Optimal Cross-Sectional Regression *

Zhipeng Liao
UCLA, Los Angeles, CA 90095 USA

Yan Liu
Purdue University, West Lafayette, IN 47906 USA

November 27, 2020

Abstract

In the context of linear-beta pricing models, we develop a new class of two-pass estimators that are available in closed form and dominate existing two-pass estimators in terms of estimation efficiency. Importantly, we map our model into the generalized method of moments (GMM) framework and show our two-pass estimator is as efficient as the optimal GMM estimator, which is known to be semiparametrically efficient in the literature. Hence, contrary to popular belief, information loss does not need to occur when we go from the more methodical GMM approach to the simple-to-implement two-pass regressors. Intuitively, our estimator improves efficiency by disentangling the impacts of idiosyncratic and systematic return innovations on pricing errors in the second-stage cross-sectional regression. As an empirical application of the new two-pass estimators, we apply our approach to current factor models and shed new light on the Fama and French (2015) versus Hou, Xue, and Zhang (2015) debate.

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

JEL Code: C14, C22.

*Current Version: November 27, 2020. First posted to SSRN: October 26, 2020. Correspondence: Zhipeng.liao@econ.ucla.edu, liu2746@purdue.edu. We have benefited from the comments of John Cochrane, Jason Donaldson, Philip Dybvig, Wayne Ferson, Jinyong Hahn, Cam Harvey, Asaf Manela, Andreas Neuhierl, David Rapach, Junbo Wang, and Guofu Zhou, as well as seminar participants at Washington University in St. Louis. We are especially grateful to Raymond Kan for comments on several versions of the paper.
1 Introduction

A fundamental issue in finance is the risk and return relationship. Investigation of this relationship has been complicated by the proliferation of anomalies (McLean and Pontiff (2016), Harvey, Liu, and Zhu (2016)) and the accompanying new evaluation techniques.\(^1\) Although different techniques may work well in different contexts, choosing among them is challenging and may lead to model dredging and eventually \(p\)-hacking (Lo and MacKinlay (1990), Harvey (2017)). In this paper, we argue for the need to recast new techniques into existing frameworks, which allows one to compare them on equal footing. We demonstrate this using the generalized method of moments (GMM) framework—one of the cornerstones of modern financial economics.

Specifically, 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 to achieve the semiparametric efficiency bound in the literature. Therefore, we provide a unified view on existing two-pass regression estimators and identify the optimal ones (hereafter, optimal cross-sectional regression (OCSR) estimators) in a well-defined sense.

Our research fills in an important gap in the literature on linear-beta pricing models. Under the assumption that idiosyncratic risks are i.i.d. normal and conditionally homoscedastic given the pricing factors, Shanken (1992) proves the equivalence between the generalized least-squares (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 non-normal, dependent and conditional heteroskedastic idiosyncratic risks. More importantly, we extend Shanken (1992)'s insight by showing 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 information loss is not necessary when one goes from the more methodical GMM approach to the simple-to-implement

From a practical point of view, we expect our OCSR to be useful to empirical researchers for several reasons. First, although Shanken (1992) shows GLS is efficient under the i.i.d. normal and conditional homoscedasticity assumptions, researchers still select cross-sectional methods in an ad hoc manner. Given the optimality of our estimator under general assumptions on the return-generating process, it may serve as the default choice in evaluating beta pricing models. Second, despite the strong theoretical appeal of the GMM framework, the GMM objective function is both high dimensional and nonlinear in model parameters and thus computationally challenging, which may explain its infrequent use in estimating linear-beta pricing models (Shanken and Zhou (2007)). By contrast, our OCSR retains the theoretical appeal of GMM but is available in closed form, and thus is straightforward to implement.

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). They 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 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 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 this bound in the same setup. This efficiency result is intuitively predicted by Ackerberg, Chen, Hahn, and Liao (2014), who show, in a generic structural economic model with unknown functions estimated through a first-step nonparametric procedure, how a semiparametric two-step GMM estimator achieves the semiparametric efficiency bound. Different from their paper, we formulate the problem in the context of linear-beta pricing models and show the sequential estimation of betas (in a first-stage regression) and risk premiums (in a second-

---

2 Alternative cross-sectional methods have also been proposed in Kim (1995) and Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2019). Different from 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.
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.

Lastly, 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 homoscedasticity, 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 non-normality, 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 achieve GMM’s optimality that improves on the parametric efficiency of the GLS estimator.

We conduct a comprehensive simulation study to examine the performance of our OCSR estimator against existing cross-sectional methods. To highlight the difference between our OCSR estimator and the GLS estimator in Shanken (1992), our simulation design attempts to capture key features of the actual data, especially conditional heteroskedasticity. Instead of imposing a parametric distribution on the data-generating process (DGP), we block bootstrap both factor and asset returns to simulate the DGP. Using a diverse set of factor models and anomaly portfolios, we document the superior performance of our estimator against alternative two-pass regressors for a $T$ (number of months) of around 500 (which roughly corresponds to the sample size of the anomaly portfolios we use): the percentage reduction in root-mean-squared error (i.e., RMSE) of our estimator compared with the two-pass GLS estimator can reach a magnitude of 10%. The reduction is even more substantial for a larger $T$ (e.g., $T = 750$). Similarly, for hypothesis tests, OCSR is more powerful than alternative two-pass estimators in terms of both original test power or the size-adjusted power, with a power gain in the neighborhood of 5%-10%.

3By contrast, our simulation results suggest finite-sample bias is usually one or two orders of magnitude smaller than RMSE, suggesting the importance of efficiency compared with bias. We thus focus on estimation efficiency and do not explicitly consider finite-sample bias adjustment (e.g., Shanken (1992)) in our paper.
Empirically, we apply our approach to shed light on the recent debate between Fama and French (FF, 2015) and Hou, Xue, and Zhang (HXZ, 2015). While both papers evaluate model performance using the time-series approach (i.e., termed the left-hand-side approach by Fama and French (2018)), we examine their performances using the alternative cross-sectional regression framework. To this end, we assume a set of $N_0$ basis portfolios are correctly priced by a given factor model, and use the two-pass cross-sectional regression to evaluate the abnormal return (i.e., alpha) of the $(N_0+1)$-th asset, that is, the candidate anomaly portfolio. Given both papers propose factor models that purport to explain the returns of a large number of assets, our assumption constitutes the minimal economic restriction on all-encompassing factor models (Fama and French (1993)).

More specifically, we take the 20 portfolios in Fama and French (2015) and the 18 portfolios in Hou, Xue, and Zhang (2015) (which are used to construct their respective factor models) as basis portfolios. After fixing the basis portfolios, we use the 156 anomaly long-short portfolios in Chen and Zimmermann (2020) as test portfolios (one by one).

Comparing time-series estimates of alphas with their cross-sectional counterparts, we find a stark contrast in performance between FF and HXZ. Whereas the time-series estimates corresponding to the 156 anomaly portfolios are highly correlated with their cross-sectional estimates for FF, this relation is substantially blurred for HXZ. This contrast leads us to examine the pricing performance of the basis portfolios, allowing us to uncover the disparity between FF and HXZ in pricing their respective basis portfolios. Whereas only two (zero) out of 20 FF basis portfolios have an alpha $t$-statistic (in absolute value) above 2.0 (3.0) when regressed on the FF model, nine (eight) out of the 18 recently publicized HXZ basis portfolios have $t$-statistics above 2.0 (3.0) against the HXZ model. An over-identification test also does not reject the FF model but strongly rejects the HXZ model.

Our result above highlights the internal conflicts of HXZ (relative to FF) in explaining its own basis portfolios, and has implications that go beyond the scope of our paper. As FF commented on HXZ: “More important, they (HXZ) are primarily concerned with explaining the returns associated with anomaly variables not used to construct their factors.” Our results provide support to FF’s comment by showing HXZ fail to explain almost half of their basis portfolios. From the perspective of comparing testing methods, we show additional insights can be gained by contrasting the time-series alpha estimates with the ones generated by cross-sectional regressions.

Given that the over-identification assumption approximately holds for FF’s basis portfolios, we focus on the FF model to compare our OCSR with alternative cross-sectional models. Among

---

estimators that have mis-specified standard errors, including the naive OLS two-pass regressor (i.e., using the identity weight matrix in the cross-sectional regression and not adjusting for first-stage beta estimation) and the usual Fama-MacBeth estimator, they lead to too many rejections that are likely false because standard errors do not adequately adjust for beta estimation.\(^5\) On the other hand, among estimators that have correctly specified standard errors but only differ in their weighting matrix in the second-stage regression, OCSR is the most powerful in detecting non-zero alphas. For example, with 156 test assets, OCSR leads to 10\% more rejections with value-weighted test portfolios and 5\% more rejections with equal-weighted test portfolios than the GLS two-pass regressor in Shanken (1992).

Our paper contributes to the large 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 our OCSR preserves GMM’s efficiency gain. The 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 (2020c).\(^6\) Different from these papers, we focus on the fixed-\(N\) scenario and aim to derive the most efficient estimator when only \(T\) goes to infinity.\(^7\)

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.\(^8\) We also focus on time-invariant betas for linear-beta pricing models. An extension that instruments betas with time-varying firm characteristics seems feasible.\(^9\) We leave these extensions to future research.

Our paper is organized as follows. In Section 2, we lay out our theoretical framework and present our main theorems. Section 3 conducts a comprehensive simulation study that compares

---

\(^5\)This is consistent with both Shanken (1992) and our simulation study.

\(^6\)Also see Bai and Zhou (2015) for a large-\(N\) extension of the usual Fama-MacBeth approach.

\(^7\)In our appendix, we provide an extension of our baseline results by allowing a slowly divergent \(N\), i.e., \(N^{1+\delta}/T \rightarrow 0\) for some \(\delta > 1\).

\(^8\)More specifically, the large-\(N\) and large-\(T\) asymptotics assume \(N \rightarrow \infty\), \(T \rightarrow \infty\), and \(N/T \rightarrow c\) for some \(c > 0\).

alternative two-pass regressors. In Section 4, we revisit the FF versus HXZ debate 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, and additional simulation results.

2 Theory

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 mis-specification of the basic linear-beta pricing model.

The linear-beta asset-pricing model is characterized by the following moment conditions:

\[ \mathbb{E} [f_t - \mu_f] = 0, \]  
\[ \mathbb{E} [(R_{i,t} - (f_t - \mu_f)\beta_i)(f_t - \mu_f)] = 0, \]  
\[ \mathbb{E} [R_{i,t} - \alpha_i - \gamma_0 - \gamma_1' \beta_i - \gamma_2' Z_{i,t}] = 0, \]

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 \( Z_{i,t} \) is a vector of \( M \) many (possibly time-varying) security characteristics.\(^{10}\)

The unknown parameters are \( \mu_f, \alpha_i, \beta_i, \) and \( \gamma \equiv (\gamma_0, \gamma_1', \gamma_2')'. \)\(^{11}\) \( \mu_f \) and \( \beta_i \) are identified by the moment conditions in (2.1) and (2.2), respectively.\(^{12}\) We assume a known subset of assets indexed by \( I_0 \subset \{1, \ldots, N\} \) exists, such that

\[ \alpha_i = 0 \text{ for any } i \in I_0. \]  

Without loss of generality, let \( I_0 \equiv \{1, \ldots, N_0\} \) and \( I_1 \equiv \{N_0 + 1, \ldots, N\} \), where \( N_0 \geq d_\gamma \) and \( d_\gamma \equiv K + M + 1. \)\(^{13}\) Given the identification of \( \mu_f \) and \( \beta_i \), the remaining unknown parameters \( \theta \equiv (\alpha_{N_0+1}, \ldots, \alpha_N, \gamma')' \) are identified under (2.3) and (2.4). In this paper, we study the optimal estimation and inference of \( \theta \) using the method of cross-sectional regression (CSR).

Several remarks on our model are worth mentioning. First, our econometric specification fully captures the risk and return relation in a linear beta-pricing model. In fact, by shutting down

\(^{10}\)Note our model uses the raw return \( R_{i,t} \), in which case, \( \gamma_0 \) 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, \( \gamma_0 \) 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.

\(^{11}\)Throughout this paper, we use \( a \equiv b \) to denote that \( a \) is defined as \( b \).

\(^{12}\)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), and Gospodinov, Kan, and Robotti (2014).

\(^{13}\)The restriction (2.4) is testable under over-identification; that is, \( N_0 > d_\gamma \). In Subsection 2.2, we develop a chi-square test for this restriction.
model mis-specification (i.e., $\alpha_i$) and firm characteristics ($Z_{i,t}$), 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).\footnote{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.} Second, following most studies on two-pass estimators (see 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.\footnote{See, for example, Roll (1977), Shanken (1987), and Kandel and Stambaugh (1987).} 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)).

Having discussed the moment restrictions, we first present the popular two-pass implementation. Let $B \equiv (\beta_1, \ldots, \beta_N)'$ and $Z_t \equiv (Z_{1,t}, \ldots, Z_{N,t})'$, which are $N \times K$ and $N \times M$ matrices, respectively. The unknown betas $B$ can be estimated by their sample analogs:

$$\hat{B} \equiv \left( \sum_{t \leq T} (R_t - \bar{R}) f_t' \right) \left( \sum_{t \leq T} (f_t - \bar{f})(f_t - \bar{f})' \right)^{-1},$$

where $R_t \equiv (R_{1,t}, \ldots, R_{N,t})'$, $\bar{R} \equiv T^{-1} \sum_{t \leq T} R_t$, and $\bar{f} \equiv T^{-1} \sum_{t \leq T} f_t$. Given $\hat{B}$, the unknown parameters $\theta \equiv (\alpha_{N_0+1}, \ldots, \alpha_N, \gamma')'$ can be estimated using the restrictions in (2.3) and (2.4). That is, $\hat{\theta}$ is the minimizer of

$$\min_\theta \left( \sum_{t \leq T} (R_t - \hat{X}_t \theta)' \bar{W} \left( \sum_{t \leq T} (R_t - \hat{X}_t \theta) \right) \right),$$

where $\bar{W}$ is an $N \times N$ real symmetric positive definite matrix, $\hat{X}_t \equiv (S_N, 1_{N \times 1}, \hat{B}, Z_t)$, and $S_N \equiv (0_{N_1 \times N_0}, I_{N_1})'$.\footnote{Throughout the paper, we use $I_d$ to denote the $d \times d$ identity matrix, $1_{d_1 \times d_2}$ and $0_{d_1 \times d_2}$ to denote the $d_1 \times d_2$ matrices of 1’s and 0’s, respectively.} Therefore,

$$\hat{\theta} = (\hat{X}' \hat{W} \hat{X})^{-1} \hat{X}' \hat{W} \bar{R},$$

where $\hat{X} \equiv T^{-1} \sum_{t \leq T} \hat{X}_t$, which is known as the weighted CSR estimator in the literature. Examples of such estimators include the weighted least-squares (WLS) estimator in Litzenberger and Ramaswamy (1979) and the GLS estimator in Shanken (1985) and Shanken (1992).

In this paper, we show that with a properly selected weighting matrix, the optimally weighted CSR (OCSR) estimator is as efficient as the optimal GMM estimator of $\theta$ obtained by a simultaneous optimally weighted GMM estimation of $\mu_f, \beta_i$, and $\theta$ using all restrictions in (2.1)-(2.4).
The unknown parameters $\mu_f$, $B$, and $\theta$ are estimated separately in the weighted CSR procedure, which is computationally attractive because both the first-step estimators of $B$ and $\mu_f$ and the second-step estimator of $\theta$ have closed-form expressions. Therefore, our results show the OCSR estimator enjoys the computational convenience without losing statistical efficiency, and inference on $\theta$ can be conducted in the optimal way using the OCSR estimator.

2.1 Asymptotic Properties of the Weighted CSR Estimator

In this subsection, we study the properties of the weighted CSR estimator when $N$ is fixed. Extension of the main results established in this subsection to the case with divergent $N$ is available in Section C of the Appendix. We first provide the conditions on the pricing factors and security characteristics.

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

Assumption 1(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 1(ii, iii) are the law of large numbers of the sample mean and the sample second moment of the pricing factors, and Assumption 1(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 Assumption 1(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 1(iv) should be considered a reasonable approximation for sorted portfolios (see, e.g., Jagannathan and Wang (2002)).

Because the CSR estimator is constructed based on moment conditions in (2.1)-(2.3), the asymptotic variance of the CSR estimator is determined by the stochastic errors of evaluating these moment conditions by their sample analogs. Specifically, the estimation error of the CSR estimator is governed by the estimation errors of $\beta_i$, which are denoted as $u_{i,t}(f_t - \mu_f)\Sigma^{-1} \gamma_1$ for $i = 1, \ldots, N$ where

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

and the pricing errors, which are denoted as $v_{i,t} \equiv R_{i,t} - \alpha_i - \gamma_0 - \gamma'_1 \beta_i - \gamma'_2 Z_{i,t}$ for $i = 1, \ldots, N$. Because $\mathbb{E}[v_{i,t}] = 0$ by (2.3), we can also write the pricing errors as

$$v_{i,t} = R_{i,t} - \mathbb{E}[R_{i,t}] - \gamma'_2 (Z_{i,t} - \mathbb{E}[Z_{i,t}])$$

(2.7)
for $i = 1, \ldots, N$. Economically, $v_{i,t}$ stands for total return innovations, adjusted for innovations in firm characteristics, whereas $u_{i,t}$ stands for idiosyncratic return innovations.

The estimation error $u_{i,t}(f_t - \mu_f)'\Sigma_f^{-1} \gamma_1$ of $\beta_i$ and the pricing error $v_{i,t}$ affect the asymptotic distribution of the CSR estimator through a linear combination,

$$\epsilon_t \equiv v_t - u_t(f_t - \mu_f)'\Sigma_f^{-1} \gamma_1,$$

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

**Assumption 2.** (i) Let $\Omega \equiv \lim_{T \to \infty} \text{Var}(T^{-1/2} \sum_{t=1}^T \epsilon_t)$, then $T^{-1/2} \sum_{t=1}^T \epsilon_t \rightarrow_d N(0, \Omega)$; (ii) $T^{-1/2} \sum_{t=1}^T u_t(1, f_t') = O_p(1)$; (iii) a non-random matrix symmetric $W$ exists such that $\hat{W} = W + o_p(1)$; and (iv) let $X = (S_N, 1_{N \times 1}, B, \mathbb{E}[Z_t])$, then the eigenvalues of $\Omega$, $\Sigma_f$, $W$, and $X'X$ are bounded from above and away from zero.

Assumption 2(i) is a central limit theorem on the partial sum $T^{-1/2} \sum_{t=1}^T \epsilon_t$, 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 $\beta_i$, we have $E[u_t] = 0_{N \times 1}$ and $E[u_t f_t'] = 0_{N \times K}$. Therefore, Assumption 2(ii) holds if the variance-covariance matrix of $T^{-1/2} \sum_{t=1}^T u_t(1, f_t')$ is bounded. Assumption 2(iii) imposes conditions on the weight matrix of the CSR estimator. The eigenvalue conditions in Assumption 2(iv) ensure the CSR estimator is $T^{-1/2}$-consistent.

**Theorem 1.** Under Assumptions 1 and 2, we have

$$T^{1/2}(\hat{\theta} - \theta) \rightarrow N(0, \Sigma(W)), \quad (2.8)$$

where $\Sigma(W) \equiv (X'WXX)^{-1}(X'W\Omega WX)(X'WX)^{-1}$. Moreover,

$$\Sigma(W) \geq (X'\Omega^{-1}X)^{-1} \quad (2.9)$$

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

The asymptotic normality of the CSR estimator is well known in the literature (see, e.g., Shanken (1992), Jagannathan and Wang (1998), Ahn and Gadarowski (1999), Jagannathan, Skoulakis, and Wang (2010) and Kan and Robotti (2012)). The optimal choice of the weight matrix in the CSR estimation has also been discussed in Kan and Robotti (2012). Theorem 1 generalizes these results to allow for security characteristics and possibly mispriced portfolios.

Theorem 1 implies $\Omega^{-1}$ is the optimal weighting matrix for the weighted CSR estimator. Therefore, the OCSR estimator can be constructed using a consistent estimator of $\Omega^{-1}$ as the weighting matrix. Because $\Omega$ is the variance-covariance matrix of $T^{-1/2} \sum_{t=1}^T \epsilon_t$, the consistent estimator
of Ω can be constructed using consistent estimators of ε_t, the sample analog of Ω when {ε_t}_t is a martingale difference or the heteroskedasticity and auto-correlation consistent (HAC) estimator when {ε_t}_t is from a weakly dependent process (see, e.g., Newey and West (1987) and Andrews (1991)).

Let ˆΩ denote the consistent estimator of Ω. The OCSR estimator is then defined as
\[
\hat{\theta}^\ast \equiv (\hat{X}'\hat{\Omega}^{-1}\hat{X})^{-1}\hat{X}'\hat{\Omega}^{-1}\hat{R}.
\] (2.10)

The consistency of ˆΩ implies Assumption 2(iii). Moreover, by Assumptions 1 and 2(ii, iv), we can show
\[
\hat{X}'\hat{\Omega}^{-1}\hat{X} = X'\Omega^{-1}X + o_p(1),
\] (2.11)
which together with Theorem 1 and the Slutsky Theorem, implies
\[
\left(\hat{X}'\hat{\Omega}^{-1}\hat{X}\right)^{1/2}T^{1/2}(\hat{\theta}^\ast - \theta) \to_d N(0, I_{d_\theta}),
\] (2.12)
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 \(d_\theta \times 1\) nonzero real vector. Then, (2.12) together with (2.11) implies
\[
\frac{T^{1/2}l'(\hat{\theta}^\ast - \theta)}{l'(\hat{X}'\hat{\Omega}^{-1}\hat{X})^{-1}l}^{1/2} \to_d N(0, 1),
\] (2.13)
which can be used to test the hypothesis on the linear combinations of θ. Let \(M\) denote any non-empty subset of \(\{1, \ldots, d_\theta\}\) with size \(m\) and let \(L_M = (l_j)_{j \in M}\) denote the \(d_\theta \times m\) selection matrix whose \(j\)th column is the \(j\)th unit vector; that is, the \(j\)th component of \(l_j\) is 1 and the remaining elements are zero. Then, by (2.12),
\[
(L_M'\hat{X}'\hat{\Omega}^{-1}\hat{X})^{-1/2}L_M'T^{1/2}(\hat{\theta}^\ast - \theta) \to_d N(0, I_m),
\]
which together with the continuous mapping Theorem implies
\[
T(\hat{\theta}^\ast - \theta)'L_M(L_M'\hat{X}'\hat{\Omega}^{-1}\hat{X})^{-1}L_M'(\hat{\theta}^\ast - \theta) \to_d \chi^2(m),
\] (2.14)
where \(\chi^2(m)\) denotes the chi-square distribution with degree of freedom \(m\). The result in (2.14) can be used to conduct joint inference of any subset of θ, such as \((\alpha_{N_0+1}, \ldots, \alpha_N)'\).

### 2.2 Specification Tests

In this subsection, we provide a specification test for the restriction (2.4). By (2.3), this restriction can be written as
\[
H_0 : \mathbb{E}[R_{i,t} - \gamma_0 - \gamma_1'\beta_i - \gamma_2'Z_{i,t}] = 0,
\] (2.15)
for \( i = 1, \ldots, N_0 \). The restrictions in (2.1) and (2.2), which are essentially the definitions of \( \mu_f \) and \( \beta_i \) 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 \( \alpha_i \) for \( i = N_0 + 1, \ldots, N \) are not involved.

Let \( \hat{X}_{0,t} \equiv (1_{N_0 \times 1}, \hat{B}, Z_{0,t}) \), where \( \hat{B} \) and \( Z_{0,t} \) denote the leading \( N_0 \times K \) and \( N_0 \times M \) submatrices of \( \hat{B} \) and \( Z_t \) respectively. The unknown parameter \( \gamma \) is estimated by

\[
\hat{\gamma} = (\hat{X}_0' \hat{\Omega}_0^{-1} \hat{X}_0)^{-1} \hat{X}_0' \hat{\Omega}_0^{-1} \hat{R}_0,
\]

where \( \hat{X}_0 \equiv T^{-1} \sum_{t \leq T} \hat{X}_{0,t}, \hat{R}_0 \equiv T^{-1} \sum_{t \leq T} R_{0,t}, \hat{\Omega}_0 \) denotes the leading \( N_0 \times N_0 \) submatrix of \( \hat{\Omega} \) and \( R_{0,t} \) denotes the leading \( N_0 \) subvector of \( R_t \). The null hypothesis (2.15) is tested using the \( J \)-test statistic, which is defined as

\[
J_T \equiv T(\hat{R}_0 - \hat{X}_0 \hat{\gamma})' \hat{\Omega}_0^{-1}(\hat{R}_0 - \hat{X}_0 \hat{\gamma}).
\]

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 \( \alpha \)

\[
\text{reject } H_0 \text{ if } J_T > \chi^2_{1-\alpha}(N_0 - d_\gamma).
\]

The above test has been proposed in the literature, see, e.g., 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.

**Theorem 2.** Suppose Assumption 1 hold. Then,

(a) Under Assumption 2, \( J_T \to_d \chi^2(N_0 - d_\gamma) \) under \( H_0 \);

(b) If we have: (i) \( \hat{R}_0 = \mathbb{E}[R_{0,t}] + o_p(1) \); (ii) \( \hat{\Omega}_0 = \Omega_1 + o_p(1) \) where \( \Omega_1 \) is a non-random symmetric positive definite matrix; and (iii) the eigenvalues of \( \Omega_1 \) and \( X_0'X_0 \) are bounded from above and away from zero, then

\[
T^{-1}J_T = \mathbb{E}[R_{0,t}'] \Omega_{N_0} \mathbb{E}[R_{0,t}] + o_p(1),
\]

where \( \Omega_{N_0} = \Omega_1^{-1} - \Omega_1^{-1}X_0(X_0' \Omega_1^{-1}X_0)^{-1}X_0' \Omega_1^{-1} \).

Theorem 2(a) shows that the test in (2.18) controls size. Theorem 2(b) derives the probability limit of the (scaled) test statistic \( J_T \) under both the null and the alternative hypotheses. Under \( H_0 \), \( \mathbb{E}[R_{0,t}] = X_0 \gamma \), and hence (2.19) implies \( T^{-1}J_T = o_p(1) \), which is consistent with the weak convergence of \( J_T \) derived in Theorem 2(a). Under the alternative hypothesis, \( \mathbb{E}[R_{0,t}] \) cannot be represented by any linear combination of \( X_0 \), and hence, we have \( \mathbb{E}[R_{0,t}'] \Omega_{N_0} \mathbb{E}[R_{0,t}] > 0 \) in general. Because \( \chi^2_{1-\alpha}(N_0 - d_\gamma) \) is a finite number, Theorem 2(b) shows the test in (2.18) is consistent as long as \( \mathbb{E}[R_{0,t}'] \Omega_{N_0} \mathbb{E}[R_{0,t}] \) is bounded away from zero.
2.3 Efficiency of the OCSR Estimator

In this subsection, we show the asymptotic optimality of \( \hat{\theta}^* \) goes far beyond the class of the CSR estimators. Specifically, we show that when the data are i.i.d., the OCSR estimator \( \hat{\theta}^* \) attains the semiparametric efficiency bound implied by (2.1)-(2.4). Therefore, no regular estimator can have smaller asymptotic variances than \( \hat{\theta}^* \). This generalizes Shanken (1992)'s parametric efficiency result (of the GLS estimator) to the semiparametric efficiency for the OCSR estimator. Moreover, without the i.i.d. assumption, we show the OCSR estimator \( \hat{\theta}^* \) is asymptotically equivalent to the optimal GMM estimator of \( \theta \). The latter is estimated simultaneously with \( \mu_f \) and \( \beta_1 \) using all the restrictions in (2.1)-(2.4) through an optimally weighted GMM. Therefore, the efficiency results established here substantially generalize the results in Shanken (1992), Jagannathan and Wang (1998), and Jagannathan and Wang (2002).

By Theorem 1, the asymptotic variance of \( \hat{\theta}^* \) with i.i.d. data is \( (X' \Omega^{-1} X)^{-1} \), where \( \Omega \equiv E[\epsilon_t \epsilon_t'] \).

To show the semiparametric efficiency of the OCSR estimator \( \hat{\theta}^* \), it is sufficient to show the semiparametric efficiency bound of \( \theta \) implied by (2.1)-(2.4) is \( X' \Omega^{-1} X \).

**Lemma 1.** Suppose the data \( \{(f_t, (R_{i,t})_{i \leq N}, (Z_{i,t})_{i \leq N})_t \} \) are i.i.d. across \( t \). Then, the semiparametric efficiency bound of \( \theta \) implied by (2.1)-(2.4) is \( X' \Omega^{-1} X \).

Lemma 1 together with Theorem 1 shows the OCSR estimator \( \hat{\theta}^* \) is semiparametrically efficient with i.i.d. data. However, financial market returns often display dependent conditional heteroskedasticity (e.g., Bollerslev, Engle, and Wooldridge (1988), Schwert and Seguin (1990)), which makes the use of this result limited. We next relax the i.i.d. assumption and show the OCSR estimator \( \hat{\theta}^* \) is as efficient as the optimal GMM estimator of \( \theta \) obtained through a simultaneous optimal GMM estimation of \( \phi \equiv (\theta', \mu'_f, \beta'_1, ..., \beta'_N)' \) using (2.1)-(2.4).

**Theorem 3.** Let \( \hat{\phi}^{\text{gmm}} \equiv (\hat{\theta}^{\text{gmm}}, \hat{\mu}^{\text{gmm}}, \hat{\beta}^{\text{gmm}}_1, ..., \hat{\beta}^{\text{gmm}}_N)' \) denote the optimal GMM estimator of \( \phi \) based on (2.1)-(2.4). Then, the asymptotic variance-covariance matrix of \( \hat{\theta}^{\text{gmm}} \) is \( (X' \Omega^{-1} X)^{-1} \).

Theorem 3 establishes the asymptotic equivalence between the OCSR estimator \( \hat{\theta}^* \) and the optimal GMM estimator \( \hat{\theta}^{\text{gmm}} \) in a general framework that represents the linear beta-pricing model. In essence, it shows the usual efficiency gain of the GMM estimator is effectively preserved by using the optimal weighting matrix in a two-pass regression.

The rest of this subsection is devoted to discussion of the connection between the OCSR estimator and the GLS estimator proposed in Shanken (1992). For ease of discussion, we assume...
the pricing factors are not traded, so our GMM moment restrictions exactly correspond to Shanken (1992)’s model.

In the usual GLS, estimation error in beta and the conditional heteroskedasticity of the idiosyncratic innovation $u_t$ (given the pricing factors) are not directly taken into account. As a result, the optimal weighting matrix in the usual GLS amounts to the covariance matrix of the idiosyncratic innovation $u_t$. As we discuss below, the efficiency of the GLS is justified under the following assumptions from Shanken (1992).

**Assumption 3.** $u_t$ is i.i.d. with $\mathbb{E}[u_t|F_T] = 0$ and $\mathbb{E}[u_t u'_t|F_T] = \Sigma_u$, where $F_T = (f_1, \ldots, f_T)$ and $\Sigma_u$ is a non-random positive definite matrix.

Under Assumption 3, $u_t(1 - (f_t - \mu_f)\Sigma_f^{-1}\gamma_1)$ forms a martingale difference array that is uncorrelated with $f_{t'} - \mu_f$ for any $t$ and $t'$. Therefore, the variance-covariance matrix $\Omega$ takes the following simplified form,

$$\Omega = \Sigma_u(1 + \gamma_1'\Sigma_f\gamma_1) + B\Sigma_f B' = \Sigma_u(1 + \gamma_1'\Sigma_f\gamma_1) + X\Sigma_f X',$$

where $\Sigma_f \equiv \text{diag}(0, \Sigma_f)$ and $\Sigma_f$ denotes the long-run variance of $T^{-1/2} \sum_{t=1}^T f_t$.\(^{17}\) Moreover, by Theorem 1, the asymptotic variance-covariance matrix of the weighted CSR estimator is

$$\Sigma(W) = (1 + \gamma_1'\Sigma_f\gamma_1)(X'\Omega^{-1}X)^{-1}(X'\Sigma_u W X)(X'W X)^{-1} + \Sigma_f^*.$$ \hspace{1cm} (2.21)

By the similar arguments for deriving (2.9), we can show that under Assumption 3,

$$\Sigma(\Sigma_u^{-1}) \leq \Sigma(W)$$ \hspace{1cm} (2.22)

for any $N \times N$ positive definite matrix $W$, where

$$\Sigma(\Sigma_u^{-1}) = (1 + \gamma_1'\Sigma_f\gamma_1)(X'\Sigma_u^{-1}X)^{-1} + \Sigma_f^*$$ \hspace{1cm} (2.23)

is the asymptotic variance-covariance matrix of the GLS estimator derived in Shanken (1992). Shanken (1992) further shows that if the joint density of $R_t$ given $F_T$ is normal, the GLS estimator is asymptotically equivalent to the maximum likelihood estimator, and hence is efficient.

Because the inequality (2.9) holds without Assumption 3, we have $(X'\Omega^{-1}X)^{-1} \leq \Sigma(\Sigma_u^{-1})$ in general, which shows the OCSR estimator is more efficient than the GLS estimator in general. However, when Assumption 3 holds, in combining (2.9) and (2.22), we obtain $(X'\Omega^{-1}X)^{-1} = \Sigma(\Sigma_u^{-1})$, which together with the parametric efficiency of the GLS estimator implies the OCSR estimator is also parametrically efficient if the joint density of $R_t$ given $F_T$ is normal.

\(^{17}\)Throughout this paper, we use $\text{diag}(A, B)$ to denote the block diagonal matrix with the square matrices $A$ and $B$ on the main diagonal.
2.4 Interpretation of the Efficiency of OCSR

In this section, we provide economic interpretation for the optimal weight in the OCSR estimator. For simplicity, assume betas are the only explanatory variables that affect returns.

Our proof for Theorem 1 permits an intuitive decomposition of the pricing error for the cross-sectional regression; that is,

\[
R - \hat{X} = v - (\hat{B} - B)\gamma_1,
\]

(2.24)

\[
= T^{-1} \sum_{t=1}^{T} \left( v_t - u_t (f_t - \mu_f)'\Sigma_f^{-1}\gamma_1 \right) + o_p(T^{-1/2}),
\]

(2.25)

\[
= T^{-1} \sum_{t=1}^{T} \left( u_t (1 - (f_t - \mu_f)'\Sigma_f^{-1}\gamma_1) + B(f_t - \mu_f) \right) + o_p(T^{-1/2}).
\]

(2.26)

Eq. (2.24) shows the pricing error can be decomposed into two parts: one part is related to \(v_t\), which captures the total innovation in returns, and the other part is given by \((\hat{B} - B)\gamma_1\), which captures estimation error in beta. One can further decompose the total innovation \(v_t\) into the idiosyncratic innovation \(u_t\) and the systematic innovation \(B(f_t - \mu_f)\), as given by eq. (2.26).

We first interpret our results under Assumption 3. We interpret the two terms in eq. (2.26) as “efficiency-relevant innovations” and “efficiency-irrelevant innovations” (under Assumption 3), for reasons as follows. For idiosyncratic innovations \(u_t\), they contribute to both first-stage beta-estimation uncertainty and second-stage cross-sectional regression uncertainty. \(u_t\) alone represents second-stage uncertainty, whereas \(-u_t (f_t - \mu_f)'\Sigma_f^{-1}\gamma_1\) 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 weighting matrix, and optimal efficiency is achieved at the variance-covariance matrix of \(u_t\) (hence, the name efficiency-relevant innovations).

For systematic innovations \(B(f_t - \mu_f)\), because of the economic restrictions of the linear-beta pricing model (in particular, \(B\gamma_1\) is linear in \(B\)), 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 produce no estimation error. 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., \(B\gamma_1\)) scales up and down in the same way as innovations (i.e., \(B(f_t - \mu_f)\)). Note, as a counterexample, the above argument breaks down if the asset-pricing model dictates a nonlinear relation between expected return and beta (e.g., \(B^{1/2}\gamma_1\),

15
where $B^{1/2}$ represents taking the square root of each element in $B$).

Combining the above two observations, because Assumption 3 allows one to decouple efficiency-relevant innovations and efficiency-irrelevant innovations (more specifically, the two innovations are uncorrelated), efficiency-irrelevant innovations (i.e., $B(f_t - \mu_f)$) can be isolated and shown not to affect the CSR estimator. Estimation efficiency only depends on efficiency-relevant innovations (i.e., $u_t(1 - (f_t - \mu_f)\Sigma_f^{-1} \gamma_1)$) and is achieved when the weighting matrix is set at $\Sigma_u^{-1}$. This explains the efficiency result in Shanken (1992).

In comparison with the two-pass GLS estimator in Shanken (1992), our OCSR can offer additional efficiency gain when our assumptions (in particular, Assumption 2) generalize Assumption 3. Several prominent features of financial market data may make our generalizations important. First, the assumption of conditional homoskedasticity in Assumption 3 may not be suitable when asset returns display non-normality and conditional heteroskedasticity (Jagannathan and Wang (1998), Jagannathan and Wang (2002)). 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 $u_t$ and higher powers of $f_t$. Note that by definition, $u_t$ is orthogonal to $f_t$ itself. But this orthogonality does not rule out potential correlations between $u_t$ 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 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 made uncorrelated, leading to a potential efficiency gain of our estimator compared with Shanken (1992) for the inference of the market risk premium.

Lastly, our estimator adjusts for serial correlations in $u_t$, $f_t$, and, more importantly, their combinations. Whereas Shanken (1992)’s GLS can be extended to cope with serial correlations in $u_t$, we show 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.

GMM is able to achieve semiparametric efficiency that guards against all of the above issues.

---

16 Nonnormality and conditional heteroskedasticity for asset returns are documented in, for example, Barone-Adesi and Talwar (1983), Bollerslev, Engle, and Wooldridge (1988), 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, Skoulikis, and Wang (2010) on the economic restriction that the assumption of conditional homoskedasticity imposes relative to the less restrictive assumptions in a GMM framework.

19 Another data feature that our OCSR guards against is the non-linear dependence between $u_t$ and $f_t$. 

18
Our new insight is to show a CSR estimator with a properly defined weighting 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 de-coupling of efficiency-relevant and efficiency-irrelevant innovations in constructing the optimal weighting matrix, we show the coupling of these two sources of innovations is essential to achieve GMM's efficiency when Assumption 3 fails to hold.

2.5 Further Discussion

2.5.1 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 result? Intuitively, one can think of an iterative GMM interpretation of the system of moment restrictions where \( \hat{\mu}_f \) and \( \hat{B} \) are estimated from moment conditions (2.1) and (2.2) first, and then fed into moment conditions (2.3) to obtain the estimate for \( \gamma \). Hence, the fact that \( \gamma \) 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 \( \gamma \) 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) does not contain separate identification information for the model parameters \( \mu_f \) and \( B \) 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 \( \mu_f \) and \( B \) as “nuisance” parameters. The efficiency gain of estimating \( \gamma \) jointly with the nuisance parameters from all three sets of moment conditions relative to estimating it through (2.2) with plug-in estimator of \( \mu_f \) and \( B \) from the just identified system (2.1) and (2.2) is vanishing asymptotically, which leads to the efficiency result of our OCSR. In Section B of the appendix, we provide a counter example which illustrates the efficiency loss of the (optimally weighted) iterative GMM estimator with an over-identified nuisance parameter.\(^{20}\)

\(^{20}\)Kan and Zhou (2001) consider two sets of parameters \( \phi_1 \) and \( \phi_2 \) with two sets of moment conditions, say, \( S_1 \) and \( S_2 \), where \( \phi_1 \) is separately identified by \( S_1 \) while \( \phi_2 \) is identified by \( S_2 \) given \( \phi_1 \). They show that if \( \phi_2 \) is just identified by \( S_2 \) given \( \phi_1 \), then the optimal GMM estimator of \( \phi_1 \) using \( S_1 \) is as efficient as the joint GMM estimator of \( \phi_1 \) using both \( S_1 \) and \( S_2 \). Although this result shares similar spirits 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 \( \mu_f \) and \( B \) can be efficiently estimated using only moment conditions in (2.1) and (2.2), as long as the parameter of interest \( \gamma \) is just identified by (2.3) given the nuisance parameters. However, our focus is on \( \gamma \), which is over-identified by (2.3). Our efficiency result applies to this parameter instead of the nuisance parameters.
The above insight allows us to easily generalize the implication of our results to other setups. For example, suppose we augment the three sets of moment restrictions with another set (called $S_1$) 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 $\mu_f$ and $B$ through (2.1), (2.2) and ($S_1$) first and then estimate $\gamma$ 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: $h(\gamma, B)$. Then our equivalence result still holds. This can be used to justify, e.g., the specification in Fama and MacBeth (1973) where risk premium is linear in squared market betas.

To place our result into the context of the literature, we relate it to two lines of research. For the first line, 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. 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.

For the second line, Shanken (1992) establishes the efficiency of the two-pass GLS estimator under his assumptions that idiosyncratic risks are i.i.d. and conditional homoscedastic given the factors. 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 GMM estimator, which is known to achieve the semiparametric efficiency bound. Our paper fills in this gap in the literature.

### 2.5.2 Implementation of OCSR

Similar to the GLS estimator, the implementation of our OCSR requires the estimation of a $N \times N$ weighting matrix, which becomes inaccurate when $N/T$ is large. This begs the question of what level of $N$ is considered suitable to implement our estimator versus alternative estimators that are potentially less demanding in terms of the estimation of the weighting matrix, e.g., the WLS.

While our theory is developed under the assumption of a fixed (or slowly divergent) $N$, the question of which level of $N$ is considered large is more empirical than theoretical. We answer

---

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

22 See Section C of the Appendix for extension of the theory to the case with divergent $N$. 

---
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 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 down weight the impact of these assets, leading to a more efficient estimation.

In our simulation study, while we do not experiment with very large \( N \)'s, theory suggests that the efficiency gain of OCSR (and GLS) should disappear at some level of \( N \).\(^{23}\) WLS and OLS may perform better instead. But in these cases, even some ad hoc shrinkage type of estimator of the error covariance matrix may go a long way in improving the efficiency of the approach. Given this is out of 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)), several features mark our simulation design. 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 that fall out of favor in empirical research. Second, to ensure our simulated 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 model as candidate models, and the 18 low-turnover anomaly sample and the 38 low-turnover and medium-turnover combined anomaly sample in Novy-Marx and Velikov (2016) as candidate samples for test assets.\(^{24}\) Our actual sample runs from July 1973 to December 2017, including 534 monthly observations.\(^{25}\) Third, our bootstrap samples keep the same level of time-series and cross-sectional dependency in the actual data, allowing our results to shed light on the performance of our approach for the actual data.

\(^{23}\)We thank instructive comments by Wayne Ferson about the impact of large \( N \) on the relative efficiency of our approach to WLS and OLS.

\(^{24}\)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.

\(^{25}\)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.
Note 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 low level of time-series dependence, our simulation study may not be the best case to 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).26

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 \( 3 \times 1 \) 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 \( 3 \times 1 \) factor realizations, and \( \hat{\varepsilon}_{p,i,t} \) is the regression residual (here, "p" stands for population). We collect the loading vectors into \( B_p \equiv (\hat{\beta}_{p,1}, \ldots, \hat{\beta}_{p,N})' \), 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 \( m (m = 1, \ldots, M, \text{ where } M = 10,000) \), 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

\[
R_m \equiv \gamma_{p,0}1_{N \times T_b} + B_p \times (F_m - (\bar{f}_p - \gamma_p)1_{1 \times T_b}) + RES_m,
\]

where \( R_m \) is a \( N \times T_b \) matrix and \( \bar{f}_p \) denotes the (original) sample mean of \( f_t \).27

---

26 Note that existing two-pass estimators such as Shanken (1992) do not take correlations in factor model residuals into account.

27 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.
To investigate the finite-sample properties of various CSR estimators, we let $\gamma_{p,0} = 0$ and $\gamma_p = \bar{f}_p$ to ensure the linear-beta pricing holds. This parametrization is slightly revised when we examine the performance of inference (i.e., hypothesis tests) based on different CSR estimators. Specifically, we reparametrize the 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 $\gamma_p = \bar{f}_p$. 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 $\gamma_{p,mkt} = a\bar{f}_{p,mkt}$, where $\bar{f}_{p,mkt}$ 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 $\gamma_p$, 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 nonzero value.

For the in-sample data $\{R_m, F_m\}$ 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 $E[u_tu_t']$, 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 as given in Assumption 2. We follow a simple approach. $v_t$ is obtained by demeaning returns. $u_t$ is obtained from asset-by-asset time-series OLS. $\mu_f$ and $\Sigma_f$ are the estimated factor mean and covariance matrix. $\gamma_1$ is taken to be the cross-sectional OLS estimate. Lastly, 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)).

For a given parameter $\gamma_{p,j}$ in $\gamma_p$ ($j = 0, mkt, smb$ 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

---

28 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:

\[
\text{Bias} = M^{-1} \sum_{m=1}^{M} \hat{\gamma}_{j,m} - \gamma_{p,j},
\]
\[
\text{RMSE} = \sqrt{M^{-1} \sum_{m=1}^{M} (\hat{\gamma}_{j,m} - \gamma_{p,j})^2},
\]
\[
\text{MAE} = M^{-1} \sum_{m=1}^{M} |\hat{\gamma}_{j,m} - \gamma_{p,j}|,
\]

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 D.1 in Appendix D for \(T = 750\). Note 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 \(\gamma_0\), 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.\(^{29}\) 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 1% for \(\gamma_{mkt}\) (i.e., \(-1\% = (0.225 - 0.228)/0.228\)) to 12% for \(\gamma_{smb}\) (i.e., \(-12\% = (0.128 - 0.146)/0.146\)). The average reduction is 7%.

Turning to Table D.1 (\(T = 750\)) in Appendix D, \(T = 750\) 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, for the modest (\(T = 500\)) to large \(T\) (\(T = 750\)) cases that we examine, OCSR seems to present

\(^{29}\)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).
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. $\gamma_0$, $\gamma_{\text{mkt}}$, $\gamma_{\text{smb}}$, $\gamma_{\text{hml}}$, $\gamma_{\text{cma}}$, and $\gamma_{\text{rmw}}$ denote the risk premiums associated with the intercept, the market factor, $\text{smb}$ (size factor), $\text{hml}$ (value factor), $\text{cma}$ (investment factor), and $\text{rmw}$ (profitability factor), respectively. Bold denotes the best performer among all methods considered.

<table>
<thead>
<tr>
<th></th>
<th>Panel A: FF 3-Factor Model, $N = 18$</th>
<th>Panel C: FF 5-Factor Model, $N = 18$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_0$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\gamma_{\text{mkt}}$</td>
<td>-0.097</td>
<td>0.005</td>
</tr>
<tr>
<td>$\gamma_{\text{smb}}$</td>
<td>0.319</td>
<td>0.278</td>
</tr>
<tr>
<td>$\gamma_{\text{hml}}$</td>
<td>0.261</td>
<td>0.220</td>
</tr>
<tr>
<td>Bias</td>
<td>OLS</td>
<td>OCSR</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.200</td>
<td>0.220</td>
</tr>
<tr>
<td>MAE</td>
<td>0.159</td>
<td>0.175</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\gamma_{\text{mkt}}$</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>$\gamma_{\text{smb}}$</td>
<td>0.339</td>
<td>0.277</td>
</tr>
<tr>
<td>$\gamma_{\text{hml}}$</td>
<td>0.261</td>
<td>0.220</td>
</tr>
<tr>
<td>Bias</td>
<td>GLS</td>
<td>GLS</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.241</td>
<td>0.232</td>
</tr>
<tr>
<td>MAE</td>
<td>0.194</td>
<td>0.165</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\gamma_{\text{mkt}}$</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>$\gamma_{\text{smb}}$</td>
<td>0.339</td>
<td>0.277</td>
</tr>
<tr>
<td>$\gamma_{\text{hml}}$</td>
<td>0.261</td>
<td>0.220</td>
</tr>
<tr>
<td>Bias</td>
<td>WLS</td>
<td>WLS</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.183</td>
<td>0.172</td>
</tr>
<tr>
<td>MAE</td>
<td>0.140</td>
<td>0.174</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\gamma_{\text{mkt}}$</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>$\gamma_{\text{smb}}$</td>
<td>0.339</td>
<td>0.277</td>
</tr>
<tr>
<td>$\gamma_{\text{hml}}$</td>
<td>0.261</td>
<td>0.220</td>
</tr>
<tr>
<td>Bias</td>
<td>OCSR</td>
<td>OCSR</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.225</td>
<td>0.199</td>
</tr>
<tr>
<td>MAE</td>
<td>0.179</td>
<td>0.174</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\gamma_{\text{mkt}}$</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>$\gamma_{\text{smb}}$</td>
<td>0.339</td>
<td>0.277</td>
</tr>
<tr>
<td>$\gamma_{\text{hml}}$</td>
<td>0.261</td>
<td>0.220</td>
</tr>
<tr>
<td>Bias</td>
<td>GLS</td>
<td>GLS</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.228</td>
<td>0.226</td>
</tr>
<tr>
<td>MAE</td>
<td>0.183</td>
<td>0.187</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\gamma_{\text{mkt}}$</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>$\gamma_{\text{smb}}$</td>
<td>0.339</td>
<td>0.277</td>
</tr>
<tr>
<td>$\gamma_{\text{hml}}$</td>
<td>0.261</td>
<td>0.220</td>
</tr>
<tr>
<td>Bias</td>
<td>WLS</td>
<td>WLS</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.248</td>
<td>0.253</td>
</tr>
<tr>
<td>MAE</td>
<td>0.198</td>
<td>0.205</td>
</tr>
</tbody>
</table>

23
Next, we study the size and power of the hypothesis test on $H_0 : \gamma_{p,j} = 0$, where $j = 0, mkt, smb, \text{or } hml$ using different CSR estimators. To perform hypothesis testing, we need to be specific about how standard errors are calculated. For GLS and WLS, we follow Theorem 1 (which is also given in Jagannathan and Wang (1998)) and use the estimated $\Sigma(W)$ to approximate the asymptotic covariance matrix, where $W$ is set to be the aforementioned GLS and WLS weight matrix, respectively, and $\Omega$ follows the long-run variance estimator $\hat{\Omega}$ as explained previously. For OCSR, asymptotic covariance matrix is given by $(X'\Omega^{-1}X)^{-1}$, which is estimated by $(\hat{X}'\hat{\Omega}^{-1}\hat{X})^{-1}$. For OLS, we distinguish between three types of OLS standard-error estimates. One is the “naive” approach that does not take beta estimation into account and simply uses OLS standard errors from the cross-sectional regression. We denote this approach as “$\text{OLS}^{1\text{stage}}$” to emphasize it is a one-step estimator. The second one is similar to how we calculate standard errors for GLS and WLS. In particular, we apply Theorem 1 and set $W$ as the identify matrix. The last one 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 $\text{OLS}^{1\text{stage}}$ 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 D (i.e., Table D.2-D.5). 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 pre-specified significance level is found and used to calculate the corresponding test power.

In Tables D.2 and D.3, $\text{OLS}^{1\text{stage}}$ 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 $\text{OLS}^{1\text{stage}}$. 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 of 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 a substantially lower power than OCSR and GLS.

---

30To save space, we focus on the Fama-French three-factor model. Results based on the Fama-French five-factor model are available upon request.

31See Harvey and Liu (2020a) for alternative ways to weight Type I and Type II errors.
For $T = 750$ (Table D.4-D.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.

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.\(^{32}\) 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.\(^{33}\) However, factors may be measured with errors (see, e.g., Roll (1977), Shanken (1987), 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

\(^{32}\)Alternatively, Barillas and Shanken (2017) and Fama and French (2018) use spanning regressions to test nested models (i.e., the RHS approach).

\(^{33}\)More specifically, for a single asset $i$, imposing the condition that risk premiums equal factor means, the following set of moment equations,

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

exactly identify the unknown parameters $\mu_f$, $\beta_i$, and $\alpha_i$, where $r_{i,t}$ denotes the excess return of asset $i$. As such, the GMM (or CSR) estimator of $\alpha_i$ equals the usual estimate from time-series regression.
inference on factor models? Given that factor models purport to explain the returns of a large number of assets, the assumption that they at least should 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, seems natural. 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 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.\footnote{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.} We term these portfolios basis assets for a given candidate factor model.

Note that one can use our over-identification tests described in Section 2.2 to directly test the above assumption on basis assets. That is exactly what we did 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 between 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 $1_N$ to capture the potential difference between the zero-beta rate and risk-free rate (because we use excess instead of gross returns for assets) and the matrix of factor loadings $\hat{B}$. We introduce another variable to capture potential model mis-specification (i.e., abnormal alpha) by asset $N$, given by $S_N = (0_{1 \times N_0}, 1)'$. The stacked matrix of regressors is given by $(S_N, 1_N, \hat{B})$. 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

$$\hat{T}A = l' \times \hat{\theta},$$

where $l = (1, 1, 0_{1 \times (d_\theta-2)})'$ (i.e., non-zero for only the first two elements), and $\hat{T}A$ stands for “total
alpha,” which is to distinguish from our alpha definition in the theory section. Although we may 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 (2.13) can be straightforwardly applied to make inference on \( \hat{TA} \).

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), whereas 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

---

35Note whether we define the estimate of alpha as \( l' \times \hat{\theta} \) or simply \( \hat{\alpha} \), \( N \) will not have a material impact on our analysis of the FF model, because the mean absolute estimate for \( \gamma_0 \) (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 \( \gamma_0 \) is not the main contributor to our results on hypothesis results reported later in Table 4.
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.

test in (2.18)). 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.
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 (2.18) for the FF model and HXZ model using their respective basis assets.

<table>
<thead>
<tr>
<th></th>
<th>FF</th>
<th>HXZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute Alpha</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.047</td>
<td>0.177</td>
</tr>
<tr>
<td>Stdev.</td>
<td>0.036</td>
<td>0.109</td>
</tr>
<tr>
<td>Max</td>
<td>0.130</td>
<td>0.395</td>
</tr>
<tr>
<td>Absolute Alpha t-stat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.107</td>
<td>2.470</td>
</tr>
<tr>
<td>Stdev.</td>
<td>1.325</td>
<td>2.730</td>
</tr>
<tr>
<td>Max</td>
<td>2.654</td>
<td>6.209</td>
</tr>
<tr>
<td>#(&gt; 2.0)</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>#(&gt; 3.0)</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Specification Tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J-stat</td>
<td>20.310</td>
<td>105.176</td>
</tr>
<tr>
<td>Degree of freedom</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>[p-value]</td>
<td>[0.121]</td>
<td>[0.000]</td>
</tr>
</tbody>
</table>

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 (2.17) does not reject 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 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 weighed in on the performance comparison of models.

Table 3: Factor-Premium Estimates: Time Series vs. 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 5-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: \( OLS^{1\text{stage}} \) 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). \( \gamma_{mkt}, \gamma_{smb}, \gamma_{hml}, \gamma_{cma}, \) and \( \gamma_{rmw} \) 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. \( \gamma_{mkt}^{HXZ}, \gamma_{smb}^{HXZ}, \gamma_{ia}, \) and \( \gamma_{roe} \) 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.

<table>
<thead>
<tr>
<th>Method</th>
<th>FF</th>
<th>HXZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>( OLS(TS) )</td>
<td>Mean</td>
<td>( \gamma_{mkt} )</td>
</tr>
<tr>
<td></td>
<td>Stdev.</td>
<td>0.185</td>
</tr>
<tr>
<td>( OLS^{1\text{stage}} )</td>
<td>Mean of Estimates</td>
<td>0.865</td>
</tr>
<tr>
<td></td>
<td>Mean of Stdev.</td>
<td>0.445</td>
</tr>
<tr>
<td>( FM )</td>
<td>Mean of Estimates</td>
<td>0.865</td>
</tr>
<tr>
<td></td>
<td>Mean of Stdev.</td>
<td>0.391</td>
</tr>
<tr>
<td>( OLS )</td>
<td>Mean of Estimates</td>
<td>0.865</td>
</tr>
<tr>
<td></td>
<td>Mean of Stdev.</td>
<td>0.302</td>
</tr>
<tr>
<td>( OCSR )</td>
<td>Mean of Estimates</td>
<td>0.794</td>
</tr>
<tr>
<td></td>
<td>Mean of Stdev.</td>
<td>0.383</td>
</tr>
<tr>
<td>( GLS )</td>
<td>Mean of Estimates</td>
<td>0.596</td>
</tr>
<tr>
<td></td>
<td>Mean of Stdev.</td>
<td>0.411</td>
</tr>
<tr>
<td>( WLS )</td>
<td>Mean of Estimates</td>
<td>0.695</td>
</tr>
<tr>
<td></td>
<td>Mean of Stdev.</td>
<td>0.451</td>
</tr>
</tbody>
</table>

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.\textsuperscript{36}

Focusing on FF in Table 3, we see 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 market betas do not help explain the returns of HXZ’s basis assets. In fact, our results in Table 2 show the returns for many basis assets in HXZ cannot be explained by any of HXZ’s factors.

Lastly, 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., “\(OLS^{1st\text{stage}}\)”), which ignores beta-estimation uncertainty, leads to a much larger number of rejections. This

\footnotesize{\textsuperscript{36}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       | Abs. Alpha | Mean | Stdev. | Abs. t-stat | Mean | Stdev. | #(|t-stat| > 2) | HXZ       | Abs. Alpha | Mean | Stdev. | Abs. t-stat | Mean | Stdev. | #(|t-stat| > 2) |
|-------------------|----------|------------|------|--------|-------------|------|--------|-----------|-----------|------------|------|--------|-------------|------|--------|-----------|
| **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  | 1.513     | 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         |      |        |             |      |        |           |

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 over the unadjusted Fama-MacBeth approach as argued in Shanken (1992). The degree of over-rejection for the naive OLS is so large that 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).

Third, among other cross-sectional methods, 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 again 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 (86) of rejections than FF (82) even with the time-series OLS. However, given the cross-sectional model for HXZ is likely mis-specified, 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 been 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., $\hat{\alpha}^T$ vs. $\hat{\alpha}_N$), whereas the time-series OLS has to lump them into the intercept. As another example, firm characteristics can be straightforwardly incorporated into OCSR to allow potential model mis-specification and enrich testable hypotheses (see, e.g., our general model specification in Section 2 as well as Jagannathan and Wang (1998)), 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, the two-pass regression can be constructed to achieve the same semiparametric efficiency bound 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 non-linear and high-dimensional GMM estimator 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


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


A Proofs

Proof of Theorem 1. Because $E[R_t] = X\theta$ by (2.3), we can write

$$R - \hat{X}\theta = R - E[R_t] + E[R_t] - \hat{X}\theta = R - E[R_t] - (\hat{X} - X)\theta = R - E[R_t] - (\hat{B} - B)\gamma_1 - (\tilde{Z} - E[Z_t])\gamma_2 = v - (\hat{B} - B)\gamma_1,$$  \quad (A.1)

where $v \equiv T^{-1}\sum_{t=1}^{T} v_t$. Therefore, we can write

$$T^{1/2}(\hat{\theta} - \theta) = (\hat{X}'\hat{W}\hat{X})^{-1}\hat{X}'\hat{W} \left[ T^{1/2}(v - (\hat{B} - B)\gamma_1) \right].$$  \quad (A.2)

Because $R_t = X\theta + B(f_t - \mu_f) + u_t$,

$$\hat{B} - B = \left( T^{-1}\sum_{t \leq T} (R_t - R)(f_t - \bar{f})' \right) \hat{\Sigma}_f^{-1} - B = T^{-1}\sum_{t \leq T} u_t(f_t - \mu_f)'\hat{\Sigma}_f^{-1} - u_T T^{-1/2} \sum_{t=1}^{T} (f_t - \mu_f)'\hat{\Sigma}_f^{-1},$$  \quad (A.3)

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

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

which together with Assumptions 1(ii) and 2(ii, iv) implies

$$\hat{B} - B = T^{-1}\sum_{t \leq T} u_t(f_t - \mu_f)'\Sigma_f^{-1} + o_p(T^{-1/2}) = O_p(T^{-1/2}).$$  \quad (A.5)

Therefore, by Assumptions 2(i) and (A.5), we have

$$v - (\hat{B} - B)\gamma_1 = T^{-1}\sum_{t \leq T} \epsilon_t + o_p(T^{-1/2}) = O_p(T^{-1/2}).$$  \quad (A.6)

Similarly, by Assumptions 1(iv) and (A.5),

$$\hat{X} = X + o_p(1).$$  \quad (A.7)

From Assumptions 2(iii, iv) and (A.7), we obtain

$$\hat{X}'\hat{W}\hat{X} = X'WX + o_p(1) \text{ and } \hat{X}'\hat{W} = X'W + o_p(1).$$  \quad (A.8)

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

$$T^{1/2}(\hat{\theta} - \theta) = (X'WX)^{-1}X'WT^{-1/2} \sum_{t \leq T} \epsilon_t + o_p(1) \rightarrow_d N(0, \Sigma(W)).$$  \quad (A.9)
which shows the first claim of the theorem.

Let \( P_N \equiv I_N - \Omega^{1/2}W\mathbf{X}(\mathbf{X}'\Omega W\mathbf{X})^{-1}\mathbf{X}'\Omega^{1/2} \). Then, we can write

\[
\mathbf{X}'\Omega^{-1}\mathbf{X} - (\Sigma(W))^{-1} = \mathbf{X}'\Omega^{-1/2}P_N\Omega^{-1/2}\mathbf{X}.
\]

Because \( P_N \) is an idempotent matrix, \( \mathbf{X}'\Omega^{-1/2}P_N\Omega^{-1/2}\mathbf{X} \) is positive semi-definite, which shows \( \mathbf{X}'\Omega^{-1}\mathbf{X} - (\Sigma(W))^{-1} \) is positive semi-definite. Because \( \mathbf{X}'\Omega^{-1}\mathbf{X} \) and \( \Sigma(W) \) are positive definite matrices, we can further deduce that \( \Sigma(W) - (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \) is positive semi-definite, which shows the second claim of the theorem. \( Q.E.D. \)

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

\[
\mathbf{R}_0 - \hat{\mathbf{X}}_0 \gamma = \mathbf{v}_0 - (\hat{\mathbf{B}}_0 - \mathbf{B}_0)\gamma_1 = T^{-1} \sum_{t \leq T} \epsilon_{0,t} + o_p(T^{-1/2}) = O_p(T^{-1/2}), \quad (A.10)
\]

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

\[
\hat{\gamma} - \gamma = (\mathbf{X}_0'\Omega^{-1}_0\mathbf{X}_0)^{-1}\mathbf{X}_0'\Omega^{-1}_0 T^{-1} \sum_{t \leq T} \epsilon_{0,t} + o_p(T^{-1/2}) = O_p(T^{-1/2}), \quad (A.11)
\]

where \( \Omega_0 \) denotes the leading \( N_0 \times N_0 \) submatrix of \( \Omega. \) Combining the results in (A.7), (A.10), and (A.11),

\[
\mathbf{R}_0 - \hat{\mathbf{X}}_0 \hat{\gamma} = \mathbf{R}_0 - \hat{\mathbf{X}}_0 \gamma - \hat{\mathbf{X}}_0(\hat{\gamma} - \gamma) = (I_{N_0} - \mathbf{X}_0(\mathbf{X}_0'\Omega_0^{-1}\mathbf{X}_0)^{-1}\mathbf{X}_0'\Omega_0^{-1})T^{-1} \sum_{t \leq T} \epsilon_{0,t} + o_p(T^{-1/2}) = \Omega_0^{1/2}M_{N_0}\Omega_0^{-1/2}T^{-1} \sum_{t \leq T} \epsilon_{0,t} + o_p(T^{-1/2}) = O_p(T^{-1/2}), \quad (A.12)
\]

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 \( \hat{\Omega}, \) Assumptions 2(i, iv), and the continuous mapping theorem implies

\[
J_T = T(\mathbf{R}_0 - \hat{\mathbf{X}}_0 \hat{\gamma})'\Omega_0^{-1}(\mathbf{R}_0 - \hat{\mathbf{X}}_0 \hat{\gamma}) + o_p(1) = \left( \Omega_0^{-1/2}T^{-1/2} \sum_{t \leq T} \epsilon_{0,t} \right) M_{N_0} \left( \Omega_0^{-1/2}T^{-1/2} \sum_{t \leq T} \epsilon_{0,t} \right) + o_p(1) \rightarrow_d \mathcal{N}_0'\mathbb{M}_{N_0}\mathcal{N}_0, \quad (A.13)
\]

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 \( \mathcal{N}_0'\mathbb{M}_{N_0}\mathcal{N}_0 \) has the same distribution as \( \chi^2(N_0 - d_\gamma). \) The claim of the theorem follows directly from (A.13).
(b) Under Assumption 1 and condition (ii) of the theorem, we can use the same arguments for showing (A.7) and (A.8) in the proof of Theorem 1 to obtain

\[ \hat{\mathbf{X}}_0 = \mathbf{X}_0 + o_p(1), \quad \hat{\mathbf{X}}_0' \Omega_0^{-1} \hat{\mathbf{X}}_0 = \mathbf{X}_0' \Omega_1^{-1} \mathbf{X}_0 + o_p(1) \quad \text{and} \quad \hat{\mathbf{X}}_0' \Omega_0^{-1} = \mathbf{X}_0' \Omega_1^{-1} + o_p(1), \]

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

\[ \mathbf{R}_0 - \hat{\mathbf{X}}_0 \hat{\gamma} = (\mathbf{I}_{N_0} - \mathbf{X}_0 (\mathbf{X}_0' \Omega_1^{-1} \mathbf{X}_0)^{-1} \mathbf{X}_0' \Omega_0^{-1}) \mathbf{E}[\mathbf{R}_0, t] + o_p(1). \quad (A.14) \]

By conditions (ii, iii) of the theorem and (A.14),

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

which proves the claim of the theorem. \( Q.E.D. \)

**Proof of Lemma 1.** Let \( Y_t \equiv (f'_t, R_{1,t}, \ldots, R_{N,t}, Z_{1,t}, \ldots, Z_{N,t})' \) and \( \phi \equiv (\theta', \delta')' \), where \( \delta \equiv (\mu'_f, \beta_1, \ldots, \beta_N)' \). Let \( A \otimes B \) denote the Kronecker product of two matrices \( A \) and \( B \). Using the restrictions in (2.4), the moment conditions in (2.1), (2.2), and (2.3) can be written as

\[
\begin{align*}
g_1(Y_t, \phi) & \equiv f_t - \mu_f, \\
g_2(Y_t, \phi) & \equiv (\mathbf{R}_t - \mathbf{B}(f_t - \mu_f)) \otimes (f_t - \mu_f), \\
g_3(Y_t, \phi) & \equiv \mathbf{R}_t - \mathbf{X}_t \theta.
\end{align*}
\]

Let \( g(Y_t, \phi) \equiv (g_1(Y_t, \phi)', g_2(Y_t, \phi)', g_3(Y_t, \phi)')' \). Then, by Chamberlain (1987), the semiparametric efficiency bound for \( \phi \) is

\[ \Omega_g^* = G' \Sigma_g^{*-1} G, \quad (A.15) \]

where \( G \equiv \mathbb{E}[\partial g(Y_t, \phi)/\partial \phi'] \) and \( \Sigma_g^{*} \equiv \mathbb{E}[g(Z_t, \phi)g(Z_t, \phi')] \). Let \( \Omega_g^{*11} \) denote the leading \( d_{\theta} \times d_{\theta} \) submatrix of \( \Omega_g^{*1} \). Because \( \theta \) is the leading \( d_{\theta} \times 1 \) subvector of \( \phi \), the semiparametric efficiency bound for \( \theta \) equals the inverse of \( \Omega_g^{*11} \). It remains to calculate \( \Omega_g^{*11} \) and show it equals \( (\mathbf{X}' \Omega_1^{-1} \mathbf{X})^{-1} \).

First, simple calculation shows

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

Let

\[
D = \begin{pmatrix}
-\mathbf{X} \theta \otimes \Sigma_f^{-1} & I_N \otimes \gamma_1' \Sigma_f^{-1} & -I_N \\
-I_K & 0_{K \times N K} & 0_{K \times N} \\
\mathbf{X} \theta \otimes \Sigma_f^{-1} & -I_N \otimes \Sigma_f^{-1} & 0_{N K \times N K}
\end{pmatrix} \quad (A.16)
\]
Because $D$ is invertible, we can write $\Omega^*_g = (DG)'(DG)^{-1}(DG)$. Because $DG = \text{diag}(X, I_{(N+1)K})$, by Lemma A.1 of Chamberlain (1987),

$$\Omega^{*11}_g = (X'(DGX)'_11X)^{-1}, \quad (A.17)$$

where $(DGX)'_11$ denotes the leading $N \times N$ submatrix of $DGX$. Let $(Dg(Y_t, \phi))_N$ denote the leading $N \times 1$ subvector of $Dg(Y_t, \phi)$. Then,

$$
(Dg(Y_t, \phi))_N = -X\theta' \gamma'_1 \Sigma^{-1} \gamma_1 g_1(Y_t, \phi) + (I_N \otimes \gamma'_1 \Sigma_f^{-1})g_2(Y_t, \phi) - g_3(Y_t, \phi)
$$

$$= -X\theta' \gamma'_1 \Sigma_f^{-1}(f_t - \mu_f) + (R_t - B(f_t - \mu_f))\gamma'_1 \Sigma_f^{-1}(f_t - \mu_f) - (R_t - X_t \theta)
$$

$$= (R_t - E[R_t] - B(f_t - \mu_f))\gamma'_1 \Sigma_f^{-1}(f_t - \mu_f) - (R_t - X_t \theta). \quad (A.18)$$

By the definition of $u_t$ and $v_t$, we deduce from (A.18) that

$$
(Dg(Y_t, \phi))_N = -v_t + u_t'(f_t - \mu_f)' \Sigma_f^{-1} \gamma_1 = -\epsilon_t, \quad (A.19)
$$

which implies

$$
(Dg(Y_t, \phi))_N = \epsilon_t, \quad (A.20)
$$

Combining the results in (A.17) and (A.20), we have $\Omega^{*11}_g = (X'\gamma^{-1}X)^{-1}$. $Q.E.D.$

**Proof of Theorem 3.** We follow the same notations defined in the proof of Lemma 1 here. By the standard arguments on showing the asymptotic properties of the GMM estimator, the variance-covariance matrix of $\phi^{*\text{gmm}}$ is $(G'\Sigma_g^{-1}G)^{-1}$, where

$$
\Sigma_g \equiv \lim_{T \to \infty} E \left[ \sum_{t_1, t_2 \leq T} g(Y_{t_1}, \phi)g(Y_{t_2}, \phi)' \right].
$$

Because $\hat{\phi}^{*\text{gmm}}$ is the leading $d_0 \times 1$ subvector of $\phi^{*\text{gmm}}$, the variance-covariance matrix of $\hat{\phi}^{*\text{gmm}}$ is the leading $d_0 \times d_0$ submatrix of $(G'\Sigma_g^{-1}G)^{-1}$. Recall the matrix $D$ defined in (A.16). Following the same arguments for showing (A.20), we deduce that

$$
(D\Sigma_g D')_{11} = \lim_{T \to \infty} E \left[ \sum_{t_1, t_2 \leq T} \epsilon_{t_1} \epsilon_{t_2}' \right] = \Omega.
$$

Therefore, by the same arguments for showing $\Omega^{*11}_g = (X'\gamma^{-1}X)^{-1}$ in the proof of Lemma 1, we can show the leading $d_0 \times d_0$ submatrix of $(G'\Sigma_g^{-1}G)^{-1}$ is $(X'\Omega^{-1}X)^{-1}$. $Q.E.D.$
B A Counter Example

Suppose that we are interested in estimating an unknown parameter \( \theta \) which is identified by the following moment conditions

\[
\begin{align*}
\mathbb{E}[Y_1 - \mu & 1_{k \times 1}] = 0_{k \times 1}, \\ \mathbb{E}[Y_2 - \mu - \theta] &= 0,
\end{align*}
\]

(B.21) (B.22)

where \( k \geq 1 \) and \( \mu \) 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

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

where \( \rho \) is a \( k \times 1 \) real vector such that \( \Sigma_\varepsilon \) is positive definition.

We compare the asymptotic variances of two GMM estimators of \( \theta \). The first one is from a joint optimal GMM estimation of \((\theta, \mu)\) using all the moment conditions in (B.21) and (B.22). The second one is obtained by an iterative GMM estimation procedure where we first obtain the optimal GMM estimator of \( \mu \) through the moment conditions in (B.21), and then plug it in (B.22) to construct the optimal GMM estimator of \( \theta \). 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 \((\theta, \mu)\) based on (B.21) and (B.22) is the inverse of the following matrix

\[
\left( \begin{array}{cc} 1_{k \times k} & 1 \\ 0_{1 \times k} & 1 \end{array} \right) \left( \begin{array}{cc} I_k & \rho' \\ \rho & 1 \end{array} \right)^{-1} \left( \begin{array}{cc} 1_{k \times 1} & 0_{k \times 1} \\ 1 & 1 \end{array} \right) = \left( \begin{array}{cc} k + (1 - 1'_{k \times 1} \rho)^2 & 1 - 1'_{k \times 1} \rho \\ 1 - 1'_{k \times 1} \rho & 1 - \rho' \rho \end{array} \right)
\]

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

\[
\text{ASV}_{\hat{\theta}^J} \equiv k + (1 - 1'_{k \times 1} \rho)^2 \left( \frac{1 - \rho' \rho}{k} \right) = 1 + k^{-1} - 2k^{-1}1'_{k \times 1} \rho + \frac{(1'_{k \times 1} \rho)^2 - k \rho' \rho}{k}. \quad (B.23)
\]

On the other hand, the optimal GMM estimator of \( \mu \) based on (B.21) is

\[
\hat{\mu}^J \equiv k^{-1}1'_{k \times 1} \bar{Y}_1
\]

where \( \bar{Y}_1 \equiv T^{-1} \sum_{t=1}^T Y_{1,t} \). Therefore, the optimal iterative GMM estimator of \( \theta \) is

\[
\hat{\theta}^J \equiv \bar{Y}_2 - \hat{\mu}^J = \bar{Y}_2 - k^{-1}1'_{k \times 1} \bar{Y}_1.
\]
From the above expression, it is easy to show that the asymptotic variance of $\hat{\theta}^J$ is

$$ASV_{\hat{\theta}^J} \equiv 1 + k^{-1} - 2k^{-1}1_{k \times 1}'\rho.$$  \hfill (B.24)

By (B.23) and (B.24), we have

$$ASV_{\hat{\theta}^J} - ASV_{\hat{\theta}^I} = k^{-1}(1_{k \times 1}'\rho)^2 - \rho'\rho \leq 0$$ \hfill (B.25)

where the equality holds if and only if $k = 1$ and/or $\rho = a1_{k \times 1}$ for some real number $a$ by the Cauchy-Schwarz inequality.

From the inequality in (B.25), 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 (B.21) and (B.22) have special dependence structure, i.e., $\rho = a1_{k \times 1}$. Under general dependence of the moment conditions in (B.21) and (B.22) (i.e., $\rho$ is not zero and is linearly independent with respect to $1_{k \times 1}$), $\hat{\theta}^J$ and $\hat{\theta}^I$ are asymptotically equivalent only when $k = 1$, i.e., the nuisance parameter $\mu$ is just identified in (B.21), which shares the same intuition of the asymptotic efficiency of our OCSR estimator. Moreover since the difference $ASV_{\hat{\theta}^J} - ASV_{\hat{\theta}^I}$ is decreasing in $k$, the larger the number of over-identification restrictions $k$, the higher the efficiency loss the iterative optimal GMM estimator will suffer.

### C 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 $N$ may go to infinity with $T$. Since the number of assets in $I_1$ may grow with $N$, the number of the unknown parameters (i.e., $\alpha_i$ for $i \in I_1$) may also diverge to infinity with $T$. We first state the conditions needed for the extension.

In the following, $\delta_j$ ($j = 1, 2, 3$) are nonnegative finite constants.

**Assumption 4.** Suppose that: (i) $\{f_t\}$ is a covariance-stationary process; (ii) $\bar{f} = \mu_f + O_p(T^{-1/2})$; (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; (iv) $\bar{Z} = T^{-1}\sum_{t \leq T} \mathbb{E}[Z_t] + O_p(N^{\delta_1}T^{-1/2})$.

Assumption 4(i) is the same as Assumption 1(i). Assumptions 4(ii, iii, iv) strengthen Assumption 1(ii, iii, iv) by providing the convergence rates of $\bar{f}$, $T^{-1}\sum_{t \leq T} f_t f_t'$ and $\bar{Z}$ to their population counterparts. Note that Assumption 4(iv) does not impose the stationary assumption on the $Z_t$ process. The factor $N^{\delta_1}$ shows up in Assumption 4(iv) because $\bar{Z}$ is an $N \times 1$ 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 $\mathbf{Z}$ are bounded from above uniformly over $N$ and $T$, one can show that Assumption 1(iv) holds with $\delta_1 = 1/2$.

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

$$T^{-1/2} \sum_{t=1}^T \epsilon_t = \Omega_T^{1/2} \mathbf{N} + o_p(1)$$

where $\Omega_T \equiv \text{Var}(T^{-1/2} \sum_{t=1}^T \epsilon_t)$; (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 $\Omega_T$, $\Sigma_f$, $W$ and $\mathbf{X}'\mathbf{X}$ are bounded from above and away from zero uniformly over $N$ and $T$ where $\mathbf{X} = (S_N, \mathbf{1}_{N \times 1}, \mathbf{B}, T^{-1} \sum_{t \leq T} \mathbb{E}[\mathbf{Z}_t])$; (v) $N^{\delta_1 + 1/2}T^{-1/2} \to 0$ 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 $T^{-1/2} \sum_{t=1}^T \epsilon_t$, which can be verified when $\{\epsilon_t\}_t$ 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 $\beta_t$, 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 $T^{1/2}$.

**Theorem 4.** Under Assumptions 4 and 5, we have

$$T^{1/2}(\hat{\theta} - \theta) = (\Sigma_T(W))^{1/2} \mathbf{N}^* + o_p(1) \quad (C.26)$$

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

$$\Sigma_T(W) \geq (\mathbf{X}'\Omega_T^{-1}\mathbf{X})^{-1} \quad (C.27)$$

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

Theorem 4 generalizes Theorem 1 to the case where both the number of assets $N$ and the number of the unknown parameters $d_\theta$ may go to infinity with $T$. Since the eigenvalues of $\Sigma_T(W)$ are bounded away from zero under Assumption 5(iv), $T^{1/2}(\hat{\theta} - \theta)$ is not asymptotically tight and hence it does not admit an asymptotic distribution. Nevertheless, (C.26) implies that the
finite sample distribution of \( T^{1/2}(\hat{\theta} - \theta) \) can still be approximated by a normal random vector with variance-covariance matrix \( \Sigma_T(W) \). Therefore, one can still conduct inference on \( \theta \) using the normal approximation (C.26).

From (C.27), it is clear that the OCSR estimator still takes the form in (2.10) and the “asymptotic” variance-covariance matrix of OCSR is \( (X'\Omega_T^{-1}X)^{-1} \). Therefore, we can use the same arguments in the proof of Theorem 3 to show that the OCSR estimator has the same “asymptotic” variance-covariance matrix of optimal GMM estimator \( \hat{\theta}^*_{gmm} \).

**Proof of Theorem 4.** By Assumptions 4(i, ii, iii),

\[
\hat{\Sigma}_f = \Sigma_f + o_p(T^{-1/2}) \tag{C.28}
\]

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

\[
\hat{B} - B = T^{-1} \sum_{t \leq T} \mathbf{u}_t (f_t - \mu_f)^\top \Sigma_f^{-1} + o_p(T^{-1/2}) = O_p(N^\delta T^{-1/2}) = o_p(1). \tag{C.29}
\]

Therefore by Assumptions 5(i, iv) and (C.29), we have

\[
\mathbf{v} - (\hat{B} - B)\gamma_1 = T^{-1} \sum_{t \leq T} \epsilon_t + o_p(T^{-1/2}) = O_p((N/T)^{1/2}). \tag{C.30}
\]

Similarly, by Assumptions 4(iv) and 5(v), and (C.29)

\[
\hat{X} = X + O_p(N^\delta T^{-1/2}) = X + o_p(1). \tag{C.31}
\]

From Assumptions 5(iii, iv, v) and (C.31), we obtain

\[
\hat{X}'\hat{W} = X'W + O_p(N^\delta T^{-1/2}) \text{ and } \hat{X}'\hat{W}\hat{X} = X'WX + O_p(N^\delta T^{-1/2}). \tag{C.32}
\]

Combining the results in (A.2), (C.28) and (C.32) and applying Assumptions 5(i, iv), we get

\[
T^{1/2}(\hat{\theta} - \theta) = (X'WX)^{-1}X'WT^{-1/2} \sum_{t \leq T} \epsilon_t + o_p(1)
= (X'WX)^{-1}X'W\Omega_T^{1/2}N + o_p(1). \tag{C.33}
\]

Since \( (X'WX)^{-1}X'W\Omega_T^{1/2}N \) has the same distribution as \( (\Sigma_T(W))^{1/2}N^* \), then claim in (C.26) follows from (C.33). The claim in (C.27) follows by the same arguments on showing (2.9) and hence is omitted. \( Q.E.D. \)

\(^{37}\text{Note that the OCSR depends on a consistent estimator of } \Omega_T. \text{ See Theorem 5 in Li and Liao (2020) for consistent variance-covariance estimator with divergent dimension.} \)
D Additional Results for the Simulation Study

Table D.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. $\gamma_0$, $\gamma_{mkt}$, $\gamma_{smb}$, $\gamma_{hml}$, $\gamma_{cma}$, and $\gamma_{rmw}$ 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.

<table>
<thead>
<tr>
<th>Panel A: FF 3-Factor Model, $N = 18$</th>
<th>Panel C: FF 3-Factor Model, $N = 38$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma_0$</td>
</tr>
<tr>
<td>True</td>
<td>0</td>
</tr>
<tr>
<td>OLS Bias</td>
<td>0.064</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.247</td>
</tr>
<tr>
<td>MAE</td>
<td>0.201</td>
</tr>
<tr>
<td>OCSR Bias</td>
<td>0.007</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.154</td>
</tr>
<tr>
<td>MAE</td>
<td>0.123</td>
</tr>
<tr>
<td>GLS</td>
<td>0.009</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.169</td>
</tr>
<tr>
<td>MAE</td>
<td>0.134</td>
</tr>
<tr>
<td>WLS</td>
<td>0.066</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.288</td>
</tr>
<tr>
<td>MAE</td>
<td>0.231</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: FF 5-Factor Model, $N = 18$</th>
<th>Panel D: FF 5-Factor Model, $N = 38$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma_0$</td>
</tr>
<tr>
<td>True</td>
<td>0</td>
</tr>
<tr>
<td>OLS Bias</td>
<td>0.008</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.191</td>
</tr>
<tr>
<td>MAE</td>
<td>0.152</td>
</tr>
<tr>
<td>OCSR Bias</td>
<td>0.015</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.137</td>
</tr>
<tr>
<td>MAE</td>
<td>0.110</td>
</tr>
<tr>
<td>GLS</td>
<td>0.017</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.158</td>
</tr>
<tr>
<td>MAE</td>
<td>0.127</td>
</tr>
<tr>
<td>WLS</td>
<td>0.007</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.169</td>
</tr>
<tr>
<td>MAE</td>
<td>0.134</td>
</tr>
</tbody>
</table>
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: $OLS^{stage}$ 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 pre-specified significance level is found and used to calculate the corresponding test power.

<table>
<thead>
<tr>
<th>Effect</th>
<th>$\gamma_0$</th>
<th>$\gamma_{mkt}$</th>
<th>$\gamma_{emb}$</th>
<th>$\gamma_{hml}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0 \times \gamma_0$</td>
<td>$\gamma_{mkt}$</td>
<td>$\gamma_{emb}$</td>
<td>$\gamma_{hml}$</td>
<td></td>
</tr>
<tr>
<td>$5%$</td>
<td>$0.126$ n.a.</td>
<td>$0.183$ n.a.</td>
<td>$0.157$ n.a.</td>
<td></td>
</tr>
<tr>
<td>$10%$</td>
<td>$0.197$ n.a.</td>
<td>$0.215$ n.a.</td>
<td>$0.200$ n.a.</td>
<td></td>
</tr>
<tr>
<td>$0.5 \times \gamma_0$</td>
<td>$\gamma_{mkt}$</td>
<td>$\gamma_{emb}$</td>
<td>$\gamma_{hml}$</td>
<td></td>
</tr>
<tr>
<td>$5%$</td>
<td>$0.077$ n.a.</td>
<td>$0.085$ n.a.</td>
<td>$0.079$ n.a.</td>
<td></td>
</tr>
<tr>
<td>$10%$</td>
<td>$0.140$ n.a.</td>
<td>$0.142$ n.a.</td>
<td>$0.137$ n.a.</td>
<td></td>
</tr>
<tr>
<td>$1.0 \times \gamma_0$</td>
<td>$\gamma_{mkt}$</td>
<td>$\gamma_{emb}$</td>
<td>$\gamma_{hml}$</td>
<td></td>
</tr>
<tr>
<td>$5%$</td>
<td>$0.277$ n.a.</td>
<td>$0.324$ n.a.</td>
<td>$0.324$ n.a.</td>
<td></td>
</tr>
<tr>
<td>$10%$</td>
<td>$0.366$ n.a.</td>
<td>$0.415$ n.a.</td>
<td>$0.415$ n.a.</td>
<td></td>
</tr>
<tr>
<td>$1.5 \times \gamma_0$</td>
<td>$\gamma_{mkt}$</td>
<td>$\gamma_{emb}$</td>
<td>$\gamma_{hml}$</td>
<td></td>
</tr>
<tr>
<td>$5%$</td>
<td>$0.523$ n.a.</td>
<td>$0.517$ n.a.</td>
<td>$0.517$ n.a.</td>
<td></td>
</tr>
<tr>
<td>$10%$</td>
<td>$0.613$ n.a.</td>
<td>$0.618$ n.a.</td>
<td>$0.618$ n.a.</td>
<td></td>
</tr>
</tbody>
</table>

| Table D.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\textsuperscript{stage} 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, OCSR is the cross-sectional OLS estimator that takes first-stage beta-estimation uncertainty into account through (2.8), OCSR\textsuperscript{Adj.} is the two-pass estimator that sets the off-diagonal elements of GLS\textsuperscript{Adj.}’s weighting matrix at zero and has standard errors calculated through (2.8) and WLS 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). We report both the original size and power (Ori.) and the size-adjusted power (Adj.), where the statistical cutoff that exactly achieves a pre-specified significance level is found and used to calculate the corresponding test power.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Sig.</th>
<th>OLS\textsuperscript{stage}</th>
<th>FM</th>
<th>OLS</th>
<th>OCSR</th>
<th>GLS</th>
<th>WLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{mkt}$</td>
<td>0.0%</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>$\gamma_{emb}$</td>
<td>0.0%</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>$\gamma_{hml}$</td>
<td>0.0%</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
</tbody>
</table>

51
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: $OLS_{stage}$ 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-premiums are used for hypothesis testing, $OCSR$ 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 pre-specified significance level is found and used to calculate the corresponding test power.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Sig.</th>
<th>$OLS_{stage}$</th>
<th>$FM$</th>
<th>OLS</th>
<th>$OCSR$</th>
<th>$GLS$</th>
<th>$WLS$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 × $\gamma_0$</td>
<td>1%</td>
<td>0.034</td>
<td>0.048</td>
<td>0.012</td>
<td>0.027</td>
<td>0.029</td>
<td>0.036</td>
</tr>
<tr>
<td>5%</td>
<td>0.109</td>
<td>0.131</td>
<td>0.051</td>
<td>0.089</td>
<td>0.094</td>
<td>0.108</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.178</td>
<td>0.202</td>
<td>0.096</td>
<td>0.155</td>
<td>0.160</td>
<td>0.177</td>
<td></td>
</tr>
<tr>
<td>0.5 × $\gamma_0$</td>
<td>1%</td>
<td>0.076</td>
<td>0.074</td>
<td>0.016</td>
<td>0.023</td>
<td>0.020</td>
<td>0.078</td>
</tr>
<tr>
<td>5%</td>
<td>0.176</td>
<td>0.171</td>
<td>0.076</td>
<td>0.089</td>
<td>0.087</td>
<td>0.178</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.254</td>
<td>0.252</td>
<td>0.130</td>
<td>0.159</td>
<td>0.265</td>
<td>0.193</td>
<td></td>
</tr>
<tr>
<td>1.0 × $\gamma_0$</td>
<td>1%</td>
<td>0.150</td>
<td>0.131</td>
<td>0.032</td>
<td>0.059</td>
<td>0.050</td>
<td>0.209</td>
</tr>
<tr>
<td>5%</td>
<td>0.296</td>
<td>0.259</td>
<td>0.134</td>
<td>0.168</td>
<td>0.166</td>
<td>0.289</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.396</td>
<td>0.350</td>
<td>0.209</td>
<td>0.263</td>
<td>0.268</td>
<td>0.487</td>
<td></td>
</tr>
<tr>
<td>1.5 × $\gamma_0$</td>
<td>1%</td>
<td>0.419</td>
<td>0.364</td>
<td>0.113</td>
<td>0.219</td>
<td>0.203</td>
<td>0.660</td>
</tr>
<tr>
<td>5%</td>
<td>0.692</td>
<td>0.489</td>
<td>0.310</td>
<td>0.448</td>
<td>0.445</td>
<td>0.813</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.899</td>
<td>0.584</td>
<td>0.429</td>
<td>0.577</td>
<td>0.584</td>
<td>0.874</td>
<td></td>
</tr>
<tr>
<td>0 × $\gamma_{mkt}$</td>
<td>1%</td>
<td>0.054</td>
<td>0.015</td>
<td>0.002</td>
<td>0.010</td>
<td>0.007</td>
<td>0.013</td>
</tr>
<tr>
<td>5%</td>
<td>0.137</td>
<td>0.058</td>
<td>0.019</td>
<td>0.051</td>
<td>0.035</td>
<td>0.052</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.211</td>
<td>0.109</td>
<td>0.051</td>
<td>0.101</td>
<td>0.075</td>
<td>0.104</td>
<td></td>
</tr>
<tr>
<td>0.5 × $\gamma_{mkt}$</td>
<td>1%</td>
<td>0.223</td>
<td>0.090</td>
<td>0.049</td>
<td>0.037</td>
<td>0.023</td>
<td>0.060</td>
</tr>
<tr>
<td>5%</td>
<td>0.358</td>
<td>0.146</td>
<td>0.132</td>
<td>0.094</td>
<td>0.169</td>
<td>0.393</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.443</td>
<td>0.234</td>
<td>0.221</td>
<td>0.171</td>
<td>0.272</td>
<td>0.514</td>
<td></td>
</tr>
<tr>
<td>1.0 × $\gamma_{mkt}$</td>
<td>1%</td>
<td>0.562</td>
<td>0.233</td>
<td>0.185</td>
<td>0.184</td>
<td>0.328</td>
<td>0.768</td>
</tr>
<tr>
<td>5%</td>
<td>0.707</td>
<td>0.429</td>
<td>0.406</td>
<td>0.410</td>
<td>0.549</td>
<td>0.917</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.772</td>
<td>0.559</td>
<td>0.545</td>
<td>0.530</td>
<td>0.669</td>
<td>0.958</td>
<td></td>
</tr>
<tr>
<td>1.5 × $\gamma_{mkt}$</td>
<td>1%</td>
<td>0.918</td>
<td>0.807</td>
<td>0.770</td>
<td>0.845</td>
<td>0.923</td>
<td>1.000</td>
</tr>
<tr>
<td>5%</td>
<td>0.963</td>
<td>0.914</td>
<td>0.908</td>
<td>0.949</td>
<td>0.972</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.978</td>
<td>0.953</td>
<td>0.947</td>
<td>0.973</td>
<td>0.986</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>0 × $\gamma_{ smb}$</td>
<td>1%</td>
<td>0.480</td>
<td>0.036</td>
<td>0.004</td>
<td>0.014</td>
<td>0.018</td>
<td>0.017</td>
</tr>
<tr>
<td>5%</td>
<td>0.595</td>
<td>0.103</td>
<td>0.023</td>
<td>0.061</td>
<td>0.068</td>
<td>0.067</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.656</td>
<td>0.173</td>
<td>0.060</td>
<td>0.115</td>
<td>0.128</td>
<td>0.122</td>
<td></td>
</tr>
<tr>
<td>0.5 × $\gamma_{ smb}$</td>
<td>1%</td>
<td>0.602</td>
<td>0.059</td>
<td>0.015</td>
<td>0.040</td>
<td>0.081</td>
<td>0.062</td>
</tr>
<tr>
<td>5%</td>
<td>0.687</td>
<td>0.142</td>
<td>0.087</td>
<td>0.074</td>
<td>0.122</td>
<td>0.250</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.736</td>
<td>0.229</td>
<td>0.161</td>
<td>0.142</td>
<td>0.204</td>
<td>0.359</td>
<td></td>
</tr>
<tr>
<td>1.0 × $\gamma_{ smb}$</td>
<td>1%</td>
<td>0.816</td>
<td>0.178</td>
<td>0.066</td>
<td>0.099</td>
<td>0.191</td>
<td>0.454</td>
</tr>
<tr>
<td>5%</td>
<td>0.862</td>
<td>0.372</td>
<td>0.237</td>
<td>0.276</td>
<td>0.377</td>
<td>0.677</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.885</td>
<td>0.499</td>
<td>0.365</td>
<td>0.406</td>
<td>0.497</td>
<td>0.782</td>
<td></td>
</tr>
<tr>
<td>1.5 × $\gamma_{ smb}$</td>
<td>1%</td>
<td>0.990</td>
<td>0.068</td>
<td>0.468</td>
<td>0.665</td>
<td>0.805</td>
<td>0.988</td>
</tr>
<tr>
<td>5%</td>
<td>0.995</td>
<td>0.841</td>
<td>0.852</td>
<td>0.755</td>
<td>0.873</td>
<td>0.926</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.996</td>
<td>0.907</td>
<td>0.905</td>
<td>0.849</td>
<td>0.937</td>
<td>0.962</td>
<td></td>
</tr>
<tr>
<td>0 × $\gamma_{ hml}$</td>
<td>1%</td>
<td>0.311</td>
<td>0.032</td>
<td>0.003</td>
<td>0.022</td>
<td>0.017</td>
<td>0.015</td>
</tr>
<tr>
<td>5%</td>
<td>0.438</td>
<td>0.101</td>
<td>0.024</td>
<td>0.078</td>
<td>0.071</td>
<td>0.066</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.512</td>
<td>0.167</td>
<td>0.059</td>
<td>0.136</td>
<td>0.127</td>
<td>0.123</td>
<td></td>
</tr>
<tr>
<td>0.5 × $\gamma_{ hml}$</td>
<td>1%</td>
<td>0.460</td>
<td>0.040</td>
<td>0.025</td>
<td>0.025</td>
<td>0.053</td>
<td>0.075</td>
</tr>
<tr>
<td>5%</td>
<td>0.569</td>
<td>0.136</td>
<td>0.099</td>
<td>0.098</td>
<td>0.159</td>
<td>0.258</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.634</td>
<td>0.220</td>
<td>0.176</td>
<td>0.171</td>
<td>0.239</td>
<td>0.359</td>
<td></td>
</tr>
<tr>
<td>1.0 × $\gamma_{ hml}$</td>
<td>1%</td>
<td>0.757</td>
<td>0.169</td>
<td>0.128</td>
<td>0.157</td>
<td>0.248</td>
<td>0.475</td>
</tr>
<tr>
<td>5%</td>
<td>0.830</td>
<td>0.384</td>
<td>0.442</td>
<td>0.314</td>
<td>0.367</td>
<td>0.477</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.883</td>
<td>0.516</td>
<td>0.548</td>
<td>0.439</td>
<td>0.497</td>
<td>0.591</td>
<td></td>
</tr>
<tr>
<td>1.5 × $\gamma_{ hml}$</td>
<td>1%</td>
<td>0.980</td>
<td>0.592</td>
<td>0.824</td>
<td>0.685</td>
<td>0.813</td>
<td>0.890</td>
</tr>
<tr>
<td>5%</td>
<td>0.992</td>
<td>0.818</td>
<td>0.822</td>
<td>0.867</td>
<td>0.944</td>
<td>0.967</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.995</td>
<td>0.897</td>
<td>0.953</td>
<td>0.921</td>
<td>0.970</td>
<td>0.981</td>
<td></td>
</tr>
</tbody>
</table>
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: $\text{OLS}^{\text{first-stage}}$ 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, $GLS$ 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 pre-specified significance level is found and used to calculate the corresponding test power.

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

<table>
<thead>
<tr>
<th>Effect</th>
<th>Sig.</th>
<th>$\text{OLS}^{\text{first-stage}}$</th>
<th>$FM$</th>
<th>$OCSR$</th>
<th>$GLS$</th>
<th>$WLS$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_mkt$</td>
<td>1%</td>
<td>0.270 n.a. 0.004 n.a. 0.001 n.a. 0.024 n.a. 0.011 n.a. 0.007 n.a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.402 n.a. 0.028 n.a. 0.010 n.a. 0.073 n.a. 0.051 n.a. 0.038 n.a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>0.483 n.a. 0.062 n.a. 0.028 n.a. 0.135 n.a. 0.102 n.a. 0.073 n.a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{smb}$</td>
<td>1%</td>
<td>0.575 n.a. 0.028 n.a. 0.003 n.a. 0.021 n.a. 0.018 n.a. 0.013 n.a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.669 n.a. 0.097 n.a. 0.021 n.a. 0.069 n.a. 0.069 n.a. 0.063 n.a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>0.729 n.a. 0.162 n.a. 0.052 n.a. 0.125 n.a. 0.127 n.a. 0.120 n.a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{hml}$</td>
<td>1%</td>
<td>0.749 0.000 0.138 0.077 0.069 0.329 0.273 0.179 0.153 0.129 0.073 0.043</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.763 0.000 0.144 0.081 0.075 0.282 0.258 0.192 0.169 0.142 0.110 0.076</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>0.786 0.219 0.227 0.143 0.129 0.199 0.442 0.397 0.321 0.281 0.241 0.211</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{rmb}$</td>
<td>1%</td>
<td>0.842 0.000 0.138 0.064 0.059 0.144 0.615 0.496 0.400 0.318 0.211 0.176</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.884 0.000 0.313 0.208 0.219 0.342 0.810 0.760 0.626 0.566 0.419 0.382</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>0.903 0.469 0.424 0.319 0.346 0.453 0.857 0.735 0.695 0.537 0.502 0.476</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{hml}$</td>
<td>1%</td>
<td>0.991 0.000 0.587 0.416 0.542 0.144 0.998 0.995 0.976 0.961 0.815 0.782</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.994 0.000 0.786 0.685 0.800 0.868 1.000 1.000 0.995 0.992 0.932 0.916</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>0.996 0.890 0.863 0.791 0.887 0.930 1.000 1.000 0.998 0.998 0.964 0.955</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>