Optimal Cross-Sectional Regression

Zhipeng Liao

UCLA, Los Angeles, CA 90095, USA

Yan Liu

Purdue University, West Lafayette, IN 47906, USA

January 7, 2022

Abstract

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

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

JEL Codes: C14, C22, G12.

*Current Version: January 7, 2022. First posted to SSRN: October 26, 2020. Send correspondence to Zhipeng Liao, UCLA Department of Economics, Los Angeles, CA 90095. Email: Zhipeng.liao@econ.ucla.edu; and Yan Liu, Krannert School of Management, Purdue University, West Lafayette, IN 47906. Email: liu2746@purdue.edu. We have benefited from the comments of Federico Bandi, John Cochrane, Jason Donaldson, Philip Dybvig, Wayne Ferson, Stefano Giglio, Bing Han, Jinyong Hahn, Kewei Hou, Yan Hong, Cam Harvey, Craig Lewis, Nengjiu Ju, Raymond Kan, Asaf Manela, Stefan Nagel, Andreas Neuhierl, James Nordlund, Jun Pan, Xuhui Pan, David Rapach, Berk Sensoy, Junbo Wang, Tan Wang, Wei Wei, Lu Zhang, Guofu Zhou, the committee of The Society for Financial Econometrics (SoFiE), as well as seminar participants at Washington University in St. Louis, Ohio State University, University of Oklahoma, City University of Hong Kong, Louisiana State University, Vanderbilt University, and Shanghai Advanced Institute of Finance (SAIF).
1 Introduction

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

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

Second, we theoretically demonstrate the optimality of our approach not only among two-pass regression estimators, somewhat surprisingly, but also over the much broader class of the generalized method of moments (GMM) type of simultaneous estimators. The GMM approach has long been argued to be the preferred approach (e.g., MacKinlay and Richardson (1991)) in estimating beta-pricing models because, intuitively, it simultaneously and potentially more efficiently uses all the information provided by the asset pricing model. For instance, betas are not only learned from return time series but also from the cross-sectional pricing restriction. However, our two-pass estimator suggests that this intuition is misguided: by appropriately weighting returns in the second-pass regression, we show that our approach is just as efficient as the (optimally weighted) GMM estimator. Our paper hence bridges the gap between the earlier literature on GMM-based asset pricing tests (e.g., MacKinlay and Richardson (1991), Zhou (1994), and Jagannathan and...
Wang (1998)) and the more popular two-pass regression literature (e.g., Shanken (1992), Gagliardi, Ossola, and Scaillet (2016), Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2019), Kim and Skoulakis (2018), and Raponi, Robotti, and Zaffaroni (2020)). We show that our new two-pass estimator, while inheriting the theoretical appeal of GMM, is computationally as attractive as leading two-pass estimators, such as the Fama-MacBeth approach.

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

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

Expanding the insights from this example, we perform a comprehensive simulation study to evaluate the performance of our new estimator. Across various specifications of the data generating process, when the \(N\) (number of test assets) over \(T\) (number of time periods) ratio is relatively small (i.e., \(\leq 10\%\)), our method consistently generates more accurate parameter estimates compared to the Fama-MacBeth regression or alternative two-pass regression techniques. The superior performance of our new estimator is manifested through both narrower confidence intervals and more

---

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

\(^2\)Our improvement over alternative two-pass estimators, for example, the estimator in Shanken (1992), is also sizable. See Table 1 for details.
powerful tests for hypotheses on the unknown risk premiums. For a larger \( N/T \), the uncertainty resulting from estimating high dimensional matrices such as the covariance matrix starts kicking in, leading to a different asymptotic theory for the two-pass regression estimator and the GMM estimator. In this case, our estimator is not optimal, and becomes less efficient compared to more robust alternatives (e.g., OLS) in our simulation studies. We quantify the efficiency gain/loss and establish the rule of thumb over the choice of \( N/T \) through a variety of asset pricing applications.

Our empirical analysis studies recent factor models proposed in Fama and French (FF, 2015).\(^3\) While FF use time series spanning regressions to evaluate factor model performance, we leverage our new two-pass estimator to develop a cross-sectional counterpart of the usual time-series alpha estimate. We offer new insights into the empirical performance of FF using cross-sectional alpha estimates.\(^4\)

We first show the FF model successfully passes the specification test, suggesting a good overall performance of the model in explaining its own basis assets (i.e., portfolios that are used to construct the FF model). We then contrast the performance between the usual time-series alpha estimates and our new cross-sectional estimates for a large number of anomaly portfolios. While time-series estimates are clearly positively correlated with their cross-sectional counterparts, large discrepancies exist for some anomalies. We also find noticeable differences between the cross-sectional risk premium estimates and the time-series factor means, suggesting the unique information the cross-sectional approach provides. Lastly, focusing on the comparison between our newly proposed estimator and existing two-pass estimators, we highlight the power of our estimator in detecting portfolios with non-zero alphas. Using a large set of anomaly portfolios and testing the hypothesis of a zero abnormal alpha, our model leads to a sizable 10% more rejections than leading alternative models.\(^5\)

One goal of our paper is to synchronize the classical GMM paradigm with current asset pricing research. Despite its strong theoretical appeal, GMM has fallen out of favor in current asset pricing research (asset pricing tests, in particular), perhaps due to the computational challenges in implementing the GMM approach (Shanken and Zhou (2007)).\(^6\) Since we want to maintain the

---

\(^3\)An earlier version of our paper also examines the Hou, Xue, and Zhang (HXZ, 2015) model. The related empirical results are available upon request.

\(^4\)While both Fama and French (2015) and Hou, Xue, and Zhang (2015) likely provide a reduced-form representation and thus an incomplete description of the cross-section of asset returns (Kozak, Nagel, and Santosh (2018) and Kozak, Nagel, and Santosh (2020)), researchers routinely compare factor models, especially those that have similar economic motivations. We provide a new approach to aid such comparisons.

\(^5\)Our results are based only on the Fama-French five-factor model because it survives the joint specification tests.

\(^6\)To appreciate the computational burden of the GMM approach in estimating a beta-pricing model, suppose we have 50 test assets and the benchmark model is a Fama-French five-factor model. This implies 250 (= 50 \( \times \) 5) betas and five risk-premium parameters to be estimated. Moreover, because the betas and the risk-premium parameters are intertwined in the GMM objective function, the optimization of the GMM objective function involves a nonlinear
theoretical appeal of GMM while also circumventing its computational burden, we propose a new two-pass approach that is highly tractable, and, at the same time, approaching GMM’s efficiency asymptotically. Because our two-pass estimator is akin to the optimally weighted GMM in beta-pricing models, it preserves three important properties of GMM that likely make it more attractive for empirical applications than alternative two-pass methods: 1. It is robust to nonnormality and conditional heteroskedasticity in asset returns (e.g., Bollerslev, Engle, and Wooldridge (1988), Schwert (1989), Schwert and Seguin (1990), and Herskovic, Kelly, Lustig, and Van Nieuwerburgh (2016)); 2. It controls for potential nonlinear dependence between return residuals and factor realizations (which is featured prominently in models that study co-skew or co-kurtosis risks; e.g., Harvey and Siddique (2000), Dittmar (2002), and Schneider, Wagner, and Zechn (2020)); and 3. It allows for serial correlations in factor returns, asset returns, and their (possibly nonlinear) combinations (Ehsani and Linnainmaa (2019)).

Our linear-beta pricing setup resembles those in MacKinlay and Richardson (1991), Shanken (1992), Jagannathan and Wang (1998). Shanken (1992) proposes a generalized least-square (GLS) two-pass estimator that improves the Fama-MacBeth approach under his assumptions. Jagannathan and Wang (1998) present the asymptotic result for a two-pass estimator with a general second-stage weighting matrix. MacKinlay and Richardson (1991) point out the potential efficiency gain from a GMM estimation of the linear-beta pricing model. Our paper reconciles these three papers in the following sense. Advancing Jagannathan and Wang (1998), we pin down the optimal second-stage weighting matrix among all choices of the weighting matrix. We show our estimator under this optimal weighting matrix collapses into Shanken’s GLS under his assumptions, but outperforms Shanken’s GLS for a general data-generating process (i.e., when his assumptions are violated). Lastly, relating to MacKinlay and Richardson (1991), we show the potential efficiency gain of the GMM implementation is fully preserved by our two-pass estimator—we prove the asymptotic equivalence between GMM and our approach.

Our paper is also related to a recent growing literature on large \( N \) (i.e., number of assets in the cross-section) inference on asset pricing models, including Gagliardini, Ossola, and Scaillet (2016), Fan, Liao, and Wang (2016), Jegadeesh, Noh, Pukhuanthong, Roll, and Wang (2019), Kim and Skoulakis (2018), Feng, Giglio, and Xiu (2020), Lettau and Pelger (2020a), Lettau and Pelger (2020b), Raponi, Robotti, and Zaffaroni (2020), Harvey and Liu (2020b), Freyberger, Neuhierl, search over a high-dimensional parameter space. As such, GMM tends to have poor finite sample performances, as documented in Shanken and Zhou (2007).

Alternative GMM-type of estimators that may also achieve GMM’s efficiency asymptotically have been proposed by, e.g., Frazier and Renault (2017). However, these estimators are not computationally as attractive as ours because they do not permit a closed-form expression. Our estimator is based on the popular two-pass implementation and is hence highly tractable analytically.
and Weber (2020), and Kim, Korajczyk, and Neuhierl (2021). In contrast to these papers, we focus on estimation efficiency for the usual fixed $N$ and large $T$ linear-beta pricing setup where the concern is over risk premium estimates for a pre-determined set of factor candidates. Despite its long tradition, this setup is still the workhorse model to estimate the impact of empirical or macro-economic factors in driving asset returns (e.g., see Croce, Marchuk, and Schlag (2019), Lin, Palazzo, and Yang (2020), and Ai, Li, Li, and Schlag (2020) for recent applications). As such, we do not attempt to explore nonlinearity (Freyberger, Neuhierl, and Weber (2020) and Kim, Korajczyk, and Neuhierl (2021)), the construction of factors themselves based on factor analysis (Lettau and Pelger (2020a) and Lettau and Pelger (2020b)), model selection for factors (Feng, Giglio, and Xiu (2020) and Harvey and Liu (2020b)), or other performance metrics such as the ex-post consistency for a diverging $N$ (Gagliardini, Ossola, and Scaillet (2016) and Raponi, Robotti, and Zaffaroni (2020)). Our research complements these papers by calling attention to estimation efficiency that influences asset pricing tests.

We also provide extensions of our model that handle the case when both $N$ and $T$ are diverging with $N/T \to 0$ and missing observations. The condition $N/T \to 0$ in this double asymptotic framework is consistent with our simulation results, which show a relatively small $N/T$ is needed to guarantee the efficiency gain offered by our approach. We therefore caution against the use of our approach to a very large $N$, especially without the theoretical investigation and the support of a thorough simulation study. In fact, in the extreme case of a divergent $N$ and a fixed (and small) $T$ (which is opposite to our model assumption), some existing estimators can achieve $N^{-1/2}$-consistent estimation (i.e., consistently estimating the ex post risk premium with a large $N$). Potential users may want to consider these alternative methods instead.

The remainder of this paper is organized as follows. In Section 2, we lay out the econometric framework and present our main theoretical results. In Section 3, we conduct a comprehensive simulation study that compares alternative two-pass regression estimators and the GMM estimator. In Section 4, we revisit FF using our cross-sectional approach. We offer concluding remarks in the final section. The appendix contains proofs for our main theoretical results and their extensions as well as additional simulation results. Additional literature review and extra simulation results are available in an On-line Supplemental Appendix.

---

8 Also see Bai and Zhou (2015) for a large-N extension of the usual Fama-MacBeth approach.
9 We do show that when the expected return-beta relation is non-linear (see Section 2.2.3), our equivalence result between our estimator and the optimally weighted GMM still holds. This result provides the basis for extending our theory to potentially non-linear beta-pricing models.
2 Theory

2.1 Optimal Cross-Sectional Regression Estimator

We present our optimal cross-sectional regression estimator in several steps. In Section 2.1.1, we introduce the moment conditions that characterize the standard linear-beta pricing model. In Section 2.1.2, we discuss the popular two-pass implementations and derive their asymptotic distributions, complementing existing research. In Section 2.1.3, we establish the optimality of one particular two-pass estimator among all two-pass regressors. Finally, in Section 2.1.4, we show our main results, which build the equivalence between the new two-pass estimator and the theoretically optimal GMM estimator.

2.1.1 Moment Restrictions Corresponding to a Linear Beta-Pricing Model

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

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

\begin{align*}
E[f_t - \mu_f] &= 0, \\
E \left[ (R_{i,t} - (f_t - \mu_f)' \beta_i)(f_t - \mu_f) \right] &= 0, \\
E[R_{i,t} - \alpha_i - \gamma_0 - \gamma_1' \beta_i - \gamma_2' Z_{i,t-1}] &= 0,
\end{align*}

for \( i \in \{1, \ldots, N\} \), where \( R_{i,t} \) denotes the return of asset \( i \) at period \( t \), \( f_t \) is a vector of \( K \) many pricing factors, and \( Z_{i,t-1} \) is a vector of \( M \) many (possibly time-varying) predetermined security characteristics.\(^{10}\) 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 that a known

\(^{10}\)Note that 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), Gospodinov, Kan, and Robotti (2014), Gospodinov, Kan, and Robotti (2017), and Gospodinov, Kan, and Robotti (2019). We also take factors as given and do not explicitly consider the issue of omitted risk factors, see, e.g., Forni, Hallin, Lippi, and Zaffaroni (2015), Gagliardini, Ossola,
subset of assets indexed by $\mathcal{I}_0 \subset \{1, \ldots, N\}$ exists, such that

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

Without loss of generality, let $\mathcal{I}_0 \equiv \{1, \ldots, N_0\}$ and $\mathcal{I}_1 \equiv \{N_0 + 1, \ldots, N\}$, where $N_0 \geq d_\gamma$ and $d_\gamma \equiv K + M + 1$ is the dimension of $\gamma$. The key unknown parameter of interest is denoted as $\theta \equiv (\alpha_{N_0+1}, \ldots, \alpha_N, \gamma')'$, which is identified through (2.3) and (2.4) given the identification of $\mu_f$ and $\beta_i$.

Several remarks on our model are worth making. First, our econometric specification fully captures the risk and return relation in a linear beta-pricing model. In fact, by shutting down model misspecification (i.e., setting $(\alpha_{N_0+1}, \ldots, \alpha_N) = 0_{1 \times (N-N_0)}$) and firm characteristics $(Z_{t,t-1})$, our moment restrictions correspond exactly to the GMM specifications in MacKinlay and Richardson (1991) and Jagannathan, Skoulakis, and Wang (2010) and are closely related to the moment restrictions in Jagannathan and Wang (2002) and the model in Shanken (1992).\(^{13}\) Second, following most studies on two-pass estimators (see an extensive discussion in Jagannathan, Schaumburg, and Zhou (2010)), our specification does not impose the condition that risk premiums must equal factor means for traded factors (i.e., $\gamma_1 = \mu_f$). As a result, we allow for measurement errors for traded factors.\(^{14}\) 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)).\(^{15}\)

### 2.1.2 Two-Pass Implementations

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

$$\hat{\mathbf{B}} \equiv \left( \sum_{t \leq T} (\mathbf{R}_t - \bar{\mathbf{R}}) f_t' \right) \left( \sum_{t \leq T} (f_t - \bar{f})(f_t - \bar{f})' \right)^{-1}, \quad (2.5)$$


\(^{13}\)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.


\(^{15}\)See Balduzzi and Robotti (2008) for another approach of inferring the risk premiums that relies on the maximum-correlation portfolios.
where $\mathbf{R}_t \equiv (R_{1,t}, \ldots, R_{N,t})'$, $\mathbf{R} \equiv T^{-1} \sum_{t \leq T} \mathbf{R}_t$, and $\bar{f} \equiv T^{-1} \sum_{t \leq T} f_t$.

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

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

where $\hat{\mathbf{W}}$ is an arbitrary $N \times N$ real symmetric positive definite matrix, $\hat{\mathbf{X}}_t \equiv (S_N, 1_{N \times 1}, \hat{\mathbf{B}}, \mathbf{Z}_{t-1})'$, $S_N \equiv (0_{N_1 \times N_0}, I_{N_1})'$, $1_{N \times 1}$ is a column vector of ones, and $N_1 \equiv N - N_0$.\(^{16}\)

Solving the least-square problem, we have

$$\hat{\theta}_{\text{CSR}} = (\hat{\mathbf{X}}' \hat{\mathbf{W}} \hat{\mathbf{X}})^{-1} \hat{\mathbf{X}}' \hat{\mathbf{W}} \mathbf{R}, \quad (2.6)$$

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

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

$$v_{i,t} \equiv R_{i,t} - \alpha_i - \gamma_0 - \gamma_1' \beta_i - \gamma_2' \mathbf{Z}_{i,t-1}, \quad (2.7)$$

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

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

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

We next provide the asymptotic distribution of the CSR estimator. Because betas are estimated, we face the “errors-in-variables” (EIV) problem. As a result, the key to deriving the asymptotic distribution is to factor in estimation errors in betas. Instead of viewing beta esti-

\(^{16}\)Throughout the paper, we use $I_d$ to denote the $d \times d$ identity matrix; we use $1_{d_1 \times d_2}$ and $0_{d_1 \times d_2}$ to denote the $d_1 \times d_2$ matrices of 1’s and 0’s, respectively.
mation errors as errors-in-variables that need to be adjusted in the second stage regression, our derivation (in Appendix A) highlights the new perspective we take on estimation errors: we show that beta estimation errors effectively introduce a new source of return innovation in the second pass regression. In particular, we first show that beta estimation uncertainty translates into a return innovation of \(-u_t(f_t - \mu_f)^\gamma_1\) at each point in time for the second stage regression. Combining this return innovation with the original total return innovation \(v_t\) that affects the regression estimate (without the errors-in-variables issue), we show the CSR estimator is essentially driven by the following return innovation:

\[
e_t \equiv v_t - u_t(f_t - \mu_f)^\gamma_1 = u_t(1 - (f_t - \mu_f)^\gamma_1) + B(f_t - \mu_f) - (Z_{t-1} - E[Z_{t-1}])\gamma_2,
\]

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

The key conditions for deriving the asymptotic distribution of the CSR estimators are presented in the following assumption.

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

Assumption 1 constitutes our main assumption. Assumption 1(i) is a central limit theorem on the partial sum \(T^{-1/2} \sum_{t \leq T} \epsilon_t\), which can be verified under some 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 \(E[u_t f_f'] = 0_{N \times K}\). Therefore, Assumption 1(ii) holds if the variance-covariance matrix of \(T^{-1/2} \sum_{t=1}^T u_t(1, f_f')\) is bounded. Assumption 1(iii) imposes conditions on the weight matrix of the CSR estimator. The eigenvalue conditions in Assumption 1(iv) ensure that the CSR estimator is \(T^{-1/2}\)-consistent.

Among the conditions in Assumption 1, note we do not make any strong assumptions on idiosyncratic risks (i.e., \(u_t\)). As such, our theory allows for potentially non-normally distributed, serially dependent, and conditionally heteroskedastic idiosyncratic risks. In contrast, Shanken (1992)’s Assumption 2 (to be presented later) makes much stronger assumptions about idiosyncratic risks. Overall, conditions in Assumption 1 are rather mild and are likely satisfied in most asset pricing applications.

The following lemma provides the asymptotic distribution of the CSR estimator.
Lemma 1. Under Assumption 1 and Assumption 3 in Appendix A, we have

\[ T^{1/2}(\hat{\theta}_{csr} - \theta) \to_d N(0, \text{Asv}(\hat{\theta}_{csr})) \]  

(2.8)

where \( \text{Asv}(\hat{\theta}_{csr}) \equiv (X'WX)^{-1}(X'\Omega WX)(X'WX)^{-1} \).

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

2.1.3 Optimal Cross-Sectional Regression (OCSR) Estimator

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

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

\[ \text{Asv}(\hat{\theta}_{csr}^*) = (X'\Omega^{-1}X)^{-1} \]  

(2.9)

and

\[ \text{Asv}(\hat{\theta}_{csr}) \geq \text{Asv}(\hat{\theta}_{csr}^*) \]  

(2.10)

where \( \text{Asv}(\hat{\theta}_{csr}) \) is defined in Lemma 1.\(^{17}\)

Proposition 1 together with Lemma 1 implies that the optimal CSR (i.e., OCSR) estimator \( \hat{\theta}_{csr}^* \) can be constructed using the inverse of a consistent estimator of \( \Omega \) as the weight matrix.\(^{18}\) Because \( \Omega \) is the asymptotic variance of \( T^{-1/2} \sum_{t=1}^{T} \epsilon_t \), a consistent estimator of \( \Omega \) can be constructed using consistent estimators of \( \epsilon_t \) and the sample analog of \( \Omega \) when \( \{\epsilon_t\} \) is a martingale difference sequence. When \( \{\epsilon_t\} \) is from a weakly dependent process (see, e.g., Newey and West (1987) and Andrews (1991)), a consistent estimator of \( \Omega \) can be constructed using the heteroskedasticity and auto-correlation consistent (HAC) estimator.

\(^{17}\)Throughout this paper, for any two real matrices \( A \) and \( B \) of the same dimensions, \( A \geq B \) means that \( A - B \) is positive semi-definite.

\(^{18}\)The optimal choice of the weight matrix in the CSR estimation has also been discussed in Kan and Robotti (2012). Proposition 1 extends their results by allowing for non-beta related explanatory variables (e.g., characteristics) and asset-specific mispricing. More importantly, we further show in the next section the optimal CSR estimator is not only optimal among two-pass estimators but also achieves the GMM efficiency bound, thus establishing its optimality across a larger class of estimators.
Similar to Shanken (1992)’s GLS estimator, the implementation of our OCSR estimator $\hat{\theta}_{csr}^*$ requires the estimation of an $N \times N$ weight matrix, which becomes inaccurate when $N/T$ is large. This raises the question of what level of $N$ is suitable for implementing our estimator versus alternative estimators that are potentially less demanding in terms of the estimation of the weight matrix, for example, the WLS estimator.\textsuperscript{19}

While our theory is developed under the assumption of a fixed (or slowly divergent) $N$, the level of $N$ considered to be large is more of an empirical than theoretical question.\textsuperscript{20} We answer this question through a comprehensive simulation study in the next section. Our evidence shows that when $N$ ranges from small (i.e., $N = 10$) to modest (i.e., $N = 50$), the efficiency gain of our OCSR estimator compared to alternative estimators, especially the WLS estimator, 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 downweight the impact of these assets, leading to a more efficient estimation.

For the case with divergent $N$, Section C.3 of the Appendix adapts our theory to the case of both diverging $N$ and $T$. While theoretically appealing, we caution against a naive application of this result in the absence of a carefully conducted simulation study. The reason is that, based on the sufficient conditions maintained in this double asymptotic framework (see, Assumption 5 and the related discussion in Section C.3), and our extensive simulation results, a small to modest $N/T$ ratio is preferred to realize the efficiency gain of our approach. As such, when $N$ diverges, we have to let $T$ diverge much faster than $N$ for our extension in Section C.3 to be empirically relevant. However, for asset pricing applications of our model, we usually do not have the luxury of having a very large $T$. Therefore, to avoid confusing readers with the requirement over $N/T$, we choose to focus on the fixed $N$ case to present our theory. We leave the potential application of our double asymptotic results in Section C.3 to future exploration.

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

\textsuperscript{19}The WLS estimator is a naive version of the GLS estimator that sets the off-diagonal elements in the GLS weighting matrix to zero; see more detailed description in Section 3.

\textsuperscript{20}See Section C.3 of the Appendix for extension of the theory to the case with divergent $N$.  

12
We investigate this issue in the next section.

### 2.1.4 Relating OCSR to GMM

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

Let $Y_t \equiv (f_t', R_t', Z_{t-1}')'$ denote the observation at period $t$. Using the restrictions in $(2.4)$, the moment functions in $(2.1)$, $(2.2)$, and $(2.3)$ can be written as

$$
g_1(Y_t, \phi) \equiv f_t - \mu_f,
$$

$$
g_2(Y_t, \phi) \equiv \left( R_t - B(f_t - \mu_f) \right) \otimes (f_t - \mu_f),
$$

$$
g_3(Y_t, \phi) \equiv R_t - X_t \theta,
$$

where $\otimes$ denotes the Kronecker product, $X_t \equiv (S_N, 1_{N \times 1}, B, Z_{t-1})$ and $\phi \equiv (\theta', \delta')'$ is the stacked parameter vector, with $\delta \equiv (\mu_1', \beta_1', \ldots, \beta_N')$. The unknown parameter $\phi$ can be estimated through the following optimal GMM procedure:

$$
\hat{\phi}_{gmm}^* \equiv \min_{\phi} \bar{g}(\phi)'(\hat{\Sigma}_g)^{-1}\bar{g}(\phi),
$$

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

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

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

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

(2.13)
among all GMM estimators of $\theta$.

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

**Proposition 2.** Suppose that $\Sigma_f$, $\Sigma^*_g$, $W$ and $X'X$ are nonsingular. Then

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

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

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

Proposition 2 is a profound result for at least two reasons. First, conceptually, a simultaneous GMM estimation appears to use more information provided by the moment conditions in (2.1), (2.2) and (2.3). For instance, we learn about betas not only from the time-series restrictions (i.e., (2.2)) but also from the linear-beta restriction (i.e., (2.3)). As a result, one would think that the GMM estimates of betas should be more efficient than the usual OLS estimates and subsequently conjecture that this efficiency gain in estimating betas would translate into efficiency gains in estimating risk premiums. We show, somewhat counterintuitively, that this conjecture is incorrect. We discuss in detail the intuition behind our equivalence result in Section 2.2.3. In short, Proposition 2 bridges the gap between two literatures on estimating beta-pricing models: One is the GMM approach (e.g., MacKinlay and Richardson (1991), Zhou (1994), and Jagannathan and Wang (2002)), and the other is the popular two-pass regression approach (e.g., Shanken (1992), Gagliardini, Ossola, and Scaillet (2016), Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2019), Kim and Skoulakis (2018), and Raponi, Robotti, and Zaffaroni (2020)).

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

2.2 Discussion

We provide an extensive discussion of our results in the context of the current literature. In Section
2.2.1, we specifically relate our estimator to the alternative two-pass estimator provided in Shanken
(1992).21 In Section 2.2.2, we offer the economic intuition for the efficiency gain of our estimator.
In Section 2.2.3, we discuss in detail what features of the linear-beta pricing model render our
equivalence result and possible generalizations of this result.

2.2.1 Comparing OCSR and Shanken (1992)’s GLS

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

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

Assumption 2. \( u_t \) is i.i.d. with \( E[u_t|F_T] = 0 \) and \( E[u_t'u'_t|F_T] = \Sigma_u \), where \( F_T = (f_1, \ldots, f_T) \) and
\( \Sigma_u \) is a non-random positive definite matrix.

Under Assumption 2, \( u_t(1 - (f_t - \mu_f)\Sigma_f^{-1}\gamma_1) \) forms a martingale difference array that is un-
correlated with \( f_{t'} - \mu_f \) for any \( t \) and \( t' \). Therefore, the variance-covariance matrix \( \Omega \) takes the

---

21Given the large body of literature on addressing the EIV biases in asset pricing tests, we review the relevant
literature in the On-line Supplemental Appendix where we also relate our paper to the recent development of asset
pricing tests in general.
following simplified form,

$$\Omega = \Sigma_u (1 + \gamma'_1 \Sigma_f \gamma_1) + B \Sigma_f B' = \Sigma_u (1 + \gamma'_1 \Sigma_f \gamma_1) + X \Sigma^*_f X',$$  \hspace{1cm} (2.15)$$

where $$\Sigma^*_f \equiv \text{diag}(0, \Sigma_f)$$ and $$\Sigma_f$$ denotes the long-run variance of $$T^{-1/2} \sum_{t=1}^{T} f_t.$$ \hspace{2em} (2.16)

Moreover, by Lemma 1, the asymptotic variance-covariance matrix of the CSR estimator becomes

$$\text{Asv}(\hat{\theta}_{csr}) = (1 + \gamma'_1 \Sigma_f \gamma_1) (X' W X)^{-1} (X' W \Sigma_u W X) (X' W X)^{-1} + \Sigma^*_f. \hspace{1cm} (2.17)$$

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.

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

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

which together with (2.16) and (2.17) implies

$$\text{Asv}(\hat{\theta}_{csr}) \geq \text{Asv}(\hat{\theta}^*_{gls}). \hspace{1cm} (2.18)$$

This shows the optimality of the GLS estimator among all CSR estimators under the assumptions in Assumption 2.

Because the inequality (2.14) holds without Assumption 2, we have $$\text{Asv}(\hat{\theta}^*_{gls}) \geq \text{Asv}(\hat{\theta}^*_{gmm})$$ in general, which, together with the asymptotic equivalence between $$\hat{\theta}^*_{gmm}$$ and $$\hat{\theta}^*_{csr}$$ established in Proposition 2, shows that the OCSR estimator $$\hat{\theta}^*_{csr}$$ is more efficient than the GLS estimator $$\hat{\theta}^*_{gls}$$ in general. However, when Assumption 2 holds, following Proposition 2 and (2.18), we obtain

$$\text{Asv}(\hat{\theta}^*_{csr}) = \text{Asv}(\hat{\theta}^*_{gls}),$$

which, together with the parametric efficiency of the GLS estimator, implies that the OCSR estimator $$\hat{\theta}^*_{csr}$$ is also parametrically efficient if the joint density of $$R_t$$ given $$F_T$$ is normal. When Assumption 2 fails, OCSR can be more efficient than the GLS estimator. We

\hspace{2em}^{22}\text{Throughout this paper, we use } \text{diag}(A, B) \text{ to denote the block diagonal matrix with the square matrices } A \text{ and } B \text{ on the main diagonal.}
discuss important scenarios in which Assumption 2 fails and how OCSR improves on the GLS estimator in the next section.

2.2.2 Interpreting the Efficiency Gain of OCSR

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

The estimation error of $\hat{\theta}_{csr}$ is mainly governed by $T^{-1/2} \sum_{t=1}^{T} \left( v_t - u_t (f_t - \mu_f)'\Sigma_f^{-1} \gamma_1 \right)$, which can be decomposed into two parts (see Appendix A): One part is related to $v_t$, 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. By (2.7),

$$T^{-1/2} \sum_{t=1}^{T} \left( v_t - u_t (f_t - \mu_f)'\Sigma_f^{-1} \gamma_1 \right) = T^{-1/2} \sum_{t=1}^{T} \left( u_t (1 - (f_t - \mu_f)'\Sigma_f^{-1} \gamma_1) \right) \text{Efficiency-relevant innovations} + \left( \hat{B} (f_t - \mu_f) \right) \text{Efficiency-irrelevant innovations}, \quad (2.19)$$

which further decomposes the total innovation $v_t$ into the idiosyncratic innovation $u_t$ and the systematic innovation $B(f_t - \mu_f)$.

We first interpret our results under Assumption 2. We interpret the two terms in Eq. (2.19) as “efficiency-relevant innovations” and “efficiency-irrelevant innovations” (under Assumption 2), for reasons as follows. Idiosyncratic innovations $u_t$ 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 weight matrix, and optimal efficiency is achieved at the variance-covariance matrix of $u_t$ (hence, the name “efficiency-relevant innovations”).

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

Combining the above two observations, efficiency-irrelevant innovations (i.e., $B(f_t - \mu_f)$) can be isolated because Assumption 2 allows us to decouple efficiency-relevant innovations and efficiency-irrelevant innovations (more specifically, the two innovations are uncorrelated). Estimation efficiency only depends on efficiency-relevant innovations (i.e., $u_t(1 - (f_t - \mu_f)'\Sigma_f^{-1}\gamma_1)$) and is achieved when the weight matrix is set at $\Sigma_u^{-1}$. This explains the efficiency result in Shanken (1992).

Unlike the GLS estimator in Shanken (1992), OCSR can offer additional efficiency gain when our assumptions (in particular, Assumption 1) generalize Assumption 2. Several prominent features of financial market data may make our generalizations important. First, the assumption of conditional homoskedasticity in Assumption 2 may not be suitable when asset returns display non-normality and conditional heteroskedasticity (Jagannathan and Wang (1998), and 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 Schneider, Wagner, and Zechner (2020); co-kurtosis risk in Dittmar (2002)) imply a non-zero correlation between market-model residuals and second (or third) powers of market returns for most assets. For these models, efficiency-relevant innovations and efficiency-irrelevant innovations cannot be uncorrelated, leading to a potential efficiency gain of our estimator compared with Shanken (1992) for the inference of the market risk premium.

Finally, our estimator adjusts for serial correlations in $u_t$, $f_t$, and, more importantly, their combinations. Whereas Shanken (1992)’s GLS can be extended to cope with serial correlations in $u_t$, we show that adjusting for serial correlations in the combined innovations (i.e., the sum of efficiency-relevant and efficient-irrelevant innovations) is crucial to achieve the asymptotic optimal-

---


24 Another data feature that our OCSR guards against is the nonlinear dependence between $u_t$ and $f_t$. 

18
ity of the optimal GMM estimator. Such an adjustment may substantially improve the estimation accuracy of a CSR estimator given the widespread time-series patterns in anomaly returns (e.g., Elsani and Linnainmaa (2019)).

GMM is able to achieve semiparametric optimality that guards against all of the above issues (i.e., conditional heteroskedasticity, non-zero dependence between $u_t$ and higher powers of $f_t$, and serial correlations). Our new insight shows that a CSR estimator with a properly defined weight matrix for the second-stage cross-sectional regression inherits the theoretical appeal of the GMM estimator and is thus also robust to the above issues. Unlike Shanken (1992)’s results that allow the decoupling of efficiency-relevant and efficiency-irrelevant innovations in constructing the optimal weight matrix, we show that mixing these two sources of innovations is essential to achieve GMM’s efficiency when Assumption 2 fails to hold.

### 2.2.3 Linear-Beta Pricing and Equivalence between OCSR and GMM

Our main result on the equivalence between OCSR and GMM deserves further discussion. First, what structure of the linear-beta relation renders this equivalence? Intuitively, one can imagine an iterative GMM interpretation of the system of moment restrictions where $\hat{\mu}_f$ and $\hat{B}$ 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) depend on $\theta$ and do 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. In Appendix B, we provide a counterexample that illustrates the efficiency loss of the (optimally weighted) iterative GMM estimator with over-identified nuisance parameters.\(^{25}\) This counter example also shows that the linearity of the moment conditions plays

\(^{25}\)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 is similar in spirit to our efficiency result, they have different implications. Lemma 1 in Kan and Zhou (2001), when applied to our setup, implies that the nuisance parameters $\mu_f$ 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.
no role in the efficiency result of the OCSR estimator.

The above insight allows us to easily generalize the implication of our results to other settings. For example, suppose we augment the three sets of moment restrictions with another set (called $S1$) that imposes additional restrictions on betas (e.g., augmenting Eq. (2.2) by assuming that idiosyncratic risks are orthogonal not only to contemporaneous factor returns but also to other lagged instruments). Then our equivalence result will not hold. More specifically, the two-pass implementation that estimates $\mu_f$ and $B$ through (2.1), (2.2), and ($S1$) first and then estimates $\gamma$ 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, for example, the specification in Fama and MacBeth (1973), where the risk premium is linear in squared market betas.

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

Within the second line of research, Shanken (1992) establishes the efficiency of the two-pass GLS estimator under Assumption 2 and joint normality of $R_t$ given $F_T$. Jagannathan and Wang (1998) develop the distribution theory for a given two-pass estimator under general assumptions on the return generating process. However, no attempt has been made so far to relate a two-pass estimator to the optimal GMM estimator, which is known to be semiparametrically efficient. Our paper addresses this gap in the literature.

### 3 A Simulation Study

One of our goals in this paper is to provide a comprehensive simulation study to systematically compare existing methods. To this end, we conduct two sets of simulation excises. In the first exercise, we examine popular factor models and test assets to have a first look of our model’s performance, as well as to illustrate our simulation design. This also corresponds to our actual
empirical analysis in the next section. In the second exercise, we systematically evaluate our model’s performance by examining a large collection of test assets and extracting factors from a factor analysis.

3.1 FF Model

Our simulation design is marked by several features. First, while we mainly focus on popular two-pass estimators that empirical researchers frequently use, we also include the computationally intensive optimally-weighted GMM approach as a performance benchmark. The optimal GMM estimator does not have an explicit expression and has to be solved using numerical methods. Therefore, its finite sample performance may be poor due to the numerical optimization errors and/or the instability caused by inverting a large dimensional variance-covariance matrix that depends on all model parameters (see related comments in Shanken and Zhou (2007)),

Second, to ensure that our simulated data-generating process (DGP) resembles the actual application, we construct bootstrap samples from the actual data of factor returns and test assets. For example, in our first simulation exercise, we consider the Fama-French three-factor and five-factor models as candidate models, and as candidate samples for test assets, we consider the 18 low-turnover anomaly sample and the 38 low-turnover and medium-turnover combined anomaly sample in Novy-Marx and Velikov (2016). Our actual sample runs from July 1973 to December 2017, including 534 monthly observations. Third, our bootstrap samples keep the same level of time-series and cross-sectional dependency in the actual data, such that our results shed light on the performance of our approach for the actual data.

Note that we focus on popular factor models in our simulation study because they correspond to our empirical application where we examine FF. Given that factors usually have a limited degree 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).

Note although our OCSR also depends on the weighting matrix, both betas and risk premiums are pre estimated and therefore fixed in constructing the weighting matrix.

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.

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 that existing two-pass estimators such as Shanken (1992) do not take time-series correlations in factor model residuals into account. Alternative approaches that infer risk premiums through latent risk factors include

27Note although our OCSR also depends on the weighting matrix, both betas and risk premiums are pre estimated and therefore fixed in constructing the weighting matrix.
28We 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.
29The 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.
30Note that existing two-pass estimators such as Shanken (1992) do not take time-series correlations in factor model residuals into account. Alternative approaches that infer risk premiums through latent risk factors include
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 \( m \) (\( m = 1, \ldots, M \), 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 \( (F_m, 3 \times T_b, \text{ which resamples from } F_p) \) and the bootstrap factor-model residuals \( (RES_m, N \times T_b, \text{ which resamples from } RES_p) \), 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

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

where \( R_m \) is a \( N \times T_b \) matrix and \( \overline{f}_p \) denotes the (original) sample mean of \( f_t \).\(^{31}\)

To investigate the finite-sample properties of various CSR estimators, we let \( \gamma_{p,0} = 0 \) and \( \gamma_p = \overline{f}_p \) to ensure the linear-beta pricing holds. This parametrization is slightly revised when we examine the performance of inference (i.e., hypothesis tests) based on different CSR estimators. Specifically, we reparametrize an individual component of \( \gamma_p = (\gamma_{p,0}, \gamma_{p,1})' \) while keeping the remaining components unchanged. For example, when we consider the two-sided test on \( H_0 : \gamma_{p,0} = 0 \) while \( \gamma_{p,0} = 0 \), we consider four different values (0, 0.1%, 0.2%, and 0.3% per month) for \( \gamma_{p,0} \) while

\(^{31}\)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.
keeping $\gamma_{p,1} = \bar{f}_p$. 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 non zero 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 $\mathbb{E}[u_t' u_t']$, is used as the weight matrix. For WLS, the off-diagonal elements of the GLS weight matrix are set to zero.

For our OCSR, we need to estimate the long-run variance $\Omega$ as given in Lemma 1. 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. Finally, given the general weak time-series dependency in financial returns, we simply set the truncation parameter at three (months) to calculate long-run variance (e.g., Wooldridge (2016), Lazarus, Lewis, Stock, and Watson (2018)).

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

\[32\] 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 that our anomaly sample runs from July 1973 to December 2017, including 534 monthly observations, so \( T = 500 \) is close to the sample size of the actual data. \( T = 750 \) is considered a “large” sample experiment, where we increase \( T = 500 \) by 50%.

Focusing on \( T = 500 \) in Table 1, simulated bias for our OCSR is usually less than 10% of the magnitude of the true parameter value (with the exception of \( \gamma_0 \), 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.\(^{33}\) 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.

GMM, which is asymptotically equivalent to OCSR, fares much worse in estimation efficiency when compared to OCSR. This highlights the attractiveness of our analytically tractable approach over the computationally demanding GMM over finite samples. In particular, while GMM occasionally performs better than other methods for some specific parameters (e.g., the estimation for \( \gamma_0 \) in Panel A), it generally underperforms OCSR by a wide margin. For instance, GMM’s RMSE for \( \gamma_{hml} \) is 0.239, which is substantially higher than all other methods, including the naive OLS.\(^{34}\)

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

---

\(^{33}\)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).

\(^{34}\)The GMM estimator is solved using the \textit{fminsearch} function with the simplex search algorithm in Matlab, which is one of the most popular methods for solving nonlinear optimization problems. Possible reasons for the poor finite sample performance of the optimal GMM estimator include the optimization error of this numerical method and instability of inverting high-dimensional matrix to obtain the optimal weight matrix in GMM. The finite sample properties of the optimal GMM estimator could be improved with a more accurate numerical method. However, this is not the goal of this paper.
case dependent. For example, for the Fama-French three-factor model with 38 test portfolios, as in Panel B, the percentage reduction in RMSE of OCSR relative to GLS ranges from 6% for $\gamma_{mkt}$ (i.e., $-6\% = (0.215 - 0.228)/0.228$) to 19% for $\gamma_{smb}$ (i.e., $-19\% = (0.118 - 0.146)/0.146$). The average reduction is 11%.

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 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 $N$.

Next, we study the size and power of the hypothesis test on $H_0: \gamma_{p,j} = 0$, where $j = 0, mkt, smb,$ or $hml$ using different CSR estimators. To perform hypothesis testing, we need to be specific about how standard errors are calculated. We use $(\hat{X}'\hat{W}'\hat{X})^{-1}(\hat{X}'\hat{W}'\hat{\Omega}\hat{W}\hat{X})(\hat{X}'\hat{W}\hat{X})^{-1}$ to approximate the asymptotic variance-covariance matrix for GLS and WLS, where $\hat{W}$ is set to be the aforementioned GLS and WLS weight matrix, respectively, and $\hat{\Omega}$ follows the long-run variance estimator as explained previously. For OCSR, the asymptotic covariance matrix is given by $(X'\hat{\Omega}^{-1}X)^{-1}$, 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, which does not take beta estimation into account and simply uses OLS standard errors from the cross-sectional regression. We denote this approach as “$OLS^{1\text{stage}}$” to emphasize that it is a one-step estimator. The second is similar to how we calculate standard errors for GLS and WLS. In particular, we set $\hat{W}$ as the identity matrix. The last type is the Fama-MacBeth standard-error estimate, where OLS is performed at each period to obtain the ex-post risk-premium estimates, and $t$-statistics are calculated based on the time-series of risk-premium estimates. Note that $OLS^{1\text{stage}}$ 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.
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. $\gamma_0$, $\gamma_{mkt}$, $\gamma_{smb}$, $\gamma_{hml}$, $\gamma_{cma}$, 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 5-Factor Model, $N = 18$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_0$</td>
<td>$\gamma_{mkt}$</td>
</tr>
<tr>
<td>True</td>
<td>0</td>
</tr>
<tr>
<td>GMM Bias</td>
<td>0.011</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.177</td>
</tr>
<tr>
<td>MAE</td>
<td>0.130</td>
</tr>
<tr>
<td>OLS Bias</td>
<td>0.097</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.319</td>
</tr>
<tr>
<td>MAE</td>
<td>0.261</td>
</tr>
<tr>
<td>OCSR Bias</td>
<td>0.012</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.200</td>
</tr>
<tr>
<td>MAE</td>
<td>0.159</td>
</tr>
<tr>
<td>GLS Bias</td>
<td>0.006</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.211</td>
</tr>
<tr>
<td>MAE</td>
<td>0.168</td>
</tr>
<tr>
<td>WLS Bias</td>
<td>0.089</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.359</td>
</tr>
<tr>
<td>MAE</td>
<td>0.291</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: FF 3-Factor Model, $N = 38$</th>
<th>Panel D: FF 5-Factor Model, $N = 38$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_0$</td>
<td>$\gamma_{mkt}$</td>
</tr>
<tr>
<td>True</td>
<td>0</td>
</tr>
<tr>
<td>GMM Bias</td>
<td>0.003</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.208</td>
</tr>
<tr>
<td>MAE</td>
<td>0.158</td>
</tr>
<tr>
<td>OLS Bias</td>
<td>0.012</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.234</td>
</tr>
<tr>
<td>MAE</td>
<td>0.186</td>
</tr>
<tr>
<td>OCSR Bias</td>
<td>0.024</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.173</td>
</tr>
<tr>
<td>MAE</td>
<td>0.140</td>
</tr>
<tr>
<td>GLS Bias</td>
<td>0.024</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.192</td>
</tr>
<tr>
<td>MAE</td>
<td>0.154</td>
</tr>
<tr>
<td>WLS Bias</td>
<td>0.010</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.213</td>
</tr>
<tr>
<td>MAE</td>
<td>0.169</td>
</tr>
</tbody>
</table>

Our results are reported in a sequence of tables in Appendix D (i.e., Table SB.1-SB.4). To save space, we focus on the Fama-French three-factor model. Results based on the Fama-French five-factor model are available upon request.
test may be powerful (under alternative hypotheses) while also being oversized (under the null hypothesis). To allow an apples-to-apples comparison across methods, we report both the original size and power (denoted as “Ori”) and size-adjusted power (denoted as “Adj”) where the statistical cutoff that exactly achieves a prespecified significance level is found and used to calculate the corresponding test power.36

In Tables SB.1 and SB.2, 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 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 among the tests we consider. The power improvement of OCSR compared with GLS—the second-best performer—is usually in the range of 5% to 10%. In comparison, the other three methods (FM, OLS, and WLS) have substantially lower power than OCSR and GLS.

For $T = 750$ (Table SB.3-SB.4), 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.

3.2 Empirical Factors

We next take a more systematic look at our model’s performance. By examining a representative cross-section of anomalies and set of factors, we aim to provide the rule of thumb regarding the cutoff $N/T$ for the empirical application of our approach.

We obtain 156 long-short anomaly portfolios in Chen and Zimmermann (2020) (also studied in our empirical analysis in the next section). Some anomalies have many missing observations from the earlier years. To obtain a balanced panel, we further condition on anomalies that have non-missing observations post January 1975 and obtain 119 anomalies in total. As a result, we have $T = 504$ monthly observations (from January 1975 to December 2016) and the maximal $N$ is taken to be $N_{\text{max}} = 119$.

We next perform a principle component (PC) analysis (Connor and Korajczyk (1988), Kozak, Nagel, and Santosh (2018)) to obtain the first $K$ dominant PCs as pricing factors that explain the bulk of anomaly return variation. We set $K$ at 3 or 5, similar to our simulation study above

36See Harvey and Liu (2020a) for alternative ways to weight Type I and Type II errors.
with the Fama-French models. PCs can be arbitrarily rotated without affecting their asset pricing implications. To facilitate the interpretation, we standardize each PC such that the standardized PC has an annualized return volatility of 15%. Note that because our test assets correspond to long-short portfolios, the first PC does not exactly correspond to the market factor (Kozak, Nagel, and Santosh (2018)).

To study the performance variation when \( N \) increases, we set \( N \) at 20, 30, 50, 75, and 100. For each given \( N \), we randomly draw (without replacement) \( M = 1,000 \) setsmenus of \( N \) assets from \( N_{\text{max}} \) assets. Fixing the asset menu, we follow the same simulation procedure as the one used above for the Fama-French models to find the simulated summary statistics. These statistics are then averaged across \( M \) simulations to generate the final statistics.\(^{37}\)

Table D.2 in Appendix D reports the simulation results. We see that for \( N = 20 \) or 30, OCSR remains the dominant estimator in terms of estimation efficiency. For \( N = 50 \), no single estimator is the clear winner: the performance ranking varies across parameters. But a closer look at the results suggest that OCSR, if not the best, has a similar performance to the best estimator, being it the OLS or GLS. When \( N \) gets even larger, the more robust OLS starts outperforming alternative estimators.

Overall, our simulation results suggest that 10% (i.e., roughly 50/504) seems to be a reasonable cutoff for \( N/T \). When \( N/T \leq 10\% \), we expect to see an efficiency gain of using our approach or at least similar performance compared to alternative methods. When \( N/T > 10\% \), potential users may want to consider OLS or other \( N^{-1/2} \)-consistent estimators as proposed by the recent literature.

4 Applications

We use our framework to provide an alternative evaluation of the factor model proposed in FF. FF relies 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.\(^{38}\) 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.\(^{39}\) However, factors may be measured with errors

\(^{37}\)Because we simulate a large number (i.e., \( 10^7 \)) of times, the simulation for the GMM approach becomes too computationally intensive. We therefore omit it from the current simulation study.

\(^{38}\)Alternatively, Barillas and Shanken (2017) and Fama and French (2018) use spanning regressions to test nested models (i.e., the RHS approach).

\(^{39}\)More specifically, for a single asset \( i \), imposing the condition that risk premiums equal factor means, the following
(see, e.g., Roll (1977), Shanken (1987), and Jagannathan, Schaumburg, and Zhou (2010)), and the
exactly identified GMM may not fully use all restrictions implied by supposedly all-encompassing
factor models. We therefore turn to the cross-sectional approach. From the perspective of a GMM
framework, we impose additional moment restrictions under mild assumptions on the candidate
factor model while relaxing the pricing restriction on factors. This leads to an over-identified
GMM, which is equivalent to our OCSR asymptotically. We use OCSR to make inference and
contrast our results with alternative two-pass estimators as well as the usual time-series approach.

FF uses a small set of test portfolios to provide support for their model. We use a comprehensive
list of anomalies constructed and publicized by Chen and Zimmermann (2020). In total, 156
distinct anomalies are included. We study both equal-weighted and value-weighted anomaly long-
short portfolios.

What additional moment restrictions can we bring in to make potentially more informative
inference on factor models? Given that factor models purport to explain the returns of a large
number of assets, it seems natural to assume that they should at least be able to explain the
sorted portfolios based on which factors are constructed from, for example, the 20 size, book-to-
market, operating profitability, and investment sorted portfolios for FF. 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 test portfolios to construct
over-identifying restrictions and test FF.\textsuperscript{40} We refer to these portfolios as basis assets for a given
candidate factor model.

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

The model we use in this section combines insights from both cross-sectional and time-series
set of moment equations,

\[
\begin{align*}
\mathbb{E}[f_{t} - \mu_f] &= 0, \\
\mathbb{E}[(r_{i,t} - (f_{t} - \mu_f)\beta_i)(f_{t} - \mu_f)] &= 0, \\
\mathbb{E}[r_{i,t} - \alpha_i - \mu_f\beta_i] &= 0,
\end{align*}
\]

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 a time-series regression.

\textsuperscript{40}Our data for both factors and factor portfolios are obtained from Ken French’s online data library.
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 the risk-free rate (because we use excess instead of gross returns for assets) and the matrix of factor loadings \( \mathbf{B} \). We introduce another variable to capture potential model misspecification (i.e., abnormal alpha) by asset \( N \), given by \( S_N \equiv (0_{1 \times N_0}, 1)' \). The stacked matrix of regressors is given by \( (S_N, 1_N, \mathbf{B}) \). Let the associated regression coefficient be

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

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

\[
\overline{TA} = l' \times \hat{\theta},
\]

where \( l = (1, 1, 0_{1 \times (d_g-2)})' \) (i.e., non-zero for only the first two elements), and \( \overline{TA} \) stands for “total alpha,” which is to distinguish from our alpha definition in the theory section.\(^{41}\) 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 (C.3) of Appendix C.1 can be straightforwardly applied to make inference on \( \overline{TA} \).

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) with estimates based on our OCSR, where Figure 1, left two panels show results for equal-weighted portfolios and right two panels for value-weighted portfolios. In each figure, 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

\(^{41}\)Note that 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.
Figure 1: **Time-Series versus Cross-Sectional Tests** For each of the 156 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 20 sorted portfolios in FF 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.
and OLS are clearly positively correlated.

Turning to the right two panels with value-weighted portfolios, while the range for alpha $t$-statistics is smaller due to less dispersion in return for value-weighted portfolios, the overall pattern is similar to equal-weighted portfolios. Time-series alpha estimates and cross-sectional estimates are positively correlated, although large discrepancies may occur for certain anomalies.

We next turn to the specification tests, which justify our cross-sectional approach. In particular, 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 through OCSR (i.e., the test in (C.8)). To highlight the result for each individual basis asset, we also report results for individual time-series regressions that project each basis asset onto the associated factor model. Table 2 reports the results.

### Table 2: FF: Specification Tests

We perform time-series regressions for the 20 basis assets for FF against the FF model. We report summary statistics for absolute alpha estimates and absolute alpha $t$-statistics. We also perform specification tests based on (C.8) for the FF model.

<table>
<thead>
<tr>
<th>FF</th>
<th>Absolute Alpha</th>
<th>Absolute Alpha $t$-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>Stdev.</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>0.130</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>1.107</td>
</tr>
<tr>
<td></td>
<td>Stdev.</td>
<td>1.325</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>2.654</td>
</tr>
<tr>
<td></td>
<td>$(&gt;2.0)$</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$(&gt;3.0)$</td>
<td>0</td>
</tr>
</tbody>
</table>

For FF, the average absolute $t$-statistic for intercept is 1.11. Only two (zero, out of 20) basis assets have an absolute $t$-statistic exceeding 2.0 (3.0). The specification test based on (C.7) does not reject the FF model ($p$-value = 0.12). Overall, specification tests largely support the hypothesis that the FF model is correctly specified, allowing us to apply our cross-sectional approach to estimate the factor risk premiums. We consider results from specification tests a reasonable sanity check for the internal consistency of factor models: factor models need to be able to explain the returns of basis assets before being taken to confront external portfolios (i.e., portfolios not used.
Table 3: Factor-Premium Estimates: Time Series versus Cross-Sectional

We report the time-series and cross-sectional factor-premium estimates. Time-series factor premiums are calculated as the time-series means of factor returns. For cross-sectional estimates, 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)). We report results for six cross-sectional estimators: $O_{LS}^{1\text{stage}}$ 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), $\gamma_{\text{mkt}}, \gamma_{\text{smb}}, \gamma_{\text{hml}}, \gamma_{\text{cma}},$ and $\gamma_{\text{rmw}}$ 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.

<table>
<thead>
<tr>
<th>Method</th>
<th>FF</th>
<th>$\gamma_{\text{mkt}}$</th>
<th>$\gamma_{\text{smb}}$</th>
<th>$\gamma_{\text{hml}}$</th>
<th>$\gamma_{\text{cma}}$</th>
<th>$\gamma_{\text{rmw}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS(TS)</td>
<td>Mean</td>
<td>0.519</td>
<td>0.250</td>
<td>0.363</td>
<td>0.279</td>
<td>0.330</td>
</tr>
<tr>
<td></td>
<td>Stdev.</td>
<td>0.185</td>
<td>0.126</td>
<td>0.118</td>
<td>0.092</td>
<td>0.083</td>
</tr>
<tr>
<td>$O_{LS}^{1\text{stage}}$</td>
<td>Mean of Estimates</td>
<td>0.865</td>
<td>0.253</td>
<td>0.257</td>
<td>0.317</td>
<td>0.204</td>
</tr>
<tr>
<td></td>
<td>Mean of Stdev.</td>
<td>0.445</td>
<td>0.140</td>
<td>0.146</td>
<td>0.126</td>
<td>0.108</td>
</tr>
<tr>
<td>FM</td>
<td>Mean of Estimates</td>
<td>0.865</td>
<td>0.253</td>
<td>0.257</td>
<td>0.317</td>
<td>0.204</td>
</tr>
<tr>
<td></td>
<td>Mean of Stdev.</td>
<td>0.391</td>
<td>0.134</td>
<td>0.132</td>
<td>0.110</td>
<td>0.096</td>
</tr>
<tr>
<td>OLS</td>
<td>Mean of Estimates</td>
<td>0.865</td>
<td>0.253</td>
<td>0.257</td>
<td>0.317</td>
<td>0.204</td>
</tr>
<tr>
<td></td>
<td>Mean of Stdev.</td>
<td>0.302</td>
<td>0.037</td>
<td>0.050</td>
<td>0.062</td>
<td>0.050</td>
</tr>
<tr>
<td>OCSR</td>
<td>Mean of Estimates</td>
<td>0.794</td>
<td>0.220</td>
<td>0.199</td>
<td>0.217</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>Mean of Stdev.</td>
<td>0.383</td>
<td>0.126</td>
<td>0.133</td>
<td>0.108</td>
<td>0.095</td>
</tr>
<tr>
<td>GLS</td>
<td>Mean of Estimates</td>
<td>0.596</td>
<td>0.232</td>
<td>0.277</td>
<td>0.293</td>
<td>0.258</td>
</tr>
<tr>
<td></td>
<td>Mean of Stdev.</td>
<td>0.411</td>
<td>0.137</td>
<td>0.145</td>
<td>0.122</td>
<td>0.104</td>
</tr>
<tr>
<td>WLS</td>
<td>Mean of Estimates</td>
<td>0.695</td>
<td>0.224</td>
<td>0.241</td>
<td>0.305</td>
<td>0.207</td>
</tr>
<tr>
<td></td>
<td>Mean of Stdev.</td>
<td>0.451</td>
<td>0.137</td>
<td>0.148</td>
<td>0.123</td>
<td>0.107</td>
</tr>
</tbody>
</table>

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.\footnote{In unreported results, we find that the HXZ model fails to explain their test portfolios. Although beyond the scope of this paper, we believe this difference in specification tests should be factored into the overall comparison of factor models.}
\footnote{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.}
One observation from Table 3 is that sometimes time-series means differ substantially from the corresponding cross-sectional estimates. If the time-series means are treated as the benchmark estimates (see, e.g., Lewellen, Nagel, and Shanken (2010)), does this imply that the cross-sectional estimates are suboptimal (i.e., less efficient)? We believe not because we do not know the underlying true parameter values. Take $HML$ as an example. Its time-series mean is 0.363, with a standard deviation of 0.118. The OCSR estimate is 0.199, representing a large deviation from the time-series mean. However, this does not necessarily imply that OCSR is inaccurate because the true underlying mean parameter may be closer to the OCSR estimate than the time-series mean (e.g., a true mean of 0.25). Another factor that may confound the comparison in risk premium estimate is the possibility of measurement error in traded factors (i.e., Roll (1977), Shanken (1987), and Jagannathan, Schaumburg, and Zhou (2010)). Time-series means may deviate from the true factor means due to measurement errors, which naturally leads to a gap in estimates between the time-series approach and our cross-sectional approach.

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%.

Finally, Table 4 provides detailed statistics on the performance of FF against both value-weighted and equal-weighted test assets. Several remarks follow. 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., $\text{OLS}^{1\text{stage}}$), which ignores beta-estimation uncertainty, leads to a much larger number of rejections. This finding is consistent with our extensive simulation evidence on the over-rejection of the naive OLS and with the results provided in Shanken (1992) and Jagannathan and Wang (1998). Our results substantiate the over-rejection concern regarding the unadjusted Fama-MacBeth approach as argued in Shanken (1992). The degree of over-rejection for the naive OLS is so large that it makes hypothesis tests for cross-sectional regressions inappropriate. We therefore follow Shanken (1992) and Jagannathan and Wang (1998) and recommend standard errors adjusted for first-stage beta estimation. In fact, a cross-sectional OLS that applies the same weighting scheme as $\text{OLS}^{1\text{stage}}$ 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 $OLS^{1\text{stage}}$) to 66 (OLS).

Table 4: Alpha Estimates: Comparing Cross-Sectional Approaches

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, 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)). 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).

<table>
<thead>
<tr>
<th>Method</th>
<th>Panel A: Equal-Weighted Anomaly Returns</th>
<th>Panel B: Value-Weighted Anomaly Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abs. Alpha &amp; t-stat #(</td>
<td>t-stat</td>
</tr>
<tr>
<td>$OLS(TS)$</td>
<td>Mean</td>
<td>Stdev.</td>
</tr>
<tr>
<td>$OLS^{1\text{stage}}$</td>
<td>0.429</td>
<td>0.465</td>
</tr>
<tr>
<td>$FM$</td>
<td>0.468</td>
<td>0.465</td>
</tr>
<tr>
<td>$OLS$</td>
<td>0.468</td>
<td>0.465</td>
</tr>
<tr>
<td>$OCSR$</td>
<td>0.445</td>
<td>0.454</td>
</tr>
<tr>
<td>$GLS$</td>
<td>0.448</td>
<td>0.461</td>
</tr>
<tr>
<td>$WLS$</td>
<td>0.455</td>
<td>0.462</td>
</tr>
<tr>
<td>$OLS(TS)$</td>
<td>Mean</td>
<td>Stdev.</td>
</tr>
<tr>
<td>$OLS^{1\text{stage}}$</td>
<td>0.382</td>
<td>0.352</td>
</tr>
<tr>
<td>$FM$</td>
<td>0.373</td>
<td>0.365</td>
</tr>
<tr>
<td>$OLS$</td>
<td>0.373</td>
<td>0.365</td>
</tr>
<tr>
<td>$OCSR$</td>
<td>0.349</td>
<td>0.326</td>
</tr>
<tr>
<td>$GLS$</td>
<td>0.350</td>
<td>0.327</td>
</tr>
<tr>
<td>$WLS$</td>
<td>0.357</td>
<td>0.350</td>
</tr>
</tbody>
</table>

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

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

To summarize, we revisit the recent FF model, for which the over-identification conditions for
basis assets approximately hold. We find substantial differences between the time-series OLS as in Fama and French (2015) and our OCSR. We also find differences in test outcomes between OCSR and other cross-sectional approaches proposed by the previous literature (e.g., GLS) and highlight the gain in test power of our approach.

Note that although our narrative has focused on the difference in results between the time-series OLS and OCSR, several other advantages of OCSR (or cross-sectional approaches in general) are worth emphasizing. For example, when the zero-beta rate is truly different from the risk-free rate, OCSR can easily distinguish between the overall alpha and the alpha in addition to the (common) zero-beta rate (i.e., $\bar{T}A$ vs. $\hat{a}_N$), whereas the time-series OLS has to lump them into a single intercept. Additionally, firm characteristics can be straightforwardly incorporated into OCSR to allow potential model misspecification and enrich testable hypotheses (see our general model specification in Section 2), whereas for the time-series OLS, allowing for firm characteristics that mainly vary in the cross-section is challenging. We leave the examination of these interesting extensions of OCSR to future research.

5 Conclusion

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

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


43
APPENDIX

A Conditions and Proofs of Main Results

In this section, we provide extra sufficient conditions and proofs of the asymptotic properties of the CSR estimators provided in Section 2. The following assumption presents the conditions on the pricing factors and security characteristics.

Assumption 3. Let $\bar{f} \equiv T^{-1} \sum_{t \leq T} f_t$ and $\bar{Z} \equiv T^{-1} \sum_{t \leq T} Z_{t-1}$. Suppose the following: (i) $\{f_t\}_t$ is a covariance-stationary process; (ii) $\bar{f} = \mu_f + o_p(1)$; (iii) $T^{-1} \sum_{t \leq T} f_t f_t' = E[f_t f_t'] + o_p(1)$; and (iv) $\bar{Z} = E[Z_{t-1}] + o_p(1)$.

Assumption 3(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 3(ii, iii) are the law of large numbers of the sample mean and the sample second moment of the pricing factors; Assumption 3(iv) is the law of large numbers on the sample mean of the security characteristics. These conditions can be verified under low-level sufficient conditions. Note that Assumption 3(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 3(iv) should be considered a reasonable approximation for sorted portfolios (see, e.g., Jagannathan and Wang (2002)).

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

$$R - \tilde{X} \theta = R - E[R_t] + E[R_t] - \tilde{X} \theta$$

$$= R - E[R_t] - (\tilde{X} - X) \theta$$

$$= R - E[R_t] - (\bar{B} - B) \gamma_1 - (\bar{Z} - E[Z_t]) \gamma_2 = \nu - (\bar{B} - B) \gamma_1,$$

(A.1)

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

$$T^{1/2}(\hat{\theta}_{csr} - \theta) = (\tilde{X}'\hat{W}\tilde{X})^{-1} \tilde{X}'\hat{W} \left[ T^{1/2}(\nu - (\bar{B} - B) \gamma_1) \right].$$

(A.2)
Because $R_t = X\theta + B(f_t - \mu_f) + u_t$,

$$
\hat{B} - B = \left( T^{-1} \sum_{t \leq T} (R_t - \bar{R})(f_t - \bar{f})' \right) \hat{\Sigma}_f^{-1} - B
= T^{-1} \sum_{t \leq T} u_t(f_t - \mu_f)' \hat{\Sigma}_f^{-1} - uT^{-1} \sum_{t \leq T} (f_t - \mu_f)' \hat{\Sigma}_f^{-1},
$$

(A.3)

where $\hat{\Sigma}_f \equiv T^{-1} \sum_{t \leq T}(f_t - \bar{f})(f_t - \bar{f})'$. By Assumptions 3(i, ii, iii),

$$
\hat{\Sigma}_f = \Sigma_f + o_p(1),
$$

(A.4)

which together with Assumptions 3(ii) and 1(ii, iv) implies

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

(A.5)

Therefore, by Assumptions 1(i) and (A.5), we have

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

(A.6)

Similarly, by Assumptions 3(iv) and (A.5),

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

(A.7)

From Assumptions 1(iii, iv) and (A.7), we obtain

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

(A.8)

Combining the results in (A.2), (A.6), and (A.8) and applying Assumptions 1(i, iv), we get

$$
T^{1/2}(\hat{\theta}_{csr} - \theta) = (X'WX)^{-1}X'WT^{-1/2} \sum_{t \leq T} \epsilon_t + o_p(1) \rightarrow_d N(0, Asv(\hat{\theta}_{csr}))
$$

(A.9)

which shows the claim of the lemma.

Q.E.D.

Proof of Proposition 1. Replacing $W$ in $Asv(\hat{\theta}_{csr})$ by $\Omega^{-1}$ obtains (2.10). Let $P_N \equiv I_N -
\[ \Omega^{1/2}W(X'W\Omega W X)^{-1}X'W\Omega^{1/2}. \]

Then, we can write

\[ (\text{Asv}(\hat{\theta}^{\ast}_{\text{csp}}))^{-1} - (\text{Asv}(\hat{\theta}^{\ast}_{\text{csp}}))^{-1} = X'\Omega^{-1/2}P_N\Omega^{-1/2}X. \]  \hspace{1cm} (A.10)

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

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

\[ \text{Asv}(\hat{\theta}^{\ast}_{\text{csp}}) = \text{Asv}(\hat{\theta}^{\ast}_{\text{gmm}}). \] \hspace{1cm} (A.11)

First, simple calculation shows

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

Let

\[ D = \begin{pmatrix} -X \theta \gamma_1' \Sigma_f^{-1} & I_N \otimes \gamma_1' \Sigma_f^{-1} & -I_N \\ -I_K & 0_{K \times NK} & 0_{K \times N} \\ X \theta \otimes \Sigma_f^{-1} & -I_N \otimes \Sigma_f^{-1} & 0_{NK \times N} \end{pmatrix}. \] \hspace{1cm} (A.12)

Because \( D \) is invertible, we can write

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

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

\[ \text{Asv}(\hat{\theta}^{\ast}_{\text{gmm}}) = S_0 \text{Asv}(\hat{\theta}^{\ast}_{\text{gmm}}) S_0' = (X'(D \Sigma_g^* D')^{-1}_{11} X)^{-1}, \] \hspace{1cm} (A.13)

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

\[
(Dg(Y_t, \phi))_N = -X\theta_1\Sigma^{-1}g_1(Y_t, \phi) + (I_N \otimes \Sigma^{-1})g_2(Y_t, \phi) - g_3(Y_t, \phi) \\
= -X\theta_1\Sigma^{-1}(f_t - \mu_f) + (R_t - B(f_t - \mu_f))\Sigma^{-1}(f_t - \mu_f) - (R_t - X_t\theta) \\
= (R_t - E[R_t] - B(f_t - \mu_f))\gamma_1\Sigma^{-1}(f_t - \mu_f) - (R_t - X_t\theta). \tag{A.14}
\]

By the definition of $u_t$ and $v_t$, we deduce from (A.14) that

\[
(Dg(Y_t, \phi))_N = -v_t + u_t(f_t - \mu_f)'\Sigma^{-1}\gamma_1, \tag{A.15}
\]

which implies

\[
(DS\Sigma'gD')_{11} = \lim_{T \to \infty} \text{Var} \left( T^{-1/2} \sum_{t=1}^{T} (v_t - u_t(f_t - \mu_f)'\Sigma^{-1}\gamma_1) \right) = \Omega. \tag{A.16}
\]

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

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

which shows (A.11). \textit{Q.E.D.}

\section*{B A Counterexample}

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

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

\[
E[Y_1 - \mu1_{k \times 1}] = 0_{k \times 1}, \tag{B.1} \\
E[Y_2 - \mu - \theta] = 0, \tag{B.2}
\]

where $k \geq 1$ and $\mu$ is a nuisance parameter. Let $\varepsilon_1 \equiv Y_1 - \mu1_{k \times 1}$ 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 definite.

We compare the asymptotic variances of two estimators of \(\theta\). The first is from a joint optimal GMM estimation of \((\theta, \mu)\) using all the moment conditions in (B.1) and (B.2). The second is obtained by an iterative GMM estimation procedure where we first obtain the optimal GMM estimator of \(\mu\) through the moment conditions in (B.1) and then plug it in (B.2) to construct the optimal two-step 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.1) and (B.2) is the inverse of the following matrix:

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

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

\[ \text{Asv}(\hat{\theta}^J) \equiv \left( k + \frac{(1 - 1'_{k \times 1}\rho)^2}{1 - \rho'\rho} \right) \frac{1 - \rho'\rho}{k} = 1 + k^{-1} - 2k^{-1}1'_{k \times 1}\rho + \frac{(1'_{k \times 1}\rho)^2}{k} - k\rho'\rho. \quad (B.3) \]

On the other hand, the optimal GMM estimator of \(\mu\) based on (B.1) is

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

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

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

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

\[ \text{Asv}(\hat{\theta}^I) \equiv 1 + k^{-1} - 2k^{-1}1'_{k \times 1}\rho. \quad (B.4) \]
By (B.3) and (B.4), we have
\[
\text{Asv}(\hat{\theta}^J) - \text{Asv}(\hat{\theta}^I) = k^{-1}(1'_{k \times 1} \rho)^2 - \rho' \rho \leq 0,
\]  
where the equality holds if and only if \( k = 1 \) or \( \rho = a1_{k \times 1} \) for some real number \( a \) by the Cauchy-Schwarz inequality.

From the inequality in (B.5), we see that in general the joint optimal GMM estimator \( \hat{\theta}^J \) dominates the iterative optimal GMM estimator. These two estimators are asymptotically equivalent in the special case where the moment conditions in (B.1) and (B.2) have special dependence structure, that is, \( \rho = a1_{k \times 1} \). Under general dependence of the moment conditions in (B.1) and (B.2) (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 asymptotic equivalent only when \( k = 1 \), that is, the nuisance parameter \( \mu \) is just identified in (B.1), which shares the same intuition of the asymptotic efficiency of our OCSR estimator.

C Some Auxiliary Results

This section contains some results which can be used to conduct inference on the unknown parameters and model specification tests. Section C.1 provides the \( t \) test and Wald test on the unknown parameter \( \theta \), and Section C.2 provides a specification test for (2.4). Section C.3 extends the asymptotic equivalence between the OCSR estimator and the optimal GMM estimator to the case where \( N \) grows with \( T \). Section C.4 considers the case with an unbalanced panel.

C.1 Inference based on OCSR

Let \( \hat{\Omega} \) denote a consistent estimator of \( \Omega \). The OCSR estimator is then defined as
\[
\hat{\theta}_{\text{csr}}^* \equiv (\hat{X}'\hat{\Omega}^{-1}\hat{X})^{-1}\hat{X}'\hat{\Omega}^{-1}\hat{R}.
\]  
(C.1)

The consistency of \( \hat{\Omega} \) implies Assumption 1(iii). Moreover, by Assumptions 1(ii, iv) and 3, we can show that
\[
\hat{X}'\hat{\Omega}^{-1}\hat{X} = X'\Omega^{-1}X + o_p(1).
\]  
(C.2)
Let \( l \) denote any \( d_\theta \times 1 \) non zero real vector. Then, (C.2) together with Lemma 1 implies
\[
T^{1/2} l'(\hat{\theta}_{csp} - \theta) \\
(l'(\hat{X}^\top \hat{\Omega}^{-1} \hat{X})^{-1} l)^{1/2} \rightarrow_d N(0, 1),
\]
which can be used to test the hypothesis on the linear combinations of \( \theta \). Let \( \mathcal{M} \) denote any non-empty subset of \( \{1, \ldots, d_\theta\} \) with size \( m \) and let \( L_{\mathcal{M}} = (l_j)_{j \in \mathcal{M}} \) denote the \( d_\theta \times m \) selection matrix whose \( j \)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 (C.2) and Lemma 1,
\[
(L_{\mathcal{M}}'(\hat{X}^\top \hat{\Omega}^{-1} \hat{X})^{-1} L_{\mathcal{M}})^{-1/2} L_{\mathcal{M}}'(\hat{\theta}_{csp} - \theta) \rightarrow_d N(0, I_m),
\]
which together with the continuous mapping Theorem implies
\[
T(\hat{\theta}_{csp} - \theta)' L_{\mathcal{M}}(L_{\mathcal{M}}'(\hat{X}^\top \hat{\Omega}^{-1} \hat{X})^{-1} L_{\mathcal{M}})^{-1} (L_{\mathcal{M}}'(\hat{\theta}_{csp} - \theta)) \rightarrow_d \chi^2(m),
\]
where \( \chi^2(m) \) denotes the chi-square distribution with degree of freedom \( m \). The result in (C.4) can be used to conduct joint inference of any subset of \( \theta \), such as \( (\alpha_{N_0+1}, \ldots, \alpha_N)' \).

C.2 Specification tests

In this section, we provide a specification test for the restriction (2.4). By (2.3), this restriction can be written as
\[
H_0 : \mathbb{E}[R_{i,t} - \gamma_0 - \gamma_1^\prime \beta_i - \gamma_2^\prime Z_{i,t}] = 0,
\]
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 \( \hat{X}_{0,t} \equiv (1_{N_0 \times 1}, \hat{B}_0, Z_{0,t-1}) \), where \( \hat{B}_0 \) and \( Z_{0,t-1} \) denote the leading \( N_0 \times K \) and \( N_0 \times M \) submatrices of \( \hat{B} \) and \( Z_{t-1} \), respectively. The unknown parameter \( \gamma \) is estimated by
\[
\hat{\gamma} \equiv (\hat{X}_{0,t}^\top \hat{\Omega}_0^{-1} \hat{X}_{0,t})^{-1} \hat{X}_{0,t}^\top \hat{\Omega}_0^{-1} \hat{R}_0,
\]
where \( \hat{X}_{0,t} \equiv T^{-1} \sum_{t \leq T} \hat{X}_{t,t} \), \( \hat{R}_0 \equiv T^{-1} \sum_{t \leq T} \hat{R}_{0,t} \), \( \hat{\Omega}_0 \) denotes the leading \( N_0 \times N_0 \) submatrix of \( \hat{\Omega} \) and \( \hat{R}_{0,t} \) denotes the leading \( N_0 \) subvector of \( \hat{R}_t \). The null hypothesis (C.5) is tested using the
The $J$-test statistic, which is defined as

$$J_T \equiv T(\widehat{R}_0 - \widehat{X}_0\hat{\gamma})'\widehat{\Omega}^{-1}_0(\widehat{R}_0 - \widehat{X}_0\hat{\gamma}).$$  \hfill (C.7)

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 $\alpha$:

reject $H_0$ if $J_T > \chi^2_{1-\alpha}(N_0 - d_\gamma).$  \hfill (C.8)

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

**Lemma 2.** Suppose Assumption 3 hold. Then,

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

(b) If we have: (i) $\widehat{R}_0 = E[R_{0,t}] + o_p(1)$; (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

$$T^{-1}J_T = E[R_{0,t}']\Pi_{N_0}E[R_{0,t}] + o_p(1),$$  \hfill (C.9)

where $\Pi_{N_0} \equiv \Omega_1^{-1} - \Omega_1^{-1}X_0(X_0'\Omega_1^{-1}X_0)^{-1}X_0'\Omega_1^{-1}$.

Lemma 2(a) shows that the test in (C.8) controls size. Lemma 2(b) derives the probability limit of the (scaled) test statistic $J_T$ under both the null and the alternative hypotheses. Under $H_0$, $E[R_{0,t}] = X_0\gamma$, and hence (C.9) implies $T^{-1}J_T = o_p(1)$, which is consistent with the weak convergence of $J_T$ derived in Lemma 2(a). Under the alternative hypothesis, $E[R_{0,t}]$ cannot be represented by any linear combination of $X_0$, and hence, we have $E[R_{0,t}']\Pi_{N_0}E[R_{0,t}] > 0$ in general. Because $\chi^2_{1-\alpha}(N_0 - d_\gamma)$ is a finite number, Lemma 2(b) shows that the test in (C.8) is consistent as long as $E[R_{0,t}']\Pi_{N_0}E[R_{0,t}]$ is bounded away from zero.

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

$$\widehat{R}_0 - \widehat{X}_0\hat{\gamma} = \mathbf{v}_0 - (\hat{B}_0 - B_0)\gamma_1 = T^{-1} \sum_{t \leq T} \epsilon_{0,t} + o_p(T^{-1/2}) = O_p(T^{-1/2}),$$  \hfill (C.10)

where $\mathbf{v}_0 \equiv T^{-1} \sum_{t \leq T} \mathbf{v}_{0,t}$, $B_0$, $\mathbf{v}_{0,t}$, and $\epsilon_{0,t}$ denote the leading $N_0 \times K$, $N_0 \times 1$, and $N_0 \times 1$...
submatrices of $B$, $v_t$, and $\epsilon_t$, respectively. Therefore, by the same arguments for showing (A.9) in the proof of Lemma 1,

$$\hat{\gamma} - \gamma = (X_0'\Omega_0^{-1}X_0)^{-1}X_0'\Omega_0^{-1}T^{-1}\sum_{t \leq T} \epsilon_{0,t} + o_p(T^{-1/2}) = O_p(T^{-1/2}),$$  \hspace{1cm} (C.11)

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

$$R_0 - \hat{X}_0\hat{\gamma} = R_0 - \hat{X}_0\gamma - \hat{X}_0(\hat{\gamma} - \gamma)$$

$$= (I_{N_0} - X_0(X_0'\Omega_0^{-1}X_0)^{-1}X_0'\Omega_0^{-1})T^{-1}\sum_{t \leq T} \epsilon_{0,t} + o_p(T^{-1/2})$$

$$= \Omega_0^{1/2}M_{N_0}\Omega_0^{-1/2}T^{-1/2}\sum_{t \leq T} \epsilon_{0,t} + o_p(T^{-1/2}) = O_p(T^{-1/2}),$$  \hspace{1cm} (C.12)

where $M_{N_0} \equiv I_{N_0} - \Omega_0^{-1/2}X_0(X_0'\Omega_0^{-1}X_0)^{-1}X_0'\Omega_0^{-1/2}$, which together with the consistency of $\hat{\Omega}$, Assumptions 1(i, iv), and the continuous mapping theorem implies

$$J_T = T(R_0 - \hat{X}_0\hat{\gamma})'\Omega_0^{-1}(R_0 - \hat{X}_0\hat{\gamma}) + o_p(1)$$

$$= \left(\Omega_0^{-1/2}T^{-1/2}\sum_{t \leq T} \epsilon_{0,t}\right)' M_{N_0} \left(\Omega_0^{-1/2}T^{-1/2}\sum_{t \leq T} \epsilon_{0,t}\right) + o_p(1) \rightarrow_d N_0'M_{N_0}N_0,$$  \hspace{1cm} (C.13)

where $N_0$ 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 $N_0'M_{N_0}N_0$ has the same distribution as $\chi^2(N_0 - d_\gamma)$. The claim in part (a) of the lemma follows directly from (C.13).

(b) Under Assumption 3 and condition (ii) of the theorem, we can use the same arguments for showing (A.7) and (A.8) in the proof of Lemma 1 to obtain

$$\hat{X}_0 = X_0 + o_p(1), \hat{X}_0'\hat{\Omega}_0^{-1}\hat{X}_0 = X_0'\Omega_1^{-1}X_0 + o_p(1)$$

and

$$\hat{X}_0'\hat{\Omega}_0^{-1} = X_0'\Omega_1^{-1} + o_p(1),$$

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

$$R_0 - \hat{X}_0\hat{\gamma} = (I_{N_0} - X_0(X_0'\Omega_1^{-1}X_0)^{-1}X_0'\Omega_1^{-1})E[R_{0,t}] + o_p(1).$$  \hspace{1cm} (C.14)
By conditions (ii, iii) of the lemma and (C.14),

\[ T^{-1}J_T = (\hat{R}_0 - \hat{X}_0\hat{\gamma})'\hat{\Sigma}_0^{-1}(\hat{R}_0 - \hat{X}_0\hat{\gamma}) = \mathbb{E}[R_{0,t}']\Pi_{N_o}\mathbb{E}[R_{0,t}] + o_p(1), \]

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

C.3 Extensions with a large \( N \)

In this section, we generalize the asymptotic normality and efficiency of the OCSR estimator established in Section 2 to the case where the number of assets \( N \) grows with \( T \). Since the number of assets in \( I_1 \) may grow with \( N \), the number of the unknown parameters (i.e., \( \alpha_i \) for \( i \in I_1 \)) also diverges 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) \( \{f_t\}_t \) is a covariance-stationary process; (ii) \( \bar{f} = \mu_f + O_p(T^{-1/2}) \); (iii) \( T^{-1}\sum_{t\leq T} f_tF_t' = \mathbb{E}[f_t f_t'] + O_p(T^{-1/2}) \), where \( \mathbb{E}[f_t f_t'] \) is bounded; and (iv) \( \bar{Z} = T^{-1}\sum_{t\leq T} \mathbb{E}[Z_{t-1}] + O_p(N^\delta_1 T^{-1/2}) \).

Assumption 4(i) is the same as Assumption 3(i). Assumptions 4(ii, iii, iv) strengthen Assumption 3(ii, iii, iv) by providing the convergence rates of \( \bar{f}, T^{-1}\sum_{t\leq T} f_tF_t' \) 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 \( Z \) is a \( N \times 1 \) random vector and \( N \) may go to infinity with \( T \). These conditions can be verified under low-level sufficient conditions. For example, when the eigenvalues of the variance-covariance matrix of \( Z \) are bounded from above uniformly over \( N \) and \( T \), one can show that Assumption 3(iv) holds with \( \delta_1 = 1/2 \).

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

\[ T^{-1/2}\sum_{t=1}^{T} \epsilon_t = \Omega_T^{1/2}N + o_p(1), \]

where \( \Omega_T \equiv \text{Var}(T^{-1/2}\sum_{t=1}^{T} \epsilon_t) \); (ii) \( T^{-1}\sum_{t=1}^{T} u_t(1, f_t') = O_p(N^{\delta_2}T^{-1/2}) \); (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 \( X'X \) are bounded from above and away from zero uniformly over \( N \) and \( T \) where \( X = (s_N, 1_{N \times 1}, B, T^{-1}\sum_{t\leq T} \mathbb{E}[Z_{t-1}]) \); and (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 an independent process (see, e.g., Theorem 10.4.10 in Pollard (2002)), or a heterogeneous dependent processes (see, e.g., Theorem 1 and Theorem 4 in Li and Liao (2020)). By the definitions of $u_{i,t}$ and $\beta_i$, we have $\mathbb{E}[u_t] = 0_{N \times 1}$ and $\mathbb{E}[u_t f_t'] = 0_{N \times K}$. Therefore, Assumption 5(ii) holds with $\delta_2 = 1/2$ if the eigenvalues of the variance-covariance matrix of $T^{-1/2} \sum_{t=1}^{T} u_t(1, f_t')$ are bounded uniformly over $N$ and $T$. Assumption 5(iii) imposes conditions on the weight matrix of the CSR estimator. The eigenvalue conditions in Assumption 5(iv) ensure the local identification of the CSR estimator. Assumption 5(v) imposes an upper bound on $N$. Since in most cases Assumptions 4(iv) and 5(ii, iii) can be verified with $\delta > 1/2$, Assumption 5(v) implies that $N$ may not go to infinity faster than $T$.

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

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

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

$$\Sigma_T(W) \geq (X'\Omega_{-1}X)^{-1},$$

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

Lemma 3 generalizes Lemma 1 to the case where both the number of assets $N$ and the number of the unknown parameters $d_\theta$ may go to infinity with $T$. Since the eigenvalues of $\Sigma_T(W)$ are bounded away from zero under Assumption 5(iv), $T^{1/2}(\hat{\theta}_{csr} - \theta)$ is not asymptotically tight, and hence it does not admit an asymptotic distribution. Nevertheless, (C.15) implies that the finite sample distribution of $T^{1/2}(\hat{\theta}_{csr} - \theta)$ can still be approximated by a normal random vector with variance-covariance matrix $\Sigma_T(W)$. Therefore, one can still conduct inference on $\theta$ using the normal approximation (C.15).

From (C.16), it is clear that the OCSR estimator still takes the form in (C.1) and the “pre-asymptotic” variance-covariance matrix of the OCSR estimator is $(X'\Omega_{-1}X)^{-1}$. Therefore, we can use the same arguments in the proof of Proposition 2 to show that the OCSR estimator has the same “asymptotic” variance-covariance matrix of optimal GMM estimator $\hat{\theta}_{gmm}^*$.\[44^\text{Note that the OCSR depends on a consistent estimator of $\Omega_T$. See Theorem 5 in Li and Liao (2020) for a consistent variance-covariance estimator with divergent dimension.}\]
Proof of Lemma 3. By Assumption 4(i, ii, iii),

\[ \tilde{\Sigma}_f = \Sigma_f + O_p(T^{-1/2}), \]  

(C.17)

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

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

(C.18)

Therefore by Assumptions 5(i, iv) and (C.18), we have

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

(C.19)

Similarly, by Assumptions 4(iv) and 5(v), and (C.18)

\[ \tilde{\mathbf{X}} = \mathbf{X} + O_p(N^\delta T^{-1/2}) = \mathbf{X} + o_p(1). \]  

(C.20)

From Assumptions 5(iii, iv, v) and (C.20), we obtain

\[ \tilde{\mathbf{X}}' \hat{\mathbf{W}} = \mathbf{X}' \mathbf{W} + O_p(N^\delta T^{-1/2}) \quad \text{and} \quad \tilde{\mathbf{X}}' \hat{\mathbf{W}} \tilde{\mathbf{X}} = \mathbf{X}' \mathbf{W} \mathbf{X} + O_p(N^\delta T^{-1/2}). \]  

(C.21)

Combining the results in (A.2), (C.17), and (C.21) and applying Assumptions 5(i, iv), we get

\[ T^{1/2}(\hat{\theta}_{car} - \theta) = (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W} T^{-1/2} \sum_{t \leq T} \epsilon_t + o_p(1) \]

\[ = (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W} \Omega_T^{1/2} \mathbf{N} + o_p(1). \]  

(C.22)

Since \((\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W} \Omega_T^{1/2} \mathbf{N}\) has the same distribution as \((\Sigma_T(W))^{1/2} \mathbf{N}^\ast\), then the claim in (C.15) follows from (C.22). The claim in (C.16) follows by the same arguments on showing (2.10) and hence is omitted. \(Q.E.D.\)

C.4 Extensions with an unbalanced panel

The asymptotic properties of the two-pass regression estimator we have investigated so far have been limited to the case that there are no missing observations on the asset returns and/or the
firms’ characteristics, and hence the data set is a balanced panel. When the number of assets \(N\) is large, there may be missing observations and the data set will take a form of unbalanced panel. In this case, the two-pass regression estimator has to be revised to account for the missing data issue.

Formally, let \(d_{i,t}\) be a binary random variable such that \(d_{i,t} = 1\) if both \(R_{i,t}\) and \(Z_{i,t-1}\) are observed at time \(t\), and \(d_{i,t} = 0\) otherwise. Let \(P(d_{i,t} = 1) = p_i\) where \(p_i\) is bounded away from zero uniformly over \(i\). Suppose that the data are missing at random in the sense that \(d_{i,t}\) is independent of \(f_t, R_{i,t}\) and \(Z_{i,t-1}\) for any \(i\) and any \(t\). Then under the missing at random assumption, the moment conditions (2.2) and (2.3) imply

\[
E[d_{i,t}(R_{i,t} - (f_t - \mu_f)')\beta_i(f_t - \mu_f)] = 0, \quad (C.23)
\]

\[
E[d_{i,t}(R_{i,t} - \alpha_i - \gamma_0 - \gamma_1'\beta_i - \gamma_2'Z_{i,t-1})] = 0, \quad (C.24)
\]

respectively. The two-pass regression estimator is then constructed using (2.1), (C.23) and (C.24).

Based on the moment condition in (C.23), the vector of betas \(\beta_i\) is estimated by

\[
\hat{\beta}_i' = \left(\sum_{t \leq T} (d_{i,t}R_{i,t} - \overline{R}_i) f_t'\right) \left(\sum_{t \leq T} d_{i,t}(f_t - \overline{f})(f_t - \overline{f})'\right)^{-1}, \quad (C.25)
\]

where \(\overline{R}_i \equiv T^{-1} \sum_{t \leq T} d_{i,t}R_{i,t}\). The matrix of betas \(B \equiv (\beta_1, \ldots, \beta_N)'\) is then estimated by \(\hat{B} \equiv (\hat{\beta}_1, \ldots, \hat{\beta}_N)'\). Given the estimators of betas, the two-pass regression estimator is constructed using (C.24) and it is the minimizer of

\[
\min_\theta \left(\sum_{t \leq T} (D_{I,t}R_t - D_{I,t}\tilde{X}_t\theta)\right)'\tilde{W} \left(\sum_{t \leq T} (D_{I,t}R_t - D_{I,t}\tilde{X}_t\theta)\right),
\]

where \(D_{I,t} = diag(d_{1,t}, \ldots, d_{N,t})\). Solving the least-square problem, we have

\[
\hat{\theta}_{csr} = (\tilde{X}'\tilde{W}\tilde{X})^{-1}\tilde{X}'\tilde{W}\tilde{R}, \quad (C.26)
\]

where \(\tilde{X} \equiv T^{-1} \sum_{t \leq T} D_{I,t}\tilde{X}_t\) and \(\tilde{R} \equiv T^{-1} \sum_{t \leq T} D_{I,t}R_t\).

Let \(\eta_{i,t} \equiv d_{i,t}R_{i,t} - E[d_{i,t}R_{i,t}] - d_{i,t}(f_t - \mu_f)'\beta_i\) for any \(i\) and any \(t\). Then \(E[\eta_{i,t}] = 0\) by the missing at random assumption. Using conditions similar to Assumptions 4 and 5, and similar

\[\text{Note that in the definition of } \hat{\beta}_i, \ d_{i,t}R_{i,t} \text{ is always observed although } R_{i,t} \text{ may not be observable for some } i \text{ and/or } t.\]
arguments in the proof of Lemma 3, we obtain

\[ T^{1/2}(\hat{\theta}_{csr} - \theta) = (X_d'WX_d)^{-1}X_d'W\left( T^{-1/2} \sum_{t \leq T} \zeta_t \right) + o_p(1) \]  

(C.27)

where \( X_d \equiv E[D_{L,t}(S_N, B, Z_{t-1})] \), \( \zeta_t \equiv D_{L,t}v_t - \eta_t(f_t - \mu_f)'\Sigma_f^{-1}\gamma_1 \) and \( \eta_t = (\eta_{1,t}, \ldots, \eta_{N,t})' \).

Let \( \Omega_T \equiv \text{Var}(T^{-1/2} \sum_{t=1}^T \zeta_t) \). Then we can use the strong approximation (i.e., condition similar to Assumption 5(i)) to show that

\[ T^{1/2}(\hat{\theta}_{csr} - \theta) = \langle \frac{1}{2} \Sigma(T(W))^{1/2} \rangle + o_p(1) \]  

(C.28)

where \( \Sigma_T(W) \equiv (X_d'WX_d)^{-1}(X_d'\Omega_TWX_d)(X_d'WX_d)^{-1} \). From (C.28), it is clear that the OCSR estimator takes the form

\[ \hat{\theta}_{csr}^* = (X_d'\tilde{\Omega})^{-1}X_d'\tilde{\Omega}^{-1}\tilde{R} \]

which is similar to (C.1), and the “pre-asymptotic” variance-covariance matrix of the OCSR estimator is \( (X_d'\tilde{\Omega}_T X_d)^{-1} \) with

\[ \tilde{\Sigma}_T(W) \geq (X_d'\tilde{\Omega}_T X_d)^{-1}, \]

for any \( N \times N \) symmetric positive definite matrix \( W \). Moreover, we can use the same arguments in the proof of Proposition 2 to show that the OCSR estimator has the same “asymptotic” variance-covariance matrix of optimal GMM estimator \( \hat{\theta}_{gmm}^* \) based on the moment conditions (2.1), (C.23) and (C.24).

We next provide the key steps for showing (C.27). Since \( \tilde{R}_t = X_t\theta + v_t \), we have

\[ \tilde{R} - \tilde{X}\theta = T^{-1} \sum_{t \leq T} D_{L,t}(\tilde{R}_t - \tilde{X}_t\theta) = T^{-1} \sum_{t \leq T} D_{L,t}v_t - D_{L}(\tilde{B} - B)\gamma_1 \]  

(C.29)

\[ \text{To avoid repetition, we only provide the key arguments for deriving (C.27), which are presented at the end of this section.} \]
where $\overline{D_I} \equiv T^{-1} \sum_{t \leq T} D_{I,t}$. By the consistency of $\tilde{X}, \tilde{W}$ and $\overline{D_I}$,

$$T^{1/2}(\hat{\theta}_{csr} - \theta) = (\tilde{X}'\tilde{W}\tilde{X})^{-1}\tilde{X}'\tilde{W}\left(T^{-1/2} \sum_{t \leq T} D_{I,t} v_t - T^{1/2}\overline{D_I}(\hat{B} - B)\gamma_1 \right)$$

$$= (X_d'W_dX_d)^{-1}X_d'W\left(T^{-1/2} \sum_{t \leq T} D_{I,t} v_t - T^{1/2}D_{I}(\hat{B} - B)\gamma_1 \right) + o_p(1)$$

(C.30)

where $D_I \equiv E[D_{I,t}]$.

To study the estimation error of $\hat{B}$, we define $\hat{\Sigma}_{f,i} \equiv T^{-1} \sum_{t \leq T} d_{i,t}(f_t - \bar{f})(f_t - \bar{f})'$. Then

$$\beta_i' = \left( T^{-1} \sum_{t \leq T} d_{i,t} R_{i,t}(f_t - \bar{f}) \right)' \hat{\Sigma}_{f,i}^{-1}$$

$$= \beta'_i + \left( T^{-1} \sum_{t \leq T} \eta_{i,t}(f_t - \mu_f) \right)' \hat{\Sigma}_{f,i}^{-1} - \left( T^{-1} \sum_{t \leq T} \eta_{i,t} (\bar{f} - \mu_f) \right)' \hat{\Sigma}_{f,i}^{-1}$$

$$+ \beta_i'(\bar{f} - \mu_f) \left( T^{-1} \sum_{t \leq T} d_{i,t}(f_t - \bar{f}) \right)' \hat{\Sigma}_{f,i}^{-1}$$

(C.31)

which together with the missing at random assumption implies that the dominating term in the estimation error $(\hat{\beta}_i - \beta_i)'$ is

$$T^{-1} \sum_{t \leq T} \eta_{i,t}(f_t - \mu_f)'\Sigma_f^{-1}.$$  

Therefore,

$$T^{1/2}(\hat{B} - B) = D_I^{-1}T^{-1/2} \sum_{t \leq T} \eta_t(f_t - \mu_f)'\Sigma_f^{-1} + o_p(1).$$

(C.32)

Using (C.30) and (C.32), we have

$$T^{1/2}(\hat{\theta}_{csr} - \theta) = (X_d'W_dX_d)^{-1}X_d'W\left(T^{-1/2} \sum_{t \leq T} (D_{I,t} v_t - \eta_t(f_t - \mu_f)'\Sigma_f^{-1}\gamma_1) \right) + o_p(1)$$

which together with the definition of $\zeta_t$ shows (C.27).
## 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>$\gamma_0$</td>
<td>$\gamma_{mkt}$</td>
</tr>
<tr>
<td>GMM Bias</td>
<td>0.010</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.141</td>
</tr>
<tr>
<td>MAE</td>
<td><strong>0.115</strong></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><strong>0.154</strong></td>
</tr>
<tr>
<td>MAE</td>
<td><strong>0.123</strong></td>
</tr>
<tr>
<td>GLS Bias</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 Bias</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>$\gamma_0$</td>
<td>$\gamma_{mkt}$</td>
</tr>
<tr>
<td>GMM Bias</td>
<td>0.024</td>
</tr>
<tr>
<td>RMSE</td>
<td><strong>0.128</strong></td>
</tr>
<tr>
<td>MAE</td>
<td><strong>0.101</strong></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 Bias</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 Bias</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>
Table D.2: Simulated Bias, RMSE, and MAE for Parameter Estimates, PC Analysis

We use 119 long-short portfolios in Chen and Zimmermann (2020) as test assets. The true factor model (either three-factor ‘PC-3’ or five-factor ‘PC-5’) is constructed from a factor analysis on the 119 portfolios. For a given $N$, we simulate $M = 1,000$ times, each time drawing $N$ distinct portfolios from the 119. We report the average statistics across the $M$ simulations.

<table>
<thead>
<tr>
<th>Panel A: PC-3 Model, $N = 20$</th>
<th>Panel B: PC-5 Model, $N = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>$\gamma$</td>
</tr>
<tr>
<td>Panel C: PC-3 Model, $N = 30$</td>
<td>Panel D: PC-5 Model, $N = 30$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$\gamma$</td>
</tr>
<tr>
<td>Panel E: PC-3 Model, $N = 50$</td>
<td>Panel F: PC-5 Model, $N = 50$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$\gamma$</td>
</tr>
<tr>
<td>Panel G: PC-3 Model, $N = 75$</td>
<td>Panel H: PC-5 Model, $N = 75$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$\gamma$</td>
</tr>
<tr>
<td>Panel I: PC-3 Model, $N = 100$</td>
<td>Panel J: PC-5 Model, $N = 100$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$\gamma$</td>
</tr>
</tbody>
</table>