Optimal Cross-Sectional Regression*

Zhipeng Liao

UCLA, Los Angeles, CA 90095, USA

Yan Liu

Purdue University, West Lafayette, IN 47906, USA

May 19, 2021

Abstract

Errors-in-variables (EIV) biases plague asset pricing tests. We offer a new perspective on addressing the EIV issue: instead of viewing EIV biases as *estimation errors* that potentially contaminate next-stage risk premium estimates, we consider them to be *return innovations* that follow a particular correlation structure. We factor this structure into our test design, yielding a new regression model that generates the most accurate risk premium estimates. We demonstrate the theoretical appeal as well as the empirical relevance of our new estimator.

Keywords: Beta uncertainty, Efficient estimation, Errors-in-variables, Factor models, Fama-MacBeth, GMM, Idiosyncratic risk, Systematic risk, Two-pass regression

JEL Codes: C14, C22, G12.

---

*Current Version: May 19, 2021. First posted to SSRN: October 26, 2020. Send correspondence to Zhipeng Liao, UCLA Department of Economics, Los Angeles, CA 90095. Email: Zhipeng.liao@econ.ucla.edu; and Yan Liu, Krannert School of Management, Purdue University, West Lafayette, IN 47906. Email: liu2746@purdue.edu. We have benefited from the comments of Federico Bandi, John Cochrane, Jason Donaldson, Philip Dybvig, Wayne Ferson, Stefano Giglio, Bing Han, Jinyong Hahn, Yan Hong, Cam Harvey, Nengjiu Ju, Raymond Kan, Asaf Manela, Stefan Nagel, Andreas Neuhierl, James Nordlund, Jun Pan, Xuhui Pan, David Rapach, Junbo Wang, Tan Wang, Wei Wei, Guofu Zhou, the committee of The Society for Financial Econometrics (SoFiE), as well as seminar participants at Washington University in St. Louis, University of Oklahoma, Louisiana State University, and Shanghai Advanced Institute of Finance (SAIF).

1 Introduction

A fundamental tenet in financial economics is that investors should be rewarded with a higher return for taking more risk. However, empirical evaluations of the risk-and-return relation are plagued by issues such as the selection of test assets (e.g., Lewellen, Nagel, and Shanken (2010), Harvey and Liu (2020b), and Giglio, Xiu, and Zhang (2021)), the multiplicity of risk factor candidates (e.g., Harvey, Liu, and Zhu (2016), Harvey and Liu (2020b), and Feng, Giglio, and Xiu (2020)), and perhaps most prominently, the "errors-in-variables" (EIV) bias. The EIV bias refers to the phenomenon that the usual two-pass regression estimator (e.g., Fama-MacBeth regression) is biased in finite samples due to estimation errors of betas in the first-pass time-series regressions. Despite various attempts, the literature has not arrived at an agreement on how to address the EIV bias in different test settings (e.g., MacKinlay and Richardson (1991), Shanken (1992), Kim (1995), Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2019), and Pukthuanthong, Roll, Wang, and Zhang (2021)).

In this paper, we propose a new two-pass regression approach to cope with the EIV bias and demonstrate its advantages over existing methods. First, our method is motivated by a new perspective on treating the EIV bias. Instead of viewing first-stage beta estimation uncertainty as *estimation errors* that need to be adjusted in the second-pass regression, we consider EIV bias as *return innovations* that follow a particular correlation structure. Adding these EIV-induced return innovations to the original idiosyncratic risks of test portfolios, our approach efficiently uses information in test portfolios to arrive at the most accurate estimates of the risk premiums.

Second, we theoretically demonstrate the optimality of our approach not only among two-pass regression estimators, somewhat surprisingly, but also over the much broader class of the generalized method of moments (GMM) type of simultaneous estimators. The GMM approach has long been argued to be the preferred approach (e.g., MacKinlay and Richardson (1991)) in estimating beta-pricing models because, intuitively, it simultaneously and potentially more efficiently uses all the information provided by the asset pricing model. For instance, betas are not only learned from return time series but also from the cross-sectional pricing restriction. However, our two-pass estimator suggests that this intuition is misguided: by appropriately weighting returns in the second-pass regression, we show that our approach is just as efficient as the (optimally weighted) GMM estimator. Our paper hence bridges the gap between the earlier literature on GMM-based asset pricing tests (e.g., MacKinlay and Richardson (1991), Zhou (1994), and Jagannathan and

Wang (1998)) and the more popular two-pass regression literature (e.g., Shanken (1992), Gagliardini, Ossola, and Scaillet (2016), Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2019), Kim and Skoulakis (2018), and Raponi, Robotti, and Zaffaroni (2020)). We show that our new two-pass estimator, while inheriting the theoretical appeal of GMM, is computationally as attractive as leading two-pass estimators, such as the Fama-MacBeth approach.

Third, we emphasize the importance of estimation efficiency (i.e., variability of the parameter estimate around its true value) as opposed to consistency, which is the predominant focus of previous literature. This shift in focus is in line with recent advances in applying machine learning techniques to asset pricing research (e.g., Gu, Kelly, and Xiu (2020), Cong, Tang, Wang, and Zhang (2020), and Chen, Pelger, and Zhu (2020)), which advocate tilting the classical bias-variance trade off toward a concern for variance rather than for bias.

To illustrate why efficiency matters, let us examine one example in our simulation study. Suppose we perform the standard Fama-MacBeth regression to test the Fama-French three-factor model using dozens of test portfolios and around 40 years of return history, mimicking a stylized application of two-pass regressions. The risk premium associated with the size factor (smb) is estimated to be 47 bps (per month). How much confidence can we put on such an estimate? It turns out the underlying true risk premium is 26 bps, indicating a 21 bps overestimation. Moreover, such a large overestimation is not an aberration—the root mean-squared estimation error (RMSE) is also 21 bps, suggesting a 32% probability of having an estimation error no less than 21 bps.[1] In comparison, estimation bias is merely 1 bp, which is more than an order of magnitude smaller than RMSE. As such, we show that for routine applications of two-pass regressions, the importance of estimation efficiency dwarfs the concern for bias. Our estimator is theoretically most efficient, shrinking the RMSE in the above example by almost 50%.[2]

Expanding the insights from this example, we perform a comprehensive simulation study to evaluate the performance of our new estimator. Across various test specifications, we show that our estimator consistently generates more accurate parameter estimates compared to the Fama-MacBeth regression or alternative two-pass regression techniques. The superior performance of our model is manifested through both narrower confidence intervals for parameter estimates and more powerful tests for the null hypothesis of a zero risk premium.

Our empirical analysis studies recent factor models proposed in Fama and French (FF, 2015) and Hou, Xue, and Zhang (HXZ, 2015). While both papers use time series spanning regressions

[1]This statement relies on the assumption that the simulated parameter values follow a normal distribution, which is approximately true for our simulation study. See Panel B in Table 1 for the example.

[2]Our improvement over alternative two-pass estimators, for example, the estimator in Shanken (1992), is also sizable. See Table 1 for details.

to evaluate factor model performance, we leverage our new two-pass estimator to develop a cross-sectional counterpart of the usual time-series alpha estimate. We offer new insights into the performance comparison between FF and HXZ using cross-sectional alpha estimates.[3]

Our first empirical finding, which is seemingly overlooked by the existing literature, is the stark contrast between how FF and HXZ use their factor models to price their own basis assets, that is, the very portfolios from which factor models are constructed. Whereas FF successfully explain their basis assets' returns and survive the joint specification test (FF's five-factor model correctly prices its test assets), HXZ's model fails to explain almost half of their 18 basis assets and is strongly rejected in specification tests. Our results thus substantiate FF's comment on HXZ's results and raise the question of whether evaluating factor models using only external anomaly portfolios (i.e., portfolios that do not belong to basis assets, as is done in HXZ) allows a fair comparison among competing factor models.[4]

Focusing on the comparison between our newly proposed estimator and existing two-pass estimators, we highlight the power of our estimator in detecting portfolios with non-zero alphas. Using a large set of anomaly portfolios and testing the hypothesis of a zero abnormal alpha, our model leads to a sizable 10% more rejections than leading alternative models.[5]

One goal of our paper is to synchronize the classical GMM paradigm with current asset pricing research. Despite its strong theoretical appeal, GMM has fallen out of favor in current asset pricing research (asset pricing tests, in particular), perhaps due to the computational challenges in implementing the GMM approach (Shanken and Zhou (2007)).[6] Since we want to maintain the theoretical appeal of GMM while also circumventing its computational burden, we propose a new two-pass approach that is highly tractable, and, at the same time, approaching GMM's efficiency asymptotically. Because our two-pass estimator is akin to the optimally weighted GMM in beta-pricing models, it preserves three important properties of GMM that likely make it more attractive for empirical applications than alternative two-pass methods: 1. It is robust to nonnormality and conditional heteroskedasticity in asset returns (e.g., Bollerslev, Engle, and Wooldridge (1988),

[3]While both Fama and French (2015) and Hou, Xue, and Zhang (2015) likely provide a reduced-form representation and thus an incomplete description of the cross-section of asset returns (Kozak, Nagel, and Santosh (2018) and Kozak, Nagel, and Santosh (2020)), researchers routinely compare factor models, especially those that have similar economic motivations. We provide a new approach to aid such comparisons.

[4]In particular, Fama and French (2015) stated, "More importantly, they (i.e., HXZ) are primarily concerned with explaining the returns associated with anomaly variables not used to construct their factors."

[5]Our results are based only on the Fama-French five-factor model because it survives the joint specification tests.

[6]To appreciate the computational burden of the GMM approach in estimating a beta-pricing model, suppose we have 50 test assets and the benchmark model is a Fama-French five-factor model. This implies 250 ($= 50 \times 5$) betas and five risk-premium parameters to be estimated. Moreover, because the betas and the risk-premium parameters are intertwined in the GMM objective function, the optimization of the GMM objective function involves a nonlinear search over a high-dimensional parameter space. As such, GMM tends to have poor finite sample performances, as documented in Shanken and Zhou (2007).

Schwert (1989), Schwert and Seguin (1990), and Herskovic, Kelly, Lustig, and Van Nieuwerburgh (2016)); 2. It controls for potential nonlinear dependence between return residuals and factor realizations (which is featured prominently in models that study co-skew or co-kurtosis risks; e.g., Harvey and Siddique (2000), Dittmar (2002), and Schneider, Wagner, and Zechner (2020)); and 3. It allows for serial correlations in factor returns, asset returns, and their (possibly nonlinear) combinations (Ehsani and Linnainmaa (2019)).[7]

Our paper is related to a recent growing literature on large N (i.e., number of assets in the cross-section) inference on asset pricing models, including Gagliardini, Ossola, and Scaillet (2016), Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2019), Kim and Skoulakis (2018), Feng, Giglio, and Xiu (2020), Raponi, Robotti, and Zaffaroni (2020), and Harvey and Liu (2020b).[8] In contrast to these papers and their focus on estimation consistency, we are mainly concerned with estimation efficiency for the usual fixed N and large T (i.e., number of time periods) scenario, which is still the popular setup where two-pass regressions are applied to (e.g., Croce, Marchuk, and Schlag (2019), Lin, Palazzo, and Yang (2020), and Ai, Li, Li, and Schlag (2020)). Nonetheless, we also propose extensions of our model to the large N scenario, which not only generalize our estimator but also provide a framework to invite thinking about enhancing estimation efficiency for the aforementioned large-N estimators.

Our paper is organized as follows. In Section 2, we lay out our theoretical framework and present our main theorems. In Section 3, we conduct a comprehensive simulation study that compares alternative two-pass regressors. In Section 4, we revisit the FF versus HXZ comparison using our cross-sectional approach. We offer concluding remarks in the final section. The appendix contains proofs for our main theoretical results and their extensions as well as additional simulation results.

2 Theory

2.1 Optimal Cross-Sectional Regression Estimator

We introduce our optimal cross-sectional regression estimator in several steps. In Section 2.1.1, we introduce the moment conditions that characterize the standard linear-beta pricing model. In Section 2.1.2, we discuss the popular two-pass implementations and derive their asymptotic

[7] Alternative GMM-type of estimators that may also achieve GMM's efficiency asymptotically have been proposed by, e.g., Frazier and Renault (2017). However, these estimators are not computationally as attractive as ours because they do not permit a closed-form expression. Our estimator is based on the popular two-pass implementation and is hence highly tractable analytically.

[8] Also see Bai and Zhou (2015) for a large-N extension of the usual Fama-MacBeth approach.

distributions, complementing existing research. In Section 2.1.3, we establish the optimality of one particular two-pass estimator among all two-pass regressors. Finally, in Section 2.1.4, we show our main results, which build the equivalence between the new two-pass estimator and the theoretically optimal GMM estimator.

2.1.1 Moment Restrictions Corresponding to a Linear Beta-Pricing Model

We first lay out our model in terms of moment restrictions. We present the most general form of our model, where we allow for both asset-specific alphas and firm characteristics to capture potential misspecification of the basic linear-beta pricing model.

The linear-beta asset-pricing model, which is deeply rooted in asset pricing theory (Merton (1973), Breeden (1979), Ross (1976), and Shanken (1987)), can be characterized by the following moment conditions:

$$\mathbb{E}[f_t - \mu_f] = 0, \tag{2.1}$$

$$\mathbb{E}\left[(R_{i,t} - (f_t - \mu_f)'\beta_i)(f_t - \mu_f)\right] = 0, \tag{2.2}$$

$$\mathbb{E}[R_{i,t} - \alpha_i - \gamma_0 - \gamma_1'\beta_i - \gamma_2' Z_{i,t-1}] = 0, \tag{2.3}$$

for $i \in \{1, \ldots, N\}$, where $R_{i,t}$ denotes the return of asset i at period t, f_t is a vector of K many pricing factors, and $Z_{i,t-1}$ is a vector of M many (possibly time-varying) predetermined security characteristics.[9] The unknown parameters are μ_f, α_i, β_i, and $\gamma \equiv (\gamma_0, \gamma_1', \gamma_2')'$.[10] μ_f and β_i are identified by the moment conditions in (2.1) and (2.2), respectively.[11] We assume that a known subset of assets indexed by $\mathcal{I}_0 \subset \{1, \ldots, N\}$ exists, such that

$$\alpha_i = 0 \text{ for any } i \in \mathcal{I}_0. \tag{2.4}$$

Without loss of generality, let $\mathcal{I}_0 \equiv \{1, \ldots, N_0\}$ and $\mathcal{I}_1 \equiv \{N_0 + 1, \ldots, N\}$, where $N_0 \geq d_\gamma$ and $d_\gamma \equiv K + M + 1$ is the dimension of γ. The key unknown parameter of interest is denoted as

[9] Note that our model uses the raw return $R_{i,t}$, in which case, γ_0 should be interpreted as the zero-beta rate, which may be different from the risk-free rate. Alternatively, we can present our model using excess return $R_{i,t} - R_{f,t}$, in which case, γ_0 would be interpreted as the difference between the zero-beta rate and the risk-free rate, which may or may not equal zero. Our econometric analysis goes through for both representations.

[10] Throughout this paper, we use $a \equiv b$ to denote that a is defined as b.

[11] We do not consider the issue of weak or spurious factors, as in Bryzgalova (2015), Kan and Zhang (1999), Kleibergen (2009), Kleibergen and Zhan (2015), Gospodinov, Kan, and Robotti (2014), Gospodinov, Kan, and Robotti (2017), and Gospodinov, Kan, and Robotti (2019). We also take factors as given and do not explicitly consider the issue of omitted risk factors, see, e.g., Forni, Hallin, Lippi, and Zaffaroni (2015), Gagliardini, Ossola, and Scaillet (2019), Giglio and Xiu (2019), and Chaieb, Langlois, and Scaillet (2020).

$\theta \equiv (\alpha_{N_0+1}, \ldots, \alpha_N, \gamma')'$, which is identified through (2.3) and (2.4) given the identification of μ_f and β_i.

Several remarks on our model are worth making. First, our econometric specification fully captures the risk and return relation in a linear beta-pricing model. In fact, by shutting down model misspecification (i.e., setting $(\alpha_{N_0+1}, \ldots, \alpha_N) = 0_{1\times(N-N_0)}$) and firm characteristics ($Z_{i,t-1}$), our moment restrictions correspond exactly to the GMM specifications in MacKinlay and Richardson (1991) and Jagannathan, Skoulakis, and Wang (2010) and are closely related to the moment restrictions in Jagannathan and Wang (2002) and the model in Shanken (1992).[12] Second, following most studies on two-pass estimators (see an extensive discussion in Jagannathan, Schaumburg, and Zhou (2010)), our specification does not impose the condition that risk premiums must equal factor means for traded factors (i.e., $\gamma_1 = \mu_f$). As a result, we allow for measurement errors for traded factors.[13] Another benefit of our setup is that we do not confine ourselves to traded factors—our framework is general enough to cope with non-traded macroeconomic factors (Jagannathan and Wang (2002)).[14]

2.1.2 Two-Pass Implementations

We next present the popular two-pass implementation. Let $\mathbf{B} \equiv (\beta_1, \ldots, \beta_N)'$ and $\mathbf{Z}_{t-1} \equiv (Z_{1,t-1}, \ldots, Z_{N,t-1})'$, which are $N \times K$ and $N \times M$ matrices, respectively. A two-pass regressor first estimates the unknown betas $\mathbf{B}$ by their sample analogs,

$$\widehat{\mathbf{B}} \equiv \left(\sum\nolimits_{t\leq T}(\mathbf{R}_t - \overline{\mathbf{R}})f_t'\right)\left(\sum\nolimits_{t\leq T}(f_t - \overline{f})(f_t - \overline{f})'\right)^{-1}, \tag{2.5}$$

where $\mathbf{R}_t \equiv (R_{1,t}, \ldots, R_{N,t})'$, $\overline{\mathbf{R}} \equiv T^{-1}\sum_{t\leq T}\mathbf{R}_t$, and $\overline{f} \equiv T^{-1}\sum_{t\leq T} f_t$.

Given $\widehat{\mathbf{B}}$, the moment condition in (2.3) can be viewed as a cross-sectional restriction. This allows us to recover the unknown parameter θ through a cross-sectional regression, for example, the Fama-MacBeth regression. However, instead of the Fama-MacBeth regression, which weights assets equally in the cross section, we present a more general weighted least-square estimator as

[12] The only difference from Jagannathan and Wang (2002) is that, consistent with MacKinlay and Richardson (1991) and Jagannathan, Skoulakis, and Wang (2010), we do not have separate moment conditions for the second moment of factor returns. Our moment restrictions are also the same as Shanken (1992) if only non-traded factors are considered. For traded factors, Shanken (1992) imposes the additional restrictions that factor means equal risk premiums.

[13] See, e.g., Roll (1977), Shanken (1987), and Kandel and Stambaugh (1987).

[14] See Balduzzi and Robotti (2008) for another approach of inferring the risk premiums that relies on the maximum-correlation portfolios.

given by the minimizer of

$$\min_{\theta}\left(\sum\nolimits_{t\le T}(\mathbf{R}_t-\widehat{\mathbf{X}}_t\theta)\right)'\widehat{W}\left(\sum\nolimits_{t\le T}(\mathbf{R}_t-\widehat{\mathbf{X}}_t\theta)\right),$$

where $\widehat{W}$ is an arbitrary $N\times N$ real symmetric positive definite matrix, $\widehat{\mathbf{X}}_t\equiv(S_N,\mathbf{1}_{N\times 1},\widehat{\mathbf{B}},\mathbf{Z}_{t-1})$, $S_N\equiv(\mathbf{0}_{N_1\times N_0},I_{N_1})'$, $\mathbf{1}_{N\times 1}$ is a column vector of ones, and $N_1\equiv N-N_0$.[15]

Solving the least-square problem, we have

$$\hat{\theta}_{csr}=(\widehat{\mathbf{X}}'\widehat{W}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}'\widehat{W}\overline{\mathbf{R}}, \tag{2.6}$$

where $\widehat{\mathbf{X}}\equiv T^{-1}\sum_{t\le T}\widehat{\mathbf{X}}_t$. Such an estimator is known as the (weighted) CSR estimator in the literature. Examples of such estimators include the weighted least-square (WLS) estimator in Litzenberger and Ramaswamy (1979) and the generalized least-square (GLS) estimator in Shanken (1985) and Shanken (1992).

To study the statistical properties of $\hat{\theta}_{csr}$, we introduce the notion of total as well as idiosyncratic return innovations. We first define *total return innovation* as

$$v_{i,t}\equiv R_{i,t}-\alpha_i-\gamma_0-\gamma_1'\beta_i-\gamma_2'Z_{i,t-1}, \tag{2.7}$$

for $i=1,\ldots,N$. Following (2.3), we have $\mathbb{E}[v_{i,t}]=0$. Idiosyncratic return innovations are defined as

$$u_{i,t}\equiv R_{i,t}-\mathbb{E}[R_{i,t}]-\beta_i'(f_t-\mu_f),$$

for $i=1,\ldots,N$. Note that we have $\mathbb{E}[u_{i,t}]=0$. By construction, idiosyncratic innovations are orthogonal to factor returns, as implied by (2.2).

We next provide the asymptotic distribution of the CSR estimator. Because betas are estimated, we face the "errors-in-variables" (EIV) problem. As a result, the key to deriving the asymptotic distribution is to factor in estimation errors in betas. Instead of viewing beta estimation errors as errors-in-variables that need to be adjusted in the second stage regression, our derivation (in Appendix A) highlights the new perspective we take on estimation errors: we show that beta estimation errors effectively introduce a new source of return innovation in the second pass regression. In particular, we first show that beta estimation uncertainty translates into a return innovation of -$\mathbf{u}_t(f_t-\mu_f)'\Sigma_f^{-1}\gamma_1$ at each point in time for the second stage regression. Com-

[15]Throughout the paper, we use I_d to denote the $d\times d$ identity matrix; we use $\mathbf{1}_{d_1\times d_2}$ and $\mathbf{0}_{d_1\times d_2}$ to denote the $d_1\times d_2$ matrices of 1's and 0's, respectively.

bining this return innovation with the original total return innovation $\mathbf{v}_t$ that affects the regression estimate (without the errors-in-variables issue), we show the CSR estimator is essentially driven by the following return innovation:

$$\boldsymbol{\epsilon}_t \equiv \mathbf{v}_t - \mathbf{u}_t(f_t - \mu_f)'\Sigma_f^{-1}\gamma_1,$$

where $\mathbf{v}_t \equiv (v_{1,t}, \ldots, v_{N,t})'$, $\mathbf{u}_t \equiv (u_{1,t}, \ldots, u_{N,t})'$ and $\Sigma_f \equiv \mathbb{E}[f_t f_t'] - \mu_f \mu_f'$.

Let $\mathbf{X} \equiv (S_N, \mathbf{1}_{N\times 1}, \mathbf{B}, \mathbb{E}[\mathbf{Z}_{t-1}])$ and W denote the probability limit of $\widehat{W}$. The following lemma provides the asymptotic distribution of the CSR estimator.

Lemma 1. *Under Assumptions 2 and 3 in Appendix A, we have*

$$T^{1/2}(\hat{\theta}_{csr} - \theta) \to_d N(0, \mathrm{Asv}(\hat{\theta}_{csr})), \tag{2.8}$$

where $\mathrm{Asv}(\hat{\theta}_{csr}) \equiv (\mathbf{X}'W\mathbf{X})^{-1}(\mathbf{X}'W\Omega W\mathbf{X})(\mathbf{X}'W\mathbf{X})^{-1}$ *and* $\Omega \equiv \lim_{T\to\infty} \mathrm{Var}(T^{-1/2}\sum_{t=1}^{T} \boldsymbol{\epsilon}_t)$.

The asymptotic normality of the CSR estimator is well established in the literature (see, e.g., Jagannathan and Wang (1998), Ahn and Gadarowski (1999), Jagannathan, Skoulakis, and Wang (2010) and Kan and Robotti (2012)). Lemma 1 generalizes these results to allow for security characteristics and possibly mispriced portfolios.

2.1.3 Optimal Cross-Sectional Regression Estimator (OCSR)

Lemma 1 implies that if the weight matrix is chosen such that $\widehat{W} \to_p \Omega^{-1}$, the asymptotic variance of the CSR estimator reduces to $(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}$. The next proposition shows that in this case, the asymptotic variance of the CSR estimator is not only simplified but also mimimized.

Proposition 1. *Let* $\hat{\theta}^*_{csr} \equiv (\widehat{\mathbf{X}}'\widehat{\Omega}^{-1}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}'\widehat{\Omega}^{-1}\overline{\mathbf{R}}$ *where* $\widehat{\Omega}$ *denotes a consistent estimator of* Ω. *Then we have*

$$\mathrm{Asv}(\hat{\theta}^*_{csr}) = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \tag{2.9}$$

and

$$\mathrm{Asv}(\hat{\theta}_{csr}) \geq \mathrm{Asv}(\hat{\theta}^*_{csr}), \tag{2.10}$$

where $\mathrm{Asv}(\hat{\theta}_{csr})$ *is defined in Lemma 1.*[16]

[16] Throughout this paper, for any two real matrices A and B of the same dimensions, $A \geq B$ means that $A - B$ is positive semi-definite.

Proposition 1 together with Lemma 1 implies that the optimal CSR estimator (i.e., OCSR) $\hat{\theta}^*_{csr}$ can be constructed using the inverse of a consistent estimator of Ω as the weight matrix.[17] Because Ω is the asymptotic variance of $T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\epsilon}_t$, a consistent estimator of Ω can be constructed using consistent estimators of $\boldsymbol{\epsilon}_t$ and the sample analog of Ω when $\{\boldsymbol{\epsilon}_t\}_t$ is a martingale difference sequence. When $\{\boldsymbol{\epsilon}_t\}_t$ is from a weakly dependent process (see, e.g., Newey and West (1987) and Andrews (1991)), a consistent estimator of Ω can be constructed using the heteroskedasticity and auto-correlation consistent (HAC) estimator. We provide more details on the implementation of OCSR in Section 2.2.5 and Section 3.

The OCSR estimator $\hat{\theta}^*_{csr}$ is computationally attractive (compared to a GMM estimator) because it has a closed-form expression. However, its optimality is limited to two-pass estimators. In general, $\hat{\theta}^*_{csr}$ may have a larger asymptotic variance than the one-step GMM estimator of θ (i.e., the one estimated jointly with $\mathbf{B}$ and μ_f through an optimal GMM procedure with all the restrictions in (2.1)–(2.4)) because the information contained in the moment conditions are not simultaneously utilized in the two-pass estimation procedure. Therefore, it is important to understand how much information loss $\hat{\theta}^*_{csr}$ generates in order to achieve its computational convenience. We investigate this issue in the next section.

2.1.4 Relating OCSR to GMM

We compare the asymptotic variance of a general CSR estimator $\hat{\theta}_{csr}$ and the optimal GMM estimator to evaluate the information loss of $\hat{\theta}_{csr}$. More specifically, we show that information loss can be avoided if one uses the optimal weighting $\widehat{W} = \widehat{\Omega}^{-1}$ in the CSR estimation, where $\widehat{\Omega}$ denotes a consistent estimator of Ω.

Let $Y_t \equiv (f_t', \mathbf{R}_t', \mathbf{Z}_t')'$ denote the observation at period t. Using the restrictions in (2.4), the moment functions in (2.1), (2.2), and (2.3) can be written as

$$\begin{aligned} g_1(Y_t,\phi) &\equiv f_t - \mu_f, \\ g_2(Y_t,\phi) &\equiv (\mathbf{R}_t - \mathbf{B}(f_t-\mu_f)) \otimes (f_t - \mu_f), \\ g_3(Y_t,\phi) &\equiv \mathbf{R}_t - \mathbf{X}_t\theta, \end{aligned}$$

where $\mathbf{X}_t \equiv (S_N, \mathbf{1}_{N\times 1}, \mathbf{B}, \mathbf{Z}_{t-1})$ and $\phi \equiv (\theta', \delta')'$ is the stacked parameter vector, with $\delta \equiv$

[17] The optimal choice of the weight matrix in the CSR estimation has also been discussed in Kan and Robotti (2012). Proposition 1 extends their results by allowing for non-beta related explanatory variables (e.g., characteristics) and asset-specific mispricing. More importantly, we further show in the next section the optimal CSR estimator is not only optimal among two-pass estimators but also achieves the GMM efficiency bound, thus establishing its optimality across a larger class of estimators..

$(\mu_f', \beta_1', \ldots, \beta_N')'$. The unknown parameter ϕ can be estimated through the following optimal GMM procedure:

$$\hat{\phi}^*_{gmm} \equiv \arg\min_{\phi} \overline{g}(\phi)'(\widehat{\Sigma}_g)^{-1}\overline{g}(\phi), \tag{2.11}$$

where $\overline{g}(\phi) \equiv T^{-1}\sum_{t=1}^{T} g(Y_t, \phi)$, $g(Y_t, \phi) \equiv (g_1(Y_t, \phi)', g_2(Y_t, \phi)', g_3(Y_t, \phi)')'$ and $\widehat{\Sigma}_g$ denotes a consistent estimator of $\Sigma_g^* \equiv \lim_{T\to\infty} \mathrm{Var}(T^{-1/2}\sum_{t=1}^{T} g(Y_t, \phi))$. By the standard arguments in the GMM literature (see, e.g., Hansen (1982) and Newey and McFadden (1994)), it is well-known that $\hat{\phi}^*_{gmm}$ has the smallest asymptotic variance-covariance,

$$\mathrm{Asv}(\hat{\phi}^*_{gmm}) \equiv (G'(\Sigma_g^*)^{-1}G)^{-1}, \text{ where } G \equiv \mathbb{E}\left[\partial g(Y_t, \phi)/\partial \phi'\right], \tag{2.12}$$

among all GMM estimators. Among all the components in $\hat{\phi}^*_{gmm}$, we select the relevant risk premium estimates through $\hat{\theta}^*_{gmm} \equiv S_\theta \hat{\phi}^*_{gmm}$, where $S_\theta \equiv (I_{d_\theta}, 0_{d_\theta \times d_\delta})$, $d_\theta \equiv N_1 + d_\gamma$ and $d_\delta \equiv (N+1)K$. Then $\hat{\theta}^*_{gmm}$ has the smallest asymptotic variance-covariance,

$$\mathrm{Asv}(\hat{\theta}^*_{gmm}) \equiv S_\theta \mathrm{Asv}(\hat{\phi}^*_{gmm}) S_\theta', \tag{2.13}$$

among all GMM estimators of θ.

The following proposition presents the formal comparison of the CSR estimator $\hat{\theta}_{csr}$ and the optimal GMM estimator $\hat{\theta}^*_{gmm}$ in terms of the asymptotic efficiency.

Proposition 2. *Suppose that Σ_f, Σ_g^*, W and $\mathbf{X}'\mathbf{X}$ are nonsingular. Then*

$$\mathrm{Asv}(\hat{\theta}_{csr}) \geq \mathrm{Asv}(\hat{\theta}^*_{gmm}), \tag{2.14}$$

where the equality is achieved if $\hat{\theta}_{csr}$ is replaced by $\hat{\theta}^*_{csr}$.

Proposition 2 shows that the CSR estimator $\hat{\theta}_{csr}$ is less efficient than the optimal GMM estimator $\hat{\theta}^*_{gmm}$ in general. However, the efficiency loss can be avoided if one uses the OCSR estimator $\hat{\theta}^*_{csr}$. Proposition 2 establishes the asymptotic equivalence between the OCSR estimator $\hat{\theta}^*_{csr}$ and the optimal GMM estimator $\hat{\theta}^*_{gmm}$ in a general framework that represents the linear beta-pricing model. In essence, it shows that the usual efficiency gain of the optimal GMM estimator is effectively preserved by using the optimal weight matrix in a two-pass regression. It thus justifies the popular two-pass regression approach from an efficiency perspective: A two-pass estimator with an appropriately chosen weight matrix can be made as efficient as the theoretically optimal GMM estimator.

Proposition 2 is a profound result for at least two reasons. First, conceptually, a simultaneous GMM estimation appears to use more information provided by the moment conditions in (2.1), (2.2) and (2.3). For instance, we learn about betas not only from the time-series restrictions (i.e., (2.2)) but also from the linear-beta restriction (i.e., (2.3)). As a result, one would think that the GMM estimates of betas should be more efficient than the usual OLS estimates and subsequently conjecture that this efficiency gain in estimating betas would translate into efficiency gains in estimating risk premiums. We show, somewhat counterintuitively, that this conjecture is incorrect. We discuss in detail the intuition behind our equivalence result in Section 2.2.4. In short, Proposition 2 bridges the gap between two literatures on estimating beta-pricing models: One is the GMM approach (e.g., MacKinlay and Richardson (1991), Zhou (1994), and Jagannathan and Wang (2002)), and the other is the popular two-pass regression approach (e.g., Shanken (1992), Gagliardini, Ossola, and Scaillet (2016), Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2019), Kim and Skoulakis (2018), and Raponi, Robotti, and Zaffaroni (2020)).

Second, it is computationally very challenging to estimate linear-beta pricing models with GMM. For example, if we use the Fama-French-Carhart four-factor model as the benchmark factor model, when we have 50 test assets, we will have 200 ($= 50 \times 4$) betas to estimate. Note that betas are intertwined with risk premium parameters in the GMM objective function, making the optimization problem highly nonlinear. As a result, the GMM estimator tends to have poor finite-sample performance (Shanken and Zhou (2007)) due to large optimization errors. Proposition 2 effectively shows that none of these computational issues is essential for achieving GMM's asymptotic optimality. Our OCSR, available in closed form, is as efficient as GMM asymptotically. For finite samples, we demonstrate the superior performance of our estimator in a simulation study in Section 3.

2.2 Discussion

We provide an extensive discussion of our results in the context of the current literature. In Section 2.2.1, we relate our method to existing approaches by providing a literature review. In Section 2.2.2, we specifically relate our estimator to the alternative two-pass estimator provided in Shanken (1992). In Section 2.2.3, we offer the economic intuition for the efficiency gain of our estimator. In Section 2.2.4, we discuss in detail what features of the linear-beta pricing model render our equivalence result and possible generalizations of this result. Finally, in Section 2.2.5, we describe our implementation of OCSR. Readers who are more interested in the empirical performance of our approach can go directly to Section 3 for a simulation study.

2.2.1 A Literature Review

Given the large body of literature on addressing the EIV biases in asset pricing tests, we review the relevant literature in Appendix B. We also relate our paper to the recent development of asset pricing tests in general.

2.2.2 Comparing OCSR and Shanken (1992)'s GLS

This section is devoted to the discussion of the connection between the OCSR estimator $\hat{\theta}^*_{csr}$ and the GLS estimator proposed in Shanken (1992). For ease of discussion, we shut down model misspecification (i.e., $(\alpha_{N_0+1}, \ldots, \alpha_N) = 0_{1\times N_1}$) and firm characteristics ($Z_{i,t-1}$), and assume the pricing factors are not traded, so our GMM moment restrictions exactly correspond to Shanken (1992)'s model.

In GLS, estimation errors in betas and the conditional heteroskedasticity of the idiosyncratic innovation $\mathbf{u}_t$ (given the pricing factors) are not directly taken into account. As a result, the optimal weight matrix in GLS amounts to the covariance matrix of the idiosyncratic innovation $\mathbf{u}_t$. As we discuss below, the efficiency of GLS is justified under the following assumptions from Shanken (1992).

Assumption 1. $\mathbf{u}_t$ *is i.i.d. with* $\mathbb{E}[\mathbf{u}_t|F_T] = 0$ *and* $\mathbb{E}[\mathbf{u}_t\mathbf{u}_t'|F_T] = \Sigma_u$*, where* $F_T = (f_1, \ldots, f_T)$ *and* Σ_u *is a non-random positive definite matrix.*

Under Assumption 1, $\mathbf{u}_t(1 - (f_t - \mu_f)'\Sigma_f^{-1}\gamma_1)$ forms a martingale difference array that is uncorrelated with $f_{t'} - \mu_f$ for any t and t'. Therefore, the variance-covariance matrix Ω takes the following simplified form,

$$\Omega = \Sigma_u(1 + \gamma_1'\Sigma_f\gamma_1) + \mathbf{B}\Sigma_{\overline{f}}\mathbf{B}' = \Sigma_u(1 + \gamma_1'\Sigma_f\gamma_1) + \mathbf{X}\Sigma^*_{\overline{f}}\mathbf{X}', \tag{2.15}$$

where $\Sigma^*_{\overline{f}} \equiv \text{diag}(0, \Sigma_{\overline{f}})$ and $\Sigma_{\overline{f}}$ denotes the long-run variance of $T^{-1/2}\sum_{t=1}^T f_t$.[18] Moreover, by Lemma 1, the asymptotic variance-covariance matrix of the CSR estimator becomes

$$\text{Asv}(\hat{\theta}_{csr}) = (1 + \gamma_1'\Sigma_f\gamma_1)(\mathbf{X}'W\mathbf{X})^{-1}(\mathbf{X}'W\Sigma_uW\mathbf{X})(\mathbf{X}'W\mathbf{X})^{-1} + \Sigma^*_{\overline{f}}. \tag{2.16}$$

The GLS estimator $\hat{\theta}^*_{gls}$ of θ, proposed in Shanken (1992), is constructed using a weight matrix $\widehat{W}$

[18]Throughout this paper, we use diag(A, B) to denote the block diagonal matrix with the square matrices A and B on the main diagonal.

in CSR such that $\widehat{W} \to_p \Sigma_u^{-1}$. Therefore, the asymptotic variance of $\hat{\theta}^*_{gls}$ takes the following form:

$$\text{Asv}(\hat{\theta}^*_{gls}) = (1 + \gamma_1' \Sigma_f \gamma_1)(\mathbf{X}' \Sigma_u^{-1} \mathbf{X})^{-1} + \Sigma_f^*. \tag{2.17}$$

Shanken (1992) further shows that if the joint density of $\mathbf{R}_t$ given F_T is normal, the GLS estimator is asymptotically equivalent to the maximum likelihood estimator and hence is efficient.

Using similar arguments for proving (2.10) in Proposition 1, we can show that

$$\mathbf{X}' \Sigma_u^{-1} \mathbf{X} \geq (\mathbf{X}' W \mathbf{X})(\mathbf{X}' W \Sigma_u W \mathbf{X})^{-1} (\mathbf{X}' W \mathbf{X}),$$

which together with (2.16) and (2.17) implies

$$\text{Asv}(\hat{\theta}_{csr}) \geq \text{Asv}(\hat{\theta}^*_{gls}). \tag{2.18}$$

This shows the optimality of the GLS estimator among all CSR estimators under the assumptions in Assumption 1.

Because the inequality (2.14) holds without Assumption 1, we have $\text{Asv}(\hat{\theta}^*_{gls}) \geq \text{Asv}(\hat{\theta}^*_{gmm})$ in general, which, together with the asymptotic equivalence between $\hat{\theta}^*_{gmm}$ and $\hat{\theta}^*_{csr}$ established in Proposition 2, shows that the OCSR estimator $\hat{\theta}^*_{csr}$ is more efficient than the GLS estimator $\hat{\theta}^*_{gls}$ in general. However, when Assumption 1 holds, following Proposition 2 and (2.18), we obtain $\text{Asv}(\hat{\theta}^*_{csr}) = \text{Asv}(\hat{\theta}^*_{gls})$, which, together with the parametric efficiency of the GLS estimator, implies that the OCSR estimator $\hat{\theta}^*_{csr}$ is also parametrically efficient if the joint density of $\mathbf{R}_t$ given F_T is normal. When Assumption 1 fails, OCSR can be more efficient than the GLS estimator. We discuss important scenarios in which Assumption 1 fails and how OCSR improves on the GLS estimator in the next section.

2.2.3 Interpreting the Efficiency Gain of OCSR

In this section, we provide an economic interpretation for the weighting scheme used by the OCSR estimator. For simplicity, let us assume that betas are the only explanatory variables that affect returns.

The estimation error of $\hat{\theta}_{csr}$ is mainly governed by $T^{-1/2} \sum_{t=1}^T \left(\mathbf{v}_t - \mathbf{u}_t (f_t - \mu_f)' \Sigma_f^{-1} \gamma_1 \right)$, which can be decomposed into two parts (see Appendix A): One part is related to $\mathbf{v}_t$, which captures the total innovation in returns, and the other part is given by $(\widehat{\mathbf{B}} - \mathbf{B})\gamma_1$, which captures estimation error in beta. By (2.7),

$$
\begin{aligned}
& T^{-1/2}\sum_{t=1}^{T}\left(\mathbf{v}_t-\mathbf{u}_t(f_t-\mu_f)'\Sigma_f^{-1}\gamma_1\right) \\
= \quad & T^{-1/2}\sum_{t=1}^{T}\left(\underbrace{\mathbf{u}_t(1-(f_t-\mu_f)'\Sigma_f^{-1}\gamma_1)}_{\text{Efficiency-relevant innovations}} \quad + \quad \underbrace{\mathbf{B}(f_t-\mu_f)}_{\text{Efficiency-irrelevant innovations}}\right), \qquad (2.19)
\end{aligned}
$$

which further decomposes the total innovation $\mathbf{v}_t$ into the idiosyncratic innovation $\mathbf{u}_t$ and the systematic innovation $\mathbf{B}(f_t-\mu_f)$.

We first interpret our results under Assumption 1. We interpret the two terms in Eq. (2.19) as "efficiency-relevant innovations" and "efficiency-irrelevant innovations" (under Assumption 1), for reasons as follows. Idiosyncratic innovations $\mathbf{u}_t$ contribute to both first-stage beta-estimation uncertainty and second-stage cross-sectional regression uncertainty. $\mathbf{u}_t$ alone represents second-stage uncertainty, whereas $-\mathbf{u}_t(f_t-\mu_f)'\Sigma_f^{-1}\gamma_1$ captures the effect of first-stage beta-estimation uncertainty, where $(f_t-\mu_f)'\Sigma_f^{-1}\gamma_1$ is the common (across assets) multiplicative effect. Hence, as in a typical GLS estimator, estimation efficiency depends on the weight matrix, and optimal efficiency is achieved at the variance-covariance matrix of $\mathbf{u}_t$ (hence, the name "efficiency-relevant innovations").

For systematic innovations $\mathbf{B}(f_t-\mu_f)$, given the economic restrictions of the linear-beta pricing model (in particular, $\mathbf{B}\gamma_1$ is linear in $\mathbf{B}$), weighting does not affect the efficiency of the CSR estimator (hence, the name "efficiency-irrelevant" innovations). Intuitively, when only systematic innovations exist, first-stage beta estimation will not produce any estimation errors. Moreover, because the expected return is assumed to be linear in beta, weighting does not affect the second-stage estimation, because the expected return (i.e., $\mathbf{B}\gamma_1$) scales up and down in the same way as innovations (i.e., $\mathbf{B}(f_t-\mu_f)$). Note that as a counterexample, the above argument breaks down if the asset-pricing model dictates a nonlinear relation between expected return and beta (e.g., $\mathbf{B}^{1/3}\gamma_1$, where $\mathbf{B}^{1/3}$ represents taking the cube root of each element in $\mathbf{B}$).

Combining the above two observations, efficiency-irrelevant innovations (i.e., $\mathbf{B}(f_t-\mu_f)$) can be isolated because Assumption 1 allows us to decouple efficiency-relevant innovations and efficiency-irrelevant innovations (more specifically, the two innovations are uncorrelated). Estimation efficiency only depends on efficiency-relevant innovations (i.e., $\mathbf{u}_t(1-(f_t-\mu_f)'\Sigma_f^{-1}\gamma_1)$) and is achieved when the weight matrix is set at Σ_u^{-1}. This explains the efficiency result in Shanken (1992).

Unlike the GLS estimator in Shanken (1992), OCSR can offer additional efficiency gain when our assumptions (in particular, Assumption 3 in Appendix A) generalize Assumption 1. Several prominent features of financial market data may make our generalizations important. First, the assumption of conditional homoskedasticity in Assumption 1 may not be suitable when asset returns display nonnormality and conditional heteroskedasticity (Jagannathan and Wang (1998), and Jagannathan and Wang (2002)).[19] Recent evidence on residual conditional heteroskedasticity includes Herskovic, Kelly, Lustig, and Van Nieuwerburgh (2016), De Nard, Ledoit, and Wolf (2018), and Engle, Ledoit, and Wolf (2019).

Second, non-zero correlations may exist between $\mathbf{u}_t$ and higher powers of f_t.[20] Note that by definition, $\mathbf{u}_t$ is orthogonal to f_t itself. But this orthogonality does not rule out potential correlations between $\mathbf{u}_t$ and higher powers of f_t. For example, models that feature co-moment risks (e.g., co-skewness risk in Harvey and Siddique (2000) and Schneider, Wagner, and Zechner (2020); co-kurtosis risk in Dittmar (2002)) imply a non-zero correlation between market-model residuals and second (or third) powers of market returns for most assets. For these models, efficiency-relevant innovations and efficiency-irrelevant innovations cannot be uncorrelated, leading to a potential efficiency gain of our estimator compared with Shanken (1992) for the inference of the market risk premium.

Finally, our estimator adjusts for serial correlations in $\mathbf{u_t}$, f_t, and, more importantly, their combinations. Whereas Shanken (1992)'s GLS can be extended to cope with serial correlations in $\mathbf{u_t}$, we show that adjusting for serial correlations in the combined innovations (i.e., the sum of efficiency-relevant and efficient-irrelevant innovations) is crucial to achieve the asymptotic optimality of the optimal GMM estimator. Such an adjustment may substantially improve the estimation accuracy of a CSR estimator given the widespread time-series patterns in anomaly returns (e.g., Ehsani and Linnainmaa (2019)).

GMM is able to achieve semiparametric efficiency that guards against all of the above issues (i.e., conditional heteroskedasticity, non-zero dependence between $\mathbf{u}_t$ and higher powers of f_t, and serial correlations). Our new insight shows that a CSR estimator with a properly defined weight matrix for the second-stage cross-sectional regression inherits the theoretical appeal of the GMM estimator and is thus also robust to the above issues. Unlike Shanken (1992)'s results that allow the decoupling of efficiency-relevant and efficiency-irrelevant innovations in constructing the optimal weight matrix, we show that mixing these two sources of innovations is essential to achieve GMM's

[19]Nonnormality and conditional heteroskedasticity for asset returns are documented in, e.g., Barone-Adesi and Talwar (1983), Bollerslev, Engle, and Wooldridge (1988), Schwert (1989), Schwert and Seguin (1990), MacKinlay and Richardson (1991), and Diebold, Lim, and Lee (1993). See extensive discussion in Jagannathan, Schaumburg, and Zhou (2010) and Jagannathan, Skoulakis, and Wang (2010) on the economic restriction that the assumption of conditional homoskedasticity imposes relative to the less restrictive assumptions in a GMM framework.

[20]Another data feature that our OCSR guards against is the nonlinear dependence between $\mathbf{u}_t$ and f_t.

efficiency when Assumption 1 fails to hold.

2.2.4 Linear-Beta Pricing and Equivalence between OCSR and GMM

Our main result on the equivalence between OCSR and GMM deserves further discussion. First, what structure of the linear-beta relation renders this equivalence? Intuitively, one can imagine an iterative GMM interpretation of the system of moment restrictions where $\hat{\mu}_f$ and $\widehat{\mathbf{B}}$ are estimated from moment conditions (2.1) and (2.2) first and then fed into moment conditions (2.3) to obtain the estimate for γ. Hence, the fact that γ only shows up in (2.3), which facilitates the iterative GMM implementation, may explain our equivalence result.

However, as our proofs in the appendix show, the above intuition is only partially correct in that the complete separation of γ from other parameters in moment conditions (2.1) and (2.2) only constitutes one of the required conditions that lead to our equivalence result. The other condition is the fact that the moment conditions in (2.3) depend on θ and do not contain separate identification information for the model parameters μ_f and $\mathbf{B}$, which are just identified by (2.1) and (2.2). Loosely speaking, the just identified system (2.1) and (2.2) allow us to think of μ_f and $\mathbf{B}$ as "nuisance" parameters. While we gain efficiency by estimating γ jointly with the nuisance parameters from all three sets of moment conditions relative to estimating it through (2.2) with plug-in estimates of μ_f and $\mathbf{B}$ from the just identified system (2.1) and (2.2), this gain is vanishing asymptotically, which leads to the efficiency result of the OCSR estimator. In Appendix C, we provide a counterexample that illustrates the efficiency loss of the (optimally weighted) iterative GMM estimator with over-identified nuisance parameters.[21] This counter example also shows that the linearity of the moment conditions plays no role in the efficiency result of the OCSR estimator.

The above insight allows us to easily generalize the implication of our results to other settings. For example, suppose we augment the three sets of moment restrictions with another set (called $S1$) that imposes additional restrictions on betas (e.g., augmenting Eq. (2.2) by assuming that idiosyncratic risks are orthogonal not only to contemporaneous factor returns but also to other lagged instruments). Then our equivalence result will not hold. More specifically, the two-pass implementation that estimates μ_f and $\mathbf{B}$ through (2.1), (2.2), and ($S1$) first and then estimates γ

[21] Kan and Zhou (2001) consider two sets of parameters ϕ_1 and ϕ_2 with two sets of moment conditions, say, $S1$ and $S2$, where ϕ_1 is separately identified by $S1$, while ϕ_2 is identified by $S2$ given ϕ_1. They show that if ϕ_2 is just identified by $S2$, given ϕ_1, then the optimal GMM estimator of ϕ_1 using $S1$ is as efficient as the joint GMM estimator of ϕ_1, using both $S1$ and $S2$. Although this result is similar in spirit to our efficiency result, they have different implications. Lemma 1 in Kan and Zhou (2001), when applied to our setup, implies that the nuisance parameters μ_f and $\mathbf{B}$ can be efficiently estimated using only moment conditions in (2.1) and (2.2), as long as the parameter of interest γ is just identified by (2.3), given the nuisance parameters. However, our focus is on γ, which is over-identified by (2.3). Our efficiency result applies to this parameter instead of the nuisance parameters.

from (2.3) will not be as efficient as the GMM estimator. As another example, suppose we keep (2.1) and (2.2) the same but alter (2.3) by specifying a general functional form of risk premium: $h(\gamma, \mathbf{B})$. Then our equivalence result still holds. This can be used to justify, for example, the specification in Fama and MacBeth (1973), where the risk premium is linear in squared market betas.

To situate our result in the context of the literature, we relate it to two lines of research. First, Kan and Zhou (1999) argue that the SDF approach is less efficient than traditional two-pass regressions. Jagannathan and Wang (2002) formulate a framework to compare the two and show the equivalence between the GMM estimates from moment restrictions of the linear-beta model and the GMM estimates from moment restrictions of the SDF (under the simplified assumption of a single-factor model). We do not consider the SDF approach.[22] Rather, we establish the precise conditions under which the GMM estimates from moment restrictions of the linear-beta model are equivalent to two-pass regressions asymptotically.

Within the second line of research, Shanken (1992) establishes the efficiency of the two-pass GLS estimator under Assumption 1 and joint normality of $\mathbf{R}_t$ given F_T. Jagannathan and Wang (1998) develop the distribution theory for a given two-pass estimator under general assumptions on the return generating process. However, no attempt has been made so far to relate a two-pass estimator to the optimal GMM estimator, which is known to be semiparametrically efficient. Our paper addresses this gap in the literature.

2.2.5 Implementation of OCSR

Similar to the GLS estimator, the implementation of our OCSR requires the estimation of an $N \times N$ weight matrix, which becomes inaccurate when N/T is large. This raises the question of what level of N is suitable for implementing our estimator versus alternative estimators that are potentially less demanding in terms of the estimation of the weight matrix, for example, the WLS estimator.[23]

While our theory is developed under the assumption of a fixed (or slowly divergent) N, the level of N considered to be large is more of an empirical than theoretical question.[24] We answer this question through a comprehensive simulation study in the next section. Our evidence shows that when N ranges from small (i.e., $N = 10$) to modest (i.e., $N = 50$), the efficiency gain of our OCSR compared to alternative estimators, especially the WLS, is large. This suggests that taking

[22] Also see Chapter 12 in Cochrane (2009) and Kan and Zhou (2001) for further discussion of this result.

[23] The WLS estimator is a naive version of the GLS estimator that sets the off-diagonal elements in the GLS weighting matrix to zero; see more detailed description in Section 3.

[24] See Section D.3 of the Appendix for extension of the theory to the case with divergent N.

cross-asset correlations into account is important. Intuitively, some test assets may have highly correlated factor model residuals with other test assets, implying that they do not provide much independent information in testing factor models. Our estimator (and GLS to some extent) allow us to downweight the impact of these assets, leading to a more efficient estimation.

In our simulation study, while we do not experiment with very large N, theory suggests that the efficiency gain of OCSR (and GLS) should disappear at some level of N.[25] WLS and OLS may perform better instead. But even in these cases, some ad hoc shrinkage estimator of the error covariance matrix may go a long way in improving the efficiency of the approach. Since this is not within the scope of our current paper, we leave it to future research.

3 A Simulation Study

One of our goals in this paper is to provide a comprehensive simulation study to systematically compare existing methods. Compared with previous studies that numerically compare beta-pricing models (e.g., Shanken and Zhou (2007)), our simulation design is marked by several features. First, we focus on popular two-pass estimators that empirical researchers frequently use and do not consider computationally intensive alternatives such as MLE- or GMM-based methods. Second, to ensure that our simulated data-generating process (DGP) resembles the actual application, we construct bootstrap samples from the actual data of factor returns and test assets. In particular, we consider the Fama-French three-factor and five-factor models as candidate models, and as candidate samples for test assets, we consider the 18 low-turnover anomaly sample and the 38 low-turnover and medium-turnover combined anomaly sample in Novy-Marx and Velikov (2016).[26] Our actual sample runs from July 1973 to December 2017, including 534 monthly observations.[27] Third, our bootstrap samples keep the same level of time-series and cross-sectional dependency in the actual data, such that our results shed light on the performance of our approach for the actual data.

Note that we focus on popular factor models in our simulation study because they correspond to our empirical application where we examine FF and HXZ. Given that factors usually have a limited degree of time-series dependence, our simulation study may not be the best case to

[25] We thank instructive comments by Wayne Ferson about the impact of large N on the relative efficiency of our approach to WLS and OLS.

[26] We refer interested readers to Novy-Marx and Velikov (2016) for the definitions of anomaly portfolios. Data on Fama-French models are obtained from Ken French's online data library.

[27] The starting date of July 1973 is constrained by the availability of several medium-turnover anomalies, as described in Novy-Marx and Velikov (2016). For low-turnover anomalies only, our sample can be extended to start from July 1963. We do not use the extended sample and simply set the starting date to be the same across medium- and low-turnover anomaly portfolios.

demonstrate the efficiency gain of our approach. For example, the efficiency gain of our estimator may be larger for non-traded macroeconomic factors that display strong serial dependence (e.g., inflation).[28]

Our simulation study is described as follows. For ease of presentation, we use the Fama-French three-factor model as an example to explain our simulation steps.

For a collection (N) of test assets, we use the full actual sample with 534 monthly observations to regress asset i's excess returns onto the three-factor model and obtain the 3×1 loading vector $\hat{\beta}_{p,i}$; that is,

$$R_{i,t} - R_{f,t} = \hat{\mu}_{p,i} + \hat{\beta}_{p,i}' f_t + \hat{\varepsilon}_{p,i,t},$$

where $\hat{\mu}_{p,i}$ is the regression intercept, $f_t = (f_{mkt,t}, f_{smb,t}, f_{hml,t})'$ is the 3×1 factor realizations, and $\hat{\varepsilon}_{p,i,t}$ is the regression residual (here, "p" stands for population). We collect the loading vectors into $\mathbf{B}_p \equiv (\hat{\beta}_{p,1}, \ldots, \hat{\beta}_{p,N})'$, which is an $N \times 3$ matrix that consists of the population factor loadings, and collect the factors into $F_p \equiv (f_1, \ldots, f_T)$, which is a $3\times T$ matrix. For asset i, let the regression residual vector be $\hat{\varepsilon}_{p,i} \equiv (\hat{\varepsilon}_{p,i,1}, \ldots, \hat{\varepsilon}_{p,i,T})'$, which is a $T \times 1$ vector. We collect the cross-section of regression residuals into $RES_p \equiv (\hat{\varepsilon}_{p,1}, \ldots, \hat{\varepsilon}_{p,N})'$, which is an $N \times T$ matrix that contains the population factor-model residuals.

We bootstrap to generate the in-sample data to study the finite-sample properties of various CSR estimators and related inference procedures. The use of bootstrap aims to capture potential cross-sectional and time-series dependency in the DGP. For bootstrap iteration m ($m = 1, \ldots, M$, where $M = 10{,}000$), we block bootstrap (i.e., Politis and Romano (1994)) time periods with a block size of 12 (i.e., 12 months). We then use the same bootstrap time periods to obtain the bootstrap factor returns (F_m, $3 \times T_b$, which resamples from F_p) and the bootstrap factor-model residuals (RES_m, $N \times T_b$, which resamples from RES_p), where T_b denotes the size of the bootstrap sample. For a given parametrization $\gamma_{p,0}$ and $\gamma_p \equiv (\gamma_{p,mkt}, \gamma_{p,smb}, \gamma_{p,hml})'$, we generate the bootstrap return panel by

$$\mathbf{R}_m \equiv \gamma_{p,0} 1_{N\times T_b} + \mathbf{B}_p \times (F_m - (\overline{f}_p - \gamma_p) 1_{1\times T_b}) + RES_m,$$

where $\mathbf{R}_m$ is a $N \times T_b$ matrix and $\overline{f}_p$ denotes the (original) sample mean of f_t.[29]

To investigate the finite-sample properties of various CSR estimators, we let $\gamma_{p,0} = 0$ and $\gamma_p = \overline{f}_p$ to ensure the linear-beta pricing holds. This parametrization is slightly revised when we

[28] Note that existing two-pass estimators such as Shanken (1992) do not take time-series correlations in factor model residuals into account. Alternative approaches that infer risk premiums through latent risk factors include Zaffaroni (2019), Kelly, Pruitt, and Su (2019), and Gu, Kelly, and Xiu (2021).

[29] Our bootstrap procedure follows the simultaneous bootstrapping approach advocated by Fama and French (2010) and Harvey and Liu (2020b). See Harvey and Liu (2020b) for a discussion of the advantage of the simultaneous bootstrapping approach over alternative bootstrapping methods.

examine the performance of inference (i.e., hypothesis tests) based on different CSR estimators. Specifically, we reparametrize an individual component of $\gamma_p = (\gamma_{p,0}, \gamma_{p,1}')'$ while keeping the remaining components unchanged. For example, when we consider the two-sided test on $H_0 : \gamma_{p,0} = 0$, we consider four different values (0, 0.1%, 0.2%, and 0.3% per month) for $\gamma_{p,0}$ while keeping $\gamma_{p,1} = \overline{f}_p$. The null clearly holds for the first reparametrization where $\gamma_{p,0}$ is set to 0, whereas the alternative holds for the other three reparametrizations where $\gamma_{p,0}$ is not zero. As another example, when we consider the two-sided test on $H_0 : \gamma_{p,mkt} = 0$, we let $\gamma_{p,mkt} = a\overline{f}_{p,mkt}$, where $\overline{f}_{p,mkt}$ denotes the (original) sample mean of $f_{mkt,t}$ and a is a multiplier that can take the value of 0, 0.5, 1, or 1.5. The remaining components of γ_p, namely, $\gamma_{0,p}$, $\gamma_{p,smb}$ and $\gamma_{p,hml}$, are unchanged. We can examine the size of various tests on $H_0 : \gamma_{p,mkt} = 0$ under the reparametrization where $\gamma_{p,mkt}$ is set to zero and the power of these tests under the reparametrization where $\gamma_{p,mkt}$ is set to a non zero value.

For the in-sample data $\{\mathbf{R}_m, F_m\}$ generated in the m-th bootstrap iteration, we consider four types of cross-sectional regression procedures: OLS (the two-pass estimator that uses the identity weighting matrix in the second-stage regression), OCSR (our approach with optimal weighting), GLS, and WLS (alternative cross-sectional approaches that are studied in Shanken (1992) and Jagannathan and Wang (1998)). For GLS, the estimated covariance matrix for factor-model residuals, that is, the estimator of $\mathbb{E}[\mathbf{u}_t\mathbf{u}_t']$, is used as the weight matrix. For WLS, the off-diagonal elements of the GLS weight matrix are set to zero.

For our OCSR, we need to estimate the long-run variance Ω as given in Lemma 1. We follow a simple approach. $\mathbf{v}_t$ is obtained by demeaning returns. $\mathbf{u}_t$ is obtained from asset-by-asset time-series OLS. μ_f and Σ_f are the estimated factor mean and covariance matrix. γ_1 is taken to be the cross-sectional OLS estimate. Finally, given the general weak time-series dependency in financial returns, we simply set the truncation parameter at three (months) to calculate long-run variance (e.g., Wooldridge (2016), Lazarus, Lewis, Stock, and Watson (2018)).[30]

For a given parameter $\gamma_{p,j}$ in γ_p ($j = 0$, mkt, smb or hml) and a given CSR estimator $\hat{\gamma}_{j,m}$ (i.e., OLS, OCSR, GLS, or WLS) in the m-th bootstrap sample, we measure estimation bias and

[30] A large body of work examines the performance of different heteroskedasticity- and autocorrelation-consistent (HAC) estimators. Examining these HAC estimators is beyond the scope of our paper. We choose something simple, following the recent advice in Wooldridge (2016) and Lazarus, Lewis, Stock, and Watson (2018). Note that whereas block size is set at 12 in our simulation to capture potential long-run dependence, we set the HAR truncation parameter conservatively to avoid data-peeking bias.

deviation with three metrics, defined as follows:

$$\begin{aligned}\text{Bias} &= M^{-1}\sum_{m=1}^{M}\hat{\gamma}_{j,m}-\gamma_{p,j},\\ \text{RMSE} &= \sqrt{M^{-1}\sum_{m=1}^{M}(\hat{\gamma}_{j,m}-\gamma_{p,j})^2},\\ \text{MAE} &= M^{-1}\sum_{m=1}^{M}|\hat{\gamma}_{j,m}-\gamma_{p,j}|,\end{aligned}$$

where RMSE and MAE stand for root mean-squared error and mean-absolute error, respectively.

We first focus on parameter estimates by reporting summary statistics on Bias, RMSE, and MAE of various CSR estimators. Table 1 reports the results for $T = 500$, and Table E.1 in Appendix E for $T = 750$. Note that our anomaly sample runs from July 1973 to December 2017, including 534 monthly observations, so $T = 500$ is close to the sample size of the actual data. $T = 750$ is considered a "large" sample experiment, where we increase $T = 500$ by 50%.

Focusing on $T = 500$ in Table 1, simulated bias for our OCSR is usually less than 10% of the magnitude of the true parameter value (with the exception of γ_0, whose true value is set at zero), and therefore does not seem to be the main contributor to estimation efficiency as measured by RMSE and MAE.[31] In terms of RMSE and MAE, OLS and WLS perform substantially worse than OCSR and GLS. In fact, WLS's performance is closer to OLS than to GLS, suggesting the importance of taking cross-asset correlations into account.

Comparing OCSR with GLS based on RMSE and MAE, OCSR stands out as the preferred method in most cases, as highlighted in bold in Table 1. The improvement of OCSR over GLS is case dependent. For example, for the Fama-French three-factor model with 38 test portfolios, as in Panel B, the percentage reduction in RMSE of OCSR relative to GLS ranges from 6% for γ_{mkt} (i.e., $-6\% = (0.215 - 0.228)/0.228$) to 19% for γ_{smb} (i.e., $-19\% = (0.118 - 0.146)/0.146$). The average reduction is 11%.

Turning to Table E.1 ($T = 750$) in Appendix E, $T = 750$ leads to further improved performance of OCSR compared with GLS, making OCSR the preferred choice in all but one specification. The average percentage reduction relative to GLS is also greater than in Table 1.

Overall, in terms of parameter estimation, our simulation results advocate the use of GLS and OCSR, given their large efficiency gain compared with OLS and WLS. Between GLS and OCSR,

[31] Note we focus on estimation efficiency both asymptotically and in relatively large samples and do not consider finite-sample bias adjustment, such as the ones considered in Shanken (1992).

for the modest ($T = 500$) to large T ($T = 750$) cases that we examine, OCSR seems to present sufficient efficiency gain relative to GLS to render it the preferred approach. We recommend the use of both in applications where it is unclear whether T can be regarded as large relative to N.

Table 1: **Simulated Bias, RMSE, and MAE for Parameter Estimates, $T = 500$.**

For a given Fama-French model (i.e., three-factor or five-factor model), we use the 18 low-turnover or the 38 low-turnover and medium-turnover anomaly sample in Novy-Marx and Velikov (2016) as test assets. γ_0, γ_{mkt}, γ_{smb}, γ_{hml}, γ_{cma}, and γ_{rmw} denote the risk premiums associated with the intercept, the market factor, *smb* (size factor), *hml* (value factor), *cma* (investment factor), and *rmw* (profitability factor), respectively. Bold denotes the best performer among all methods considered.

		Panel A: FF 3-Factor Model, $N = 18$				**Panel C: FF 5-Factor Model, $N = 18$**					
		γ_0	γ_{mkt}	γ_{smb}	γ_{hml}	γ_0	γ_{mkt}	γ_{smb}	γ_{hml}	γ_{cma}	γ_{rmw}
	True	0	0.595	0.257	0.339	0	0.595	0.257	0.339	0.287	0.323
OLS	Bias	0.097	-0.089	0.006	-0.048	0.005	0.005	0.013	-0.013	0.007	-0.026
	RMSE	0.319	0.339	0.195	0.200	0.278	0.277	0.190	0.182	0.167	0.270
	MAE	0.261	0.276	0.156	0.163	0.220	0.219	0.151	0.145	0.132	0.217
OCSR	Bias	0.012	0.009	0.004	-0.018	0.032	0.005	0.012	-0.006	0.008	-0.024
	RMSE	**0.200**	**0.237**	**0.128**	**0.170**	**0.220**	**0.231**	**0.128**	**0.178**	**0.146**	**0.191**
	MAE	**0.159**	0.194	**0.101**	**0.134**	**0.175**	0.186	**0.101**	**0.140**	**0.114**	**0.150**
GLS	Bias	0.006	0.006	0.005	-0.022	0.029	-0.016	0.007	-0.012	-0.002	-0.030
	RMSE	0.211	0.242	0.152	0.185	0.222	0.232	0.151	0.190	0.161	0.209
	MAE	0.168	**0.193**	0.122	0.148	0.177	**0.185**	0.121	0.151	0.127	0.167
WLS	Bias	0.089	-0.078	0.000	-0.049	0.058	-0.045	0.006	-0.010	0.002	-0.042
	RMSE	0.359	0.378	0.186	0.197	0.307	0.306	0.178	0.180	0.171	0.266
	MAE	0.291	0.306	0.149	0.161	0.247	0.245	0.142	0.142	0.135	0.216

		Panel B: FF 3-Factor Model, $N = 38$				**Panel D: FF 5-Factor Model, $N = 38$**					
		γ_0	γ_{mkt}	γ_{smb}	γ_{hml}	γ_0	γ_{mkt}	γ_{smb}	γ_{hml}	γ_{cma}	γ_{rmw}
	True	0	0.595	0.257	0.339	0	0.595	0.257	0.339	0.287	0.323
OLS	Bias	0.012	-0.014	0.012	-0.028	0.056	-0.049	-0.003	-0.011	-0.013	-0.038
	RMSE	0.234	0.242	0.209	0.185	0.245	0.239	0.181	0.214	0.173	0.241
	MAE	0.186	0.193	0.167	0.149	0.200	0.195	0.144	0.170	0.137	0.195
OCSR	Bias	0.024	0.015	0.015	-0.020	0.085	-0.034	0.003	-0.014	-0.005	-0.046
	RMSE	**0.173**	**0.215**	**0.118**	**0.155**	**0.199**	**0.212**	**0.120**	**0.172**	**0.136**	**0.162**
	MAE	**0.140**	**0.169**	**0.092**	**0.124**	**0.174**	**0.169**	**0.094**	**0.135**	**0.106**	**0.129**
GLS	Bias	0.024	-0.021	-0.001	-0.019	0.058	-0.054	-0.003	-0.003	-0.012	-0.041
	RMSE	0.192	0.228	0.146	0.172	0.207	0.226	0.142	0.176	0.150	0.178
	MAE	0.154	0.183	0.117	0.137	0.174	0.187	0.114	0.140	0.118	0.145
WLS	Bias	0.010	-0.003	0.002	-0.023	0.061	-0.054	-0.003	-0.008	-0.006	-0.051
	RMSE	0.213	0.248	0.178	0.181	0.238	0.253	0.167	0.213	0.171	0.251
	MAE	0.169	0.198	0.141	0.145	0.196	0.205	0.133	0.169	0.136	0.206

Next, we study the size and power of the hypothesis test on $H_0 : \gamma_{p,j} = 0$, where $j = 0$, *mkt*, *smb*, or *hml* using different CSR estimators. To perform hypothesis testing, we need to be specific about how standard errors are calculated. We use $(\widehat{\mathbf{X}}'\widehat{W}\widehat{\mathbf{X}})^{-1}(\widehat{\mathbf{X}}'\widehat{W}\widehat{\Omega}\widehat{W}\widehat{\mathbf{X}})(\widehat{\mathbf{X}}'\widehat{W}\widehat{\mathbf{X}})^{-1}$ to approximate the asymptotic variance-covariance matrix for GLS and WLS, where $\widehat{W}$ is set to be the aforementioned GLS and WLS weight matrix, respectively, and $\widehat{\Omega}$ follows the long-run variance estimator as explained previously. For OCSR, the asymptotic covariance matrix is given by $(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}$, which is estimated by $(\widehat{\mathbf{X}}'\widehat{\Omega}^{-1}\widehat{\mathbf{X}})^{-1}$. For OLS, we distinguish between three types of OLS standard-error estimates. One is the "naive" approach, which does not take beta estimation into account and simply uses OLS standard errors from the cross-sectional regression. We denote this approach as "OLS^{1stage}" to emphasize that it is a one-step estimator. The second is similar to how we calculate standard errors for GLS and WLS. In particular, we set $\widehat{W}$ as the identity matrix. The last type is the Fama-MacBeth standard-error estimate, where OLS is performed at each period to obtain the ex-post risk-premium estimates, and t-statistics are calculated based on the time-series of risk-premium estimates. Note that OLS^{1stage} is likely a strawman benchmark. We still include it in our analysis, because researchers sometimes apply the two-pass OLS without adjusting for beta uncertainty. Knowing to what degree this estimator is biased in a realistic simulation study is thus interesting.

Our results are reported in a sequence of tables in Appendix E (i.e., Table E.2-E.5).[32] A test may be powerful (under alternative hypotheses) while also being oversized (under the null hypothesis). To allow an apples-to-apples comparison across methods, we report both the original size and power (denoted as "Ori") and size-adjusted power (denoted as "Adj") where the statistical cutoff that exactly achieves a prespecified significance level is found and used to calculate the corresponding test power.[33]

In Tables E.2 and E.3, OLS^{1stage} is severely oversized, which is not surprising because, by ignoring beta uncertainty, standard errors are severely underestimated, leading to too many false rejections under the null. Other methods also seem to be somewhat oversized under the null but with a magnitude much smaller than OLS^{1stage}. Comparing test power with the size-adjusted power (so differences in test size across methods are taken into account), OCSR stands out as the most powerful among the tests we consider. The power improvement of OCSR compared with GLS—the second-best performer—is usually in the range of 5% to 10%. In comparison, the other three methods (FM, OLS, and WLS) have substantially lower power than OCSR and GLS.

For $T = 750$ (Table E.4-E.5), all methods have improved performance. The issue of oversized tests is mitigated, while OCSR continues to dominate others in terms of size-adjusted test power. Overall, consistent with our results on parameter estimates discussed previously, OCSR is the preferred model in terms of test power (both the original and the size-adjusted power) and is comparable to other methods in terms of test size.

[32]To save space, we focus on the Fama-French three-factor model. Results based on the Fama-French five-factor model are available upon request.

[33]See Harvey and Liu (2020a) for alternative ways to weight Type I and Type II errors.

4 Applications

We use our framework to sort out recent factor models proposed in FF and HXZ. Both papers rely on time-series regressions that judge a model's performance based on the intercepts (i.e., alphas) for a set of test assets. Fama and French (2018) term this the left-hand-side (LHS) approach.[34] We aim to provide alternative inference for the LHS approach using our new framework.

In particular, time-series tests can be mapped into an exactly identified GMM under the assumption that risk premiums equal factor means.[35] However, factors may be measured with errors (see, e.g., Roll (1977), Shanken (1987), and Jagannathan, Schaumburg, and Zhou (2010)), and the exactly identified GMM may not fully use all restrictions implied by supposedly all-encompassing factor models. We therefore turn to the cross-sectional approach. From the perspective of a GMM framework, we impose additional moment restrictions under mild assumptions on the candidate factor model while relaxing the pricing restriction on factors. This leads to an over-identified GMM, which is equivalent to our OCSR asymptotically. We use OCSR to make inference and contrast our results with alternative two-pass estimators as well as the usual time-series approach.

FF and HXZ use their own test portfolios to provide support for their respective models. To level the playing field, we use a comprehensive list of anomalies constructed and publicized by Chen and Zimmermann (2020). In total, 156 distinct anomalies are included. We study both equal-weighted and value-weighted anomaly long-short portfolios.

What additional moment restrictions can we bring in to make potentially more informative inference on factor models? Given that factor models purport to explain the returns of a large number of assets, it seems natural to assume that they should at least be able to explain the sorted portfolios based on which factors are constructed from, for example, the 20 size, book-to-market, operating profitability, and investment sorted portfolios for FF; and the 18 size, investment-to-assets, and return on equity triple sorted portfolios in HXZ. This assumption imposes mild economic restrictions on candidate factor models and poses a reasonable hurdle for candidate factor models (that aim to explain the cross-section of expected returns) to surpass (e.g., Fama

[34] Alternatively, Barillas and Shanken (2017) and Fama and French (2018) use spanning regressions to test nested models (i.e., the RHS approach).

[35] More specifically, for a single asset i, imposing the condition that risk premiums equal factor means, the following set of moment equations,

$$\begin{aligned} \mathbb{E}[f_t - \mu_f] &= 0, \\ \mathbb{E}\left[(r_{i,t} - (f_t - \mu_f)'\beta_i)(f_t - \mu_f)\right] &= 0, \\ \mathbb{E}[r_{i,t} - \alpha_i - \mu_f'\beta_i] &= 0, \end{aligned}$$

exactly identify the unknown parameters μ_f, β_i, and α_i, where $r_{i,t}$ denotes the excess return of asset i. As such, the GMM (or CSR) estimator of α_i equals the usual estimate from a time-series regression.

and French (1993)). We therefore make this assumption and use the 20 and 18 test portfolios to construct over-identifying restrictions and test FF and HXZ, respectively.[36] We refer to these portfolios as basis assets for a given candidate factor model.

Note that one can use the over-identification tests proposed in Appendix D.2 to directly test the above assumption on basis assets, which is exactly what we do later in this section. However, from the perspective of an exploratory data analysis, it is interesting to contrast the usual time-series estimates with the cross-sectional estimates. A large discrepancy among them indicates potential problems with the over-identification assumption, which leads to more rigorous over-identification tests. We therefore follow this route to present our empirical findings.

The model we use in this section combines insights from both cross-sectional and time-series regressions as discussed previously. In particular, suppose N_0 basis assets exist (e.g., the 20 sorted portfolios used in the construction of FF). Let the N-th asset be the asset to be tested, where $N = N_0 + 1$. In cross-sectional regressions, we have the usual regressor of $\mathbf{1}_N$ to capture the potential difference between the zero-beta rate and the risk-free rate (because we use excess instead of gross returns for assets) and the matrix of factor loadings $\widehat{\mathbf{B}}$. We introduce another variable to capture potential model misspecification (i.e., abnormal alpha) by asset N, given by $S_N \equiv (\mathbf{0}_{1\times N_0}, 1)'$. The stacked matrix of regressors is given by $(S_N, \mathbf{1}_N, \widehat{\mathbf{B}})$. Let the associated regression coefficient be

$$\hat{\theta} = (\hat{\alpha}_N, \hat{\gamma}_0, \hat{\gamma}_1')',$$

where $\hat{\alpha}_N$ (scalar) is the slope coefficient for S_N, $\hat{\gamma}_0$ (scalar) is the estimated zero-beta rate minus the risk-free rate, and $\hat{\gamma}_1$ is the vector of factor-premium estimates. Our final estimate for alpha, which is the cross-sectional counterpart to the time-series alpha estimate, is given by

$$\widehat{TA} = l' \times \hat{\theta}, \tag{4.1}$$

where $l = (1, 1, \mathbf{0}_{1\times(d_\theta-2)})'$ (i.e., non-zero for only the first two elements), and $\widehat{TA}$ stands for "total alpha," which is to distinguish from our alpha definition in the theory section.[37] Although we may

[36] Our data for both factors and factor portfolios are obtained from Ken French's online data library and Lu Zhang's recently publicized data library on investment CAPM.

[37] Note that whether we define the estimate of alpha as $l' \times \widehat{\theta}$ or simply $\widehat{\alpha}_N$ will not have a material impact on our analysis of the FF model, because the mean absolute estimate for γ_0 (across 156 regressions) is 0.08% (per month), with a standard deviation of 0.08%. Hence, the difference between the zero-beta rate and the risk-free rate is economically small for the FF model. By contrast, the corresponding numbers are 0.67% and 0.11% for HXZ, suggesting a large discrepancy between the zero-beta rate and the risk-free rate. Given our emphasis on the FF model for which the over-identification test is not rejected, the estimate of γ_0 is not the main contributor to our results on hypothesis results reported later in Table 4.

use alpha and total alpha interchangeably in our follow-up discussion, our precise definition follows Eq. (4.1) for the rest of this section. Because $l' \times \hat{\theta}$ is a linear transformation of the cross-sectional estimates $\hat{\theta}$, our theory developed in (D.4) of Appendix D.1 can be straightforwardly applied to make inference on $\widehat{TA}$.

We take a preliminary look at our results by contrasting the time-series OLS estimated alphas and alpha t-statistics (i.e., the approach taken by FF and HXZ) with estimates based on our OCSR, where Figure 1 shows results for value-weighted portfolios and Figure 2 for equal-weighted portfolios. In each figure and for each panel, the solid curve shows the OLS alphas (alpha t-statistics) in ascending order, whereas the dashed line plots the corresponding OCSR alphas (alpha t-statistics).

Focusing on the left two panels in Figure 1 (i.e., FF), OCSR alphas and alpha t-statistics center around their OLS counterparts across anomalies. A discrepancy exists for some anomalies. For example, focusing on t-statistics, the dashed curve sometimes differs from the solid curve by a magnitude of 2.0, suggesting a large difference in t-statistics. But overall, estimates from OCSR and OLS are clearly positively correlated.

Turning to the right two panels in Figure 1 (i.e., HXZ), although OCSR estimates still center around OLS estimates, their positive correlation is blurred by the high level of dispersion for the OCSR estimates around their OLS counterparts. Indeed, focusing on t-statistics, several anomalies that are deemed significant based on the solid curve (i.e., absolute OLS t-statistic exceeding 2.0) have OCSR t-statistics that fall in a neighborhood of zero, rendering these anomalies insignificant with our estimator. The large discrepancy between OCSR and OLS is even more pronounced for equal-weighted portfolios, as shown in Figure 2.

What is causing this difference between FF and HXZ in creating the discrepancy between time-series OLS and our OCSR? Because our OCSR (corresponding to an over-identified GMM) augments the time-series OLS (corresponding to an exactly identified GMM) with additional moment restrictions, these moment restrictions may not hold in the data. To this end, we examine the over-identification assumptions for the basis assets in FF and HXZ through OCSR (i.e., the test in (D.9)). To highlight the result for each individual basis asset, we also report results for individual time-series regressions that project each basis asset onto the associated factor model. Table 2 reports the results.

We see a sharp contrast between FF and HXZ in explaining their own basis assets. For FF, the average absolute t-statistic for intercept is 1.11, whereas for HXZ, it is 2.47. Only two (zero, out of 20) basis assets have an absolute t-statistic exceeding 2.0 (3.0). By contrast, the corresponding number is nine (eight, out of 18) for HXZ. The specification test based on (D.8) does not reject

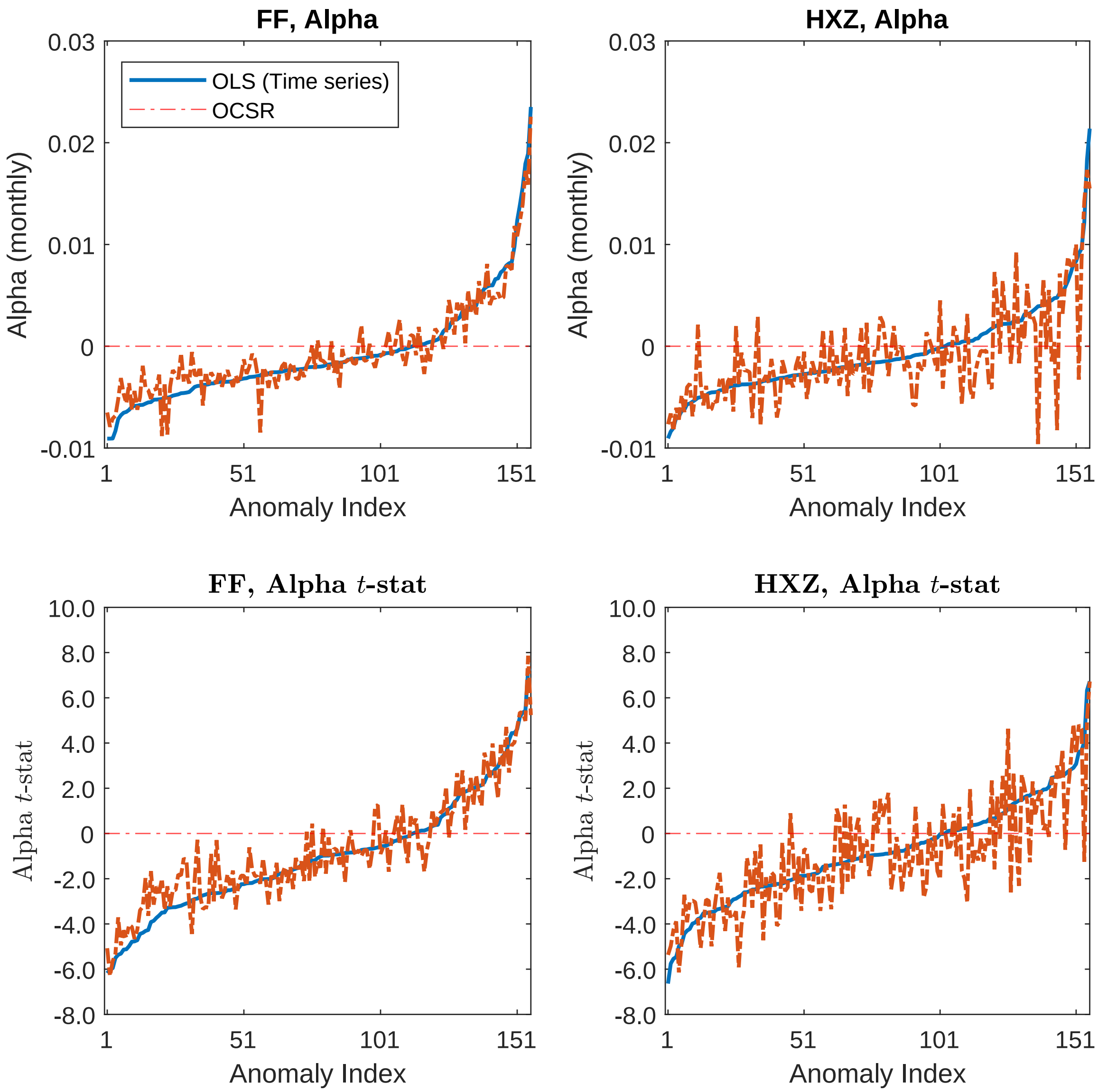

Figure 1: **Time-Series and Cross-Sectional Tests with Value-Weighted Test Portfolios.** For each of the 156 value-weighted long-short anomaly portfolios in Chen and Zimmermann (2020), we perform both the usual time-series OLS and our OCSR to estimate anomaly alpha. For OCSR, we use the respective 20 sorted portfolios (for FF) and the 18 sorted portfolios (for HXZ) as basis assets and estimate one anomaly alpha at a time. We report the (total) alpha estimates as well as the corresponding t-statistics. For ease of presentation, we sort anomalies by their time-series alpha estimates (or alpha t-statistics) in ascending order.

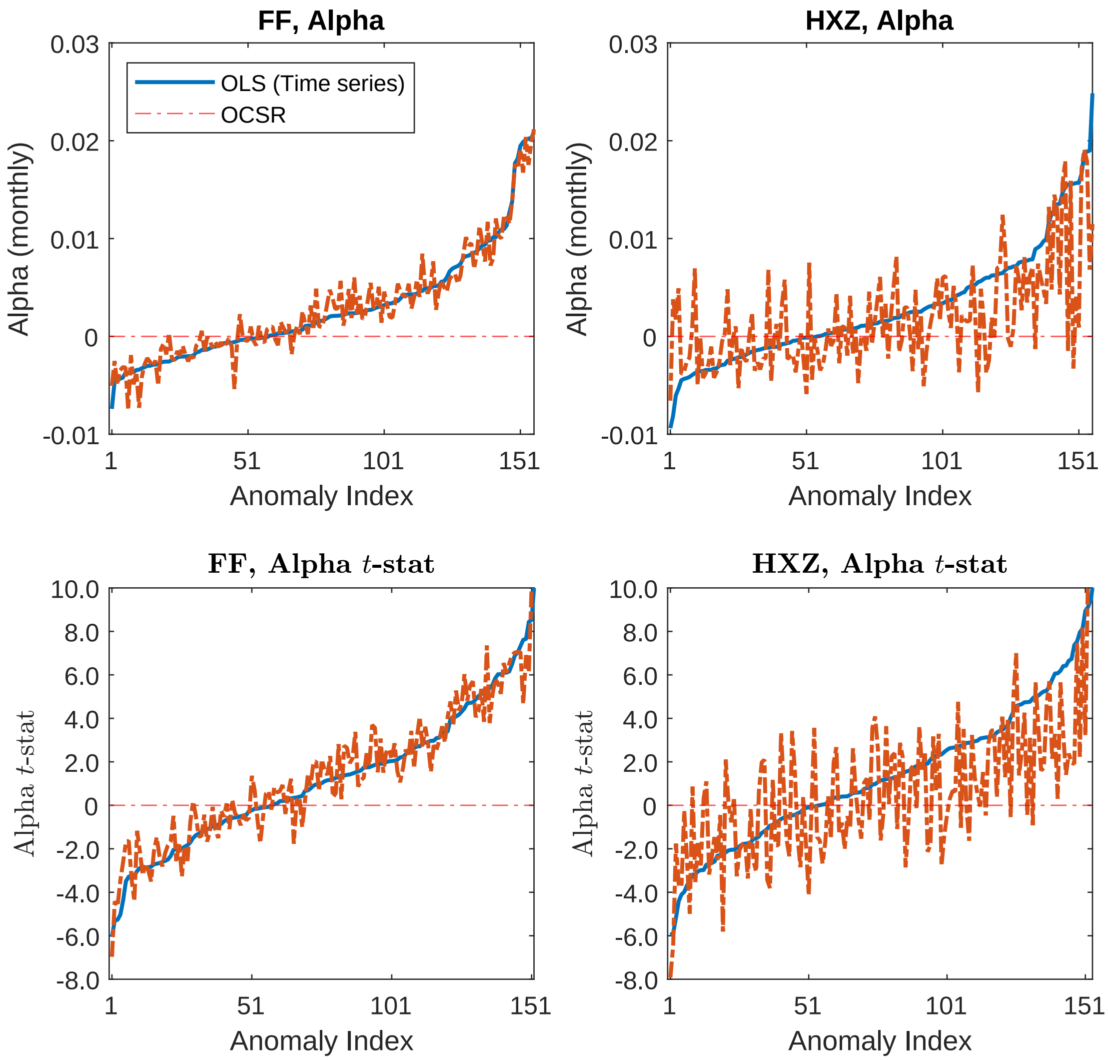

Figure 2: **Time-Series and Cross-Sectional Tests with Equal-Weighted Test Portfolios.** For each of the 156 equal-weighted long-short anomaly portfolios in Chen and Zimmermann (2020), we perform both the usual time-series OLS and our OCSR to estimate anomaly alpha. For OCSR, we use the respective 20 sorted portfolios (for FF) and the 18 sorted portfolios (for HXZ) as basis assets and estimate one anomaly alpha at a time. We report the (total) alpha estimates as well as the corresponding t-statistics. For ease of presentation, we sort anomalies by their time-series alpha estimates (or alpha t-statistics) in ascending order.

Table 2: **FF and HXZ: Explaining Their Respective Basis Assets**

We perform time-series regressions for the 20 (18) basis assets for FF (HXZ) against the FF (HXZ) model. We report summary statistics for absolute alpha estimates and absolute alpha t-statistics. We also perform specification tests based on (D.9) for the FF model and HXZ model using their respective basis assets.

	FF	HXZ
Absolute Alpha		
Mean	0.047	0.177
Stdev.	0.036	0.109
Max	0.130	0.395
Absolute Alpha t-stat		
Mean	1.107	2.470
Stdev.	1.325	2.730
Max	2.654	6.209
$\#(> 2.0)$	2	9
$\#(> 3.0)$	0	8
Specification Tests		
J-stat	20.310	105.176
Degree of freedom	14	13
[p-value]	[0.121]	[0.000]

the FF model (p-value $= 0.12$) but does so firmly for the HXZ model (p-value $= 0.00$).

The issue we uncover for HXZ goes beyond the theme of our paper. If factor models purport to explain the cross-section of expected returns, a reasonable sanity check that imposes minimal economic restrictions is to require that candidate models explain the returns of basis assets based on which factors are constructed. Given that HXZ clearly fails this check, whereas FF broadly survives, additional model-comparison exercises that seek to further compare HXZ and FF based on additional anomaly portfolios seem to put FF at a disadvantage.

Regardless of the general implication of the above finding, our results highlight the additional insights gained through our OCSR using over-identified pricing restrictions. Through contrasting our OCSR with the usual time-series OLS, we not only can identify differences in estimates due to the differences in the underlying economic assumptions we make (i.e., perfectly measured factors and no over-identification restrictions for time-series OLS vs. potentially imperfectly measured factors and over-identification restrictions based on basis assets) but may also detect failures in over-identification conditions. Such failures should be taken into count when comparing alternative factor models.

Table 3: **Factor-Premium Estimates: Time Series versus Cross-Sectional**

We report the time-series and cross-sectional factor-premium estimates. Time-series factor premiums are calculated as the time-series means of factor returns. For cross-sectional estimates and for a given model (e.g., the FF five-factor model), we perform 156 sets of cross-sectional regressions, each set using the same 20 basis assets and one anomaly portfolio (from Chen and Zimmermann (2020)). For the HXZ four-factor model, we use its associated 18 basis assets. We report results for six cross-sectional estimators: OLS^{1stage} corresponds to the cross-sectional OLS estimator that ignores first-stage beta-estimation uncertainty, FM is the Fama-MacBeth estimator where risk premiums are estimated for each period and simple t-statistics for the time series of risk-premium estimates are used for hypothesis testing, OLS is the cross-sectional OLS estimator that takes first-stage beta-estimation uncertainty into account through (2.8), $OCSR$ is our proposed estimator, GLS is the two-pass estimator that uses the variance-covariance matrix for factor-model residuals as the weighting matrix and has standard errors calculated through (2.8), and WLS is the two-pass estimator that sets the off-diagonal elements of GLS's weighting matrix at zero and has standard errors calculated through (2.8). γ_{mkt}, γ_{smb}, γ_{hml}, γ_{cma}, and γ_{rmw} denote the risk premiums associated with the market factor, *smb* (size factor), *hml* (value factor), *cma* (investment factor), and *rmw* (profitability factor) for the FF five-factor model. γ_{mkt}^{HXZ}, γ_{smb}^{HXZ}, γ_{ia}, and γ_{roe} denote the risk premiums associated with HXZ's market factor, *smb* (size factor), *ia* (investment), and *roe* (return on equity) for the HXZ four-factor model.

Method		FF					HXZ			
		γ_{mkt}	γ_{smb}	γ_{hml}	γ_{rmw}	γ_{cma}	γ_{mkt}^{HXZ}	γ_{smb}^{HXZ}	γ_{ia}	γ_{roe}
$OLS(TS)$	Mean	0.519	0.250	0.363	0.279	0.330	0.503	0.309	0.410	0.543
	Stdev.	0.185	0.126	0.118	0.092	0.083	0.185	0.126	0.077	0.104
OLS^{1stage}	Mean of Estimates	0.865	0.253	0.257	0.317	0.204	0.212	0.298	0.324	0.487
	Mean of Stdev.	0.445	0.140	0.146	0.126	0.108	0.396	0.135	0.091	0.116
FM	Mean of Estimates	0.865	0.253	0.257	0.317	0.204	0.212	0.298	0.324	0.487
	Mean of Stdev.	0.391	0.134	0.132	0.110	0.096	0.396	0.135	0.091	0.116
OLS	Mean of Estimates	0.865	0.253	0.257	0.317	0.204	0.212	0.298	0.324	0.487
	Mean of Stdev.	0.302	0.037	0.050	0.062	0.050	0.856	0.101	0.112	0.120
$OCSR$	Mean of Estimates	0.794	0.220	0.199	0.217	0.192	-0.030	0.263	0.232	0.475
	Mean of Stdev.	0.383	0.126	0.133	0.108	0.095	0.347	0.133	0.086	0.111
GLS	Mean of Estimates	0.596	0.232	0.277	0.293	0.258	-0.019	0.296	0.395	0.538
	Mean of Stdev.	0.411	0.137	0.145	0.122	0.104	0.344	0.135	0.088	0.113
WLS	Mean of Estimates	0.695	0.224	0.241	0.305	0.207	0.201	0.339	0.334	0.512
	Mean of Stdev.	0.451	0.137	0.148	0.123	0.107	0.376	0.136	0.090	0.117

Given the inability of HXZ to explain its basis assets, we exert caution in interpreting our results for HXZ in our followup analysis. We next report the factor-premium estimates for OCSR and contrast them with the time-series estimates from the time-series OLS, as well as other cross-sectional approaches. The results are reported in Table 3.[38]

Focusing on FF in Table 3, we see that risk-premium estimates based on time series have large standard errors. For example, whereas the mean market risk premium is estimated at 0.519% (per month), the associated standard error is 0.185%. Using cross-sectional regressions, risk-premium estimates can differ from time-series estimates. For example, market risk premium is estimated at 0.794% for OCSR, which is somewhat higher than the time-series estimate of 0.519% but still falls within the 95% confidence band based on the time-series estimates (it is about 1.5 standard error above the mean estimate). We consider our cross-sectional estimates largely consistent with the time-series estimates (see, e.g., Lewellen, Nagel, and Shanken (2010)). Large differences also exist among cross-sectional methods, highlighting the impact of different weighting schemes. For example, cross-sectional OLS generates an estimate of 0.865% for the market risk premium, whereas the GLS-implied market risk premium is 0.596%.

Turning to HXZ, the failure of over-identification conditions (equivalently, the model is misspecified) is manifested through the large discrepancy between time-series OLS estimates and cross-sectional estimates. In particular, the estimate of the market risk premium for cross-sectional methods is only around 0.200% (OLS and WLS) or even negative (OCSR and GLS). This finding implies that market betas do not help explain the returns of HXZ's basis assets. In fact, our results in Table 2 show that the returns for many basis assets in HXZ cannot be explained by any of HXZ's factors.

[38]With 156 anomalies, we run 156 regressions for cross-sectional approaches. Because only one asset differs across the 156 regressions, risk-premium estimates are similar across regressions. We therefore report the average estimate and the average standard deviation of the estimate. In addition, anomalies can be value-weighted or equal-weighted. But because the set of basis assets is the same across regressions, whether the single test asset is value-weighted or equal-weighted does not cause much difference in the premium estimates. We therefore only report results based on value-weighted anomalies.

Table 4: **Alpha Estimates: Time-Series versus Cross-Sectional**

We report summary statistics on alpha estimates. $OLS(TS)$ corresponds to the usual time-series alpha estimates associated with a factor model. For cross-sectional estimates and for a given model (e.g., the FF five-factor model), we perform 156 sets of cross-sectional regressions, each set using the same 20 basis assets and one anomaly portfolio (from Chen and Zimmermann (2020)). For the HXZ four-factor model, we use its associated 18 basis assets. We report results for six cross-sectional estimators: OLS^{1stage} corresponds to the cross-sectional OLS estimator that ignores first-stage beta-estimation uncertainty; FM is the Fama-MacBeth estimator, where risk premiums are estimated for each period and simple t-statistics for the time series of risk-premium estimates are used for hypothesis testing; OLS is the cross-sectional OLS estimator that takes first-stage beta-estimation uncertainty into account through (2.8); $OCSR$ is our proposed estimator; GLS is the two-pass estimator that uses the variance-covariance matrix for factor-model residuals as the weighting matrix and has standard errors calculated through (2.8); and WLS is the two-pass estimator that sets the off-diagonal elements of GLS's weighting matrix at zero and has standard errors calculated through (2.8).

Method	FF					HXZ				
	Abs. Alpha		Abs. t-stat		$\#(\lvert t\text{-stat}\rvert > 2)$	Abs. Alpha		Abs. t-stat		$\#(\lvert t\text{-stat}\rvert > 2)$
	Mean	Stdev.	Mean	Stdev.		Mean	Stdev.	Mean	Stdev.	
Panel A: Value-Weighted Anomaly Returns										
$OLS(TS)$	0.382	0.352	2.264	1.628	79	0.321	0.285	1.888	1.411	64
OLS^{1stage}	0.373	0.365	4.541	4.110	115	0.309	0.275	1.153	1.103	15
FM	0.373	0.365	2.205	1.628	76	0.309	0.275	1.951	1.469	68
OLS	0.373	0.365	1.945	1.452	66	0.309	0.275	1.714	1.319	53
$OCSR$	0.349	0.326	2.105	1.482	71	0.361	0.286	2.122	1.432	74
GLS	0.350	0.327	1.998	1.389	64	0.303	0.256	1.759	1.313	60
WLS	0.357	0.350	1.880	1.410	61	0.315	0.274	1.753	1.334	56
Panel B: Equal-Weighted Anomaly Returns										
$OLS(TS)$	0.429	0.465	2.911	2.985	82	0.458	0.476	2.891	2.701	86
OLS^{1stage}	0.468	0.465	6.119	6.182	118	0.442	0.466	1.675	1.777	39
FM	0.468	0.465	3.163	3.073	93	0.442	0.466	2.997	2.961	87
OLS	0.468	0.465	2.655	2.712	74	0.442	0.466	2.376	2.475	71
$OCSR$	0.445	0.454	3.006	3.110	84	0.397	0.381	2.667	2.825	78
GLS	0.448	0.461	2.832	2.941	80	0.424	0.442	2.499	2.569	78
WLS	0.455	0.462	2.599	2.736	71	0.444	0.465	2.382	2.444	72

Finally, Table 4 provides detailed statistics on the performance of FF and HXZ against both value-weighted and equal-weighted test assets. Several remarks follow. We focus on FF to interpret our results. First, time-series OLS and OCSR can lead to a substantial difference in the testing outcome. For example, under value-weighted test assets, 79 anomalies do not survive the 2.0 t-statistic cutoff under time-series OLS. The corresponding number for OCSR is 71. Therefore, fewer strategies are rejected when pricing restrictions are imposed on not only FF factors but also on FF basis assets.

Second, among cross-sectional approaches, the naive cross-sectional OLS (i.e., "OLS^{1stage}"), which ignores beta-estimation uncertainty, leads to a much larger number of rejections. This finding is consistent with our extensive simulation evidence on the over-rejection of the naive OLS and with the results provided in Shanken (1992) and Jagannathan and Wang (1998). Our results substantiate the over-rejection concern regarding the unadjusted Fama-MacBeth approach as argued in Shanken (1992). The degree of over-rejection for the naive OLS is so large that it makes hypothesis tests for cross-sectional regressions inappropriate. We therefore follow Shanken (1992) and Jagannathan and Wang (1998) and recommend standard errors adjusted for first-stage beta estimation. In fact, a cross-sectional OLS that applies the same weighting scheme as OLS^{1stage} but uses adjusted standard errors seems to be able to reduce the number of false rejections substantially: the number of rejections goes down from 115 (for OLS^{1stage}) to 66 (OLS).

Similar to OLS^{1stage}, the usual Fama-MacBeth estimator (FM) also tends to over reject (albeit to a lesser extent compared to OLS^{1stage}), consistent with the evidence in our simulation study. This is again because FM has incorrectly specified standard errors for risk premium estimates.

Third, among cross-sectional methods that have correctly specified standard errors (i.e., OLS, OCSR, GLS, and WLS), our OCSR seems most powerful in detecting abnormal alphas. For example, for value-weighted test portfolios, OCSR identifies 71 rejections, whereas GLS (WLS) identifies 64 (61). This is consistent with our simulation evidence showing the power of OCSR in comparison with other cross-sectional methods.

Finally, note that whereas HXZ results in fewer rejections than FF, that is, 64 rejections (HXZ) versus 79 (FF) for value-weighted portfolios (consistent with the evidence in Hou, Xue, and Zhang (2015)), this result is fragile for the following reasons: (1) Using cross-sectional methods, HXZ leads to a larger number (i.e., 74) of rejections than FF (71); and (2) Under equal weighting, HXZ leads to a larger number of rejections (86) than FF (82) even with the time-series OLS. However, given that the cross-sectional model for HXZ is likely misspecified, we do not place a high weight on the cross-sectional estimates for HXZ.

To summarize, we revisit the recent FF versus HXZ debate using our OCSR. For HXZ, the contrast in results between the usual time-series OLS and OCSR allows us to uncover its inadequacy in pricing its own basis assets, which casts doubts on HXZ as an all-encompassing factor model. For FF, for which over-identification conditions for basis assets approximately hold, we find substantial differences between the time-series OLS as in Fama and French (2015) and our OCSR. We also find differences in test outcomes between OCSR and other cross-sectional approaches proposed by the previous literature (e.g., GLS) and highlight the gain in test power of our approach.

Note that although our narrative has focused on the difference in results between the time-series OLS and OCSR, several other advantages of OCSR (or cross-sectional approaches in general) are worth emphasizing. For example, when the zero-beta rate is truly different from the risk-free rate, OCSR can easily distinguish between the overall alpha and the alpha in addition to the (common) zero-beta rate (i.e., $\widehat{TA}$ vs. $\widehat{\alpha}_N$), whereas the time-series OLS has to lump them into a single intercept. Additionally, firm characteristics can be straightforwardly incorporated into OCSR to allow potential model misspecification and enrich testable hypotheses (see our general model specification in Section 2), whereas for the time-series OLS, allowing for firm characteristics that mainly vary in the cross-section is challenging. We leave the examination of these interesting extensions of OCSR (not limited to the comparison of FF and HXZ) to future research.

5 Conclusion

Our paper builds a strong link between the GMM approach and the popular two-pass regression approach. We show, in the context of linear-beta pricing models, that the two-pass regression can be constructed to achieve the same asymptotic efficiency as the optimally weighted GMM estimator. Hence, the sequential nature of the two-pass estimator does not make it inherently suboptimal compared with the one-step GMM approach. On the other hand, the challenge in implementing the nonlinear and large-dimensional GMM estimation can be surpassed by using the tractable two-pass estimator, which will likely facilitate the application of our approach in empirical research.

Our general idea of mapping two-pass estimators into the well-studied GMM framework is useful for future research. For example, it may help digest recently proposed cross-sectional estimators that are best suited to conduct inference with a large cross-section of individual stocks or anomaly portfolios. Although we focus on the fixed- (or slowly divergent-) N and large-T asymptotics for our estimator, deriving the large-N and large-T asymptotics to address the challenge of having a large cross-section is also possible. We plan to explore these interesting extensions in future research.

References

Ahn, S. C., and C. Gadarowski (1999): “Two-pass cross-sectional regression of factor pricing models: Minimum distance approach,” *Available at SSRN 191970*.

Ai, H., J. E. Li, K. Li, and C. Schlag (2020): “The collateralizability premium,” *The Review of Financial Studies*, 33(12), 5821–5855.

Andrews, D. W. K. (1991): “Heteroskedasticity and autocorrelation consistent covariance matrix estimation,” *Econometrica*, 59(3), 817–58.

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

Balduzzi, P., and C. Robotti (2008): “Mimicking portfolios, economic risk premia, and tests of multi-beta models,” *Journal of Business & Economic Statistics*, 26(3), 354–368.

Barillas, F., and J. Shanken (2017): “Which alpha?,” *The Review of Financial Studies*, 30(4), 1316–1338.

Barone-Adesi, G., and P. P. Talwar (1983): “Market models and heteroscedasticity of residual security returns,” *Journal of Business & Economic Statistics*, 1(2), 163–168.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988): “A capital asset pricing model with time-varying covariances,” *Journal of political Economy*, 96(1), 116–131.

Breeden, D. T. (1979): “An intertemporal asset pricing model with stochastic consumption and investment opportunities,” *Journal of Financial Economics*, 7(3), 265–296.

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

Chaieb, I., H. Langlois, and O. Scaillet (2020): “Factors and risk premia in individual international stock returns,” *Forthcoming, Journal of Financial Economics*, (18-04).

Chamberlain, G. (1987): “Asymptotic efficiency in estimation with conditional moment restrictions,” *Journal of Econometrics*, 34(3), 305–334.

Chen, A. Y., and T. Zimmermann (2020): “Publication bias and the cross-section of stock returns,” *The Review of Asset Pricing Studies*, 10(2), 249–289.

Chen, L., M. Pelger, and J. Zhu (2020): "Deep learning in asset pricing," *Available at SSRN 3350138.*

Christopherson, J. A., W. E. Ferson, and D. A. Glassman (1998): "Conditioning manager alphas on economic information: Another look at the persistence of performance," *The Review of Financial Studies*, 11(1), 111–142.

Cochrane, J. H. (2009): *Asset Pricing: Revised Edition.* Princeton University Press.

Cong, L. W., K. Tang, J. Wang, and Y. Zhang (2020): "AlphaPortfolio for investment and economically interpretable AI," *Available at SSRN 3554486.*

Croce, M. M., T. Marchuk, and C. Schlag (2019): "The leading premium," Discussion paper, National Bureau of Economic Research.

Davidson, J. (1994): *Stochastic Limit Theory.* Oxford University Press.

De Nard, G., O. Ledoit, and M. Wolf (2018): "Factor models for portfolio selection in large dimensions: The good, the better and the ugly," *Journal of Financial Econometrics.*

Diebold, F. X., S. C. Lim, and C. J. Lee (1993): "A note on conditional heteroskedasticity in the market model," *Journal of Accounting, Auditing & Finance*, 8(2), 141–150.

Dittmar, R. F. (2002): "Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns," *The Journal of Finance*, 57(1), 369–403.

Ehsani, S., and J. T. Linnainmaa (2019): "Factor momentum and the momentum factor," Discussion paper, National Bureau of Economic Research.

Engle, R. F., O. Ledoit, and M. Wolf (2019): "Large dynamic covariance matrices," *Journal of Business & Economic Statistics*, 37(2), 363–375.

Fama, E. F., and K. R. French (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33(1), 3–56.

——— (2010): "Luck versus skill in the cross-section of mutual fund returns," *The Journal of Finance*, 65(5), 1915–1947.

——— (2015): "A five-factor asset pricing model," *Journal of financial economics*, 116(1), 1–22.

——— (2018): "Choosing factors," *Journal of financial economics*, 128(2), 234–252.

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

Feng, G., S. Giglio, and D. Xiu (2020): "Taming the factor zoo: A test of new factors," *The Journal of Finance*, 75(3), 1327–1370.

Ferson, W. E., S. Sarkissian, and T. Simin (2006): "Asset pricing models with conditional betas and alphas: The effects of data snooping and spurious regression," Discussion paper, National Bureau of Economic Research.

Ferson, W. E., and R. W. Schadt (1996): "Measuring fund strategy and performance in changing economic conditions," *The Journal of Finance*, 51(2), 425–461.

Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2015): "Dynamic factor models with infinite-dimensional factor spaces: One-sided representations," *Journal of Econometrics*, 185(2), 359–371.

Frazier, D. T., and E. Renault (2017): "Efficient two-step estimation via targeting," *Journal of Econometrics*, 201(2), 212–227.

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

Gagliardini, P., E. Ossola, and O. Scaillet (2016): "Time-varying risk premium in large cross-sectional equity data sets," *Econometrica*, 84(3), 985–1046.

——— (2019): "A diagnostic criterion for approximate factor structure," *Journal of Econometrics*, 212(2), 503–521.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989): "A test of the efficiency of a given portfolio," *Econometrica: Journal of the Econometric Society*, pp. 1121–1152.

Giglio, S., and D. Xiu (2019): "Asset pricing with omitted factors," *Chicago Booth Research Paper*, (16-21).

Giglio, S., D. Xiu, and D. Zhang (2021): "Test assets and weak factors," *Chicago Booth Research Paper*.

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

——— (2017): "Spurious inference in reduced-rank asset-pricing models," *Econometrica*, 85(5), 1613–1628.

——— (2019): "Too good to be true? Fallacies in evaluating risk factor models," *Journal of Financial Economics*, 132(2), 451–471.

Gu, S., B. Kelly, and D. Xiu (2020): "Empirical asset pricing via machine learning," *The Review of Financial Studies*, 33(5), 2223–2273.

——— (2021): "Autoencoder asset pricing models," *Journal of Econometrics*, 222(1), 429–450.

Hall, P., and C. Heyde (1980): *Martingale Limit Theory and its Application*, Probability and Mathematical Statistics: A Series of Monographs and Textbooks. Academic Press.

Hansen, L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50(4), 1029–1054.

Harvey, C. R., and Y. Liu (2020a): "False (and missed) discoveries in financial economics," *The Journal of Finance*.

——— (2020b): "Luck versus Skill in the Cross-Section of Mutual Fund Returns: Reexamining the Evidence," *Available at SSRN 3623537*.

Harvey, C. R., Y. Liu, and H. Zhu (2016): "... And the cross-section of expected returns," *Review of Financial Studies*, 29, 5–68.

Harvey, C. R., and A. Siddique (2000): "Conditional skewness in asset pricing tests," *The Journal of Finance*, 55(3), 1263–1295.

Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2016): "The common factor in idiosyncratic volatility: Quantitative asset pricing implications," *Journal of Financial Economics*, 119(2), 249–283.

Hou, K., C. Xue, and L. Zhang (2015): "Digesting anomalies: An investment approach," *The Review of Financial Studies*, 28(3), 650–705.

Huang, D., J. Li, and G. Zhou (2019): "Shrinking factor dimension: A reduced-rank approach," *Available at SSRN 3205697*.

Jagannathan, R., E. Schaumburg, and G. Zhou (2010): "Cross-sectional asset pricing tests," *Annual Review of Financial Economics*, 2(1), 49–74.

Jagannathan, R., G. Skoulakis, and Z. Wang (2010): "The analysis of the cross-section of security returns," in *Handbook of Financial Econometrics: Applications*, pp. 73–134. Elsevier.

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

——— (2002): "Empirical evaluation of asset-pricing models: a comparison of the SDF and beta methods," *The Journal of Finance*, 57(5), 2337–2367.

Jegadeesh, N., J. Noh, K. Pukthuanthong, R. Roll, and J. Wang (2019): "Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation," *Journal of Financial Economics*, 133(2), 273–298.

Kan, R., and C. Robotti (2012): "Evaluation of asset pricing models using two-pass cross-sectional regressions," in *Handbook of Computational Finance*, pp. 223–251. Springer.

Kan, R., C. Robotti, and J. Shanken (2013): "Pricing model performance and the two-pass cross-sectional regression methodology," *The Journal of Finance*, 68(6), 2617–2649.

Kan, R., and C. Zhang (1999): "Two-pass tests of asset pricing models with useless factors," *the Journal of Finance*, 54(1), 203–235.

Kan, R., and G. Zhou (1999): "A critique of the stochastic discount factor methodology," *The Journal of Finance*, 54(4), 1221–1248.

——— (2001): "Empirical asset pricing: The beta method versus the stochastic discount factor method," *Working paper*.

Kandel, S., and R. F. Stambaugh (1987): "On correlations and inferences about mean-variance efficiency," *Journal of Financial Economics*, 18(1), 61–90.

——— (1995): "Portfolio inefficiency and the cross-section of expected returns," *The Journal of Finance*, 50(1), 157–184.

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

Kim, D. (1995): "The errors in the variables problem in the cross-section of expected stock returns," *The Journal of Finance*, 50(5), 1605–1634.

Kim, S., and G. Skoulakis (2018): "Ex-post risk premia estimation and asset pricing tests using large cross sections: The regression-calibration approach," *Journal of Econometrics*, 204(2), 159–188.

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

Kleibergen, F., and Z. Zhan (2015): "Unexplained factors and their effects on second pass R-squared's," *Journal of Econometrics*, 189(1), 101–116.

Kozak, S., S. Nagel, and S. Santosh (2018): "Interpreting factor models," *The Journal of Finance*, 73(3), 1183–1223.

——— (2020): "Shrinking the cross-section," *Journal of Financial Economics*, 135(2), 271–292.

Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson (2018): "HAR inference: Recommendations for practice," *Journal of Business & Economic Statistics*, 36(4), 541–559.

Lettau, M., and M. Pelger (2020): "Factors that fit the time series and cross-section of stock returns," *The Review of Financial Studies*, 33(5), 2274–2325.

Lewellen, J., S. Nagel, and J. Shanken (2010): "A skeptical appraisal of asset pricing tests," *Journal of Financial Economics*, 96(2), 175–194.

Li, J., and Z. Liao (2020): "Uniform Nonparametric Inference for Time Series," *Journal of Econometrics*, 219(1), 38–51.

Lin, X., B. Palazzo, and F. Yang (2020): "The risks of old capital age: Asset pricing implications of technology adoption," *Journal of Monetary Economics*, 115, 145–161.

Litzenberger, R. H., and K. Ramaswamy (1979): "The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence," *Journal of Financial Economics*, 7(2), 163–195.

MacKinlay, A. C., and M. P. Richardson (1991): "Using generalized method of moments to test mean-variance efficiency," *The Journal of Finance*, 46(2), 511–527.

Merton, R. C. (1973): "An intertemporal capital asset pricing model," *Econometrica: Journal of the Econometric Society*, pp. 867–887.

Newey, W. K., and D. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 4, 2111–2245.

Newey, W. K., and K. D. West (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55(3), 703–08.

Novy-Marx, R., and M. Velikov (2016): "A taxonomy of anomalies and their trading costs," *The Review of Financial Studies*, 29(1), 104–147.

Politis, D. N., and J. P. Romano (1994): "The stationary bootstrap," *Journal of the American Statistical Association*, 89(428), 1303–1313.

Pollard, D. (2002): *A User's Guide to Measure Theoretic Probability*, vol. 8 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

Pukthuanthong, K., R. Roll, J. Wang, and T. Zhang (2021): "Testing asset pricing model with non-traded factors: A new method to resolve (measurement/econometric) issues in factor-mimicking portfolio," *Working Paper*.

Raponi, V., C. Robotti, and P. Zaffaroni (2020): "Testing beta-pricing models using large cross-sections," *The Review of Financial Studies*, 33(6), 2796–2842.

Roll, R. (1977): "A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory," *Journal of Financial Economics*, 4(2), 129–176.

Ross, S. (1976): "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, 13(3), 341–360.

Schneider, P., C. Wagner, and J. Zechner (2020): "Low-risk anomalies?," *The Journal of Finance*, 75(5), 2673–2718.

Schwert, G. W. (1989): "Why does stock market volatility change over time?," *Journal of Finance*, 44(5), 1115–1153.

Schwert, G. W., and P. J. Seguin (1990): "Heteroskedasticity in stock returns," *Journal of Finance*, 45(4), 1129–1155.

Shanken, J. (1985): "Multivariate tests of the zero-beta CAPM," *Journal of Financial Economics*, 14(3), 327–348.

——— (1987): "Multivariate proxies and asset pricing relations: Living with the Roll critique," *Journal of Financial Economics*, 18(1), 91–110.

——— (1990): "Intertemporal asset pricing: An empirical investigation," *Journal of Econometrics*, 45(1-2), 99–120.

——— (1992): "On the estimation of beta-pricing models," *The Review of Financial Studies*, 5(1), 1–33.

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

Wooldridge, J. M. (2016): *Introductory Econometrics: A Modern Approach*. Nelson Education.

Zaffaroni, P. (2019): "Factor models for asset pricing," *Available at SSRN 3398169*.

Zhou, G. (1994): "Analytical GMM tests: Asset pricing with time-varying risk premiums," *The Review of Financial Studies*, 7(4), 687–709.

A Conditions and Proofs of Main Results

In this section, we provide the sufficient conditions and proofs of the asymptotic properties of the CSR estimators provided in Section 2. We first present the conditions on the pricing factors and security characteristics.

Assumption 2. *Let* $\overline{f} \equiv T^{-1}\sum_{t\leq T} f_t$ *and* $\overline{\mathbf{Z}} \equiv T^{-1}\sum_{t\leq T} \mathbf{Z}_{t-1}$*. Suppose the following: (i)* $\{f_t\}_t$ *is a covariance-stationary process; (ii)* $\overline{f} = \mu_f + o_p(1)$*; (iii)* $T^{-1}\sum_{t\leq T} f_t f_t' = \mathbb{E}[f_t f_t'] + o_p(1)$*; and (iv)* $\overline{\mathbf{Z}} = \mathbb{E}[\mathbf{Z}_{t-1}] + o_p(1)$*.*

Assumption 2(i) requires that the pricing factors are covariance-stationary across time, which is commonly imposed in the literature. Under this condition, the first and the second moments of f_t are time invariant. Assumptions 2(ii, iii) are the law of large numbers of the sample mean and the sample second moment of the pricing factors; Assumption 2(iv) is the law of large numbers on the sample mean of the security characteristics. These conditions can be verified under low-level sufficient conditions. Note that Assumption 2(iv) may not hold for individual stocks, because security characteristics at the stock level may not be covariance stationary (e.g., firm size has a trend). Instead, Assumption 2(iv) should be considered a reasonable approximation for sorted portfolios (see, e.g., Jagannathan and Wang (2002)).

Assumption 3. *(i) Let* $\Omega \equiv \lim_{T\to\infty} \mathrm{Var}(T^{-1/2}\sum_{t\leq T} \boldsymbol{\epsilon}_t)$*, then* $T^{-1/2}\sum_{t\leq T} \boldsymbol{\epsilon}_t \to_d N(0, \Omega)$*; (ii)* $T^{-1/2}\sum_{t\leq T} \mathbf{u}_t(1, f_t') = O_p(1)$*; (iii) a non-random matrix symmetric* W *exists such that* $\widehat{W} = W + o_p(1)$*; and (iv) let* $\mathbf{X} \equiv (S_N, \mathbf{1}_{N\times 1}, \mathbf{B}, \mathbb{E}[\mathbf{Z}_{t-1}])$*, then the eigenvalues of* Ω*,* Σ_f*,* W*, and* $\mathbf{X}'\mathbf{X}$ *are bounded from above and away from zero.*

Assumption 3(i) is a central limit theorem on the partial sum $T^{-1/2}\sum_{t\leq T} \boldsymbol{\epsilon}_t$, which can be verified under low-level sufficient conditions (see, e.g., Hall and Heyde (1980) and Davidson (1994)). By the definitions of $u_{i,t}$ and β_i, we have $\mathbb{E}[\mathbf{u}_t] = \mathbf{0}_{N\times 1}$ and $\mathbb{E}[\mathbf{u}_t f_t'] = \mathbf{0}_{N\times K}$. Therefore, Assumption 3(ii) holds if the variance-covariance matrix of $T^{-1/2}\sum_{t=1}^{T} \mathbf{u}_t(1, f_t')$ is bounded. Assumption 3(iii) imposes conditions on the weight matrix of the CSR estimator. The eigenvalue conditions in Assumption 3(iv) ensure that the CSR estimator is $T^{-1/2}$-consistent.

Proof of lemma 1. Because $\mathbb{E}[\mathbf{R}_t] = \mathbf{X}\theta$ by (2.3) where $\theta \equiv (\alpha_{N_0+1}, \ldots, \alpha_N, \gamma')'$, we can write

$$
\begin{aligned}
\overline{\mathbf{R}} - \widehat{\mathbf{X}}\theta &= \overline{\mathbf{R}} - \mathbb{E}[\mathbf{R}_t] + \mathbb{E}[\mathbf{R}_t] - \widehat{\mathbf{X}}\theta \\
&= \overline{\mathbf{R}} - \mathbb{E}[\mathbf{R}_t] - (\widehat{\mathbf{X}} - \mathbf{X})\theta \\
&= \overline{\mathbf{R}} - \mathbb{E}[\mathbf{R}_t] - (\widehat{\mathbf{B}} - \mathbf{B})\gamma_1 - (\overline{\mathbf{Z}} - \mathbb{E}[\mathbf{Z}_t])\gamma_2 = \overline{\mathbf{v}} - (\widehat{\mathbf{B}} - \mathbf{B})\gamma_1, \qquad (A.1)
\end{aligned}
$$

where $\overline{\mathbf{v}} \equiv T^{-1}\sum_{t=1}^{T}\mathbf{v}_t$ and the last equality follows from (2.3) and (2.7). Therefore, we can write

$$T^{1/2}(\hat{\theta}_{csr} - \theta) = (\widehat{\mathbf{X}}'\widehat{W}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}'\widehat{W}\left[T^{1/2}(\overline{\mathbf{v}} - (\widehat{\mathbf{B}} - \mathbf{B})\gamma_1)\right]. \tag{A.2}$$

Because $\mathbf{R}_t = \mathbf{X}\theta + \mathbf{B}(f_t - \mu_f) + \mathbf{u}_t$,

$$\begin{aligned} \widehat{\mathbf{B}} - \mathbf{B} &= \left(T^{-1}\sum_{t\leq T}(\mathbf{R}_t - \overline{\mathbf{R}})(f_t - \overline{f})'\right)\widehat{\Sigma}_f^{-1} - \mathbf{B} \\ &= T^{-1}\sum_{t\leq T}\mathbf{u}_t(f_t - \mu_f)'\widehat{\Sigma}_f^{-1} - \overline{\mathbf{u}}T^{-1}\sum_{t\leq T}(f_t - \mu_f)'\widehat{\Sigma}_f^{-1}, \end{aligned} \tag{A.3}$$

where $\widehat{\Sigma}_f \equiv T^{-1}\sum_{t\leq T}(f_t - \overline{f})(f_t - \overline{f})'$. By Assumptions 2(i, ii, iii),

$$\widehat{\Sigma}_f = \Sigma_f + o_p(1), \tag{A.4}$$

which together with Assumptions 2(ii) and 3(ii, iv) implies

$$\widehat{\mathbf{B}} - \mathbf{B} = T^{-1}\sum_{t\leq T}\mathbf{u}_t(f_t - \mu_f)'\Sigma_f^{-1} + o_p(T^{-1/2}) = O_p(T^{-1/2}). \tag{A.5}$$

Therefore, by Assumptions 3(i) and (A.5), we have

$$\overline{\mathbf{v}} - (\widehat{\mathbf{B}} - \mathbf{B})\gamma_1 = T^{-1}\sum_{t\leq T}\boldsymbol{\epsilon}_t + o_p(T^{-1/2}) = O_p(T^{-1/2}). \tag{A.6}$$

Similarly, by Assumptions 2(iv) and (A.5),

$$\widehat{\mathbf{X}} = \mathbf{X} + o_p(1). \tag{A.7}$$

From Assumptions 3(iii, iv) and (A.7), we obtain

$$\widehat{\mathbf{X}}'\widehat{W}\widehat{\mathbf{X}} = \mathbf{X}'W\mathbf{X} + o_p(1) \text{ and } \widehat{\mathbf{X}}'\widehat{W} = \mathbf{X}'W + o_p(1). \tag{A.8}$$

Combining the results in (A.2), (A.6), and (A.8) and applying Assumptions 3(i, iv), we get

$$T^{1/2}(\hat{\theta}_{csr} - \theta) = (\mathbf{X}'W\mathbf{X})^{-1}\mathbf{X}'WT^{-1/2}\sum_{t\leq T}\boldsymbol{\epsilon}_t + o_p(1) \rightarrow_d N(0, \text{Asv}(\hat{\theta}_{csr})) \tag{A.9}$$

which shows the claim of the lemma. *Q.E.D.*

Proof of Proposition 1. Replacing W in $\text{Asv}(\hat{\theta}_{csr})$ by Ω^{-1} obtains (2.10). Let $P_N \equiv I_N - \Omega^{1/2}W\mathbf{X}(\mathbf{X}'W\Omega W\mathbf{X})^{-1}\mathbf{X}'W\Omega^{1/2}$. Then, we can write

$$(\text{Asv}(\hat{\theta}^*_{csr}))^{-1} - (\text{Asv}(\hat{\theta}_{csr}))^{-1} = \mathbf{X}'\Omega^{-1/2}P_N\Omega^{-1/2}\mathbf{X}. \tag{A.10}$$

Since P_N is an idempotent matrix, $\mathbf{X}'\Omega^{-1/2}P_N\Omega^{-1/2}\mathbf{X}$ is positive semi-definite which together with (A.10) shows that $(\text{Asv}(\hat{\theta}^*_{csr}))^{-1} - (\text{Asv}(\hat{\theta}_{csr}))^{-1}$ is positive semi-definite. Because $(\text{Asv}(\hat{\theta}^*_{csr}))^{-1}$ and $(\text{Asv}(\hat{\theta}_{csr}))^{-1}$ are positive definite matrices, we can further deduce that $\text{Asv}(\hat{\theta}_{csr}) - \text{Asv}(\hat{\theta}^*_{csr})$ is positive semi-definite, that is, $\text{Asv}(\hat{\theta}_{csr}) \geq \text{Asv}(\hat{\theta}^*_{csr})$. *Q.E.D.*

Proof of Proposition 2. In view of (2.10) in Proposition 1, it is sufficient to show that

$$\text{Asv}(\hat{\theta}^*_{csr}) = \text{Asv}(\hat{\theta}^*_{gmm}). \tag{A.11}$$

First, simple calculation shows

$$G = \begin{pmatrix} 0_{K\times d_\theta} & -I_K & 0_{K\times NK} \\ 0_{NK\times d_\theta} & -\mathbf{X}\theta \otimes I_K & -I_N \otimes \Sigma_f \\ -\mathbf{X} & 0_{N\times K} & -I_N \otimes \gamma_1' \end{pmatrix}$$

where $A \otimes B$ denotes the Kronecker product of two matrices A and B.

Let

$$D = \begin{pmatrix} -\mathbf{X}\theta\gamma_1'\Sigma_f^{-1} & I_N \otimes \gamma_1'\Sigma_f^{-1} & -I_N \\ -I_K & 0_{K\times NK} & 0_{K\times N} \\ \mathbf{X}\theta \otimes \Sigma_f^{-1} & -I_N \otimes \Sigma_f^{-1} & 0_{NK\times N} \end{pmatrix}. \tag{A.12}$$

Because D is invertible, we can write

$$(\text{Asv}(\hat{\theta}^*_{gmm}))^{-1} = (DG)'(D\Sigma_g^* D')^{-1}(DG).$$

Because $DG = \text{diag}(\mathbf{X}, I_{(N+1)K})$, by Lemma A.1 of Chamberlain (1987),

$$\text{Asv}(\hat{\theta}^*_{gmm}) = S_\theta \text{Asv}(\hat{\phi}^*_{gmm}) S'_\theta = (\mathbf{X}'(D\Sigma^*_g D')^{-1}_{11}\mathbf{X})^{-1}, \tag{A.13}$$

where $(D\Sigma^*_g D')_{11}$ denotes the leading $N \times N$ submatrix of $D\Sigma^*_g D'$. Let $(Dg(Y_t, \phi))_N$ denote the leading $N \times 1$ subvector of $Dg(Y_t, \phi)$. Then,

$$\begin{aligned}
(Dg(Y_t, \phi))_N &= -\mathbf{X}\theta\gamma'_1\Sigma_f^{-1}g_1(Y_t, \phi) + (I_N \otimes \gamma'_1\Sigma_f^{-1})g_2(Y_t, \phi) - g_3(Y_t, \phi) \\
&= -\mathbf{X}\theta\gamma'_1\Sigma_f^{-1}(f_t - \mu_f) + (\mathbf{R}_t - \mathbf{B}(f_t - \mu_f))\gamma'_1\Sigma_f^{-1}(f_t - \mu_f) - (\mathbf{R}_t - \mathbf{X}_t\theta) \\
&= (\mathbf{R}_t - \mathbb{E}[\mathbf{R}_t] - \mathbf{B}(f_t - \mu_f))\gamma'_1\Sigma_f^{-1}(f_t - \mu_f) - (\mathbf{R}_t - \mathbf{X}_t\theta). \qquad \text{(A.14)}
\end{aligned}$$

By the definition of $\mathbf{u}_t$ and $\mathbf{v}_t$, we deduce from (A.14) that

$$(Dg(Y_t, \phi))_N = -\mathbf{v}_t + \mathbf{u}_t(f_t - \mu_f)'\Sigma_f^{-1}\gamma_1, \tag{A.15}$$

which implies

$$(D\Sigma^*_g D')_{11} = \lim_{T\to\infty} \text{Var}\left(T^{-1/2}\sum\nolimits_{t=1}^T (\mathbf{v}_t - \mathbf{u}_t(f_t - \mu_f)'\Sigma_f^{-1}\gamma_1)\right) = \Omega. \tag{A.16}$$

Combining the results in (2.9), (A.13) and (A.16), we have

$$\text{Asv}(\hat{\theta}^*_{gmm}) = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} = \text{Asv}(\hat{\theta}^*_{csr})$$

which shows (A.11). *Q.E.D.*

B A Literature Review

In the context of the classical linear-beta pricing framework, we construct the set of moment restrictions that are related to existing two-pass cross-sectional regression (e.g., the Fama-MacBeth regression) estimators. We then develop a new class of two-pass estimators that dominate existing estimators in terms of estimation efficiency. Furthermore, we establish the asymptotic equivalence between our estimators and the optimally weighted GMM estimators, which are known in the literature to achieve the semiparametric efficiency bound. Therefore, we provide a unified view on existing two-pass regression estimators and identify those that are optimal, i.e., the OCSR estimators, in a well-defined sense.

Our research addresses an important gap in the literature on linear-beta pricing models. Under the assumption that idiosyncratic risks are i.i.d. normal and conditionally homoskedastic given the pricing factors, Shanken (1992) proves the equivalence between the GLS two-pass regression estimator and the maximum-likelihood estimator, demonstrating that the GLS estimator is asymptotically efficient in the parametric framework. We improve on Shanken (1992) by characterizing the OCSR estimator among two-pass regression estimators with possibly nonnormal, dependent and conditional heteroskedastic idiosyncratic risks. More importantly, we extend Shanken (1992)'s insight by showing that the optimality of our OCSR goes far beyond the class of two-pass regressors. It actually extends to the realm of GMM estimators. As such, we establish the legitimacy of the usual two-pass regression approach by showing that information loss is not necessary when one goes from the more methodical GMM approach to the simple-to-implement two-pass regressions.[39]

Our theory unfolds in several steps. First, we introduce the set of moment restrictions that are closely related to the pricing framework in Jagannathan and Wang (1998) and the maximum-likelihood framework in Shanken (1992). These moment restrictions characterize the linear-beta pricing model and naturally map into the usual two-pass regression implementation. Working within the class of two-pass regressors, we derive the asymptotic distribution of a generic cross-sectional regression estimator with an arbitrary weighting matrix and characterize the optimal weighting matrix that generates the smallest asymptotic covariance matrix (in a matrix sense). Jagannathan and Wang (1998) establish the asymptotic distribution of a general cross-sectional regression estimator. Our results thus extend Jagannathan and Wang (1998) by: (1) allowing for potential mispricing of certain assets and additional characteristics; and (2) pinning down the optimal weighting matrix for the two-pass estimator.

We then examine the efficiency of our estimator and link it to the optimal GMM estimator based on the system of moment restrictions. We prove that our OCSR estimator and the optimal GMM estimator share the same asymptotic variance-covariance matrix and hence are asymptotically equivalent. Because the optimal GMM estimator achieves the semiparametric efficiency bound, our OCSR estimator attains the same bound. As a result, we show that the sequential estimation of betas (in a first-stage regression) and risk premiums (in a second-stage regression) does not lead to efficiency loss compared with the one-step GMM estimator. In general, our results provide a strong theoretical foundation for the two-pass regression approach, extending Shanken (1992)'s insight on the connection between the two-pass GLS estimator and the maximum-likelihood estimator.

[39] Alternative cross-sectional methods have also been proposed in Kim (1995) and Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2019). Unlike our paper, these methods focus on the N-consistency of the proposed estimators and do not examine the estimation efficiency for the risk premium parameters.

Finally, we provide an economic interpretation of the weight matrix used by our OCSR estimator. Our general formulation allows us to decompose the pricing error in the second-stage regression into two parts. One part multiplies idiosyncratic risks by a common multiplicative factor that captures first-stage beta-estimation uncertainty. The second part is related to systematic innovations that do not influence beta uncertainty. Under Shanken (1992)'s assumptions of i.i.d. and conditional homoskedasticity, the two parts are uncorrelated, allowing one to separately identify the impact on estimation efficiency by idiosyncratic and systematic risks. Given the linear-beta structure, systematic risks alone do not affect estimation efficiency, leading to Shanken (1992)'s results that only the residual variance-covariance matrix affects the efficiency of the two-pass estimator. In our framework, where we allow for nonnormality, general dependence and conditional heteroskedasticity, we can no longer separately identify the contribution of idiosyncratic and systematic risks in affecting estimation efficiency. In fact, this non-separability is the key for our OCSR estimator to achieving GMM's optimality, which improves on the parametric efficiency of the GLS estimator.

Our paper contributes to the large body of literature on testing beta-pricing models (e.g., Gibbons, Ross, and Shanken (1989), Kandel and Stambaugh (1995), Shanken (1992), Jagannathan and Wang (1998), Shanken and Zhou (2007), Kan, Robotti, and Shanken (2013)). We propose a new two-pass regressor that improves on existing methods in terms of estimation efficiency. Our paper is also related to several GMM implementations that aim to provide robust inference on beta-pricing models, including MacKinlay and Richardson (1991), Zhou (1994), and Jagannathan and Wang (2002). Under a general set of moment restrictions, we show that our OCSR preserves GMM's efficiency gain.

More recent literature proposes new methods to cope with the large N scenario, such as Gagliardini, Ossola, and Scaillet (2016), Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2019), Kim and Skoulakis (2018), Feng, Giglio, and Xiu (2020), Raponi, Robotti, and Zaffaroni (2020), and Harvey and Liu (2020b). Unlike these papers, we focus on the fixed-N scenario and aim to derive the most efficient estimator when only T goes to infinity.[40]

Another theme of the recent literature is to develop methods that explicitly address the proliferation of anomalies (Gagliardini, Ossola, and Scaillet (2016), Barillas and Shanken (2017), Kim and Skoulakis (2018), Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2019), Huang, Li, and Zhou (2019), Giglio and Xiu (2019), Kelly, Pruitt, and Su (2019), Lettau and Pelger (2020), Freyberger, Neuhierl, and Weber (2020), Feng, Giglio, and Xiu (2020), and **?**). While these papers

[40]We provide an extension of our baseline results by allowing a slowly divergent N, i.e., $N^{1+\delta}/T \to 0$ for some $\delta > 1$.

emphasize the selection of factors, we focus on the most efficient use of test portfolios in estimating the risk premiums for a pre-selected set of factors. Our approach is thus more relevant for evaluating the performance of a fixed and small menu of candidate factors; see recent examples in Croce, Marchuk, and Schlag (2019), Lin, Palazzo, and Yang (2020), and Ai, Li, Li, and Schlag (2020).

Our idea of embedding two-pass regressors into the GMM framework can be extended along several dimensions. Although we focus on fixed-N and large-T asymptotics, deriving the large-N and large-T asymptotics is also possible by explicitly taking into account the impact of a large cross-section on the asymptotic distribution.[41] We also focus on time-invariant betas for linear-beta pricing models. An extension that instruments betas with time-varying firm characteristics seems feasible.[42] We leave these extensions to future research.

C A Counterexample

In this section, we provide a simple example to show that the asymptotic equivalence between the optimal two-step estimator, such as the OCSR estimator, and the optimal one-step GMM estimator comes from the just identification of the unknown nuisance parameters estimated in the first step. The linear structure of the moment conditions plays no role for the asymptotic equivalence result.

Suppose that we are interested in estimating an unknown parameter θ which is identified by the following moment conditions

$$\mathbb{E}[Y_1 - \mu 1_{k\times 1}] = 0_{k\times 1}, \tag{C.1}$$

$$\mathbb{E}[Y_2 - \mu - \theta] = 0, \tag{C.2}$$

where $k \geq 1$ and μ is a nuisance parameter. Let $\varepsilon_1 \equiv Y_1 - \mu 1_{k\times 1}$ and $\varepsilon_2 \equiv Y_2 - \mu - \theta$, where the long-run variance-covariance matrix of the partial sum of $(\varepsilon_1', \varepsilon_2')'$ is

$$\Sigma_\varepsilon \equiv \begin{pmatrix} I_k & \rho \\ \rho' & 1 \end{pmatrix},$$

[41] More specifically, the large-N and large-T asymptotics assume $N \to \infty$, $T \to \infty$, and $N/T \to c$ for some $c > 0$.

[42] See, e.g., Shanken (1990), Ferson and Schadt (1996), Christopherson, Ferson, and Glassman (1998), and Ferson, Sarkissian, and Simin (2006).

where ρ is a $k \times 1$ real vector such that Σ_ε is positive definite.

We compare the asymptotic variances of two estimators of θ. The first is from a joint optimal GMM estimation of (θ, μ) using all the moment conditions in (C.1) and (C.2). The second is obtained by an iterative GMM estimation procedure where we first obtain the optimal GMM estimator of μ through the moment conditions in (C.1) and then plug it in (C.2) to construct the optimal two-step GMM estimator of θ. We denote the first GMM estimator as $\hat{\theta}^J$ and the second GMM estimator as $\hat{\theta}^I$.

The asymptotic variance-covariance matrix of the joint optimal GMM estimator of (θ, μ) based on (C.1) and (C.2) is the inverse of the following matrix:

$$\begin{pmatrix} 1_{1\times k} & 1 \\ 0_{1\times k} & 1 \end{pmatrix} \begin{pmatrix} I_k & \rho \\ \rho' & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1_{k\times 1} & 0_{k\times 1} \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} k + \frac{(1-1'_{k\times 1}\rho)^2}{1-\rho'\rho} & \frac{1-1'_{k\times 1}\rho}{1-\rho'\rho} \\ \frac{1-1'_{k\times 1}\rho}{1-\rho'\rho} & \frac{1}{1-\rho'\rho} \end{pmatrix},$$

which implies that the asymptotic variance of $\hat{\theta}^J$ is

$$\text{Asv}(\hat{\theta}^J) \equiv \left(k + \frac{(1-1'_{k\times 1}\rho)^2}{1-\rho'\rho}\right) \frac{1-\rho'\rho}{k} = 1 + k^{-1} - 2k^{-1}1'_{k\times 1}\rho + \frac{(1'_{k\times 1}\rho)^2 - k\rho'\rho}{k}. \tag{C.3}$$

On the other hand, the optimal GMM estimator of μ based on (C.1) is

$$\hat{\mu}^J \equiv k^{-1}1'_{k\times 1}\overline{Y}_1$$

where $\overline{Y}_1 \equiv T^{-1}\sum_{t\leq T} Y_{1,t}$. Therefore, the optimal iterative GMM estimator of θ is

$$\hat{\theta}^I \equiv \overline{Y}_2 - \hat{\mu}^J = \overline{Y}_2 - k^{-1}1'_{k\times 1}\overline{Y}_1.$$

From the above expression, it is easy to show that the asymptotic variance of $\hat{\theta}^I$ is

$$\text{Asv}(\hat{\theta}^I) \equiv 1 + k^{-1} - 2k^{-1}1'_{k\times 1}\rho. \tag{C.4}$$

By (C.3) and (C.4), we have

$$\text{Asv}(\hat{\theta}^J) - \text{Asv}(\hat{\theta}^I) = k^{-1}(1'_{k\times 1}\rho)^2 - \rho'\rho \leq 0, \tag{C.5}$$

where the equality holds if and only if $k = 1$ or $\rho = a 1_{k\times 1}$ for some real number a by the Cauchy-Schwarz inequality.

From the inequality in (C.5), we see that in general the joint optimal GMM estimator $\hat{\theta}^J$ dominates the iterative optimal GMM estimator. These two estimators are asymptotically equivalent in the special case where the moment conditions in (C.1) and (C.2) have special dependence structure, that is, $\rho = a1_{k\times 1}$. Under general dependence of the moment conditions in (C.1) and (C.2) (i.e., ρ is not zero and is linearly independent with respect to $1_{k\times 1}$), $\hat{\theta}^J$ and $\hat{\theta}^I$ are asymptotic equivalent only when $k = 1$, that is, the nuisance parameter μ is just identified in (C.1), which shares the same intuition of the asymptotic efficiency of our OCSR estimator.

D Some Auxiliary Results

This section contains some results which can be used to conduct inference on the unknown parameters and model specification tests. Section D.1 provides the t test and Wald test on the unknown parameter θ, and Section D.2 provides a specification test for (2.4). Section D.3 extends the asymptotic equivalence between the OCSR estimator and the optimal GMM estimator to the case where N grows with T.

D.1 Inference based on OCSR

Let $\widehat{\Omega}$ denote a consistent estimator of Ω. The OCSR estimator is then defined as

$$\hat{\theta}^*_{csr} \equiv (\widehat{\mathbf{X}}'\widehat{\Omega}^{-1}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}'\widehat{\Omega}^{-1}\overline{\mathbf{R}}. \tag{D.1}$$

The consistency of $\widehat{\Omega}$ implies Assumption 3(iii). Moreover, by Assumptions 2 and 3(ii, iv), we can show that

$$\widehat{\mathbf{X}}'\widehat{\Omega}^{-1}\widehat{\mathbf{X}} = \mathbf{X}'\Omega^{-1}\mathbf{X} + o_p(1), \tag{D.2}$$

which together with Theorem 1 and the Slutsky Theorem, implies

$$(\widehat{\mathbf{X}}'\widehat{\Omega}^{-1}\widehat{\mathbf{X}})^{1/2} T^{1/2}(\hat{\theta}^*_{csr} - \theta) \to_d N(0, I_{d_\theta}), \tag{D.3}$$

where $d_\theta \equiv N_1 + K + M + 1$. The above result can be directly applied to conduct inference on θ.

Let l denote any $d_\theta \times 1$ non zero real vector. Then, (D.2) together with (D.3) implies

$$\frac{T^{1/2} l'(\hat{\theta}^*_{csr} - \theta)}{(l'(\widehat{\mathbf{X}}'\widehat{\Omega}^{-1}\widehat{\mathbf{X}})^{-1} l)^{1/2}} \to_d N(0,1), \tag{D.4}$$

which can be used to test the hypothesis on the linear combinations of θ. Let $\mathcal{M}$ denote any non-empty subset of $\{1, \ldots, d_\theta\}$ with size m and let $L_{\mathcal{M}} = (l_j)_{j \in \mathcal{M}}$ denote the $d_\theta \times m$ selection matrix whose jth column is the jth unit vector; that is, the jth component of l_j is 1 and the remaining elements are zero. Then, by (D.2) and (D.3),

$$(L'_{\mathcal{M}}(\widehat{\mathbf{X}}'\widehat{\Omega}^{-1}\widehat{\mathbf{X}})^{-1} L_{\mathcal{M}})^{-1/2} L'_{\mathcal{M}} T^{1/2}(\hat{\theta}^*_{csr} - \theta) \to_d N(0, I_m),$$

which together with the continuous mapping Theorem implies

$$T(\hat{\theta}^*_{csr} - \theta)' L_{\mathcal{M}}(L'_{\mathcal{M}}(\widehat{\mathbf{X}}'\widehat{\Omega}^{-1}\widehat{\mathbf{X}})^{-1} L_{\mathcal{M}})^{-1} L'_{\mathcal{M}}(\hat{\theta}^*_{csr} - \theta) \to_d \chi^2(m), \tag{D.5}$$

where $\chi^2(m)$ denotes the chi-square distribution with degree of freedom m. The result in (D.5) can be used to conduct joint inference of any subset of θ, such as $(\alpha_{N_0+1}, \ldots, \alpha_N)'$.

D.2 Specification test

In this section, we provide a specification test for the restriction (2.4). By (2.3), this restriction can be written as

$$H_0 : \mathbb{E}[R_{i,t} - \gamma_0 - \gamma_1'\beta_i - \gamma_2' Z_{i,t}] = 0, \tag{D.6}$$

for $i = 1, \ldots, N_0$. The restrictions in (2.1) and (2.2), which are essentially the definitions of μ_f and β_i for $i = 1, \ldots, N_0$, are maintained under both the null and the alternative hypotheses. Because the focus is on testing (2.4), estimation and inference of α_i for $i = N_0 + 1, \ldots, N$ are not involved.

Let $\widehat{\mathbf{X}}_{0,t} \equiv (\mathbf{1}_{N_0 \times 1}, \widehat{\mathbf{B}}_0, \mathbf{Z}_{0,t-1})$, where $\widehat{\mathbf{B}}_0$ and $\mathbf{Z}_{0,t-1}$ denote the leading $N_0 \times K$ and $N_0 \times M$ submatrices of $\widehat{\mathbf{B}}$ and $\mathbf{Z}_{t-1}$, respectively. The unknown parameter γ is estimated by

$$\hat{\gamma} \equiv (\widehat{\mathbf{X}}_0'\widehat{\Omega}_0^{-1}\widehat{\mathbf{X}}_0)^{-1}\widehat{\mathbf{X}}_0'\widehat{\Omega}_0^{-1}\overline{\mathbf{R}}_0, \tag{D.7}$$

where $\widehat{\mathbf{X}}_0 \equiv T^{-1}\sum_{t \le T}\widehat{\mathbf{X}}_{0,t}$, $\overline{\mathbf{R}}_0 \equiv T^{-1}\sum_{t \le T}\mathbf{R}_{0,t}$, $\widehat{\Omega}_0$ denotes the leading $N_0 \times N_0$ submatrix of $\widehat{\Omega}$ and $\mathbf{R}_{0,t}$ denotes the leading N_0 subvector of $\mathbf{R}_t$. The null hypothesis (D.6) is tested using the

J-test statistic, which is defined as

$$J_T \equiv T(\overline{\mathbf{R}}_0 - \widehat{\mathbf{X}}_0\widehat{\gamma})'\widehat{\Omega}_0^{-1}(\overline{\mathbf{R}}_0 - \widehat{\mathbf{X}}_0\widehat{\gamma}). \tag{D.8}$$

Let $\chi^2_{1-\alpha}(N_0 - d_\gamma)$ denote the $1-\alpha$ quantile of $\chi^2(N_0 - d_\gamma)$. We consider the following test at the significance level α:

$$\text{reject } H_0 \text{ if } J_T > \chi^2_{1-\alpha}(N_0 - d_\gamma). \tag{D.9}$$

The above test has been proposed in the literature (see, for example, Kan and Robotti (2012)). We next provide the asymptotic properties of J_T under both the null and the alternative hypothesis, which complements the existing results in the literature.

Lemma 2. *Suppose Assumption 2 hold. Then,*

Theorem 1. *(a) Under Assumption 3, $J_T \to_d \chi^2(N_0 - d_\gamma)$ under H_0;*

(b) If we have: (i) $\overline{\mathbf{R}}_0 = \mathbb{E}[\mathbf{R}_{0,t}] + o_p(1)$; (ii) $\hat{\Omega}_0 = \Omega_1 + o_p(1)$ where Ω_1 is a non-random symmetric positive definite matrix; and (iii) the eigenvalues of Ω_1 and $\mathbf{X}_0'\mathbf{X}_0$ are bounded from above and away from zero, then

$$T^{-1}J_T = \mathbb{E}[\mathbf{R}_{0,t}']\Pi_{N_0}\mathbb{E}[\mathbf{R}_{0,t}] + o_p(1), \tag{D.10}$$

where $\Pi_{N_0} \equiv \Omega_1^{-1} - \Omega_1^{-1}\mathbf{X}_0(\mathbf{X}_0'\Omega_1^{-1}\mathbf{X}_0)^{-1}\mathbf{X}_0'\Omega_1^{-1}$.

Lemma 2(a) shows that the test in (D.9) controls size. Lemma 2(b) derives the probability limit of the (scaled) test statistic J_T under both the null and the alternative hypotheses. Under H_0, $\mathbb{E}[\mathbf{R}_{0,t}] = \mathbf{X}_0\gamma$, and hence (D.10) implies $T^{-1}J_T = o_p(1)$, which is consistent with the weak convergence of J_T derived in Lemma 2(a). Under the alternative hypothesis, $\mathbb{E}[\mathbf{R}_{0,t}]$ cannot be represented by any linear combination of $\mathbf{X}_0$, and hence, we have $\mathbb{E}[\mathbf{R}_{0,t}']\Pi_{N_0}\mathbb{E}[\mathbf{R}_{0,t}] > 0$ in general. Because $\chi^2_{1-\alpha}(N_0 - d_\gamma)$ is a finite number, Lemma 2(b) shows that the test in (D.9) is consistent as long as $\mathbb{E}[\mathbf{R}_{0,t}']\Pi_{N_0}\mathbb{E}[\mathbf{R}_{0,t}]$ is bounded away from zero.

Proof of Lemma 2. (a) Under Assumptions 2 and 3, we can use the same arguments in showing (A.1) and (A.6) in the proof of Theorem 1 to obtain

$$\overline{\mathbf{R}}_0 - \widehat{\mathbf{X}}_0\gamma = \overline{\mathbf{v}}_0 - (\widehat{\mathbf{B}}_0 - \mathbf{B}_0)\gamma_1 = T^{-1}\sum_{t\leq T}\boldsymbol{\epsilon}_{0,t} + o_p(T^{-1/2}) = O_p(T^{-1/2}), \tag{D.11}$$

where $\overline{\mathbf{v}}_0 \equiv T^{-1}\sum_{t\leq T}\mathbf{v}_{0,t}$, $\mathbf{B}_0$, $\mathbf{v}_{0,t}$, and $\boldsymbol{\epsilon}_{0,t}$ denote the leading $N_0 \times K$, $N_0 \times 1$, and $N_0 \times 1$ submatrices of $\mathbf{B}$, $\mathbf{v}_t$, and $\boldsymbol{\epsilon}_t$, respectively. Therefore, by the same arguments for showing (A.9) in the proof of Theorem 1,

$$\hat{\gamma} - \gamma = (\mathbf{X}_0'\Omega_0^{-1}\mathbf{X}_0)^{-1}\mathbf{X}_0'\Omega_0^{-1}T^{-1}\sum_{t\leq T}\boldsymbol{\epsilon}_{0,t} + o_p(T^{-1/2}) = O_p(T^{-1/2}), \tag{D.12}$$

where Ω_0 denotes the leading $N_0 \times N_0$ submatrix of Ω. Combining the results in (A.7), (D.11), and (D.12),

$$\begin{aligned}
\overline{\mathbf{R}}_0 - \widehat{\mathbf{X}}_0\hat{\gamma} &= \overline{\mathbf{R}}_0 - \widehat{\mathbf{X}}_0\gamma - \widehat{\mathbf{X}}_0(\hat{\gamma} - \gamma) \\
&= (I_{N_0} - \mathbf{X}_0(\mathbf{X}_0'\Omega_0^{-1}\mathbf{X}_0)^{-1}\mathbf{X}_0'\Omega_0^{-1})T^{-1}\sum_{t\leq T}\boldsymbol{\epsilon}_{0,t} + o_p(T^{-1/2}) \\
&= \Omega_0^{1/2}M_{N_0}\Omega_0^{-1/2}T^{-1}\sum_{t\leq T}\boldsymbol{\epsilon}_{0,t} + o_p(T^{-1/2}) = O_p(T^{-1/2}),
\end{aligned} \tag{D.13}$$

where $M_{N_0} \equiv I_{N_0} - \Omega_0^{-1/2}\mathbf{X}_0(\mathbf{X}_0'\Omega_0^{-1}\mathbf{X}_0)^{-1}\mathbf{X}_0'\Omega_0^{-1/2}$, which together with the consistency of $\widehat{\Omega}$, Assumptions 3(i, iv), and the continuous mapping theorem implies

$$\begin{aligned}
J_T &= T(\overline{\mathbf{R}}_0 - \widehat{\mathbf{X}}_0\hat{\gamma})'\Omega_0^{-1}(\overline{\mathbf{R}}_0 - \widehat{\mathbf{X}}_0\hat{\gamma}) + o_p(1) \\
&= \left(\Omega_0^{-1/2}T^{-1/2}\sum_{t\leq T}\boldsymbol{\epsilon}_{0,t}\right) M_{N_0} \left(\Omega_0^{-1/2}T^{-1/2}\sum_{t\leq T}\boldsymbol{\epsilon}_{0,t}\right) + o_p(1) \rightarrow_d \mathcal{N}_0'M_{N_0}\mathcal{N}_0,
\end{aligned} \tag{D.14}$$

where $\mathcal{N}_0$ denotes a $N_0 \times 1$ standard normal random vector. Because M_{N_0} is an idempotent matrix with rank $N_0 - d_\gamma$, the random variable $\mathcal{N}_0'M_{N_0}\mathcal{N}_0$ has the same distribution as $\chi^2(N_0 - d_\gamma)$. The claim in part (a) of the lemma follows directly from (D.14).

(b) Under Assumption 2 and condition (ii) of the theorem, we can use the same arguments for showing (A.7) and (A.8) in the proof of Lemma 1 to obtain

$$\widehat{\mathbf{X}}_0 = \mathbf{X}_0 + o_p(1),\ \widehat{\mathbf{X}}_0'\widehat{\Omega}_0^{-1}\widehat{\mathbf{X}}_0 = \mathbf{X}_0'\Omega_1^{-1}\mathbf{X}_0 + o_p(1) \text{ and } \widehat{\mathbf{X}}_0'\widehat{\Omega}_0^{-1} = \mathbf{X}_0'\Omega_1^{-1} + o_p(1),$$

which together with conditions (i, iii) of the lemma shows that

$$\overline{\mathbf{R}}_0 - \widehat{\mathbf{X}}_0\hat{\gamma} = (I_{N_0} - \mathbf{X}_0(\mathbf{X}_0'\Omega_1^{-1}\mathbf{X}_0)^{-1}\mathbf{X}_0'\Omega_1^{-1})\mathbb{E}[\mathbf{R}_{0,t}] + o_p(1). \tag{D.15}$$

By conditions (ii, iii) of the lemma and (D.15),

$$T^{-1}J_T = (\overline{\mathbf{R}}_0 - \widehat{\mathbf{X}}_0\hat{\gamma})'\widehat{\Omega}_0^{-1}(\overline{\mathbf{R}}_0 - \widehat{\mathbf{X}}_0\hat{\gamma}) = \mathbb{E}[\mathbf{R}'_{0,t}]\Pi_{N_0}\mathbb{E}[\mathbf{R}_{0,t}] + o_p(1),$$

which proves the claim in part (b) of the lemma. *Q.E.D.*

D.3 Extension

In this section, we generalize the asymptotic normality and efficiency of the OCSR estimator established in Section 2 to the case where the number of assets N grows with T. Since the number of assets in $\mathcal{I}_1$ may grow with N, the number of the unknown parameters (i.e., α_i for $i \in \mathcal{I}_1$) may also diverge to infinity with T. We first state the conditions needed for the extension. In the following, δ_j $(j = 1, 2, 3)$ are nonnegative finite constants.

Assumption 4. *Suppose that: (i) $\{f_t\}_t$ is a covariance-stationary process; (ii) $\overline{f} = \mu_f + O_p(T^{-1/2})$; (iii) $T^{-1}\sum_{t\leq T} f_t f_t' = \mathbb{E}[f_t f_t'] + O_p(T^{-1/2})$, where $\mathbb{E}[f_t f_t']$ is bounded; and (iv) $\overline{\mathbf{Z}} = T^{-1}\sum_{t\leq T}\mathbb{E}[\mathbf{Z}_{t-1}] + O_p(N^{\delta_1}T^{-1/2})$.*

Assumption 4(i) is the same as Assumption 2(i). Assumptions 4(ii, iii, iv) strengthen Assumption 2(ii, iii, iv) by providing the convergence rates of $\overline{f}$, $T^{-1}\sum_{t\leq T} f_t f_t'$ and $\overline{\mathbf{Z}}$ to their population counterparts. Note that Assumption 4(iv) does not impose the stationary assumption on the $\mathbf{Z}_t$ process. The factor N^{δ_1} shows up in Assumption 4(iv) because $\overline{\mathbf{Z}}$ is a $N \times 1$ random vector and N may go to infinity with T. These conditions can be verified under low-level sufficient conditions. For example, when the eigenvalues of the variance-covariance matrix of $\overline{\mathbf{Z}}$ are bounded from above uniformly over N and T, one can show that Assumption 2(iv) holds with $\delta_1 = 1/2$.

Assumption 5. *(i) There exists a $N \times 1$ standard normal random vector $\mathcal{N}$ such that*

$$T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\epsilon}_t = \Omega_T^{1/2}\mathcal{N} + o_p(1),$$

where $\Omega_T \equiv \mathrm{Var}(T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\epsilon}_t)$; (ii) $T^{-1}\sum_{t=1}^{T}\mathbf{u}_t(1, f_t') = O_p(N^{\delta_2}T^{-1/2})$; (iii) there exists a non-random symmetric matrix W such that $\hat{W} = W + O_p(N^{\delta_3}T^{-1/2})$; (iv) the eigenvalues of Ω_T, Σ_f, W and $\mathbf{X}'\mathbf{X}$ are bounded from above and away from zero uniformly over N and T where $\mathbf{X} = (S_N, \mathbf{1}_{N\times 1}, \mathbf{B}, T^{-1}\sum_{t\leq T}\mathbb{E}[\mathbf{Z}_{t-1}])$; and (v) $N^{\delta+1/2}T^{-1/2} \to 0$ as $N, T \to \infty$ where $\delta \equiv \max_{j=1,2,3}\delta_j$.

Assumption 5(i) is a high-dimensional central limit theorem on the partial sum $T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\epsilon}_t$, which can be verified when $\{\boldsymbol{\epsilon}_t\}_t$ is from independent processes (see, e.g., Theorem 10.4.10 in Pollard (2002)), from martingale difference arrays, or from heterogeneous dependent processes (see, e.g., Theorem 1 and Theorem 4 in Li and Liao (2020)). By the definitions of $u_{i,t}$ and β_i, we have $\mathbb{E}[\mathbf{u}_t] = \mathbf{0}_{N\times 1}$ and $\mathbb{E}[\mathbf{u}_t f_t'] = \mathbf{0}_{N\times K}$. Therefore, Assumption 5(ii) holds with $\delta_2 = 1/2$ if the eigenvalues of the variance-covariance matrix of $T^{-1/2}\sum_{t=1}^{T}\mathbf{u}_t(1, f_t')$ are bounded uniformly over N and T. Assumption 5(iii) imposes conditions on the weight matrix of the CSR estimator. The eigenvalue conditions in Assumption 5(iv) ensure the local identification of the CSR estimator. Assumption 5(v) imposes an upper bound on N. Since in most cases Assumptions 4(iv) and 5(ii, iii) can be verified with $\delta > 1/2$, Assumption 5(v) implies that N may not go to infinity faster than T.

Lemma 3. *Under Assumptions 4 and 5, we have*

$$T^{1/2}(\hat{\theta}_{csr} - \theta) = (\Sigma_T(W))^{1/2}\mathcal{N}^* + o_p(1), \tag{D.16}$$

where $\Sigma_T(W) \equiv (\mathbf{X}'W\mathbf{X})^{-1}(\mathbf{X}'W\Omega_T W\mathbf{X})(\mathbf{X}'W\mathbf{X})^{-1}$ *and* $\mathcal{N}^*$ *denotes a* $d_\theta \times 1$ *standard normal random vector. Moreover,*

$$\Sigma_T(W) \geq (\mathbf{X}'\Omega_T^{-1}\mathbf{X})^{-1}, \tag{D.17}$$

for any $N \times N$ *symmetric positive definite matrix* W.

Lemma 3 generalizes Lemma 1 to the case where both the number of assets N and the number of the unknown parameters d_θ may go to infinity with T. Since the eigenvalues of $\Sigma_T(W)$ are bounded away from zero under Assumption 5(iv), $T^{1/2}(\hat{\theta}_{csr} - \theta)$ is not asymptotically tight, and hence it does not admit an asymptotic distribution. Nevertheless, (D.16) implies that the finite sample distribution of $T^{1/2}(\hat{\theta}_{csr} - \theta)$ can still be approximated by a normal random vector with variance-covariance matrix $\Sigma_T(W)$. Therefore, one can still conduct inference on θ using the normal approximation (D.16).

From (D.17), it is clear that the OCSR estimator still takes the form in (D.1) and the "pre-asymptotic" variance-covariance matrix of the OCSR estimator is $(\mathbf{X}'\Omega_T^{-1}\mathbf{X})^{-1}$.[43] Therefore, we can use the same arguments in the proof of Proposition 2 to show that the OCSR estimator has the same "asymptotic" variance-covariance matrix of optimal GMM estimator $\hat{\theta}^*_{gmm}$.

[43] Note that the OCSR depends on a consistent estimator of Ω_T. See Theorem 5 in Li and Liao (2020) for a consistent variance-covariance estimator with divergent dimension.

Proof of Lemma 3. By Assumption 4(i, ii, iii),

$$\widehat{\Sigma}_f = \Sigma_f + O_p(T^{-1/2}), \tag{D.18}$$

which together with (A.3) in the proof of Lemma 1, Assumptions 4(ii) and 5(ii, iv, v) implies that

$$\widehat{\mathbf{B}} - \mathbf{B} = T^{-1}\sum_{t\leq T}\mathbf{u}_t(f_t-\mu_f)'\Sigma_f^{-1} + o_p(T^{-1/2}) = O_p(N^{\delta}T^{-1/2}) = o_p(1). \tag{D.19}$$

Therefore by Assumptions 5(i, iv) and (D.19), we have

$$\overline{\mathbf{v}} - (\widehat{\mathbf{B}} - \mathbf{B})\gamma_1 = T^{-1}\sum_{t\leq T}\boldsymbol{\epsilon}_t + o_p(T^{-1/2}) = O_p((N/T)^{1/2}). \tag{D.20}$$

Similarly, by Assumptions 4(iv) and 5(v), and (D.19)

$$\widehat{\mathbf{X}} = \mathbf{X} + O_p(N^{\delta}T^{-1/2}) = \mathbf{X} + o_p(1). \tag{D.21}$$

From Assumptions 5(iii, iv, v) and (D.21), we obtain

$$\widehat{\mathbf{X}}'\widehat{W} = \mathbf{X}'W + O_p(N^{\delta}T^{-1/2}) \text{ and } \widehat{\mathbf{X}}'\widehat{W}\widehat{\mathbf{X}} = \mathbf{X}'W\mathbf{X} + O_p(N^{\delta}T^{-1/2}). \tag{D.22}$$

Combining the results in (A.2), (D.18), and (D.22) and applying Assumptions 5(i, iv), we get

$$\begin{aligned} T^{1/2}(\hat{\theta}_{csr} - \theta) &= (\mathbf{X}'W\mathbf{X})^{-1}\mathbf{X}'WT^{-1/2}\sum_{t\leq T}\boldsymbol{\epsilon}_t + o_p(1) \\ &= (\mathbf{X}'W\mathbf{X})^{-1}\mathbf{X}'W\Omega_T^{1/2}\mathcal{N} + o_p(1). \end{aligned} \tag{D.23}$$

Since $(\mathbf{X}'W\mathbf{X})^{-1}\mathbf{X}'W\Omega_T^{1/2}\mathcal{N}$ has the same distribution as $(\Sigma_T(W))^{1/2}\mathcal{N}^*$, then the claim in (D.16) follows from (D.23). The claim in (D.17) follows by the same arguments on showing (2.10) and hence is omitted. *Q.E.D.*

E Additional Results for the Simulation Study

Table E.1: **Simulated Bias, RMSE, and MAE for Parameter Estimates, $T = 750$**

For a given Fama-French model (i.e., three-factor or five-factor model), we use the 18 low-turnover or the 38 low-turnover and medium-turnover anomaly sample in Novy-Marx and Velikov (2016) as test assets. γ_0, γ_{mkt}, γ_{smb}, γ_{hml}, γ_{cma}, and γ_{rmw} denote the risk premiums associated with the intercept, the market factor, *smb* (size factor), *hml* (value factor), *cma* (investment factor), and *rmw* (profitability factor), respectively. Bold denotes the best performer among all methods considered.

		Panel A: FF 3-Factor Model, $N = 18$				**Panel C: FF 3-Factor Model, $N = 38$**					
		γ_0	γ_{mkt}	γ_{smb}	γ_{hml}	γ_0	γ_{mkt}	γ_{smb}	γ_{hml}	γ_{cma}	γ_{rmw}
	True	0	0.595	0.257	0.339	0	0.595	0.257	0.339	0.287	0.323
OLS	Bias	0.064	-0.060	0.004	-0.030	0.006	0.002	0.010	-0.010	0.005	-0.013
	RMSE	0.247	0.261	0.161	0.165	0.218	0.220	0.155	0.147	0.132	0.223
	MAE	0.201	0.211	0.128	0.133	0.173	0.175	0.123	0.117	0.105	0.177
OCSR	Bias	0.007	0.005	0.002	-0.011	0.022	0.002	0.008	-0.006	0.005	-0.008
	RMSE	**0.154**	**0.181**	**0.101**	**0.136**	**0.170**	**0.181**	**0.100**	**0.140**	**0.114**	**0.150**
	MAE	**0.123**	**0.142**	**0.079**	**0.106**	**0.137**	0.150	**0.078**	**0.110**	**0.090**	**0.116**
GLS	Bias	0.009	0.001	0.004	-0.015	0.016	-0.007	0.002	-0.009	0.004	-0.018
	RMSE	0.169	0.194	0.125	0.152	0.175	0.187	0.123	0.155	0.127	0.170
	MAE	0.134	0.155	0.099	0.121	0.140	**0.149**	0.098	0.124	0.101	0.136
WLS	Bias	0.066	-0.056	0.000	-0.032	0.033	-0.024	0.002	-0.007	0.004	-0.026
	RMSE	0.288	0.307	0.144	0.157	0.244	0.248	0.145	0.148	0.138	0.219
	MAE	0.231	0.247	0.115	0.127	0.194	0.197	0.116	0.118	0.109	0.175
		Panel B: FF 5-Factor Model, $N = 18$				**Panel D: FF 5-Factor Model, $N = 38$**					
		γ_0	γ_{mkt}	γ_{smb}	γ_{hml}	γ_0	γ_{mkt}	γ_{smb}	γ_{hml}	γ_{cma}	γ_{rmw}
	True	0	0.595	0.257	0.339	0	0.595	0.257	0.339	0.287	0.323
OLS	Bias	0.008	-0.008	0.007	-0.021	0.039	-0.033	-0.003	-0.005	-0.006	-0.029
	RMSE	0.191	0.196	0.172	0.150	0.196	0.195	0.149	0.174	0.140	0.200
	MAE	0.152	0.156	0.137	0.121	0.159	0.157	0.119	0.138	0.111	0.160
OCSR	Bias	0.015	0.009	0.010	-0.014	0.051	-0.021	0.004	-0.006	-0.000	-0.029
	RMSE	**0.137**	**0.163**	**0.087**	**0.119**	**0.155**	**0.162**	**0.091**	**0.128**	**0.101**	**0.119**
	MAE	**0.110**	**0.127**	**0.068**	**0.094**	**0.130**	**0.129**	**0.070**	**0.099**	**0.078**	**0.094**
GLS	Bias	0.017	-0.016	-0.002	-0.012	0.038	-0.037	-0.003	-0.000	-0.005	-0.031
	RMSE	0.158	0.189	0.121	0.138	0.171	0.189	0.116	0.145	0.119	0.146
	MAE	0.127	0.151	0.097	0.111	0.140	0.153	0.093	0.115	0.095	0.118
WLS	Bias	0.007	-0.003	0.001	-0.016	0.041	-0.037	-0.001	-0.003	-0.002	-0.036
	RMSE	0.169	0.201	0.146	0.149	0.196	0.209	0.135	0.174	0.140	0.212
	MAE	0.134	0.160	0.117	0.119	0.159	0.169	0.108	0.138	0.111	0.172

Table E.2: **Simulated Size and Power for Hypothesis Tests, $N = 18$, $T = 500$**

We block bootstrap (with a block length of 12 months) to simulate the data-generating process under various parameterizations. We simulate $M = 10,000$ times to estimate test size and test power. We consider six methods: OLS^{1stage} corresponds to the cross-sectional OLS estimator that ignores first-stage beta-estimation uncertainty; FM is the Fama-MacBeth estimator where risk premiums are estimated for each period and simple t-statistics for the time-series of risk-premium estimates are used for hypothesis testing; OLS is the cross-sectional OLS estimator that takes first-stage beta-estimation uncertainty into account through (2.8),;$OCSR$ is our proposed estimator; GLS is the two-pass estimator that uses the variance-covariance matrix for factor-model residuals as the weighting matrix and has standard errors calculated through (2.8); and WLS is the two-pass estimator that sets the off-diagonal elements of GLS's weighting matrix at zero and has standard errors calculated through (2.8). We report both the original size and power (*Ori.*) and the size-adjusted power (*Adj.*), where the statistical cutoff that exactly achieves a prespecified significance level is found and used to calculate the corresponding test power.

Effect	*Sig.*	OLS^{1stage}		FM		OLS		$OCSR$		GLS		WLS	
	level	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.
$0 \times \hat{\gamma}_0$	1%	0.046	n.a.	0.054	n.a.	0.039	n.a.	0.043	n.a.	0.038	n.a.	0.043	n.a.
	5%	0.126	n.a.	0.138	n.a.	0.111	n.a.	0.120	n.a.	0.112	n.a.	0.115	n.a.
	10%	0.197	n.a.	0.215	n.a.	0.179	n.a.	0.190	n.a.	0.178	n.a.	0.188	n.a.
$0.5 \times \hat{\gamma}_0$	1%	0.077	0.020	0.085	0.019	0.062	0.019	0.080	0.022	0.073	0.023	0.074	0.020
	5%	0.184	0.083	0.188	0.079	0.163	0.077	0.185	0.092	0.172	0.092	0.174	0.085
	10%	0.267	0.154	0.270	0.142	0.242	0.148	0.263	0.161	0.247	0.159	0.250	0.153
$1.0 \times \hat{\gamma}_0$	1%	0.140	0.042	0.138	0.037	0.113	0.042	0.184	0.068	0.168	0.063	0.115	0.043
	5%	0.277	0.149	0.270	0.130	0.244	0.133	0.336	0.206	0.301	0.198	0.242	0.129
	10%	0.366	0.242	0.364	0.215	0.338	0.226	0.429	0.304	0.393	0.288	0.335	0.214
$1.5 \times \hat{\gamma}_0$	1%	0.342	0.159	0.333	0.129	0.292	0.137	0.510	0.290	0.456	0.265	0.286	0.128
	5%	0.523	0.355	0.517	0.320	0.477	0.327	0.688	0.539	0.638	0.504	0.473	0.306
	10%	0.613	0.479	0.618	0.445	0.587	0.452	0.769	0.661	0.727	0.622	0.576	0.437
$0 \times \hat{\gamma}_{mkt}$	1%	0.071	n.a.	0.013	n.a.	0.011	n.a.	0.014	n.a.	0.008	n.a.	0.014	n.a.
	5%	0.161	n.a.	0.054	n.a.	0.049	n.a.	0.063	n.a.	0.038	n.a.	0.062	n.a.
	10%	0.236	n.a.	0.107	n.a.	0.099	n.a.	0.117	n.a.	0.082	n.a.	0.112	n.a.
$0.5 \times \hat{\gamma}_{mkt}$	1%	0.169	0.049	0.042	0.035	0.047	0.045	0.150	0.121	0.101	0.116	0.051	0.039
	5%	0.287	0.135	0.133	0.124	0.142	0.143	0.308	0.274	0.250	0.283	0.138	0.122
	10%	0.371	0.210	0.211	0.200	0.222	0.222	0.416	0.390	0.357	0.390	0.220	0.200
$1.0 \times \hat{\gamma}_{mkt}$	1%	0.422	0.201	0.194	0.167	0.201	0.192	0.581	0.527	0.487	0.519	0.207	0.171
	5%	0.570	0.372	0.393	0.376	0.396	0.399	0.778	0.746	0.715	0.752	0.394	0.366
	10%	0.644	0.481	0.512	0.499	0.511	0.512	0.859	0.843	0.816	0.841	0.502	0.481
$1.5 \times \hat{\gamma}_{mkt}$	1%	0.831	0.638	0.773	0.744	0.757	0.748	0.996	0.995	0.988	0.990	0.744	0.702
	5%	0.905	0.795	0.893	0.885	0.884	0.885	1.000	0.999	0.997	0.998	0.873	0.859
	10%	0.930	0.862	0.930	0.925	0.928	0.928	1.000	1.000	0.999	0.999	0.917	0.910
$0 \times \hat{\gamma}_{smb}$	1%	0.476	n.a.	0.035	n.a.	0.018	n.a.	0.016	n.a.	0.018	n.a.	0.017	n.a.
	5%	0.591	n.a.	0.104	n.a.	0.072	n.a.	0.071	n.a.	0.070	n.a.	0.068	n.a.
	10%	0.652	n.a.	0.177	n.a.	0.130	n.a.	0.128	n.a.	0.131	n.a.	0.126	n.a.
$0.5 \times \hat{\gamma}_{smb}$	1%	0.567	0.000	0.058	0.019	0.038	0.023	0.081	0.056	0.063	0.043	0.043	0.027
	5%	0.663	0.114	0.169	0.084	0.126	0.094	0.202	0.166	0.177	0.140	0.131	0.098
	10%	0.714	0.183	0.258	0.164	0.206	0.165	0.294	0.249	0.270	0.227	0.212	0.180
$1.0 \times \hat{\gamma}_{smb}$	1%	0.748	0.000	0.189	0.081	0.143	0.101	0.313	0.258	0.259	0.204	0.161	0.120
	5%	0.808	0.264	0.379	0.243	0.321	0.265	0.532	0.479	0.470	0.414	0.333	0.282
	10%	0.837	0.379	0.487	0.371	0.435	0.384	0.645	0.599	0.589	0.540	0.449	0.404
$1.5 \times \hat{\gamma}_{smb}$	1%	0.973	0.000	0.702	0.497	0.651	0.560	0.912	0.883	0.863	0.818	0.681	0.608
	5%	0.983	0.710	0.866	0.764	0.832	0.793	0.973	0.964	0.950	0.933	0.853	0.816
	10%	0.988	0.818	0.918	0.860	0.896	0.868	0.986	0.981	0.975	0.967	0.909	0.888
$0 \times \hat{\gamma}_{hml}$	1%	0.320	n.a.	0.031	n.a.	0.019	n.a.	0.029	n.a.	0.020	n.a.	0.016	n.a.
	5%	0.443	n.a.	0.101	n.a.	0.076	n.a.	0.093	n.a.	0.073	n.a.	0.070	n.a.
	10%	0.521	n.a.	0.167	n.a.	0.134	n.a.	0.165	n.a.	0.137	n.a.	0.127	n.a.
$0.5 \times \hat{\gamma}_{hml}$	1%	0.410	0.034	0.061	0.019	0.045	0.027	0.090	0.039	0.064	0.036	0.048	0.034
	5%	0.524	0.097	0.157	0.088	0.135	0.096	0.217	0.134	0.177	0.131	0.140	0.109
	10%	0.590	0.167	0.243	0.257	0.210	0.169	0.311	0.228	0.263	0.214	0.213	0.178
$1.0 \times \hat{\gamma}_{hml}$	1%	0.631	0.114	0.226	0.106	0.183	0.128	0.319	0.189	0.251	0.177	0.188	0.148
	5%	0.721	0.262	0.415	0.290	0.368	0.304	0.529	0.406	0.457	0.386	0.362	0.314
	10%	0.767	0.374	0.520	0.415	0.480	0.419	0.635	0.543	0.569	0.510	0.475	0.430
$1.5 \times \hat{\gamma}_{hml}$	1%	0.935	0.444	0.803	0.648	0.743	0.659	0.905	0.821	0.861	0.796	0.736	0.685
	5%	0.966	0.669	0.913	0.851	0.888	0.847	0.967	0.937	0.952	0.930	0.876	0.848
	10%	0.977	0.774	0.947	0.913	0.928	0.908	0.983	0.969	0.974	0.962	0.926	0.909

Table E.3: **Simulated Size and Power for Hypothesis Tests, $N = 38$, $T = 500$**

We block bootstrap (with a block length of 12 months) to simulate the data-generating process under various parameterizations. We simulate $M = 10,000$ times to estimate test size and test power. We consider six methods: OLS^{1stage} corresponds to the cross-sectional OLS estimator that ignores first-stage beta-estimation uncertainty; FM is the Fama-MacBeth estimator where risk premiums are estimated for each period and simple t-statistics for the time-series of risk-premium estimates are used for hypothesis testing; OLS is the cross-sectional OLS estimator that takes first-stage beta-estimation uncertainty into account through (2.8); $OCSR$ is our proposed estimator; GLS is the two-pass estimator that uses the variance-covariance matrix for factor-model residuals as the weighting matrix and has standard errors calculated through (2.8); and WLS is the two-pass estimator that sets the off-diagonal elements of GLS's weighting matrix at zero and has standard errors calculated through (2.8). We report both the original size and power ($Ori.$) and the size-adjusted power ($Adj.$), where the statistical cutoff that exactly achieves a prespecified significance level is found and used to calculate the corresponding test power.

Effect	*Sig.*	OLS^{1stage}		FM		OLS		$OCSR$		GLS		WLS	
	level	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.
$0 \times \hat{\gamma}_0$	1%	0.308	n.a.	0.038	n.a.	0.031	n.a.	0.137	n.a.	0.102	n.a.	0.042	n.a.
	5%	0.432	n.a.	0.108	n.a.	0.091	n.a.	0.249	n.a.	0.205	n.a.	0.110	n.a.
	10%	0.508	n.a.	0.173	n.a.	0.150	n.a.	0.314	n.a.	0.288	n.a.	0.176	n.a.
$0.5 \times \hat{\gamma}_0$	1%	0.367	0.022	0.064	0.018	0.052	0.022	0.242	0.034	0.185	0.028	0.076	0.024
	5%	0.487	0.080	0.153	0.081	0.136	0.078	0.372	0.116	0.304	0.107	0.172	0.091
	10%	0.559	0.141	0.225	0.144	0.207	0.149	0.454	0.193	0.384	0.182	0.251	0.159
$1.0 \times \hat{\gamma}_0$	1%	0.481	0.058	0.123	0.044	0.101	0.045	0.409	0.088	0.323	0.069	0.155	0.064
	5%	0.595	0.157	0.242	0.153	0.224	0.147	0.553	0.234	0.471	0.207	0.295	0.180
	10%	0.660	0.240	0.331	0.230	0.313	0.241	0.626	0.347	0.549	0.320	0.389	0.278
$1.5 \times \hat{\gamma}_0$	1%	0.769	0.220	0.366	0.169	0.320	0.188	0.800	0.351	0.691	0.274	0.477	0.272
	5%	0.838	0.428	0.556	0.414	0.516	0.400	0.884	0.623	0.813	0.555	0.660	0.518
	10%	0.872	0.544	0.652	0.537	0.618	0.536	0.919	0.749	0.861	0.687	0.746	0.643
$0 \times \hat{\gamma}_{mkt}$	1%	0.262	n.a.	0.003	n.a.	0.006	n.a.	0.034	n.a.	0.011	n.a.	0.007	n.a.
	5%	0.396	n.a.	0.020	n.a.	0.031	n.a.	0.105	n.a.	0.054	n.a.	0.035	n.a.
	10%	0.472	n.a.	0.052	n.a.	0.070	n.a.	0.175	n.a.	0.100	n.a.	0.073	n.a.
$0.5 \times \hat{\gamma}_{mkt}$	1%	0.574	0.101	0.041	0.106	0.076	0.111	0.257	0.116	0.118	0.111	0.095	0.120
	5%	0.679	0.260	0.154	0.250	0.203	0.261	0.443	0.308	0.270	0.258	0.240	0.287
	10%	0.733	0.373	0.257	0.367	0.306	0.369	0.543	0.429	0.384	0.382	0.348	0.399
$1.0 \times \hat{\gamma}_{mkt}$	1%	0.892	0.419	0.303	0.486	0.416	0.498	0.764	0.544	0.535	0.520	0.484	0.537
	5%	0.938	0.662	0.568	0.686	0.653	0.718	0.897	0.814	0.742	0.731	0.709	0.758
	10%	0.951	0.766	0.692	0.791	0.760	0.809	0.940	0.890	0.824	0.823	0.808	0.843
$1.5 \times \hat{\gamma}_{mkt}$	1%	0.999	0.927	0.948	0.982	0.971	0.982	1.000	0.998	0.992	0.991	0.991	0.994
	5%	1.000	0.983	0.990	0.997	0.993	0.996	1.000	1.000	0.998	0.998	0.999	1.000
	10%	1.000	0.994	0.997	0.999	0.996	0.998	1.000	1.000	0.999	0.999	1.000	1.000
$0 \times \hat{\gamma}_{smb}$	1%	0.578	n.a.	0.028	n.a.	0.017	n.a.	0.030	n.a.	0.017	n.a.	0.015	n.a.
	5%	0.671	n.a.	0.094	n.a.	0.062	n.a.	0.097	n.a.	0.073	n.a.	0.067	n.a.
	10%	0.720	n.a.	0.162	n.a.	0.117	n.a.	0.158	n.a.	0.129	n.a.	0.125	n.a.
$0.5 \times \hat{\gamma}_{smb}$	1%	0.639	0.000	0.051	0.019	0.032	0.017	0.144	0.065	0.067	0.047	0.043	0.032
	5%	0.720	0.000	0.154	0.087	0.117	0.095	0.293	0.192	0.182	0.140	0.133	0.108
	10%	0.765	0.190	0.239	0.161	0.192	0.172	0.399	0.298	0.277	0.231	0.210	0.181
$1.0 \times \hat{\gamma}_{smb}$	1%	0.785	0.000	0.146	0.065	0.110	0.073	0.461	0.285	0.257	0.208	0.135	0.109
	5%	0.838	0.000	0.323	0.215	0.271	0.232	0.670	0.541	0.470	0.402	0.309	0.264
	10%	0.863	0.384	0.433	0.334	0.384	0.353	0.760	0.675	0.587	0.537	0.427	0.382
$1.5 \times \hat{\gamma}_{smb}$	1%	0.971	0.000	0.592	0.414	0.553	0.466	0.967	0.916	0.876	0.839	0.616	0.562
	5%	0.980	0.000	0.792	0.690	0.763	0.732	0.991	0.980	0.958	0.941	0.811	0.774
	10%	0.984	0.790	0.863	0.801	0.847	0.831	0.996	0.991	0.979	0.970	0.880	0.856
$0 \times \hat{\gamma}_{hml}$	1%	0.378	n.a.	0.021	n.a.	0.014	n.a.	0.044	n.a.	0.014	n.a.	0.012	n.a.
	5%	0.500	n.a.	0.077	n.a.	0.057	n.a.	0.115	n.a.	0.061	n.a.	0.060	n.a.
	10%	0.572	n.a.	0.132	n.a.	0.108	n.a.	0.185	n.a.	0.117	n.a.	0.107	n.a.
$0.5 \times \hat{\gamma}_{hml}$	1%	0.516	0.042	0.063	0.038	0.049	0.041	0.131	0.038	0.061	0.045	0.054	0.048
	5%	0.622	0.131	0.163	0.118	0.146	0.133	0.269	0.147	0.172	0.150	0.147	0.129
	10%	0.676	0.217	0.251	0.202	0.230	0.219	0.363	0.242	0.262	0.234	0.228	0.217
$1.0 \times \hat{\gamma}_{hml}$	1%	0.785	0.170	0.265	0.191	0.220	0.189	0.439	0.220	0.289	0.243	0.232	0.219
	5%	0.847	0.361	0.463	0.389	0.417	0.393	0.635	0.470	0.513	0.477	0.437	0.408
	10%	0.878	0.492	0.580	0.516	0.536	0.521	0.727	0.607	0.626	0.593	0.557	0.541
$1.5 \times \hat{\gamma}_{hml}$	1%	0.990	0.608	0.869	0.814	0.812	0.783	0.959	0.870	0.905	0.877	0.832	0.823
	5%	0.995	0.825	0.954	0.932	0.931	0.924	0.986	0.965	0.974	0.968	0.941	0.931
	10%	0.997	0.910	0.976	0.965	0.966	0.964	0.992	0.983	0.988	0.984	0.971	0.968

Table E.4: **Simulated Size and Power for Hypothesis Tests, $N = 18$, $T = 750$**

We block bootstrap (with a block length of 12 months) to simulate the data-generating process under various parameterizations. We simulate $M = 10,000$ times to estimate test size and test power. We consider six methods: OLS^{1stage} corresponds to the cross-sectional OLS estimator that ignores first-stage beta-estimation uncertainty; FM is the Fama-MacBeth estimator where risk premiums are estimated for each period and simple t-statistics for the time-series of risk-premium estimates are used for hypothesis testing; OLS is the cross-sectional OLS estimator that takes first-stage beta-estimation uncertainty into account through (2.8); $OCSR$ is our proposed estimator; GLS is the two-pass estimator that uses the variance-covariance matrix for factor-model residuals as the weighting matrix and has standard errors calculated through (2.8); and WLS is the two-pass estimator that sets the off-diagonal elements of GLS's weighting matrix at zero and has standard errors calculated through (2.8). We report both the original size and power ($Ori.$) and the size-adjusted power ($Adj.$), where the statistical cutoff that exactly achieves a prespecified significance level is found and used to calculate the corresponding test power.

Effect	*Sig.*	OLS^{1stage}		FM		OLS		$OCSR$		GLS		WLS	
	level	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.
$0 \times \hat{\gamma}_0$	1%	0.034	n.a.	0.048	n.a.	0.012	n.a.	0.027	n.a.	0.029	n.a.	0.036	n.a.
	5%	0.109	n.a.	0.131	n.a.	0.051	n.a.	0.089	n.a.	0.094	n.a.	0.108	n.a.
	10%	0.178	n.a.	0.202	n.a.	0.096	n.a.	0.155	n.a.	0.160	n.a.	0.177	n.a.
$0.5 \times \hat{\gamma}_0$	1%	0.076	0.030	0.074	0.016	0.023	0.020	0.078	0.039	0.074	0.036	0.068	0.019
	5%	0.176	0.100	0.171	0.076	0.089	0.087	0.178	0.121	0.176	0.112	0.165	0.087
	10%	0.254	0.166	0.252	0.130	0.156	0.159	0.265	0.193	0.258	0.185	0.242	0.155
$1.0 \times \hat{\gamma}_0$	1%	0.150	0.068	0.131	0.032	0.054	0.050	0.209	0.126	0.195	0.113	0.127	0.044
	5%	0.296	0.195	0.259	0.134	0.168	0.166	0.384	0.289	0.354	0.258	0.276	0.163
	10%	0.396	0.284	0.350	0.209	0.263	0.268	0.487	0.402	0.457	0.366	0.373	0.263
$1.5 \times \hat{\gamma}_0$	1%	0.419	0.264	0.305	0.113	0.219	0.203	0.660	0.524	0.598	0.467	0.376	0.186
	5%	0.602	0.479	0.489	0.310	0.448	0.445	0.813	0.742	0.768	0.682	0.582	0.428
	10%	0.689	0.588	0.593	0.429	0.577	0.584	0.874	0.824	0.839	0.778	0.683	0.567
$0 \times \hat{\gamma}_{mkt}$	1%	0.054	n.a.	0.015	n.a.	0.002	n.a.	0.010	n.a.	0.007	n.a.	0.013	n.a.
	5%	0.137	n.a.	0.058	n.a.	0.019	n.a.	0.051	n.a.	0.035	n.a.	0.052	n.a.
	10%	0.211	n.a.	0.109	n.a.	0.051	n.a.	0.101	n.a.	0.075	n.a.	0.104	n.a.
$0.5 \times \hat{\gamma}_{mkt}$	1%	0.223	0.090	0.049	0.037	0.023	0.060	0.201	0.203	0.154	0.190	0.076	0.066
	5%	0.358	0.215	0.146	0.132	0.094	0.169	0.393	0.392	0.349	0.402	0.194	0.187
	10%	0.443	0.311	0.234	0.221	0.171	0.272	0.514	0.513	0.469	0.516	0.290	0.285
$1.0 \times \hat{\gamma}_{mkt}$	1%	0.562	0.355	0.223	0.185	0.184	0.328	0.768	0.771	0.702	0.753	0.350	0.325
	5%	0.707	0.553	0.429	0.406	0.410	0.549	0.917	0.916	0.877	0.904	0.565	0.556
	10%	0.772	0.659	0.545	0.530	0.552	0.669	0.958	0.957	0.932	0.946	0.673	0.668
$1.5 \times \hat{\gamma}_{mkt}$	1%	0.918	0.813	0.807	0.770	0.845	0.923	1.000	1.000	1.000	1.000	0.919	0.911
	5%	0.963	0.914	0.919	0.908	0.949	0.972	1.000	1.000	1.000	1.000	0.970	0.968
	10%	0.978	0.953	0.951	0.947	0.973	0.986	1.000	1.000	1.000	1.000	0.985	0.985
$0 \times \hat{\gamma}_{smb}$	1%	0.480	n.a.	0.036	n.a.	0.004	n.a.	0.014	n.a.	0.018	n.a.	0.017	n.a.
	5%	0.595	n.a.	0.103	n.a.	0.023	n.a.	0.061	n.a.	0.068	n.a.	0.067	n.a.
	10%	0.656	n.a.	0.173	n.a.	0.060	n.a.	0.115	n.a.	0.128	n.a.	0.122	n.a.
$0.5 \times \hat{\gamma}_{smb}$	1%	0.602	0.000	0.059	0.016	0.015	0.040	0.105	0.081	0.096	0.062	0.057	0.034
	5%	0.687	0.142	0.165	0.087	0.074	0.122	0.250	0.223	0.231	0.191	0.165	0.134
	10%	0.736	0.229	0.251	0.161	0.142	0.204	0.359	0.331	0.330	0.289	0.259	0.222
$1.0 \times \hat{\gamma}_{smb}$	1%	0.816	0.000	0.178	0.066	0.099	0.191	0.454	0.400	0.391	0.309	0.247	0.183
	5%	0.862	0.372	0.370	0.237	0.276	0.377	0.677	0.649	0.618	0.563	0.456	0.403
	10%	0.885	0.499	0.485	0.365	0.406	0.497	0.782	0.756	0.725	0.687	0.573	0.533
$1.5 \times \hat{\gamma}_{smb}$	1%	0.990	0.000	0.689	0.468	0.665	0.805	0.988	0.981	0.967	0.945	0.862	0.802
	5%	0.995	0.841	0.852	0.755	0.873	0.926	0.998	0.997	0.993	0.990	0.949	0.935
	10%	0.996	0.907	0.905	0.849	0.937	0.962	0.999	0.999	0.997	0.996	0.973	0.966
$0 \times \hat{\gamma}_{hml}$	1%	0.311	n.a.	0.032	n.a.	0.003	n.a.	0.022	n.a.	0.017	n.a.	0.015	n.a.
	5%	0.438	n.a.	0.101	n.a.	0.024	n.a.	0.078	n.a.	0.071	n.a.	0.066	n.a.
	10%	0.512	n.a.	0.167	n.a.	0.059	n.a.	0.136	n.a.	0.127	n.a.	0.123	n.a.
$0.5 \times \hat{\gamma}_{hml}$	1%	0.460	0.040	0.068	0.025	0.025	0.053	0.112	0.054	0.092	0.059	0.071	0.053
	5%	0.569	0.136	0.177	0.099	0.098	0.159	0.258	0.194	0.223	0.177	0.179	0.153
	10%	0.634	0.220	0.266	0.176	0.171	0.239	0.359	0.302	0.319	0.276	0.268	0.236
$1.0 \times \hat{\gamma}_{hml}$	1%	0.757	0.169	0.249	0.128	0.157	0.248	0.475	0.323	0.401	0.314	0.309	0.251
	5%	0.830	0.384	0.442	0.314	0.367	0.477	0.685	0.611	0.624	0.564	0.524	0.483
	10%	0.863	0.516	0.548	0.439	0.497	0.591	0.777	0.727	0.727	0.690	0.640	0.601
$1.5 \times \hat{\gamma}_{hml}$	1%	0.980	0.592	0.824	0.685	0.813	0.890	0.984	0.962	0.972	0.951	0.913	0.887
	5%	0.992	0.818	0.922	0.867	0.944	0.967	0.996	0.994	0.993	0.991	0.973	0.965
	10%	0.995	0.897	0.953	0.921	0.970	0.981	0.999	0.997	0.997	0.996	0.986	0.982

Table E.5: **Simulated Size and Power for Hypothesis Tests,** $N = 38$**,** $T = 750$

We block bootstrap (with a block length of 12 months) to simulate the data-generating process under various parameterizations. We simulate $M = 10,000$ times to estimate test size and test power. We consider six methods: OLS^{1stage} corresponds to the cross-sectional OLS estimator that ignores first-stage beta-estimation uncertainty; FM is the Fama-MacBeth estimator where risk premiums are estimated for each period and simple t-statistics for the time-series of risk-premium estimates are used for hypothesis testing; OLS is the cross-sectional OLS estimator that takes first-stage beta-estimation uncertainty into account through (2.8); $OCSR$ is our proposed estimator; GLS is the two-pass estimator that uses the variance-covariance matrix for factor-model residuals as the weighting matrix and has standard errors calculated through (2.8); and WLS is the two-pass estimator that sets the off-diagonal elements of GLS's weighting matrix at zero and has standard errors calculated through (2.8). We report both the original size and power ($Ori.$) and the size-adjusted power ($Adj.$), where the statistical cutoff that exactly achieves a prespecified significance level is found and used to calculate the corresponding test power.

Effect	*Sig.*	OLS^{1stage}		FM		OLS		$OCSR$		GLS		WLS	
	level	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.	Ori.	Adj.
$0\hat{\gamma}_0$	1%	0.303	n.a.	0.042	n.a.	0.008	n.a.	0.097	n.a.	0.091	n.a.	0.034	n.a.
	5%	0.433	n.a.	0.114	n.a.	0.037	n.a.	0.201	n.a.	0.194	n.a.	0.105	n.a.
	10%	0.512	n.a.	0.182	n.a.	0.081	n.a.	0.280	n.a.	0.270	n.a.	0.172	n.a.
$0.5\hat{\gamma}_0$	1%	0.380	0.026	0.059	0.018	0.016	0.021	0.219	0.048	0.182	0.043	0.080	0.033
	5%	0.511	0.090	0.152	0.070	0.068	0.085	0.350	0.139	0.308	0.122	0.186	0.105
	10%	0.581	0.156	0.227	0.136	0.127	0.153	0.435	0.222	0.392	0.194	0.269	0.181
$1.0\hat{\gamma}_0$	1%	0.544	0.084	0.117	0.040	0.060	0.071	0.450	0.149	0.366	0.123	0.204	0.109
	5%	0.648	0.210	0.246	0.135	0.163	0.192	0.604	0.326	0.522	0.274	0.370	0.248
	10%	0.705	0.312	0.335	0.224	0.256	0.286	0.683	0.457	0.605	0.385	0.463	0.363
$1.5\hat{\gamma}_0$	1%	0.850	0.350	0.352	0.176	0.275	0.311	0.892	0.593	0.796	0.480	0.624	0.450
	5%	0.901	0.555	0.537	0.384	0.505	0.551	0.948	0.818	0.892	0.709	0.788	0.678
	10%	0.923	0.672	0.630	0.509	0.637	0.669	0.967	0.895	0.927	0.808	0.854	0.782
$0\hat{\gamma}_{mkt}$	1%	0.270	n.a.	0.004	n.a.	0.001	n.a.	0.022	n.a.	0.011	n.a.	0.007	n.a.
	5%	0.402	n.a.	0.028	n.a.	0.010	n.a.	0.073	n.a.	0.051	n.a.	0.038	n.a.
	10%	0.483	n.a.	0.062	n.a.	0.028	n.a.	0.135	n.a.	0.102	n.a.	0.073	n.a.
$0.5\hat{\gamma}_{mkt}$	1%	0.671	0.140	0.050	0.091	0.044	0.166	0.325	0.230	0.184	0.172	0.150	0.186
	5%	0.765	0.352	0.166	0.231	0.165	0.357	0.534	0.469	0.373	0.367	0.333	0.381
	10%	0.809	0.471	0.260	0.339	0.274	0.472	0.651	0.593	0.492	0.491	0.459	0.515
$1.0\hat{\gamma}_{mkt}$	1%	0.960	0.578	0.328	0.444	0.408	0.694	0.918	0.858	0.744	0.731	0.706	0.751
	5%	0.977	0.816	0.579	0.672	0.693	0.868	0.980	0.968	0.890	0.886	0.879	0.902
	10%	0.985	0.886	0.704	0.775	0.814	0.922	0.986	0.978	0.935	0.935	0.933	0.950
$1.5\hat{\gamma}_{mkt}$	1%	1.000	0.978	0.944	0.971	0.991	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	5%	1.000	0.997	0.987	0.994	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10%	1.000	0.999	0.995	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$0\hat{\gamma}_{smb}$	1%	0.575	n.a.	0.028	n.a.	0.003	n.a.	0.021	n.a.	0.018	n.a.	0.013	n.a.
	5%	0.669	n.a.	0.097	n.a.	0.021	n.a.	0.069	n.a.	0.069	n.a.	0.063	n.a.
	10%	0.720	n.a.	0.162	n.a.	0.052	n.a.	0.125	n.a.	0.127	n.a.	0.120	n.a.
$0.5\hat{\gamma}_{smb}$	1%	0.671	0.000	0.044	0.016	0.012	0.041	0.154	0.094	0.089	0.057	0.052	0.040
	5%	0.749	0.000	0.138	0.077	0.069	0.126	0.329	0.273	0.223	0.179	0.153	0.129
	10%	0.786	0.219	0.227	0.143	0.129	0.199	0.442	0.397	0.321	0.281	0.241	0.211
$1.0\hat{\gamma}_{smb}$	1%	0.842	0.000	0.138	0.064	0.059	0.144	0.615	0.496	0.400	0.318	0.211	0.176
	5%	0.884	0.000	0.313	0.208	0.219	0.342	0.810	0.760	0.626	0.566	0.419	0.382
	10%	0.903	0.469	0.424	0.319	0.346	0.453	0.881	0.857	0.735	0.695	0.537	0.502
$1.5\hat{\gamma}_{smb}$	1%	0.991	0.000	0.587	0.416	0.542	0.144	0.998	0.995	0.976	0.961	0.815	0.782
	5%	0.994	0.000	0.786	0.685	0.800	0.886	1.000	1.000	0.995	0.992	0.932	0.916
	10%	0.996	0.890	0.863	0.791	0.887	0.930	1.000	1.000	0.998	0.998	0.964	0.955
$0\hat{\gamma}_{hml}$	1%	0.381	n.a.	0.023	n.a.	0.002	n.a.	0.025	n.a.	0.013	n.a.	0.012	n.a.
	5%	0.505	n.a.	0.081	n.a.	0.018	n.a.	0.082	n.a.	0.061	n.a.	0.054	n.a.
	10%	0.575	n.a.	0.145	n.a.	0.047	n.a.	0.143	n.a.	0.115	n.a.	0.104	n.a.
$0.5\hat{\gamma}_{hml}$	1%	0.581	0.068	0.070	0.037	0.025	0.074	0.151	0.083	0.092	0.080	0.076	0.068
	5%	0.680	0.184	0.181	0.128	0.104	0.192	0.317	0.241	0.231	0.201	0.205	0.197
	10%	0.727	0.275	0.272	0.211	0.186	0.286	0.427	0.353	0.336	0.310	0.304	0.298
$1.0\hat{\gamma}_{hml}$	1%	0.882	0.280	0.275	0.184	0.178	0.354	0.608	0.464	0.448	0.411	0.375	0.354
	5%	0.924	0.512	0.494	0.407	0.427	0.584	0.796	0.726	0.674	0.640	0.604	0.593
	10%	0.940	0.639	0.605	0.534	0.571	0.689	0.865	0.820	0.771	0.750	0.720	0.714
$1.5\hat{\gamma}_{hml}$	1%	0.999	0.812	0.877	0.811	0.879	0.956	0.997	0.992	0.989	0.986	0.961	0.956
	5%	1.000	0.942	0.956	0.932	0.971	0.989	0.999	0.999	0.998	0.997	0.990	0.990
	10%	1.000	0.974	0.978	0.966	0.988	0.994	1.000	0.999	0.999	0.999	0.996	0.996