window time-series regressions of asset returns on the observed factors. The most popular implementation uses rolling window regressions, often with a choice of a five-year window. In a second step, the estimated factor exposures, based on data up to the previous period, are ordered and used to group assets into portfolios. These portfolios then represent assets with a similar degree of exposure to the risk factors, and the degree of return differential for differently exposed assets is used to assess the compensation for bearing this common risk. Most frequently this is achieved by differencing the portfolio returns from the two most extreme portfolios. Finally, an average over time of these return differentials is taken to infer whether the risk is priced unconditionally—whether the portfolio earns systematic (and significant) excess returns. Notwithstanding the simple and intuitive nature of the methodology, little is known of the formal properties of this estimator and its associated inference procedures.

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

---

[1]See, for example, Jagannathan and Wang (1998), Chen and Kan (2004), Shanken and Zhou (2007), Kleibergen (2009), Ang et al. (2020), Gospodinov et al. (2014), Adrian et al. (2015), Bai and Zhou (2015), Bryzgalova (2015), Gagliardini et al. (2016), Chordia et al. (2017), Kleibergen et al. (2019), 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. (2021b), Eisdorfer et al. (2022), Goldberg and Nozawa (2021), Fan et al. (2022b), Chen et al. (2023), and Chen et al. (2024).

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

This paper provides a host of new methodological and theoretical results. First, we introduce conditions that ensure validity and asymptotic normality of the beta-sorted portfolios estimator. We characterize precise conditions on the size of the window, H, of the first-stage rolling regression estimator, and the number of portfolios, J, of the second-stage estimator, relative to the sample sizes n and T. We show that the rate of convergence of the estimator depends on the value of beta. For beta values closer to zero the rate of convergence is faster and is slower otherwise; in fact, for values of beta away from zero we show that the rate of convergence of the estimator is only $\sqrt{T}$, despite an effective sample size of the order nT, reflecting specific properties of the setting of interest. However, we also show that certain features of expected returns such as the discrete second derivative—which represents a butterfly spread trade—can be estimated with higher precision through faster rates of convergence for all values of beta, namely, $\sqrt{nT/J}$ for a single risk factor. This result also accommodates more powerful tests of the null hypothesis of no-arbitrage.

In addition, we uncover some limitations along with layers of nuance in the current empirical practice employing the beta-sorted portfolios methodology. First, as with all nonparametric estimators, the choice of tuning parameters, H and J, are key to successful performance and are dependent on the sample sizes n and T. In contrast, empirical practice often chooses window length in the first step and total portfolios in the second step irrespective of the sample size at hand. Second, we show how valid inference depends critically on the object of interest. Using our framework, we distinguish between the population and sample average of conditional expected returns and argue that the latter object should be the estimand of interest. Moreover, when the focus is on the sample average of conditional expected returns, we show that the widely-used Fama and MacBeth (1973) variance estimator is not consistent, in general, but instead is upward biased. However, we show that the Fama-MacBeth variance estimator still leads to valid, albeit possibly

conservative, inference.

To address this limitation, we propose a new variance estimator and provide an empirical implementation which also produces valid, albeit possibly conservative, inference. That said, in our empirical application we show that our new variance estimator provides much sharper inference than the Fama-MacBeth variance estimator. Finally, we also provide results on the limitations of the beta-sorted portfolio estimator for a fixed time period. We show that differential returns for each period, often used as inputs for assessing the time-series properties of conditional expected returns, are contaminated by an additional term when risk factors are serially correlated. However, we demonstrate that some features of conditional expected returns for a single time period can be illuminated by utilizing butterfly spread trades.

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

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

---

[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 smoothly-varying coefficients nonparametric regression, and therefore prior results are not applicable to the settings considered herein.

[4]See, for example, Goyal (2012), Nagel (2013), Gospodinov and Robotti (2013), or Gagliardini 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), Fan et al. (2016), Kelly et al. (2019), Connor et al. (2021), Fan et al. (2022a), and Borri et al. (2024).

nest our setup. For example, the classical beta-sorted portfolio estimator implies a data-generating process that does not (necessarily) exclude arbitrage opportunities and supposes risk exposures which are smoothly-varying. Furthermore, we show that valid estimation and inference can be achieved without requiring an assumption of the functional form of the conditional expectation of the risk factors. This is in contrast to the existing literature (e.g., Adrian et al. 2015, Gagliardini et al. 2016) where such an assumption is utilized.

Our paper is also related to the financial econometrics literature on nonparametric estimation and inference. In particular, the two steps of the beta-sorted portfolio algorithm align individually with Ang and Kristensen (2012), who study kernel regression estimators of time-varying alphas and betas, and Cattaneo 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. Furthermore, we offer a host of new results characterizing the properties of beta-sorted portfolios.

We demonstrate the practical usefulness of our estimation and inference results for beta-sorted portfolios in a substantive empirical application with a novel methodological contribution. More precisely, we introduce a new risk factor—a measure of the business credit cycle—and show that it is strongly predictive of both the cross-section and time-series behavior of U.S. stock returns. We also show the effectiveness of our new variance estimator as inference is much more informative relative to inference employing the widely used Fama-MacBeth variance estimator.

In summary, this paper makes a number of contributions to understanding the underlying foundational properties and practical use of beta-sorted portfolios: we introduce an econometric framework to study the beta-sorted portfolio estimator and clarify its properties (Section 2); we provide asymptotic theory for the estimator accommodating the nonparametric first and second steps (Sections 3 and 4); we characterize the properties of the commonly-used Fama-MacBeth variance estimator along with a new plug-in variance estimator (Section 4.3); we provide results for joint inference across multiple values of beta, including a new test of no-arbitrage restrictions in Section 5; and we introduce a novel risk factor and demonstrate its desirable properties in an empirical application (Section 6). Proofs of our theoretical results are relegated to a Supplemental Appendix

4

(hereafter, SA) to streamline the exposition, and may be of independent technical interest.

1.1   Notation and conventions

For a constant $k \in N$ and a vector $v = (v_1, \ldots, v_d)^\top \in \mathbb{R}^d$, we denote $|v|_k = (\sum_{i=1}^d |v_i|^k)^{1/k}$, and $|v|_\infty = \max_{1 \leq i \leq d} |v_i|$. For a random variable V, let $\|V\|_q = (\mathbb{E}[|V|^q])^{1/q}$. We set $(a_n : n \geq 1)$ and $(b_n : n \geq 1)$ to be positive number sequences. We write $a_n = O(b_n)$ or $a_n \lesssim b_n$ (resp. $a_n \asymp b_n$) if there exists a positive constant C such that $a_n/b_n \leq C$ (resp. $1/C \leq a_n/b_n \leq C$) for all large n, and we denote $a_n = o(b_n)$ (resp. $a_n \sim b_n$), if $a_n/b_n \to 0$ (resp. $a_n/b_n \to C$). $\text{plim}_{n \to \infty} X_n = X$ means that $X_n \to_\mathbb{P} X$. $\to_\mathcal{L}$ denotes convergence in law. Define $X_n = O_\mathbb{P}(a_n)$: $\exists N_\varepsilon > 0, \delta_\varepsilon > 0$ such that $\mathbb{P}(|X_n| \geq \delta_\varepsilon) \leq \varepsilon$ $\forall n > N_\varepsilon$. Define $X_n = o_\mathbb{P}(a_n)$: $\forall \varepsilon, \delta > 0$ $\exists N_{\varepsilon,\delta} > 0$ such that $\mathbb{P}(|X_n| \geq \delta) \leq \varepsilon$ $\forall n > N_{\varepsilon,\delta}$. We also let $X_n \lesssim_\mathbb{P} a_n$ to mean $X_n = O_\mathbb{P}(a_n)$; furthermore, if $X_n \lesssim_\mathbb{P} a_n$ and $a_n \lesssim_\mathbb{P} X_n$ then we write $X_n \asymp_\mathbb{P} a_n$. Limits are taken as $n, T, J, H \to \infty$ unless otherwise stated explicitly. Set $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$.

Following Wu (2005), Chen et al. (2021a), and Han and Wu (2023), we consider the following measure of time-series dependence throughout.

Definition 1.1. *The (n, T)-varying time series $Y_{\bullet,n,T} := (Y_{t,n,T} : t \in \mathbb{Z}, n \geq 1)$ is said to be a nonlinear time series system (NTSS) if $Y_{t,n,T} = g_{t,n,T}(\cdots, \xi_{t-2}, \xi_{t-1}, \xi_t)$, where the $(\xi_t : t \in \mathbb{Z})$ are independent and identically distributed (i.i.d.) random vectors.*

Let $Y_{\bullet,n,T}^*(\ell)$ be a NTSS with $Y_{t,n,T}^*(\ell)$ a random variable with $\xi_{t-\ell}$ replaced by $\xi_{t-\ell}^*$, where ξ_t^* is an *i.i.d.* copy of ξ_t for each $t \in \mathbb{Z}$. Then, the dependence adjusted norm of $Y_{\bullet,n,T}$ is

$$\Theta_{n,T}(Y_{\bullet,n,T}; q, v) := \sup_{m \geq 0}(m+1)^v \sum_{\ell=m}^\infty \max_{1 \leq t \leq T} (\mathbb{E}[|Y_{t,n,T} - Y_{t,n,T}^*(\ell)|^q])^{1/q},$$

with $v \geq 0$ and $q \geq 1$. When the $Y_{\bullet,n,T}$ does not depend on n and T, we simplify the notation to $\Theta_T(Y_\bullet; q, v)$ for the possibly nonstationary NTSS $Y_\bullet := (Y_t : t \in \mathbb{Z})$, where $Y_t = g_t(\cdots, \xi_{t-2}, \xi_{t-1}, \xi_t)$ with the function $g_t(\cdot)$ no longer a function of n and T. Furthermore, if $g_t(\cdot)$ is not a function of t, then $Y_\bullet$ is a stationary NTSS and we write $\Theta(Y_\bullet; q, v)$.

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 discuss the relevant properties of the model with respect to the potential presence of arbitrage opportunities.

2.1 Modeling returns

Let R_{it} denote the return of asset i at time t, 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,

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

where α_{it} and β_{it} are smoothly-varying, random coefficients with $\alpha_{it}, \beta_{it} \in \mathcal{F}_{t-1}$, ε_{it} is an idiosyncratic error term, and $\mathcal{F}_t$ is the information set up to time t. To be more precise, we define the sigma field,

$$\mathcal{F}_t = \sigma\left(\{f_s\}_{s=1}^t, \{\varepsilon_{it}\}_{i=1,s=1}^{n_t,t}, \{\aleph_{1,i}\}_{i=1}^{n_t}, \{\aleph_{2,s}\}_{s=1}^t, \{\aleph_{3,is}\}_{i=1,s=1}^{n_t,t}\right), \tag{2.2}$$

where $\aleph_{1,i}$, $\aleph_{2,t}$, and $\aleph_{3,it}$ are vectors of auxiliary random variables, which are possibly observed, and will be discussed further in later sections. Since α_{it} and β_{it} are both $\mathcal{F}_{t-1}$-measurable we have that $\mathbb{E}[R_{it} | \mathcal{F}_{t-1}, f_t] = \alpha_{it} + \beta_{it}^\top f_t$. The sigma field $\mathcal{F}_t$ will, in general, depend on n and T but we suppress this dependence for notational convenience. 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 $t = 1, \cdots, T$). For an alternative example of a random coefficient model tailored to a financial application, see Barras et al. (2022).

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

$$\mathbb{E}(R_{it} | \mathcal{F}_{t-1}) = \mu_0(\beta_{it}; \mathcal{G}_{t-1}) =: \mu_t(\beta_{it}), \qquad \mathcal{G}_t = \sigma(\{f_s\}_{s=1}^t, \{\aleph_{2,s}\}_{s=1}^t). \tag{2.3}$$

Equation (2.3) 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 i-invariant random variables. This restriction captures the notion that nonzero expected returns must reflect the compensation investors require for exposure (as measured by β_{it}) to the risk factors, f_t. We will make precise assumptions on $\mu_t(\cdot)$ in later sections but, loosely, one can think of $\mu_t(\cdot)$ as being a sufficiently smooth random function of β_{it}.

Using equations (2.1) and (2.3) we obtain,

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

which clarifies the restriction on α_{it} implied by equation (2.3). Equation (2.3) is a nonparametric analog to a reduced-rank restriction in parametric models as it imposes cross-equation restrictions between the linear coefficients in equation (2.1).

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

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

We assume throughout that R_{it} represents excess returns, but $\mu_t(0)$ may be interpreted as the zero-beta rate at time t in the case when R_{it} represents raw returns. 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 R_{it} represents excess returns then the no-arbitrage restriction implies that $\mu_t(\beta_{it}) = \beta_{it}^\top \lambda_t$ for some λ_t (Gagliardini et al., 2016). Our model nests, but does not require, the imposition of the absence of arbitrage opportunities so that

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

The presence of this additional term representing the deviation from no-arbitrage restrictions can be

motivated by appealing to structural models which feature violations of the law of one price. Such a setup as in equation (2.5) could arise, for example, in the margin-constraints model of Garleanu and Pedersen (2011) under the assumption that the security's margin is a nonlinear function of its (past) beta.

2.2 Relation to the two-step estimator

Throughout the paper we set $d = 1$ to simplify the notation and the exposition. All results can be generalized to the $d > 1$ case. To see why equation (2.5) rationalizes the beta-sorted portfolio algorithm, let us revisit the standard two steps:

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

$$\left(\widehat{\alpha}_{it_0}, \widehat{\beta}_{it_0}\right)^{\top} = \left(\sum_{s=1}^{H} X_{t_0-s} X_{t_0-s}^{\top}\right)^{-1} \left(\sum_{s=1}^{H} X_{t_0-s} R_{i(t_0-s)}\right), \tag{2.6}$$

where $X_t = (1, f_t)^{\top}$ and H is the window length. This construction purposely does not have "look-ahead bias", as neither the estimators of $\widehat{\alpha}_{it_0}$ or $\widehat{\beta}_{it_0}$ use data from time t_0 (or after) in their construction (a "leave-one-out" estimator). For each cross-sectional unit, this estimation of the time-varying random coefficients can be interpreted as a leave-one-out kernel regression of equation (2.1) using a uniform kernel and a choice of bandwidth h which satisfies $H = \lfloor Th \rfloor$, where $\lfloor . \rfloor$ denotes the floor function. In the SA we provide all proofs for an arbitrary one-sided kernel which generalizes the results we present in the following sections. ∎

Step 2: Sorting portfolios using estimated β_{it}. To see that this step comprises a cross-sectional nonparametric estimation, observe that, for fixed t, equation (2.3) is the conditional mean of interest. Now, to cement intuition, assume (temporarily) that β_{it} is observed *and* takes on a finite number of values, one of which is $\bar{b}$. Then, to estimate $\mathbb{E}[R_{it}|\beta_{it} = \bar{b}]$ we need only calculate the sample mean of returns for assets i for which $\beta_{it} = \bar{b}$:

$$\frac{1}{\#\{i : \beta_{it} = \bar{b}\}} \sum_{i:\beta_{it}=\bar{b}} R_{it}. \tag{2.7}$$

8

When β_{it} can take on a continuum of values, we can estimate $\mathbb{E}[R_{it}|\beta_{it} = \bar{b}]$ by averaging returns for those assets i such that β_{it} is "near" $\bar{b}$. Forming portfolios is one way to operationalize this local averaging approach. Then, the portfolio return representing assets with values of β_{it} closest to $\bar{b}$ is our estimate for this conditional expectation.

We define $\mathcal{B} = [\beta_l, \beta_u]$, with β_l and β_u fixed constants, as the support of the possible realizations of β_{it} across i and t. Since, in practice, β_{it} is unobserved, we may obtain estimates $\widehat{\beta}_{it}$ from Step 1 and, for each $t = 1, \ldots, T$, we can define the beta-adaptive partition of $\mathcal{B}$ as

$$\widehat{P}_{jt} = [\widehat{\beta}_{(\lfloor n_t(j-1)/J_t\rfloor)t}, \widehat{\beta}_{(\lfloor n_t j/J_t\rfloor)t}), \qquad j = 0, \ldots J_t - 1$$

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

where $\widehat{\beta}_{(\ell)t}$ denotes the ℓth order statistic of the estimated betas in the first step, across i for fixed t, i.e., the order statistics of $\{\widehat{\beta}_{it} : i = 1, \ldots, n_t\}$, and where we set $\widehat{\beta}_{(0)t} = \beta_l$ and $\widehat{\beta}_{(n_t)t} = \beta_u$ for simplicity. The intervals, $\widehat{P}_{jt}$, allow us to partition assets into portfolios based on their estimated beta, $\widehat{\beta}_{it}$. The number of portfolios J_t, and their random structure (i.e., breakpoint positions based on estimated β_{it}), vary for each time period but we assume that $J_t \asymp J$. Finally, we can construct portfolio returns by averaging the returns of the assets in each portfolio. ∎

Given the two-step construction outlined above, suppose we would like to estimate $\mathbb{E}[R_{it}|\beta_{it} = \beta_l]$. The estimator is simply the portfolio return for the first portfolio, $\widehat{P}_{1t}$, which we write as $\widehat{\mu}_t(\beta_l)$. Similarly, we can estimate $\mathbb{E}[R_{it}|\beta_{it} = \beta_u]$ with the portfolio return for the last portfolio, $\widehat{P}_{J_t t}$, as $\widehat{\mu}_t(\beta_u)$. We can then average the differential portfolio returns for these two extreme portfolios across all available time periods as,

$$\widehat{\mu}(\beta_u) - \widehat{\mu}(\beta_l) = \frac{1}{T-H} \sum_{t=H+1}^{T} [\widehat{\mu}_t(\beta_u) - \widehat{\mu}_t(\beta_l)]. \tag{2.8}$$

This is exactly the estimator that is used in practice.

A few comments are in order. First, the above two steps are completely in line with the empirical finance literature. Importantly, at no point in the two-step algorithm is there a requirement to estimate the conditional expectation of the risk factors, $\mathbb{E}[f_t|\mathcal{F}_{t-1}]$, and so the researcher remains agnostic about the dynamics of these risk factors. We will revisit this issue in later sections. Second,

the practice of using moving-window regressions to accommodate time variation in β_{it} suggests a slowly-varying coefficient model as previously used in finance applications such as in Ang and Kristensen (2012) and Adrian et al. (2015). However, in contrast to these previous formulations, we do not condition on the realizations of the random processes α_{it} and β_{it}. Instead, we retain the randomness in these objects so that the second-stage beta-sorted portfolio estimator can have a well-defined limit as $n, T \rightarrow \infty$. Third, an alternative to the smoothly-varying coefficients approach is to specify β_{it} as a function of individual characteristics and possibly also of economy-wide variables (see, for example, Gagliardini et al., 2020, and references therein). Our approach can accommodate such settings by modifying the rolling window regressions (kernel regressions) appropriately.

Although we are motivated by empirical practice, which has been replicated exactly in equation (2.8), we introduce and work with the more general estimation approach,

$$\widehat{\mu}(\beta) = \frac{1}{T - H} \sum_{t=H+1}^{T} \widehat{\mu}_t(\beta). \tag{2.9}$$

Let $\widehat{P}_{j_t^\star t}$ be the portfolio that contains the value β at time t. Then, $\widehat{\mu}_t(\beta)$ is simply the return of $\widehat{P}_{j_t^\star t}$. Importantly, the value of $j_t^\star$ will generally change over time. For example, it may be that assets with values of β near $1/2$ fall in the sixth portfolio at times and the fifth portfolio at other times and so on. Thus, this more general estimation approach does not constitute spurious generality. The conventional implementation of beta-sorted portfolios relies on a constant choice of $J_t = J \ \forall t$ and so averages J portfolios across all time periods. However, if the cross-sectional distribution of the β_{it} are changing over time then there is no guarantee that a chosen portfolio represents assets with sufficiently similar betas. Therefore, the conventional estimator will be, in general, both more biased and more variable than the estimator given in equation (2.9), all else equal. This is of special importance when we are interested in expected returns for intermediate values of betas and also in situations where tests of monotonicity or shape restrictions are of interest. We discuss these issues further in later sections.

3 First step: rolling regressions

The first step in the estimation procedure involves a sequence of rolling window time-series regressions. To establish consistency (and, in the SA, asymptotic normality) of $\widehat{\beta}_{it}$ we require technical, but relatively standard, assumptions on the underlying data generating process. We first restrict the behavior of the idiosyncratic error terms, ε_{it}, in equation (2.1).

Assumption 1 (Idiosyncratic errors)**.** *For each i, $\varepsilon_{i\bullet}$ is a NTSS satisfying equation (2.1) with $\mathbb{E}[\varepsilon_{it}|\mathcal{F}_{t-1}, f_t] = 0$, $\max_{1 \leq t \leq T} \max_{1 \leq i \leq n_t} \mathbb{E}[|\varepsilon_{it}|^q|\mathcal{F}_{t-1}, f_t] \lesssim 1$ for $q > 8$, and $\min_{1 \leq t \leq T} \min_{1 \leq i \leq n_t} \mathbb{E}[\varepsilon_{it}^2|\mathcal{F}_{t-1}, f_t] \gtrsim 1$. In addition, $\Theta_T(\varepsilon_{i\bullet}; q, v) \lesssim 1$ for some $2v > 1/2 - 2/q$.*

Assumption 1 imposes moment conditions on the idiosyncratic error term, ε_{it} along with regularity conditions controlling the rate of decay of the time-series dependence. We define $\mathbb{E}[\varepsilon_{it}^2|\mathcal{F}_{t-1}, f_t] = \sigma_t^2$ which Assumption 1 ensures is bounded and bounded away from zero. It is important to emphasize that for the first step we need only impose that equation (2.1) holds (without imposing equation (2.3)). We now characterize the behavior of the risk factors, f_t.

Assumption 2 (Factors)**.** *$f_t = \tau(t/T) + z_t$, where $\tau(\cdot)$ is twice differentiable on $[0, 1]$, and $z_\bullet$ is a stationary NTSS, where $\mathbb{E}[z_t] = 0$, $\mathbb{E}[|z_t|^q] \lesssim 1$ with $q > 8$, $\max_{1 \leq t \leq T} \mathbb{V}\mathrm{ar}[z_t|\mathcal{F}_{t-1}] \lesssim 1$, and $\min_{1 \leq t \leq T} \mathbb{V}\mathrm{ar}[z_t|\mathcal{F}_{t-1}] \gtrsim 1$. In addition, $\Theta(z_\bullet; q, v) \lesssim 1$ for some $v > 1/2 - 2/q$.*

Assumption 2 imposes some structure on the time series properties of the factor f_t but is quite general and allows for certain forms of nonstationary behavior. We could relax some of these assumptions to allow for even more complex time-series properties at the expense of more detailed notations and proofs. We next restrict the behavior of α_{it} and β_{it}.

Assumption 3 (Varying coefficients)**.** *Let $\alpha_{it} = \alpha_i(t/T; \mathcal{F}_{t-1})$ and $\beta_{it} = \beta_i(t/T; \mathcal{F}_{t-1})$ where $\alpha_i(\cdot)$ and $\beta_i(\cdot)$ are nonrandom functions which are bounded for all i and satisfy, for any $t, t' \in [H+1, T]$, $|\alpha_{it} - \alpha_{it'}| \leq C_\alpha |t - t'|/T$ and $|\beta_{it} - \beta_{it'}| \leq C_\beta |t - t'|/T$, where C_α, C_β are two nonrandom, positive, and bounded constants uniformly in i, t, and t'. Further, Let $h = (H+1)/T$, where $H = H(n, T)$.*

Assumption 3 ensures that the alphas and betas, although random, are sufficiently smooth over time (i.e., satisfying a Lipschitz-type condition). This is the formal assumption which justifies the common empirical approach of using rolling window regressions.

To provide intuition for our proof approach recall that $X_t = (1, f_t)^\top$ so that for a fixed time period $t_0 \in [H + 1, T]$ we can rewrite equation (2.1) as

$$R_{it_0} = X_{t_0}^\top(\alpha_{it_0}, \beta_{it_0}) + \varepsilon_{it_0}.$$

Then, by Assumption (1) and since α_{it} and β_{it} are measurable with respect to $\mathcal{F}_{t-1}$, α_{it_0} and β_{it_0} can be identified as

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

We can use the estimator from equation (2.6) to obtain $(\widehat{\alpha}_{it_0}, \widehat{\beta}_{it_0})$. In order to accommodate the random coefficients we exploit the fact that $\sum_{s=1}^H X_{t_0-s} X_{t_0-s}^\top$ and $\sum_{s=1}^H X_{t_0-s} R_{i(t_0-s)}$ are close, in the appropriate sense, to $\sum_{s=1}^H \mathbb{E}[X_{t_0-s} X_{t_0-s}^\top | \mathcal{F}_{t_0-1}]$ and $\sum_{s=1}^H \mathbb{E}[X_{t_0-s} R_{i(t_0-s)} | \mathcal{F}_{t_0-1}]$. This follows because their difference are summands of martingale difference sequences.

We first provide a (uniform) consistency result of our estimator $\widehat{\beta}_{it_0}$ over i and t_0. This result generalizes the time varying coefficient analyses in Zhang and Wu (2012). We require this result to precisely control the effect of estimating β_{it} in the first step when entering the second-step estimator. We establish this consistency on a compact interval of a trimmed support $[H + 1, T]$.

Theorem 3.1 (First-step estimator). *Suppose Assumptions 1–3 hold, and $\max_{1 \le t \le T} n_t \lesssim n$. If $\log(nT)/(Th) \to 0$ and $T^{2/q-1}n^{2/q}/h \to 0$, then*

$$\sup_{\lfloor Th \rfloor \le t_0 \le T} \max_{1 \le i \le n_{t_0}} \left| \widehat{\beta}_{it_0} - \beta_{it_0} \right| \lesssim_\mathbb{P} \sqrt{\frac{\log(nT)}{Th}} + h =: \mathsf{R}_{nT}.$$

Theorem 3.1 provides uniform (over t and i) rates of convergence for the first-stage rolling-window (kernel) estimators of the betas. Naturally, these rates depend on n, T, and H but are also directly dependent on q which represents the number of bounded moments of the idiosyncratic error term and the observed factors. For sufficiently large q, consistency of the rolling regression estimator can attain the usual nonparametric (optimal) convergence rate. In the SA we also provide conditions that ensure asymptotic normality of $\widehat{\beta}_{it}$, which may be of independent interest.

4 Second step: beta sorts

The second step of the estimation procedure is to sort assets into portfolios based on their value of $\widehat{\beta}_{it}$, which is obtained from the procedure described in the previous section. Then, returns for each portfolio are constructed and used to assess the relation between exposure to the risk factors and subsequent asset returns. In this section, we formalize the properties of this estimator and provide conditions for its validity. We also discuss the importance of the cross-sectional restrictions implied by equation (2.4) for our main results. Before proceeding, however, we will first introduce two different objects of interest which will be useful for clarifying our results, and discussing empirical practice more broadly.

Let us first define the sample average conditional expected returns (SACER) as:

$$\bar{\mu}_T(\beta; H) = \frac{1}{T - H} \sum_{t=H+1}^{T} \mu_t(\beta). \tag{4.1}$$

We can compare the SACER directly to equation (2.9) to observe that this would, at first glance, be the natural candidate for our "estimand". However, it is important to point out that the SACER is a sample average of random quantities and, hence, is 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. To do so we need the following assumption:

Assumption P (PACER). *There exists a non-random function $\mu(\cdot)$ such that $\sup_{\beta \in \mathcal{B}} |T^{-1} \sum_{t=1}^{T} \mu_t(\beta) - \mu(\beta)| = o_{\mathbb{P}}(1)$.*

Assumption P imposes an ergodicity condition on the sample average of conditional expected returns. When $H/T = o(1)$ (as we impose) then Assumption P also implies that $\sup_{\beta \in \mathcal{B}} |\bar{\mu}_T(\beta; H) - \mu(\beta)| = o_{\mathbb{P}}(1)$ so that the PACER can be thought of as the probability limit of the SACER. In the special case where $\mu_t(\cdot) = \mu(\cdot) \ \forall t$, then the SACER and PACER are equal.

The SACER and the PACER represent two distinct objects of interest. To make things concrete, suppose that $\mu_t(\cdot)$ is random only through a finite collection of strictly stationary state variables for the economy, say, $\mathsf{S}_{t-1} \in \mathcal{G}_{t-1}$. Then, when we study the SACER we are learning about average conditional expected returns for the realizations of these state variables over a specific time period only. Equation (2.3) 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. Consequently, for any particular sample, it may be that the SACER and PACER are "far away" from each other.

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 μ_t 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, and the presence of unusual events (e.g., financial crises or natural disasters). Conversely, interpretation of the PACER will generally hinge on the ergodic distribution for $\mu_t(\cdot)$ which itself depends on how the (subset of) economy-wide state variables drive 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 is necessarily at least as challenging as inference on the SACER since, in practice, we only observe a finite sample. We formalize this intuition in Theorem 4.3. Despite our preference for the SACER, in Section 4.4 we provide a further discussion of issues related to inference on the PACER.

4.1   Cross-sectional restrictions

We impose the following assumption to justify the second step of the standard empirical approach.

Assumption 4 (Cross-sectional restrictions)**.** *Equation* (2.3) *holds, and* $\mathbb{E}[f_t|\mathcal{F}_{t-1}] = \mathbb{E}[f_t|\mathcal{G}_{t-1}]$ *for* $t = 1, 2, \ldots, T$.

Assumption 4 formally imposes the economic restrictions on conditional expected returns discussed in Section 2. We restrict the conditional expectation of the factors to be a function of random variables in the smaller sigma field $\mathcal{G}_t$. Taken together, Assumption 4 provides the foundation for the validity of the beta-sorted portfolio estimator. Let us first define systematic realized returns as

$$M_t(\beta_{it}) = \alpha_{it} + \beta_{it}f_t = \mu_t(\beta_{it}) + \beta_{it}(f_t - \mathbb{E}[f_t|\mathcal{G}_{t-1}]), \tag{4.2}$$

14

where the second equality follows by Assumption 4. Next, to provide intuition, temporarily assume that the β_{it} are observed. Then, using equation (4.2), our model can be written as

$$R_{it} = \mu_t(\beta_{it}) + \beta_{it}(f_t - \mathbb{E}[f_t|\mathcal{G}_{t-1}]) + \varepsilon_{it} = M_t(\beta_{it}) + \varepsilon_{it}, \qquad (4.3)$$

and, under Assumption 1, we have that $\mathbb{E}(\varepsilon_{it}|\mathcal{F}_{t-1}, f_t) = 0$ (recall that $\beta_{it} \in \mathcal{F}_{t-1}$). The second equality in equation (4.3) makes clear that, for a fixed time period, we can only nonparametrically estimate the unknown function $M_t(\cdot)$ rather than the direct object of interest $\mu_t(\cdot)$. However,

$$\frac{1}{T-H} \sum_{t=H+1}^{T} M_t(\beta) = \frac{1}{T-H} \sum_{t=H+1}^{T} \mu_t(\beta) + \frac{1}{T-H} \sum_{t=H+1}^{T} \beta(f_t - \mathbb{E}[f_t|\mathcal{G}_{t-1}]). \qquad (4.4)$$

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

Assumption 4 allows us to highlight another important issue. As discussed in Section 3, the first-stage estimator requires smoothly-varying random coefficients α_{it} and β_{it}. However, the cross-sectional restrictions in Assumption 4 impose additional structure on these random coefficients. The combination then implies restrictions on the functional form of conditional expected returns. For example, if $\mathbb{E}[f_t|\mathcal{G}_{t-1}]$ is constant for all t, then, for α_{it} to be smooth over time, we could have $\mu_t(\cdot)$ be constant over time or, alternatively, a smooth random function over time. However, consider a more realistic example. Let $\mathsf{S}_t \in \mathcal{G}_t$ again be a vector of strictly stationary state variables, but further assume $\mathbb{E}[f_t|\mathcal{G}_{t-1}] = \zeta + \Lambda \mathsf{S}_{t-1}$, $\Lambda \neq 0$ (as in, e.g., Adrian et al. 2015 or Gagliardini et al. 2016). Then, since equation (2.4) holds by Assumption 4, we *must* have that $\mu_t(\cdot)$ varies over time or else we violate Assumption 3 and the first-stage estimator of β_{it} will not be consistent in general. Moreover, when $\mu_t(\cdot)$ varies over time, the SACER and PACER will not be equal, and this directly affects the interpretation of the estimation and inference results in any particular

empirical application; see Theorem 4.3 and associated discussion. Results such as these underscore the importance of providing a formal framework for interpreting the beta-sorted portfolio estimator.

4.2 Estimation and inference

The remainder of this section presents our main results, culminating in the asymptotic normality of the beta-sorted portfolio estimator under appropriate conditions. We first restrict the cross-sectional behavior of the idiosyncratic error term, ε_{it}.

Assumption 5 (Idiosyncratic errors). *For each $t = 1, \ldots, T$, conditional on $(\mathcal{F}_{t-1}, f_t)$, $\{\varepsilon_{it} : i = 1, \ldots, n_t\}$ are independent and identically distributed, where $n \lesssim \min_{1 \leq t \leq T} n_t \leq \max_{1 \leq t \leq T} n_t \lesssim n$.*

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 in the innovations over time as a function of the factors f_t and the variables in $\aleph_{2,t}$.

We next impose restrictions on the β_{it}, which serve as the covariates in the second-stage non-parametric regression.

Assumption 6 (Cross-sectional betas). *For each $t = 1, \ldots, T$, conditional on $\mathcal{G}_{t-1}$, $\{\beta_{it} : i = 1, \ldots, n_t\}$ are independent and identically distributed. $F_{\beta,t}(b) = \mathbb{P}[\beta_{it} \leq b | \mathcal{G}_{t-1}]$ is twice continuously differentiable on its support $\mathcal{B} = [\beta_l, \beta_u]$, where $f_{\beta,t}(b) = \frac{d}{db} F_{\beta,t}(b)$ is bounded away from zero uniformly on $b \in \mathcal{B}$ and $t = 1, 2, \ldots, T$. Further, $J \lesssim \min_{1 \leq t \leq T} J_t \leq \max_{1 \leq t \leq T} J_t \lesssim J$, where $J_t = J_t(n_t, T)$.*

Assumption 6 imposes regularity conditions on the conditional distribution of the β_{it} ensuring that it is sufficiently well behaved. Specifically, the assumption ensures that the partitioning estimator is well defined with the probability of empty portfolios vanishing asymptotically. Furthermore, if we define $\Phi_{i,j,t} = \mathbb{1}(F_{\beta,t}^{-1}((j-1)/J_t) \leq \beta_{it} < F_{\beta,t}^{-1}(j/J_t))$ then $q_{jt} = \mathbb{E}[\Phi_{i,j,t} | \mathcal{G}_{t-1}]$ is of the order J^{-1}. The conditional *i.i.d.* assumption in Assumption 6 is similar to that of Assumption 5 and is quite general allowing, for example, a nonlinear factor structure in β_{it}. As a concrete example, let $\tilde{\aleph}_{1,i}$ and $\tilde{\aleph}_{3,i(t-1)}$ be selected *i.i.d.* variables from $\aleph_{1,i}$ and $\aleph_{3,i(t-1)}$, re-

spectively. Then, $\beta_{it} = g_{\beta,0}(\frac{t}{T}; g_{\beta,1}(\tilde{\aleph}_{1,i}, \tilde{\aleph}_{3,i(t-1)})^\top g_{\beta,2}(\aleph_{2,t-1}, \ldots, \aleph_{2,t-k_\aleph}, f_{t-2}, \ldots, f_{t-k_f}))$ for sufficiently smooth $g_{\beta,0}(\cdot)$, measurable (non-constant) $g_{\beta,1}(\cdot)$ and $g_{\beta,2}(\cdot)$, and fixed lag lengths $k_\aleph$ and k_f, satisfies Assumption 6.

Finally, the last assumption we require is that our object of interest, $\mu_t(\beta)$, is sufficiently smooth in β to accommodate nonparametric estimation.

Assumption 7 (Smoothness). $\mu_t(\beta)$ *is differentiable with bounded derivative uniformly in* $\beta \in \mathcal{B}$ *and* $t = 1, 2, \ldots, T$.

This assumption is standard in the nonparametric literature and rules out discontinuities and other pathologies that would invalidate standard nonparametric estimation approaches. As discussed in Section 2, portfolio sorting can be interpreted as a nonparametric estimator of a conditional mean. In order to operationalize the estimator let us define $F_{\widehat{\beta},n,t}(\cdot) = \frac{1}{n_t} \sum_{i=1}^{n_t} \mathbb{1}(\widehat{\beta}_{it} \leq \cdot)$ and $F_{\widehat{\beta},n,t}^{-1}$ as the empirical CDF and empirical quantile function for the estimates of the β_{it} (recall that $\widehat{\beta}_{(s)t} = F_{\widehat{\beta},n,t}^{-1}(s/n_t)$ for $s = 1, \ldots, n_t$), respectively. Then we can define

$$\widehat{\Phi}_{i,j,t} = \mathbb{1}\left(F_{\widehat{\beta},n,t}^{-1}((j-1)/J_t) \leq \widehat{\beta}_{it} < F_{\widehat{\beta},n,t}^{-1}(j/J_t)\right). \tag{4.5}$$

In words, $\widehat{\Phi}_{i,j,t}$ is a binary variable which takes on the value of one when asset i is in portfolio j at time t, and zero otherwise. We can stack these binary variables into the $J_t \times n_t$ matrix $\widehat{\Phi}_t$ and obtain

$$\widehat{\mu}_t(\beta) = \widehat{p}_t(\beta)^\top \widehat{a}_t = \sum_{j=1}^{J_t} \widehat{p}_{jt}(\beta) \widehat{a}_{jt}, \tag{4.6}$$

where $\widehat{p}_t(\beta) = (\widehat{p}_{1t}(\beta), \ldots, \widehat{p}_{J_t t}(\beta))^\top$ with $\widehat{p}_{jt}(\beta) = \mathbb{1}(\beta \in \widehat{P}_{jt})$, $\widehat{a}_t = (\widehat{\Phi}_t \widehat{\Phi}_t^\top)^{-1} \widehat{\Phi}_t R_t = (\widehat{a}_{1t}, \cdots, \widehat{a}_{J_t t})^\top$, and $R_t = (R_{1t}, \ldots, R_{n_t t})^\top$. Equation (4.6) shows that $\widehat{\mu}_t(\beta)$ requires two inputs. The first input is which (unique) portfolio the evaluation point, β, resides in. In equation (4.6) this translates to the choice of $j^\star$ which gives $\widehat{p}_{j^\star t}(\beta) = 1$. The second input is the return on the $j^\star$th portfolio. This is simply given by the $j^\star$th element of $\widehat{a}_t$.

In small samples there is always the possibility that some portfolios are empty so that the inverse of $\left(\widehat{\Phi}_t \widehat{\Phi}_t^\top / n_t\right)$ may not exist. However, under our assumptions, the results in the SA show that $\left(\widehat{\Phi}_t \widehat{\Phi}_t^\top / n_t\right)$ exists and is finite with probability approaching one. Throughout, we assume we are on the event that this inverse exists but we suppress this from the notation and main results for

simplicity of exposition. The following theorem characterizes the leading terms of the beta-sorted estimator.

Lemma 4.1 (Leading term linearization)**.** *Suppose Assumptions 1–7 and the conditions in Theorem 3.1 hold. In addition, assume that* $\mathsf{R}_{nT}^2 \log(nT) \to 0$ *and* $J^2 \log(nT)/n \to 0$*. Then,*

$$\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H) = \frac{1}{T-H} \sum_{t=H+1}^{T} \widehat{p}_t(\beta)^\top Q_t^{-1} \frac{1}{n_t} \sum_{i=1}^{n_t} \Phi_{i,t}\varepsilon_{it} \tag{4.7}$$

$$+ \frac{1}{T-H} \sum_{t=H+1}^{T} \widehat{p}_t(\beta)^\top Q_t^{-1} \frac{1}{n_t} \sum_{i=1}^{n_t} \Phi_{i,t}\beta_{it}(f_t - \mathbb{E}[f_t|\mathcal{G}_{t-1}]) \tag{4.8}$$

$$+ \mathscr{B}(\beta) + \mathscr{R}(\beta),$$

where $\Phi_{i,t} = (\Phi_{i,1,t}, \dots, \Phi_{i,J_t,t})^\top$*,* Q_t *is a diagonal matrix with elements* $\{q_{jt} = \mathbb{E}[\Phi_{i,j,t}|\mathcal{G}_{t-1}] : j = 1, \dots, J_t\}$*, the bias satisfies*

$$\mathscr{B}(\beta) = \frac{1}{T-H} \sum_{t=H+1}^{T} \mathscr{B}_t(\beta), \qquad \mathscr{B}_t(\beta) = \widehat{p}_t(\beta)^\top (\widehat{\Phi}_t \widehat{\Phi}_t^\top)^{-1} \sum_{i=1}^{n_t} \widehat{\Phi}_{i,t}\big(\mu_t(\beta_{it}) - \mu_t(\beta)\big) \lesssim_{\mathbb{P}} J^{-1},$$

where $\widehat{\Phi}_{i,t} = (\widehat{\Phi}_{i,1,t}, \dots, \widehat{\Phi}_{i,J_t,t})^\top$*, and the remainder satisfies*

$$\mathscr{R}(\beta) = \begin{cases} o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right) & \text{if } \beta \neq 0 \\ o_{\mathbb{P}}\left(\sqrt{\frac{J}{nT} + \frac{1}{TJ^2}}\right) & \text{if } \beta = 0 \end{cases},$$

with $\mathscr{R}(\beta)$ *defined explicitly in the SA.*

Lemma 4.1 introduces some key properties of the grand mean estimator, $\widehat{\mu}(\beta)$. First, under our assumptions, we may ignore the generated errors of the first-stage estimation of the β_{it} when analyzing the second stage portfolio sorting estimator. Second, the theorem shows that when estimating the SACER the leading term is comprised of two elements: the first term, equation (4.7), would appear in any generic nonparametric problem whereas the second term, equation (4.8), is specific to the asset pricing setup. Importantly, the second term is of the order $O_{\mathbb{P}}(T^{-1/2})$ representing the summation of the product of the conditional beta and the deviation of the factor from its conditional mean, $(f_t - \mathbb{E}[f_t|\mathcal{G}_{t-1}])$. Thus, despite an approximate sample size of nT in concert with a nonparametric procedure with tuning parameter J, the grand mean estimator, for

all values of β except zero, achieves only a $\sqrt{T}$ rate of convergence. That said, for β evaluated at zero, there is a discontinuity and the second term becomes degenerate and the first term dominates leading to a faster rate of convergence, namely, $O_{\mathbb{P}}\left(\sqrt{\frac{J}{nT} + \frac{1}{TJ^2}}\right)$.[5] This is reflected in the remainder term, $\mathscr{R}(\beta)$, in Lemma 4.1. Finally, we see that the bias of the estimator at time t, $\mathscr{B}_t(\beta)$, and also for the grand mean, $\mathscr{B}(\beta)$, are of the order $O_{\mathbb{P}}(J^{-1})$. In words, the sample average across time does not alter the order of the bias since $\widehat{\mu}(\beta)$ is a sample average of nonparametric estimates taken one cross-section at a time.

Remark 4.2. As discussed above, for fixed t we can only consistently estimate $M_t(\beta)$ not $\mu_t(\beta)$. However, we can affect the asymptotic properties by focusing on specific features of the unknown function, $\mu_t(\beta)$. By taking a discrete second derivative we can simplify the asymptotic properties since

$$\mathrm{BFLY}_t(\beta_1, \beta_2, \beta_3; M_t) = M_t(\beta_1) - 2M_t(\beta_2) + M_t(\beta_3) = \mu_t(\beta_1) - 2\mu_t(\beta_2) + \mu_t(\beta_3), \qquad (4.9)$$

whenever $\beta_1 - \beta_2 = \beta_2 - \beta_3$ with three distinct points $\beta_1, \beta_2, \beta_3 \in \mathcal{B}$. This object can be interpreted as a "butterfly" trade where one goes long one unit of each of two assets (one with β_1 and one with β_3) and short two units of an asset (with β_2). By results provided in the SA, we can show that

$$\mathrm{BFLY}_t(\beta_1, \beta_2, \beta_3; \widehat{\mu}_t) - \mathrm{BFLY}_t(\beta_1, \beta_2, \beta_3; \mu_t) = o_{\mathbb{P}}(1), \qquad (4.10)$$

and

$$\frac{1}{T-H} \sum_{t=H+1}^{T} \left(\mathrm{BFLY}_t(\beta_1, \beta_2, \beta_3; \widehat{\mu}_t) - \mathrm{BFLY}_t(\beta_1, \beta_2, \beta_3; \mu_t)\right) = o_{\mathbb{P}}(1). \qquad (4.11)$$

Moreover, both of these estimators achieve the faster rate of convergence, $O_{\mathbb{P}}(\sqrt{J/nT})$. This imposes the additional rate restriction $nT/J^3 \to 0$ which ensures sufficient undersmoothing for the "butterfly" estimator. Under the assumption of the absence of arbitrage opportunities, the expression in equation (4.9) is equal to zero for all β_1, β_2, and β_3 satisfying $\beta_1 - \beta_2 = \beta_2 - \beta_3$. Thus, this statistic can serve as the foundation to conduct inference on whether the no-arbitrage restriction holds (see Section 5.3 below).

---

[5]To see that this rate of convergence is faster, note that $J/n \to 0$ is required to ensure that the probability of an empty portfolio is vanishing asymptotically.

Next we provide a central limit theorem for $\widehat{\mu}(\beta)$ which allows us to conduct inference on the estimator of the SACER. Our result employs a martingale central limit theorem (Hall and Heyde, 2014) combined with the dependence structure introduced earlier.

Theorem 4.3 (Central limit theorem). *Suppose Assumptions 1-7 hold along with the rate restrictions on h and T given in Theorem 3.1. In addition, assume that*

$$\min\{J^{-1}, J^2/n\}T^{-1/2} \max_{1\leq t\leq T} \max_{1\leq i\leq n_t} \Theta_{n,T}((\widehat{p}_\bullet(\beta)^\top Q_\bullet^{-1}\Phi_{i,\bullet}\varepsilon_{i\bullet})^2; q, v) \to 0$$

$$\min\{n/J, J^2\}T^{-1/2}\Theta_{n,T}((\widehat{p}_\bullet(\beta)^\top Q_\bullet^{-1}\mathbb{E}[\Phi_{i,\bullet}\beta_{i\bullet}|\mathcal{G}_{t-1}]f_\bullet)^2; q, v) \to 0$$

for $2v > 1/2 - 1/q$. Then,

$$\frac{\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H) - \mathscr{B}(\beta)}{\sqrt{\mathbb{E}[\sigma_\varepsilon^2(\beta) + \sigma_f^2(\beta)]}} \to_\mathcal{L} \mathsf{N}(0, 1), \tag{4.12}$$

where

$$\sigma_\varepsilon^2(\beta) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \sum_{j=1}^{J_t} n_t^{-2} \sum_{i=1}^{n_t} \widehat{p}_{j,t}(\beta) q_{jt}^{-2}\mathbb{E}(\Phi_{i,j,t}\varepsilon_{it}^2|\mathcal{G}_{t-1}) \asymp_\mathbb{P} \frac{J}{nT},$$

$$\sigma_f^2(\beta) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \sum_{j=1}^{J_t} \widehat{p}_{j,t}(\beta) q_{jt}^{-2}\mathbb{E}(\Phi_{i,j,t}\beta_{it}|\mathcal{G}_{t-1})^2(f_t - \mathbb{E}[f_t|\mathcal{G}_{t-1}])^2 \asymp_\mathbb{P} \begin{cases} \frac{1}{T} & \text{if } \beta \neq 0 \\ \frac{1}{TJ^2} & \text{if } \beta = 0 \end{cases}.$$

Theorem 4.3 characterizes the limiting distribution of the beta-sorted portfolio estimator when centered at the SACER (plus the smoothing bias, $\mathscr{B}(\beta)$). A few remarks are in order. First, despite the differential rates of convergence in the numerator depending on the value of β shown in Lemma 4.1, once properly scaled, the limiting distribution is unaffected. This comes about because the denominator adapts to the appropriate rate of convergence. Second, Theorem 4.3 provides the basis of a feasible inference procedure based on a new plug-in variance estimator, discussed in Section 4.3. Third, we require restrictions on the dependence properties of the data, using our concept of time-series dependence introduced in Section 1.1, to ensure the convergence in distribution holds.

Finally, to provide some intuition for the rate conditions given in Theorem 4.3 we give the following example for $\beta \neq 0$. Let $n = T^{\gamma_1}$, $h = T^{-\gamma_2}$ and $J = T^{\gamma_3}$, for $\gamma_1, \gamma_2, \gamma_3 > 0$, $\gamma_2 < 1$ and

$\gamma_3 < \gamma_1$. If we ignore log terms, the rate conditions in the above theorem require that $2\gamma_1 + q\gamma_2 < q - 2$ and $1 < 2\gamma_3 < \gamma_1$. Thus, even for large γ_1 (e.g., $\gamma_1 = 2$), our rate requirements conform with a choice of bandwidth that produces meaningful undersmoothing in the first step relative to, for example, the MSE optimal rate of $T^{-1/3}$. Intuitively, this comes about because variation in the first-stage estimates is beneficial for estimating $\mu_t(\cdot)$ since β_{it} are the regressors in the second-stage nonparametric estimation problem. As a consequence the trade-off between bias and variance that would arise if the first stage was considered in isolation is no longer the case in the combined procedure: the cost of bias relative to variance rises.

Remark 4.4 (De-biasing the Small Exposure Portfolio). The results presented in Lemma 4.1 and Theorem 4.3 make clear that the portfolio which includes the value $\beta = 0$ is special relative to all other portfolios. The rate of convergence accelerates to $\sqrt{\frac{J}{nT} + \frac{1}{TJ^2}}$ which amplifies the smoothing bias in relative terms. By Theorem 4.3, the rate condition $\sqrt{T}J^{-1} \to 0$ enables sufficient undersmoothing when $\beta \neq 0$. However, when $\beta = 0$, this rate condition is insufficient and the bias of the estimator must be further reduced to employ the asymptotic normality results in Theorem 4.3. In our empirical application in Section 6 we de-bias this portolio by fitting a linear regression of returns on betas within the bin rather than a constant fit.

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 is constructed as

$$\widehat{\sigma}_{\mathsf{FM}}^2(\beta) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \left(\widehat{\mu}_t(\beta) - \widehat{\mu}(\beta) \right)^2. \tag{4.13}$$

Thus, $\widehat{\mu}_t(\beta)$ for $t = H+1, \ldots, T$ serves as the sample "observations" and $\widehat{\mu}(\beta)$ serves as the sample "mean."

The second feasible inference procedure is based on a new plug-in (PI) variance estimator which can be constructed using the results in Lemma 4.1 and Theorem 4.3. To see the logic of our

approach, first consider the (unrealistic) case where we observe ε_{it} and $\mathbb{E}[f_t|\mathcal{G}_{t-1}]$. We can exploit the fact that the two leading terms shown in Theorem 4.1 are uncorrelated and so the natural plug-in variance estimator is

$$\widetilde{\sigma}^2_{\texttt{PI}}(\beta) = \widetilde{\sigma}^2_{f,\texttt{PI}}(\beta) + \widetilde{\sigma}^2_{\varepsilon,\texttt{PI}}(\beta),$$

where

$$\widetilde{\sigma}^2_{f,\texttt{PI}}(\beta) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \sum_{j=1}^{J_t} \frac{1}{n_t^2 \widehat{q}_{jt}^2} \left(\sum_{i=1}^{n_t} \widehat{p}_{j,t}(\beta)\widehat{p}_{j,t}(\widehat{\beta}_{it})\widehat{\beta}_{it} \right)^2 (f_t - \mathbb{E}[f_t|\mathcal{G}_{t-1}])^2,$$

$$\widetilde{\sigma}^2_{\varepsilon,\texttt{PI}}(\beta) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \sum_{j=1}^{J_t} \frac{1}{n_t^2 \widehat{q}_{jt}^2} \sum_{i=1}^{n_t} \widehat{p}_{j,t}(\beta)\widehat{p}_{j,t}(\widehat{\beta}_{it})\varepsilon_{it}^2,$$

and $\widehat{q}_{jt} = n_t^{-1} \sum_{i=1}^{n_t} \widehat{\Phi}_{i,j,t}$. Of course, $\widetilde{\sigma}^2_{\texttt{PI}}(\beta)$ is infeasible. As a feasible alternative, consider

$$\widehat{\sigma}^2_{\texttt{PI}}(\beta) = \widehat{\sigma}^2_{f,\texttt{PI}}(\beta) + \widehat{\sigma}^2_{\varepsilon,\texttt{PI}}(\beta), \tag{4.14}$$

where

$$\widehat{\sigma}^2_{f,\texttt{PI}}(\beta) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \sum_{j=1}^{J_t} \frac{1}{n_t^2 \widehat{q}_{jt}^2} \left(\sum_{i=1}^{n_t} \widehat{p}_{j,t}(\beta)\widehat{p}_{j,t}(\widehat{\beta}_{it})\widehat{\beta}_{it} \right)^2 (f_t - \widehat{\mathbb{E}[f_t|\mathcal{G}_{t-1}]})^2 \tag{4.15}$$

$$\widehat{\sigma}^2_{\varepsilon,\texttt{PI}}(\beta) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \sum_{j=1}^{J_t} \frac{1}{n_t^2 \widehat{q}_{jt}^2} \sum_{i=1}^{n_t} \widehat{p}_{j,t}(\beta)\widehat{p}_{j,t}(\widehat{\beta}_{it})\widehat{\varepsilon}_{it}^2. \tag{4.16}$$

Here $\widehat{\varepsilon}_{it} = R_{it} - \widehat{\mu}_t(\beta_{it})$ and $\widehat{\mathbb{E}[f_t|\mathcal{G}_{t-1}]}$ is a feasible estimator 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 $\widehat{\mathbb{E}[f_t|\mathcal{G}_{t-1}]} = h_{t-1}(\widehat{\vartheta})$ 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. The estimator, $\widehat{\mathbb{E}[f_t|\mathcal{G}_{t-1}]}$, may not be correctly specified but we will demonstrate in Theorem 4.3 below that we can still conduct valid, albeit possibly conservative, inference.

Finally, it is important to emphasize that, just as for the beta-sorted portfolio estimator itself (see Section 2), both the FM and PI variance estimators require an accounting, at each time t, of

which portfolio β resides in.

Before stating the pertinent properties of the two variance estimators, it will be useful to define the following object

$$\sigma_\mu^2(\beta) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \left(\mu_t(\beta) - \bar{\mu}_T(\beta; H)\right)^2, \tag{4.17}$$

which is the sample variance of the SACER. In addition, define the following (infeasible) version of the FM variance estimator where the individual elements have had their smoothing bias removed,

$$\widetilde{\sigma}_{\texttt{FM}}^2(\beta) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \left(\widehat{\mu}_t(\beta) - \mathscr{B}_t(\beta) - \left(\widehat{\mu}(\beta) - \mathscr{B}(\beta)\right)\right)^2. \tag{4.18}$$

We then have the following result.

Theorem 4.5 (Variance estimation). *Suppose Assumptions 1–7 hold along with the rate restrictions on h and T given in Theorem 4.3. Then,*

(i) if $\beta \neq 0$

$$\left|\widehat{\sigma}_{\texttt{FM}}^2(\beta) - \sigma_\varepsilon^2(\beta) - \sigma_f^2(\beta) - \sigma_\mu^2(\beta)\right| = o_{\mathbb{P}}\left(\frac{1}{T}\right),$$

and, if $\mathbb{E}[\sigma_\mu^2(\beta)] = o\left(\frac{J}{nT} + \frac{1}{TJ^2}\right)$, then for $\beta = 0$

$$\left|\widetilde{\sigma}_{\texttt{FM}}^2(\beta) - \sigma_\varepsilon^2(\beta) - \sigma_f^2(\beta) - \sigma_\mu^2(\beta)\right| = o_{\mathbb{P}}\left(\frac{J}{Tn} + \frac{1}{TJ^2}\right);$$

(ii) Assume $\max_{H+1 \leq t \leq T} \mathbb{E}\left[(\mathbb{E}[f_t|\mathcal{G}_{t-1}] - \mathbb{E}[\widehat{f_t|\mathcal{G}_{t-1}}])^2\right]$ is $o(T)$ if $\beta \neq 0$ and $o\left(\frac{J^2T}{n^2} + \frac{T}{J^4}\right)$ if $\beta = 0$. Then, for $\beta \in \mathcal{B}$,

$$\left|\widehat{\sigma}_{\varepsilon,\texttt{PI}}^2(\beta) - \sigma_\varepsilon^2(\beta)\right| + \left|\widehat{\sigma}_{f,\texttt{PI}}^2(\beta) - \sigma_f^2(\beta) - s_{nT}(\beta)\right| = \begin{cases} o_{\mathbb{P}}\left(\frac{1}{T}\right) & \text{if } \beta \neq 0 \\ o_{\mathbb{P}}\left(\frac{J}{nT} + \frac{1}{TJ^2}\right) & \text{if } \beta = 0 \end{cases},$$

where $s_{nT}(\beta) \geq 0$ is defined in Section SA-2.5 of the SA.

Theorem 4.5 characterizes the (differential) asymptotic properties of the PI and the FM variance estimators. The intuition behind these results is specific to each of the variance estimators. For the PI variance estimator, consistency can be achieved when the specification of $\mathbb{E}[\widehat{f_t|\mathcal{G}_{t-1}}]$ is correct. However, even if this estimator is misspecified $\widehat{\sigma}_{\texttt{PI}}(\beta)$ will be biased upward only. This

conservativeness arises by the standard minimum mean-square error property of the conditional expectation with s_{NT} representing the contribution arising from the misspecification. For the FM variance estimator, we can understand the result from the following decomposition of the summands of the estimator:

$$\widehat{\mu}_t(\beta) - \frac{1}{T}\sum_{t=1}^{T}\widehat{\mu}_t(\beta) = (\widehat{\mu}_t(\beta) - \mu_t(\beta)) + [\mu_t(\beta) - \frac{1}{T}\sum_{t=1}^{T}\mu_t(\beta)] - \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\mu}_t(\beta) - \mu_t(\beta)\right).$$

It is the first term that plays the role of capturing the contribution of the variance from $\sigma_f^2(\cdot)+\sigma_\varepsilon^2(\cdot)$, while the second term captures the contribution of $\sigma_\mu^2(\cdot)$. The third term, instead, is asymptotically negligible. When $\beta = 0$, in order to use the FM variance estimator, the bias from each $\widehat{\mu}_t(\beta)$ must be removed. This is why we replace $\widehat{\sigma}_{\text{FM}}^2(\beta)$ by $\widetilde{\sigma}_{\text{FM}}^2(\beta)$. As discussed in Remark 4.4, in practice for $\widehat{\sigma}_{\text{FM}}^2(\beta)$ we de-bias the portfolio which contains $\beta = 0$, so the associated FM variance estimator will inherit this adjustment.

We can use the results of Theorem 4.5 as the basis for our feasible inference procedures.

Corollary 4.5.1. *Let the Assumptions of Theorem 4.5 hold. Define* $\widehat{L}_T(\beta,\widehat{\sigma}) = \widehat{\mu}(\beta) - \widehat{\sigma}(\beta)\mathsf{q}_{1-\alpha}^{\mathsf{N}}/\sqrt{T}$, *and* $\widehat{U}_T(\beta,\widehat{\sigma}) = \widehat{\mu}(\beta) + \widehat{\sigma}(\beta)\mathsf{q}_{1-\alpha}^{\mathsf{N}}/\sqrt{T}$ *where* $\mathsf{q}_{1-\alpha}^{\mathsf{N}}$ *is the* $1 - \alpha$ *quantile of a standard Gaussian variable. Then, for fixed* $\beta \in \mathcal{B}$ *and a pre-specified nominal level* 2α,

$$\liminf_{N,T\to\infty} \mathbb{P}\Big(\bar{\mu}_T(\beta;H) \in [\widehat{L}_T(\beta,\widehat{\sigma}_{\text{PI}}), \widehat{U}_T(\beta,\widehat{\sigma}_{\text{PI}})]\Big) \geq 1 - 2\alpha, \tag{4.19}$$

$$\liminf_{N,T\to\infty} \mathbb{P}\Big(\bar{\mu}_T(\beta;H) \in [\widehat{L}_T(\beta,\widehat{\sigma}_{\text{FM}}), \widehat{U}_T(\beta,\widehat{\sigma}_{\text{FM}})]\Big) \geq 1 - 2\alpha. \tag{4.20}$$

Since the SACER is a random object, we can use the results in Corollary 4.5.1 to form prediction intervals for $\bar{\mu}_T(\beta;H)$. With either choice of variance estimator, inference is asymptotically valid but may be conservative. When using the PI variance estimator we obtain asymptotic coverage of exactly $1 - 2\alpha$ when the conditional expectation is specified correctly, whereas for the FM variance estimator this occurs when $\mu_t(\beta)$ does not vary over time.

Remark 4.6. Theorem 4.5 also implies that we can construct an asymptotic lower bound for the sample variance of conditional expected returns, $\sigma_\mu^2(\beta)$, by taking the difference between $\widehat{\sigma}_{\text{FM}}^2(\beta)$

and $\widehat{\sigma}^2_{\texttt{PI}}(\beta)$. Under the assumptions in Theorem 4.5, we have that for $\beta \in \mathcal{B}$,

$$\widehat{\sigma}^2_{\texttt{FM}}(\beta) - \widehat{\sigma}^2_{\texttt{PI}}(\beta) = \sigma^2_\mu(\beta) - s_{nT}(\beta) + o_\mathbb{P}\Big(\frac{1}{T}\Big). \tag{4.21}$$

Thus, through the medium of the two alternative variance estimators we can learn about other features of conditional expected returns. Although this only provides an estimate of a lower bound on $\sigma^2_\mu(\beta)$, in our empirical application, we find that this difference is relatively large and so this bound can be informative.

Remark 4.7. Theorem 4.5 also implies that valid inference on the SACER can be achieved without the need to stipulate the form of the conditional expectation of the risk factors, $\mathbb{E}[f_t|\mathcal{G}_{t-1}]$. This stands in contrast to existing approaches, for example, Adrian et al. (2015) and Gagliardini et al. (2016), where a (first-order) Markovian structure is imposed. In practice, specifying the correct functional form including the appropriate conditioning variables for the risk factor dynamics is a challenge. This is an advantage of the estimation approach we study.

4.4 Feasible inference on the PACER

Although we have argued that the SACER should be the preferred estimand for the beta-sorted portfolio estimator, it is illuminating to compare the conditions required to obtain results for this alternative estimand. First note that the beta-sorted portfolio estimator is consistent for the PACER so long as Assumption P holds (i.e., the PACER exists):

$$\widehat{\mu}(\beta) - \mu(\beta) = \big(\widehat{\mu}(\beta) - \bar{\mu}_T(\beta; H)\big) + \big(\bar{\mu}_T(\beta; H) - \mu(\beta)\big). \tag{4.22}$$

The first term is $o_\mathbb{P}(1)$ under the assumptions given in Section 4 and the second term is $o_\mathbb{P}(1)$ if Assumption P also holds.

Although consistency is ensured with only a mild strengthening of the assumptions necessary for our results for the SACER, asymptotic normality requires substantially more structure on the properties of the data. This is because, for the PACER, the martingale difference property no long holds and so our previous approach to obtaining asymptotic results cannot be used. By equation (4.22) it follows that the variance when estimating the SACER is smaller (by the magnitude $\sigma^2_\mu(\beta)$;

see Theorem 4.3) than when estimating the PACER with equality only when $\mu_t(\beta)$ is constant over time. Furthermore, by Theorem 4.5 the FM variance estimator will consistently estimate this larger variance under appropriate regularity conditions.

Establishing asymptotic theory and feasible inference procedures for the PACER is challenging because of the highly nonlinear time series dependence and complicated nonparametric structure of the beta-sorted portfolio estimator, and therefore we leave this task for future work.

5   Joint Inference

In our main results we have laid a foundational framework and associated assumptions which ensure consistency, asymptotic normality, and feasible inference for the beta-sorted portfolios estimator of the SACER. These results immediately imply valid procedures for joint inference over finitely many values of β. In this section we discuss how to conduct joint inference for the SACER over multiple beta values along with how to formally test for the presence of profitable high-low and "butterfly" trading strategies.

In order to construct joint intervals for multiple values of β we require estimators of the covariance of $\widehat{\mu}(\beta_1)$ and $\widehat{\mu}(\beta_2)$ for $\beta_1, \beta_2 \in \mathcal{B}$. For the FM variance estimator we have,

$$\widehat{\sigma}_{\text{FM}}(\beta_1, \beta_2) = \frac{1}{(T-H)^2} \sum_{t=H+1}^{T} \left(\widehat{\mu}_t(\beta_1) - \widehat{\mu}(\beta_1)\right)\left(\widehat{\mu}_t(\beta_2) - \widehat{\mu}(\beta_2)\right),$$

whereas for the PI variance estimator we can use,

$\widehat{\sigma}_{\text{PI}}(\beta_1, \beta_2)$

$$= \frac{1}{T-H} \sum_{t=H+1}^{T} \sum_{j_1=1}^{J_t} \sum_{j_2=1}^{J_t} n_t^{-2} \widehat{q}_{j_1 t}^{-1} \widehat{q}_{j_2 t}^{-1} \left(\sum_{i=1}^{n_t} \widehat{p}_{j_1,t}(\beta_1) \widehat{p}_{j_2,t}(\beta_2) \widehat{p}_{j_1,t}(\widehat{\beta}_{it}) \widehat{p}_{j_2,t}(\widehat{\beta}_{it}) \widehat{\beta}_{it}^2 \right) (f_t - \mathbb{E}(\widehat{f_t | \mathcal{G}_{t-1}}))^2$$

$$+ \frac{1}{T-H} \sum_{t=H+1}^{T} \sum_{j_1=1}^{J_t} \sum_{j_2=1}^{J_t} n_t^{-2} \widehat{q}_{j_1 t}^{-1} \widehat{q}_{j_2 t}^{-1} \sum_{i=1}^{n_t} \widehat{p}_{j_1,t}(\beta_1) \widehat{p}_{j_2,t}(\beta_2) \widehat{p}_{j_1,t}(\widehat{\beta}_{it}) \widehat{p}_{j_2,t}(\widehat{\beta}_{it}) \widehat{\varepsilon}_{it}^2.$$

The results we present are valid for either the FM or PI approaches and so for simplicity we present the results with the generic notation $\widehat{\sigma}(\beta_1, \beta_2)$ as a stand-in for either estimator with $\widehat{\sigma}(\beta, \beta) = \widehat{\sigma}^2(\beta)$.

Finally, the following notation will be used throughout this section. Let $\mathcal{B}_G = \{b_1, b_2, \ldots, b_G\} \subset \mathcal{B}$ be a finite collection of G evaluation points of β. Our theoretical results imply the existence of a G-variate Gaussian limiting distribution with the covariance matrix, Σ_G, following from Theorem 4.3. More precisely, let $\widehat{\mu}_G = (\widehat{\mu}(b_1), \ldots, \widehat{\mu}(b_G))^\top$ and $\mu_G = (\mu(b_1), \ldots, \mu(b_G))^\top$, so that $(\widehat{\mu}_G - \mu_G) \sim_a \Sigma_G^{1/2} Z$ where $Z \sim \mathcal{N}(0, I_G)$.

5.1 Joint inference for the SACER

A valid joint prediction interval for SACER can be constructed via $[\widehat{L}_T(\beta), \widehat{U}_T(\beta)]$ where $\widehat{L}_T(\beta) = \widehat{\mu}(\beta) - \widehat{\sigma}(\beta)\mathsf{q}_{1-\alpha}/\sqrt{T}$, and $\widehat{U}_T(\beta) = \widehat{\mu}(\beta) + \widehat{\sigma}(\beta)\mathsf{q}_{1-\alpha}/\sqrt{T}$, and $\mathsf{q}_{1-\alpha}$ is obtained as the $1-\alpha$ quantile of the distribution of $\left|\Sigma_{G,\text{corr}}^{1/2} Z\right|_\infty$ where $\Sigma_{G,\text{corr}} = \text{diag}(\Sigma_G)^{-1/2}\Sigma_G\text{diag}(\Sigma_G)^{-1/2}$ and $\text{diag}(A)$ denotes a diagonal matrix with diagonal entries the same as the diagonal entries of A. In practice, we replace Σ_G with $\widehat{\Sigma}_G$ using either the FM or PI covariance estimator introduced above. Then, with a pre-specified coverage level $1 > 2\alpha > 0$,

$$\liminf_{N,T\to\infty} \mathbb{P}\Big(\bar{\mu}_T(\beta; H) \in [\widehat{L}_T(\beta), \widehat{U}_T(\beta)], \text{ for all } \beta \in \mathcal{B}_G\Big) \geq 1 - 2\alpha.$$

Intuitively, if 0 is not contained in this joint prediction interval then we can reject the null that there exists values of $\beta \in \mathcal{G}$ which do not earn (expected) returns for the exposure to the factor. We provide an algorithm to implement the joint inference procedure below.

Algorithm 1 Joint inference for the SACER.

Require: $n_t, T \geq 0$

1: Obtain (i, k) elements of $\widehat{\Sigma}_G$ as $\widehat{\sigma}(b_i, b_k)$ using the G grid points in $\mathcal{B}_G$.

2: Simulate standard normal random variables $Z^{(s)}$ of $G \times 1$ dimension for $s = 1, \cdots, S$ times, where S is the number of draws.

3: Construct $\widetilde{Z}^{(s)} = \widehat{\Sigma}_{G,\text{corr}}^{1/2} Z^{(s)}$, where $\widehat{\Sigma}_{G,\text{corr}} = \text{diag}(\widehat{\Sigma}_G)^{-1/2}\widehat{\Sigma}_G \text{diag}(\widehat{\Sigma}_G)^{-1/2}$ and $\widehat{\Sigma}_{G,\text{corr}}^{1/2}$ is the symmetric square root of $\widehat{\Sigma}_{G,\text{corr}}$.

4: Obtain the $1 - \alpha$ quantile of $|\widetilde{Z}|_\infty$ from the S simulated draws and denote as $\widehat{\mathsf{q}}_{1-\alpha}$.

5: Create the joint prediction interval by $[\widehat{L}_T(b_i), \widehat{U}_T(b_i)]$, where $\widehat{L}_T(b_i) = \widehat{\mu}(b_i) - \widehat{\sigma}(b_i)\widehat{\mathsf{q}}_{1-\alpha}/\sqrt{T}$ and $\widehat{U}_T(b_i) = \widehat{\mu}(b_i) - \widehat{\sigma}(b_i)\widehat{\mathsf{q}}_{1-\alpha}/\sqrt{T}$ for $i = 1, \ldots, G$.

5.2 The generalized high-minus-low strategy

The most popular object of interest in the empirical finance literature is to compare the time-average of returns from the two extreme portfolios (i.e., the portfolios which encompass the evaluation points β_l and β_u) as discussed in Section 2. The goal is to assess whether a long-short portfolio trading strategy earns statistically significant returns, i.e., has a nonzero unconditional risk premium. However, we can use our general framework and new theoretical results to formulate a more effective inference procedure to assess the properties of expected returns. Rather than focus on $|\mu(\beta_u) - \mu(\beta_l)|$ we instead consider $\max_{\beta \in \mathcal{B}_G} \mu(\beta) - \min_{\beta \in \mathcal{B}_G} \mu(\beta)$. In words, we study the most profitable long-short strategy available. We refer to this as the generalized high-minus low strategy. In the special case when $\mu(\beta)$ is monotonic (and the grid $\mathcal{B}_G$ includes β_l and β_u), then the two expressions are equivalent. Thus, we nest the popular high minus low portfolio inference approach as we test for the presence of *any* profitable long-short strategy in $\mathcal{B}_G$.

This class of high-minus-low statistics can be re-expressed in the following form,

$$
\max_{\beta \in \mathcal{B}_G} \widehat{\mu}(\beta) - \min_{\beta \in \mathcal{B}_G} \widehat{\mu}(\beta) = \max_{\beta \in \mathcal{B}_G} \widehat{\mu}(\beta) + \max_{\beta \in \mathcal{B}_G} \big(-\widehat{\mu}(\beta) \big)
$$
$$
= \max_{\beta \in \mathcal{B}_G} \Big(\frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} \widehat{p}_{j,t}(\beta) \widehat{a}_{jt} \Big) + \max_{\beta \in \mathcal{B}_G} \Big(-\frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} \widehat{p}_{j,t}(\beta) \widehat{a}_{jt} \Big),
$$

where we denote $\beta_{\max}$ as the point attaining $\max_{\beta \in \mathcal{B}_G}(\frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} \widehat{p}_{j,t}(\beta) \widehat{a}_{jt})$ and $\beta_{\min}$ as the point attaining $\max_{\beta \in \mathcal{B}_G}(-\frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} \widehat{p}_{j,t}(\beta) \widehat{a}_{jt})$. Algorithm 2 below shows how to construct a joint prediction interval $[\widehat{L}_T, \widehat{U}_T]$. Intuitively, if 0 is outside of this interval then there exists at least one profitable long-short strategy across a pair of values in $\mathcal{B}_G$.

5.3 The "butterfly" strategy

We introduce one final joint inference procedure which is motivated by a "butterfly" trading strategy corresponding to a discrete second derivative (see Remark 4.2; the long-short trading strategy of the previous subsection can be thought of as a discrete first derivative). Furthermore, under the assumption of the absence of arbitrage opportunities, then the "butterfly" trading strategy we formulate should have zero (conditional) expected returns. Thus, we can also directly conduct inference on whether the data are consistent with a no-arbitrage assumption. In words, under

Algorithm 2 Inference for the generalized high-minus-low strategy.

Require: $n_t, T \geq 0$

1: Obtain (i,k) elements of $\widehat{\Sigma}_G$ as $\widehat{\sigma}(b_i, b_k)$ using the G grid points in $\mathcal{B}_G$.

2: Simulate standard normal random variables $Z^{(s)}$ of $G \times 1$ dimension for $s = 1, \cdots, S$ times, where S is the number of draws.

3: Obtain $\widehat{\mu}(\beta_{\max}) = \max_{\beta \in \mathcal{B}_G} T^{-1} \sum_{t=1}^T \widehat{\mu}_t(\beta)$ and $\widehat{\mu}(\beta_{\min}) = \min_{\beta \in \mathcal{B}_G} T^{-1} \sum_{t=1}^T \widehat{\mu}_t(\beta)$. Let $\widehat{\mu}_{\mathrm{GHL}} = \widehat{\mu}(\beta_{\max}) - \widehat{\mu}(\beta_{\min})$.

4: Construct $\widetilde{Z}^{(s)} = \widehat{\Sigma}_G^{1/2} Z^{(s)}$, where $\widehat{\Sigma}_G^{1/2}$ is the symmetric square root of $\widehat{\Sigma}_G$, and obtain $\widetilde{Z}_{\beta_{\max}}^{(s)} - \widetilde{Z}_{\beta_{\min}}^{(s)}$ using the evaluation points corresponding to $\beta_{\max}$ and $\beta_{\min}$.

5: Obtain the $1 - \alpha$ quantile of $\left| \widetilde{Z}_{\beta_{\max}}^{(s)} - \widetilde{Z}_{\beta_{\min}}^{(s)} \right|$ from the simulated draws, and denote as $\widehat{\mathsf{q}}_{1-\alpha}$. (note: $\widetilde{Z}_{\beta_{\min}}^{(s)}, \widetilde{Z}_{\beta_{\max}}^{(s)}$ are the Gaussian limits corresponding to $\beta_{\max}$ and β_{min}, respectively.)

6: Create the prediction interval $[\widehat{L}_T, \widehat{U}_T]$, where $\widehat{L}_T = \widehat{\mu}_{\mathrm{GHL}} - \widehat{\mathsf{q}}_{1-\alpha}/\sqrt{T}$ and $\widehat{U}_T = \widehat{\mu}_{\mathrm{GHL}} + \widehat{\mathsf{q}}_{1-\alpha}/\sqrt{T}$.

---

no-arbitrage, there does not exist an ex-ante profitable trade where one goes long one unit of each of two assets (one with β_1 and one with β_3) and short two units of an asset (with β_2). Naturally, we will focus on the estimator:

$$\max_{\beta_1 + \beta_3 = 2\beta_2} \mathrm{BFLY}(\beta_1, \beta_2, \beta_3; \widehat{\mu}) = \max_{\beta_1 + \beta_3 = 2\beta_2} \frac{1}{T} \sum_{t=1}^T \mathrm{BFLY}_t(\beta_1, \beta_2, \beta_3; \widehat{\mu}_t).$$

To construct asymptotically valid prediction intervals we can utilize the results of Lemma 4.1 which yields the following leading term expansion,

$$\mathrm{BFLY}(\beta_1, \beta_2, \beta_3; \widehat{\mu}) - \mathrm{BFLY}(\beta_1, \beta_2, \beta_3; \mu)$$
$$= \frac{1}{T} \sum_{t=1}^T \frac{1}{n_t} \sum_{i=1}^{n_t} \sum_{j_t=1}^{J_t} \mathrm{BFLY}_t(\beta_1, \beta_2, \beta_3; \widehat{p}_{jt}) q_{jt}^{-1} \Phi_{i,j,t} \varepsilon_{it} + O_{\mathbb{P}}(J^{-1}).$$

As we discussed in Remark 4.2, the "butterfly" estimator enjoys a better rate of convergence for all values of β, namely, $\sqrt{nT}/\sqrt{J}$ and so we require an alternative variance estimator to form prediction intervals. The asymptotic variance of $T^{-1} \sum_{t=1}^T \sqrt{Tn_t/J_t} \cdot \mathrm{BFLY}_t(\beta_1, \beta_2, \beta_3; \widehat{\mu})$ may be

estimated by

$$\widehat{\sigma}^2_{\text{DiD}}(\beta_1, \beta_2, \beta_3) = \frac{1}{T} \sum_{t=1}^{T} \widehat{\sigma}^2_{\text{DiD},t}(\beta_1, \beta_2, \beta_3),$$

where

$$\widehat{\sigma}^2_{\text{DiD},t}(\beta_1, \beta_2, \beta_3) = \frac{1}{n_t J_t} \sum_{i=1}^{n_t} \sum_{j=1}^{J_t} \text{BFLY}_t(\beta_1, \beta_2, \beta_3; \widehat{p}_{jt})^2 \widehat{q}_{jt}^{-2} \widehat{\Phi}_{i,j,t} \widehat{\varepsilon}^2_{it}.$$

Let $\beta_{1,2,3}$ be an abbreviation for $\beta_1, \beta_2, \beta_3$ and $\beta'_{1,2,3}(\neq \beta_{1,2,3})$ as an abbreviation for $\beta'_1, \beta'_2, \beta'_3$. We define the corresponding covariance estimator as

$$\widehat{\sigma}_{\text{DiD}}(\beta_{1,2,3}, \beta'_{1,2,3}) = \frac{1}{T} \sum_{t=1}^{T} \widehat{\sigma}_{\text{DiD},t}(\beta_{1,2,3}, \beta'_{1,2,3}),$$

where

$$\widehat{\sigma}_{\text{DiD},t}(\beta_{1,2,3}, \beta'_{1,2,3}) = \frac{1}{n_t J_t} \sum_{i=1}^{n_t} \sum_{j=1}^{J_t} \text{BFLY}_t(\beta_1, \beta_2, \beta_3; \widehat{p}_{jt}) \text{BFLY}_t(\beta'_1, \beta'_2, \beta'_3; \widehat{p}_{jt}) \widehat{q}_{jt}^{-2} \widehat{\Phi}_{i,j,t} \widehat{\varepsilon}^2_{it}.$$

For an equal-spaced choice of gridpoints, $\mathcal{B}_G$, we can then collect all ordered triplets (b_1, b_2, b_3) which satisfy $2b_2 = b_1 + b_3$ ($b_1 \neq b_3$). Then, we can construct the $G^\star \times G^\star$ estimated variance-covariance matrix $\widehat{\Sigma}_{\text{DiD}}$ by using $\widehat{\sigma}^2_{\text{DiD}}(b_1, b_2, b_3)$ for the diagonal elements and $\widehat{\sigma}_{\text{DiD}}(b_{1,2,3}, b'_{1,2,3})$ for the off-diagonal elements.[6]

Using Algorithm 3 below we can construct the appropriate prediction interval. Intuitively, if zero is outside the prediction interval, then there exists at least one "butterfly" strategy based on $\mathcal{B}_G$ which earns nonzero expected returns. Furthermore, if zero is outside of the interval then we cannot reject the hypothesis that the no-arbitrage condition holds in the data.

Remark 5.1. The grand mean allows for inference on *unconditional* risk premia but we would also like to accommodate inference on *conditional* risk premia. For example, a risk factor may be associated with a significant risk premium only in certain time periods. Conversely, the conditional risk premium may be zero under some conditions but not unconditionally. Drawing inferences about

---

[6]When G is even, then $G^\star = G(G-2)/4$ and when G is odd, then $G^\star = (G-1)^2/4$.

Algorithm 3 Inference for the "butterfly" strategy.

Require: $n_t, T \geq 0$

1: Choose an equally spaced grid, $\mathcal{B}_G$. Obtain $\widehat{\Sigma}_{\mathrm{DiD}}$ using $\widehat{\sigma}^2_{\mathrm{DiD}}(b_1, b_2, b_3)$ and $\widehat{\sigma}_{\mathrm{DiD}}(b_{1,2,3}, b'_{1,2,3})$.

2: Simulate standard normal random variables $Z^{(s)}$ of $G^\star \times 1$ dimension for $s = 1, \cdots, S$ times, where S is the number of draws.

3: Obtain $\widehat{\mu}(b_i)$ for all G gridpoints $b_i \in \mathcal{B}_G$. Find $\beta^{\max}_{1,2,3}$ which satisfies $\beta^{\max}_{1,2,3} = \max_{b_1, b_2, b_3 \in \mathcal{B}_G, 2b_2 = b_1 + b_3} \mathrm{BFLY}(b_1, b_2, b_3; \widehat{\mu})$.

4: Construct $\widetilde{Z}^{(s)} = \widehat{\Sigma}^{1/2}_{\mathrm{DiD,corr}} Z^{(s)}$, where $\widehat{\Sigma}_{\mathrm{DiD,corr}} = \mathrm{diag}(\widehat{\Sigma}_{\mathrm{DiD}})^{-1/2} \widehat{\Sigma}_{\mathrm{DiD}} \, \mathrm{diag}(\widehat{\Sigma}_{\mathrm{DiD}})^{-1/2}$ and $\widehat{\Sigma}^{1/2}_{\mathrm{DiD,corr}}$ is the symmetric square root of $\widehat{\Sigma}_{\mathrm{DiD,corr}}$.

5: Obtain the $1 - \alpha$ quantile of $|\widetilde{Z}|_\infty$ from the S simulated draws and denote as $\widehat{\mathsf{q}}_{1-\alpha}$.

6: Create the prediction interval $[\widehat{L}_T, \widehat{U}_T]$, where $\widehat{L}_T = \mathrm{BFLY}(\beta^{\max}_1, \beta^{\max}_2, \beta^{\max}_3; \widehat{\mu}) - \widehat{\sigma}_{\mathrm{DiD}}(\beta^{\max}_1, \beta^{\max}_2, \beta^{\max}_3) \widehat{\mathsf{q}}_{1-\alpha} T^{-3/2} \sum_{t=1}^{T} \sqrt{J_t/n_t}$ and $\widehat{U}_T = \mathrm{BFLY}(\beta^{\max}_1, \beta^{\max}_2, \beta^{\max}_3; \widehat{\mu}) + \widehat{\sigma}_{\mathrm{DiD}}(\beta^{\max}_1, \beta^{\max}_2, \beta^{\max}_3) \widehat{\mathsf{q}}_{1-\alpha} T^{-3/2} \sum_{t=1}^{T} \sqrt{J_t/n_t}$.

---

conditional risk premia can provide additional information to understand the economic mechanisms underpinning the risk-return trade-off. Without assuming a functional form for the conditional mean of the risk factors, there are limitations about what we can infer about $\mu_t(\beta)$. But we can characterize some features of conditional expected returns at time t.

For a given time period t, consider $\max_{\beta_1 + \beta_3 = 2\beta_2} \mathrm{BFLY}_t(\beta_1, \beta_2, \beta_3; \widehat{\mu}_t)$. We can interpret the prediction interval for this object in two ways. First, if zero is outside of the interval then we find evidence in favor of arbitrage opportunities at time t. Second, if zero is outside of the interval, then we find evidence that the risk factor has nonzero conditional expected returns at time t, even though we cannot consistently estimate $\widehat{\mu}_t(\beta)$. We can follow similar steps as in Algorithm 3 to construct the associated time t prediction interval (see Algorithm 4 below).

6   Empirical Application

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

Our risk factor is a new measure of the business credit cycle. The business credit cycle is commonly evaluated by means of ratios of credit aggregates to measures of output. Although

---
Algorithm 4 Inference for the fixed-t "butterfly" strategy.

Require: $n_t, T \geq 0$

1: Choose an equally spaced grid, $\mathcal{B}_G$. Obtain $\widehat{\Sigma}_{\mathrm{DiD},t}$ using $\widehat{\sigma}^2_{\mathrm{DiD},t}(b_1, b_2, b_3)$ and $\widehat{\sigma}_{\mathrm{DiD},t}(b_{1,2,3}, b'_{1,2,3})$ using the G gridpoints in $\mathcal{B}_G$.

2: Simulate standard normal random variables $Z^{(s)}$ of $G^\star \times 1$ dimension for $s = 1, \cdots, S$ times, where S is the number of draws.

3: Obtain $\widehat{\mu}_t(b_i)$ for all G gridpoints $b_i \in \mathcal{B}_G$. Find $\beta^{\max,t}_{1,2,3}$ which satisfies $\beta^{\max,t}_{1,2,3} = \max_{b_1, b_2, b_3 \in \mathcal{B}_G, 2b_2 = b_1 + b_3} \mathrm{BFLY}_t(b_1, b_2, b_3; \widehat{\mu}_t)$.

4: Construct $\widetilde{Z}^{(s)} = \widehat{\Sigma}^{1/2}_{\mathrm{DiD},t,\mathrm{corr}} Z^{(s)}$, where $\widehat{\Sigma}_{\mathrm{DiD},t,\mathrm{corr}} = \mathrm{diag}(\widehat{\Sigma}_{\mathrm{DiD},t})^{-1/2} \widehat{\Sigma}_{\mathrm{DiD},t} \, \mathrm{diag}(\widehat{\Sigma}_{\mathrm{DiD},t})^{-1/2}$ and $\widehat{\Sigma}^{1/2}_{\mathrm{DiD},t,\mathrm{corr}}$ is the symmetric square root of $\widehat{\Sigma}_{\mathrm{DiD},t,\mathrm{corr}}$.

5: Obtain the $1 - \alpha$ quantile of $|\widetilde{Z}|_\infty$ from the S simulated draws and denote as $\widehat{\mathsf{q}}_{1-\alpha}$.

6: Create the prediction interval $[\widehat{L}_t, \widehat{U}_t]$, where $\widehat{L}_t = \mathrm{BFLY}_t(\beta^{\max,t}_1, \beta^{\max,t}_2, \beta^{\max,t}_3; \widehat{\mu}_t) - \widehat{\sigma}_{\mathrm{DiD},t}(\beta^{\max,t}_1, \beta^{\max,t}_2, \beta^{\max,t}_3)\widehat{\mathsf{q}}_{1-\alpha}\sqrt{J_t/n_t}$ and $\widehat{U}_t = \mathrm{BFLY}_t(\beta^{\max,t}_1, \beta^{\max,t}_2, \beta^{\max,t}_3; \widehat{\mu}) + \widehat{\sigma}_{\mathrm{DiD},t}(\beta^{\max,t}_1, \beta^{\max,t}_2, \beta^{\max,t}_3)\widehat{\mathsf{q}}_{1-\alpha}\sqrt{J_t/n_t}$.

---

theoretically appealing, a drawback to these approaches is that it is difficult to parse out movements in credit ratios that are arising from composition changes in the aggregates as compared to all other movements. Here we take a different approach. We rely on the Federal Reserve's Senior Loan Officer Opinion Survey[7] (SLOOS) as our proxy for the "credit" portion of the ratio and the ISM Manufacturing Index as our measure of the "output" portion. This has three distinct advantages. First, as the SLOOS and ISM are both diffusion indices, they have uniform behavior across their history even in the face of changes in the structure of the economy. Second, they are much more timely than credit aggregates and national accounts data which tend to be released with a substantial lag. Third, they are not subject to revision. Thus we have a timely factor which we can evaluate in real time with no look-ahead bias.

Our factor is simply constructed as

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

where SLOOS_t is the net percentage of large domestic banks tightening standards for commercial and industrial loans to all firms and ISM_t is the ISM index.[8] Although both the SLOOS and the

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

[8]The Senior Loan Officer Opinion Survey is currently conducted on a quarterly basis. To construct a monthly

ISM are diffusion indices, they are scaled differently and so the affine transformation of the SLOOS is implemented so that they are both on the same scale (between 0 and 100). To understand why this is (the inverse of) a credit-to-output type measure note that a fall in the SLOOS corresponds to easier lending standards (higher credit growth) and a fall in the ISM to less output. Thus, when the CCW variable is low, credit-to-output is high. A similar logic applies for when the CCW variable is high. Our factor is available starting in January 1965 when the first SLOOS was implemented.

As a preliminary check for the validity of our factor we assess its ability to predict future market returns. Specifically, an implication of our setup (see equation (2.1)) is that if the factor is serially correlated then lagged values should be predictive of future equity returns. To show this, we consider the standard predictive regression setup and run predictive regressions of the form,

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

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

We can now investigate how our risk factor performs in explaining the cross-section of equity returns. We implement our estimators as described in Sections 3 and 4. We use monthly data from the Center for Research in Security Prices (CRSP) over the sample period January 1926 to December 2019. We restrict these data to those firms listed on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), or Nasdaq and use only returns on common shares

---

series we keep the SLOOS value constant until a new value is available. For the period from 1984m1 through 1990m1, the credit standards question was not included in the SLOOS. For this period we use as a replacement the net willingness to make consumer installment loans by large domestic banks.

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

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

(i.e., CRSP share code 10 or 11). To deal with delisting returns we follow the procedure described in Bali et al. (2016). When forming market equity we use quotes when closing prices are not available and set to missing all observations with 0 shares outstanding. For our risk factor we use a measure of the business credit cycle described in equation (6.1). We utilize five-year rolling regressions to estimate betas and we choose the number of portfolios as $J_t = J_1 \cdot \left(\frac{n_t}{\max_{1 \le t \le T} n_t}\right)^{\frac{1}{2}}$ where $J_1 = 10$. The latter choice can be motivated by appealing to Cattaneo et al. (2020a) as the optimal choice of portfolios under the simplifying assumption that all β_{it} were known.

Figure 1 presents our estimate of the grand mean, $\widehat{\mu}(\beta)$ in the black line. There is a clear downward slope in the relationship between β and expected returns – although it does not appear to be linear. To address the differential behavior of the beta-sorted portfolio estimator for the portfolio containing $\beta = 0$, we fit a linear regression in this portfolio only and use constant fits for all other portfolios. The grey vertical lines in Figure 1 depict confidence intervals at each selected point in the support of β. The top chart in Figure 1 uses the PI variance estimator we introduced in equation (4.16) whereas the bottom chart uses the FM variance estimator. To

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

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

implement our plug-in variance estimator we use an (expanding window) AR(1) specification in our risk factor. We can clearly see the difference in the precision for drawing inferences from the data. The confidence intervals based on our new PI variance estimator are substantially shorter than those of the FM variance estimator. Although we showed in Section 4.3 that both of these estimators are conservative, in general, for inference on the SACER, the FM is much larger in this application. By Remark 4.6 we can use the difference between the FM variance estimate and the PI variance estimate as an estimate of the lower bound on $\sigma^2_\mu(\beta)$. Inspecting Figure 1 we see that this lower bound can be informative as the length of the 95% confidence interval based on the FM variance estimator are at least three times as long as that based on the PI variance estimator.

We can see the difference between the two variance estimators even more clearly in Table 3 where we present the point estimate for selected values of β along with lower and upper bounds for confidence intervals constructed with the two different variance estimators. The results are striking. Across all values of β and for both nominal coverage rates, the confidence intervals

Figure 1: **Inference on Expected Returns.** This figure shows the grand mean estimate, $\widehat{\mu}(\beta) = T^{-1} \sum_{t=1}^{T} \widehat{\mu}_t(\beta)$ (black line) with associated confidence intervals (grey vertical lines). The top chart constructs confidence intervals using the plug-in variance estimator introduced in equation (4.16) while the bottom chart uses the Fama-MacBeth variance estimator. The nominal coverage is 95%. The sample period is 1965m1–2019m12.

(a) *Plug-In Variance Estimator*
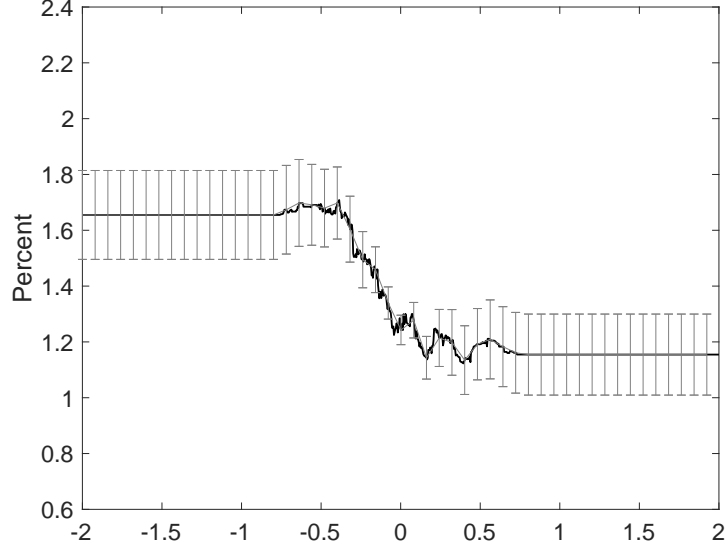
(b) *Fama-MacBeth Variance Estimator*
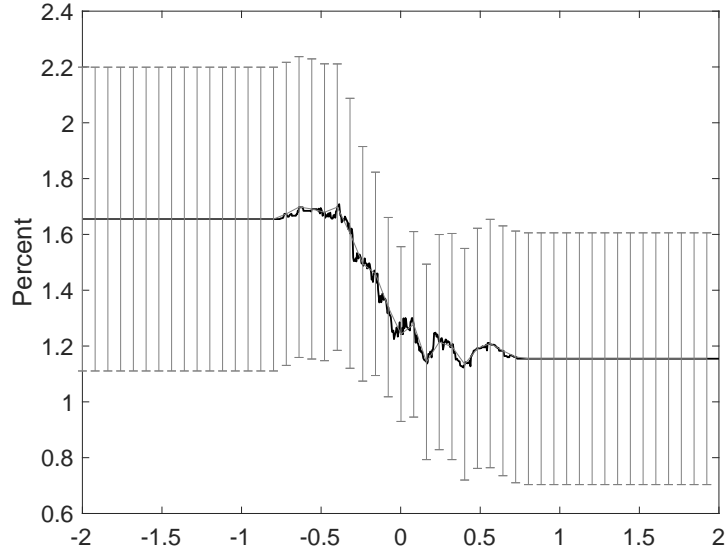
formed using our plug-in variance estimator are approximately 30% of the length of those using the FM variance estimator. We can also illustrate the difference by considering confidence intervals on the high-minus-low estimator, $\widehat{\mu}(\beta_u) - \widehat{\mu}(\beta_l)$, which has a point estimate of 0.5. When using the PI

Table 3: **Inference on Expected Returns**: This table presents the grand mean estimate, $\widehat{\mu}(\beta) = T^{-1}\sum_{t=1}^{T}\widehat{\mu}_t(\beta)$ along with upper and lower bounds for nominal coverage of 95% and 99%. Confidence intervals constructed using the plug-in variance estimator introduced in equation (4.16) are denoted by "PI-LB" and "PI-UB" whereas those using the Fama-MacBeth variance estimator are denoted by "FM-LB" and "FM-UB".The sample period is 1965m1–2019m12.

β	$\widehat{\mu}(\beta)$	90% Coverage				95% Coverage			
		PI-LB	PI-UB	FM-LB	FM-UB	PI-LB	PI-UB	FM-LB	FM-UB
-1.00	1.65	1.50	1.81	1.11	2.20	1.47	1.84	1.01	2.30
-0.50	1.69	1.55	1.83	1.16	2.22	1.52	1.86	1.05	2.33
-0.25	1.52	1.41	1.62	1.09	1.94	1.39	1.64	1.01	2.03
0.00	1.26	1.19	1.32	0.89	1.63	1.17	1.34	0.77	1.74
0.25	1.22	1.11	1.32	0.83	1.60	1.09	1.34	0.76	1.68
0.50	1.19	1.06	1.33	0.76	1.62	1.03	1.35	0.68	1.70
1.00	1.15	1.01	1.30	0.70	1.61	0.98	1.33	0.62	1.69

variance estimator the 95% confidence interval is $[0.14, 0.86]$ whereas for the FM variance estimator it lengthens to $[-0.69, 1.69]$.

7 Conclusion

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

We provide a comprehensive statistical framework which rationalizes this commonly-used estimator. Armed with this foundation we study the theoretical properties of beta-sorted portfolios linking directly to the choice of estimation window in the first step and the number of portfolios in the second step which serves as the tuning parameters for each nonparametric estimator. We introduce conditions that ensure consistency and asymptotic normality for a single cross-section and for the grand mean estimator. We also introduce a new variance estimator and characterize the properties of the ubiquitous FM variance estimator. We provide joint inference procedures for

Figure 2: **Joint Inference on Expected Returns.** This figure shows the grand mean estimate, $\widehat{\mu}(\beta) = T^{-1} \sum_{t=1}^{T} \widehat{\mu}_t(\beta)$ (black line) with associated joint prediction intervals (shaded area). The top chart constructs prediction intervals using the plug-in variance estimator introduced in equation (4.16) while the bottom chart uses the Fama-MacBeth variance estimator. The nominal coverage is 95%. The sample period is 1965m1–2019m12.

(a) *Plug-In Variance Estimator*
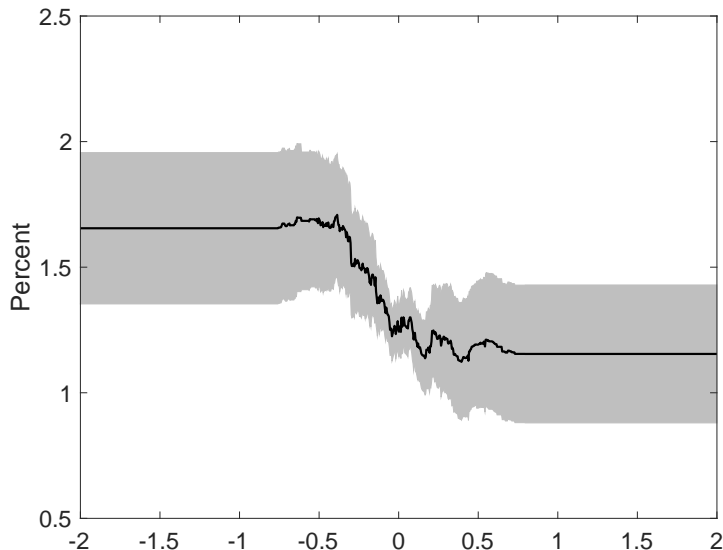
(b) *Fama-MacBeth Variance Estimator*
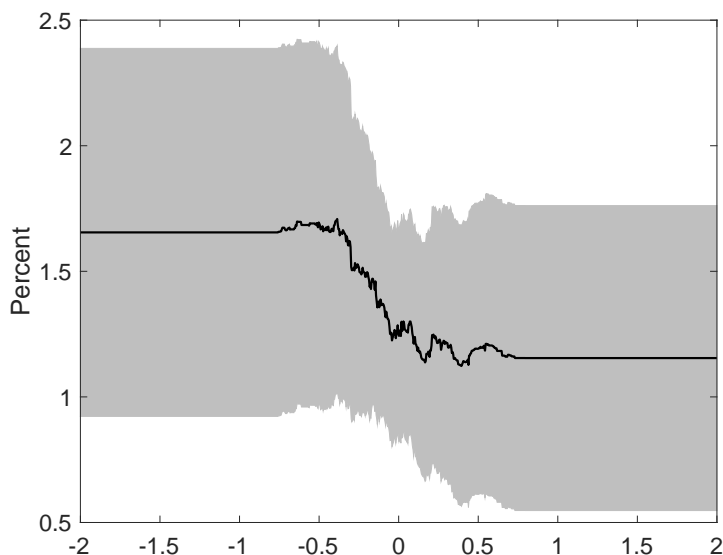
different hypotheses of interest in financial applications including tests of the no-arbitrage assumption. Finally, we also discover some limitations of current practices and provide new guidance on appropriate implementation and interpretation of empirical results.

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.

Bai, J., Zhou, G., 2015. Fama–macbeth two-pass regressions: Improving risk premia estimates. Finance Research Letters 15, 31–40.

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.

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 No. 32365, NBER.

Bryzgalova, S., 2015. Spurious factors in linear asset pricing models, working paper.

Cattaneo, M. D., Crump, R. K., Farrell, M. H., Feng, Y., 2024. On binscatter. American Economic Review Forthcoming.

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.

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

Chen, L., Smetanina, E., Wu, W. B., 2021a. Estimation of nonstationary nonparametric regression model with multiplicative structure. Econometrics Journal 25, 176–214.

Chen, R., Kan, R., 2004. Finite sample analysis of two-pass cross-sectional regressions, working paper.

Chen, Y., Han, B., Pan, J., 2021b. 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., 2024. Investor sentiment and the pricing of macro risks for hedge funds. Management Science Forthcoming.

Chordia, T., Goyal, A., Shanken, J. A., 2017. Cross-sectional asset pricing with individual stocks: Betas versus characteristics, working paper.

Cochrane, J., 2005. Asset Pricing. Princeton University Press, Princeton, revised edition.

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

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

Connor, G., Li, S., Linton, O. B., 2021. A dynamic semiparametric characteristics-based model for optimal portfolio selection. Research Paper 21-1, Michael J. Brennan Irish Finance Working Paper Series.

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

Crump, R. K., Luck, S., 2023. The sloos and economic downturns, working paper.

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, J., Ke, Z. T., Liao, Y., Neuhierl, A., 2022a. Structural deep learning in conditional asset pricing, working paper.

Fan, J., Liao, Y., Wang, W., 2016. Projected principal component analysis in factor models. Annals of Statistics 44, 219–254.

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

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. 7, chap. 3, pp. 219–282.

Garleanu, 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.

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

Hall, P., Heyde, C. C., 2014. Martingale limit theory and its application. Academic press.

Han, F., Wu, W. B., 2023. Probability inequalities for high-dimensional time series under a triangular array framework. In: *Springer Handbook of Engineering Statistics*, Springer, pp. 849–863.

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.

Kleibergen, F., Lingwei, K., Zhan, Z., 2019. Identification robust testing of risk premia in finite samples. Journal of Financial Econometrics 21, 263–297.

Lown, C., 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., 2002. Credit effects in the monetary mechanism. FRBNY Economic Policy Review 8, 217–235.

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.

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

Raponi, V., Robotti, C., Zaffaroni, P., 2020. Testing Beta-Pricing Models Using Large Cross-Sections. The Review of Financial Studies 33, 2796–2842.

Schreft, S. L., Owens, R. E., 1991. Survey evidence of tighter credit conditions: What does it mean? Federal Reserve Bank of Richmond Economic Quarterly 77, 29–34.

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.

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

Wu, W. B., 2005. Nonlinear system theory: Another look at dependence. Proceedings of the National Academy of Sciences 102, 14150–14154.

Zhang, T., Wu, W. B., 2012. Inference of time-varying regression models. Annals of Statistics 40, 1376–1402.