Testing Ex-post Implications of Asset Pricing Models using Individual Stocks*

Soohun Kim[†]
Georgia Institute of Technology

Georgios Skoulakis[‡]
University of British Columbia

August 11, 2017

*We are grateful to Markus Baldauf, Murray Carlson, Adlai Fisher, Neil Galpin, Will Gornall, Ravi Jagannathan, Raymond Kan, Hernan Ortiz-Molina, Elena Pikulina, Chen Xue, seminar participants at KAIST, the University of British Columbia, and HEC Montreal, as well as conference participants at the Society for Financial Econometrics 2016 Conference (Hong Kong), the Financial Econometrics & Empirical Asset Pricing 2016 Conference (Lancaster), the European Finance Association 2016 Meeting (Oslo), the Northern Finance Association 2016 Meeting (Mont Tremblant), and the Midwest Finance Association 2017 Meeting (Chicago) for their comments and suggestions. We also thank Chen Xue for providing factor data. Any errors are our responsibility. Comments and suggestions are welcome at `soohun.kim@scheller.gatech.edu` and `georgios.skoulakis@sauder.ubc.ca`.

†Finance Department, Scheller College of Business, Georgia Institute of Technology, Atlanta, GA 30308; Email: `soohun.kim@scheller.gatech.edu`.

‡Finance Department, Sauder School of Business, University of British Columbia, Vancouver, BC, Canada V6T 1Z2; Email: `georgios.skoulakis@sauder.ubc.ca`.

Testing Ex-post Implications of Asset Pricing Models using Individual Stocks

Abstract

This paper develops an over-identified IV approach that uses past beta estimates and firm characteristics as instruments for estimating ex-post risk premia while addressing the error-in-variables problem in the two-pass cross-sectional regression method. The approach is developed in the context of large cross sections of individual stocks and short time series. We establish the N-consistency of the resulting IV ex-post risk premia estimator and obtain its asymptotic distribution along with an estimator of its asymptotic variance-covariance matrix. These results are then used to develop new tests for asset pricing model implications. Empirically, we examine a number of popular asset pricing models and find support for the recent q-factor model proposed by Hou, Xue, and Zhang (2015).

Keywords: Error-in-variables problem, instrumental variables, individual stocks, N-consistent ex-post risk premia estimator, asset pricing tests.

JEL Classification Codes: C36, C55, C58, G12.

1 Introduction

Asset pricing models suggest that an asset's average return should be related to its exposure to systematic risk. Models differ in the factors they identify as sources of relevant systematic risk. A typical model identifies a small number of pervasive risk factors and postulates that the average return on an asset is a linear function of the factor betas. The quest for the identification of relevant risk factors at the theoretical level can be traced back to the works of Sharpe (1964), Lintner (1965) and Mossin (1966) on the CAPM and Ross (1976) on the APT. On the empirical front, a long line of research on the evaluation of such models has been developed, starting with Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973).

One important aspect of empirical evaluation of an asset pricing model involves determining the cross section of test assets. On standard approach in the literature, introduced by Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973), is to perform the asset pricing tests on a small number of portfolios. Indeed, following Fama and French (1992), it has become standard practice to sort stocks according to some firm characteristic in order to form sets of portfolios, typically deciles, that are subsequently used as test assets. However, as Lewellen, Nagel, and Shanken (2010) and Daniel and Titman (2012) demonstrate, inference regarding the performance of an asset pricing model crucially depends on the choice of test assets.The method used to form the test portfolios could indeed affect the inference results in undesirable ways. As Roll (1977) points out, in the process of forming portfolios, important mispricing in individual stocks can be averaged out within portfolios, making it harder to reject the wrong model. Lo and MacKinlay (1990) are concerned about the exact opposite error: if stocks are grouped into portfolios with respect to attributes already observed to be related to average returns, the correct model may be rejected too often. In a recent contribution to the literature, Kogan and Tian (2015) question the standard practice used in the literature to form portfolio deciles by sorting firms on various characteristics, construct factors as long-short portfolio spreads, and finally using the portfolio deciles as test assets. They point out that, by searching through the firm characteristics known to be associated with substantial spreads in stock returns, it is easy to construct seemingly successful empirical factor pricing models and argue that factor model mining can be a serious concern.

Motivated by this findings, we develop a framework for estimating and evaluating asset pricing factor models using large cross sections of individual stock return data, instead of employing portfolios as test assets, as originally suggested by Litzenberger and Ramaswamy (1979). We only consider short time horizons so that we can evaluate the implications of asset pricing models locally in time, and, hence, we restrict ourselves to factors that are traded portfolio returns or spreads focusing on ex-post risk premia (see Shanken (1992)).

The existing methodological literature on the estimation and evaluation of asset pricing models mainly focuses the case in which the time-series sample size, T, is large while the size of the cross section of test assets, N, is small. This scenario is suitable when portfolios, as opposed to individual stocks, are used as test assets.[1] The analysis of linear asset pricing factor models when the number of test assets N is large has been the subject of a few recent papers. Gagliardini, Ossola, and Scaillet (2012) extend the two-pass cross-sectional methodology to the case of a conditional factor model incorporating firm characteristics. Their asymptotic theory, based on N and T jointly increasing to infinity at suitable rates, facilitates studying time varying risk premia. Chordia, Goyal, and Shanken (2015), building on Shanken (1992), use bias-corrected risk premia estimates in a context with individual stocks and time variation in the betas through macroeconomic variables and firm characteristics. Their focus is the relative contribution of betas and characteristics in explaining cross-sectional differences in conditional expected returns. More closely related to our paper is the recent paper by Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2015) which employs an instrumental variable approach to deal with the EIV problem in the risk premia estimation using individual stocks, where the instruments are betas estimated over separate time periods. However, they do not offer an estimator of the variance-covariance matrix of the risk premia estimator. Instead, they resort to the original Fama-MacBeth approach for computing standard errors and test statistics, as used in the large T case, and, hence, ignore the error-in-variables (EIV) problem in the estimation of the variance-covariance matrix (Shanken (1992) and Jagannathan and Wang (1998)).

We contribute to the extant literature by developing an instrumental variable-generalized method of moments (IV-GMM) approach for estimating ex-post risk premia when the number of assets, N, tends to infinity while the time-series length T is fixed. In the standard two-pass procedure used for estimating risk premia, the second step is a regression of average returns on estimated betas. As explained in Section 6 in Shanken (1992), when T is fixed and N tends to infinity, the orthogonality condition required for consistency in the second pass is *not* satisfied rendering the two-pass CSR estimator inconsistent. This is a manifestation of the well-known EIV problem which emerges from using beta estimates instead of the true betas. Our approach uses past beta estimates and firm characteristics as instrumental variables in order to deal with the EIV problem. We establish that the overidentified IV-GMM ex-post risk estimator is N-

[1]The long list of related papers includes, among others, Gibbons (1982), Shanken (1985), Connor and Korajczyk (1988), Lehmann and Modest (1988), Gibbons, Ross, and Shanken (1989), Harvey (1989), Lo and MacKinlay (1990), Zhou (1991), Shanken (1992), Connor and Korajczyk (1993), Zhou (1993), Zhou (1994), Berk (1995), Hansen and Jagannathan (1997), Ghysels (1998), Jagannathan and Wang (1998), Kan and Zhou (1999), Jagannathan and Wang (2002), Chen and Kan (2004), Lewellen and Nagel (2006), Shanken and Zhou (2007), Kan and Robotti (2009), Hou and Kimmel (2010), Lewellen, Nagel, and Shanken (2010), Nagel and Singleton (2011), Ang and Kristensen (2012), Kan, Gospodinov, and Robotti (2013) and Kan, Robotti, and Shanken (2013).

consistent and show that it asymptotically follows a normal distribution. Finally, incorporating a cluster structure for idiosyncratic shock correlations, we obtain an N-consistent estimator of the asymptotic variance-covariance matrix which we use to develop statistics for testing asset pricing model implications. There are significant differences between our paper and the aforementioned paper by Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2015), on which we elaborate in Section 2. Importantly, in addition to beta estimates from past periods, we use firm characteristics as additional instruments. Furthermore, we provide a fully operational asymptotic theory for the IV-GMM estimator that we use to build asset pricing tests.

Litzenberger and Ramaswamy (1979) initiated the line of research on N-consistent risk premia estimators which was continued by Shanken (1992) and Jagannathan, Skoulakis, and Wang (2010). However, these papers do not provide inference tools as they do not address the issue of the sampling distribution of risk premia estimators. The recent paper by Kim and Skoulakis (2015), using the regression-calibration approach, obtain the asymptotic distribution of the ex-post risk premia estimator along with an estimator of its variance-covariance matrix based on which they construct asset pricing tests. In this paper, we offer an alternative approach based on the use of instrumental variables for addressing the error-in-variables problem.

We examine the performance of the IV-GMM ex-post risk premia estimator in a number of Monte Carlo simulation experiments. In our empirical investigation, we use the IV-GMM estimator to test the implications of four popular asset pricing model: the CAPM, the Fama and French (1993) three-factor model (FF3), the Hou, Xue, and Zhang (2015) four-factor model (HZX4) and the Fama and French (2015) five-factor model (FF5). To make them relevant for our empirical exercise, we calibrate our simulations to the CAPM, the FF3 model and the HXZ4 model. The simulation results clearly show the significant bias reduction in the cross-sectional regression intercept and ex-post risk premia estimates achieved by the IV-GMM approach and the good performance of our asset pricing tests for relevant sample sizes. Empirically, we find that the CAPM, the FF3 model and the FF5 model are mostly rejected by the IV-GMM test statistics in our sample. In contrast, we find evidence in favor of the HXZ4 model, for which we find strong support in five out of eight periods under all alternative clustering schemes.

The rest of the paper is organized as follows. In Section 2, we describe the general econometric framework and develop the IV-GMM ex-post risk premia estimator using past beta estimates and firm characteristics as instruments. We further establish the N-consistency of the IV-GMM estimator, obtain its asymptotic distribution, provide an estimator of its asymptotic variance-covariance matrix and develop novel asset pricing tests. In Section 3, we provide Monte Carlo evidence on the finite sample behavior of the IV-GMM estimator and the associated tests. Section 4 presents empirical evidence on four popular asset pricing models. Finally, Section 5 concludes. Proofs are collected in the Appendix and additional results are delegated to the Online Appendix.

2 Econometric Framework

2.1 Model specification

Consider an economy with N traded assets and K factors. Each factor is assumed to be a portfolio return spread. Let $\mathbf{r}_t = [\ r_{1,t} \ \cdots \ r_{N,t}\ ]'$ be the vector of returns of the N traded assets in excess of the risk-free return and $\mathbf{f}_t = [\ f_{1,t} \ \cdots \ f_{K,t}\ ]'$ be the vector of factor realizations at time t. We assume that data are available over times 1 through T, where T is finite and fixed, and formally consider the case in which the number of assets, N, tends to infinity. Given that the time-series sample size is fixed, the uncertainty about the factors cannot be resolved. Hence, our analysis is conducted conditionally on the factor realizations.

We refer to the periods covering times 1 through τ_1 and $\tau_1 + 1$ through $T = \tau_1 + \tau_2$ as the pretesting and testing periods, respectively. That is, τ_1 and τ_2, that are fixed throughout our analysis, are the pretesting and testing time-series sample sizes. We are interested in testing the implications of an asset pricing model over the period from time $\tau_1 + 1$ through $T = \tau_1 + \tau_2$.

The expectations of the excess return $\mathbf{r}_t$ and the factor $\mathbf{f}_t$ are denoted by $\boldsymbol{\mu}_r = \mathrm{E}[\mathbf{r}_t]$ and $\boldsymbol{\mu}_f = \mathrm{E}[\mathbf{f}_t]$, respectively. Furthermore, the $K \times K$ factor variance-covariance matrix is denoted by $\boldsymbol{\Sigma}_f = \mathrm{E}[(\mathbf{f}_t - \boldsymbol{\mu}_f)\,(\mathbf{f}_t - \boldsymbol{\mu}_f)']$, while the $N \times K$ excess return-factor covariance matrix is denoted by $\boldsymbol{\Sigma}_{rf} = \mathrm{E}[(\mathbf{r}_t - \boldsymbol{\mu}_r)(\mathbf{f}_t - \boldsymbol{\mu}_f)']$. The $N \times K$ beta matrix is then defined by

$$\mathbf{B} = [\ \boldsymbol{\beta}_1 \ \cdots \ \boldsymbol{\beta}_N\ ]' = \boldsymbol{\Sigma}_{rf}\boldsymbol{\Sigma}_f^{-1}, \tag{1}$$

where $\boldsymbol{\beta}_i$ denotes the beta vector for the i-th asset, $i = 1, \dots, N$. Given that the factors comprising $\mathbf{f}_t$ belong to the return space, the risk premia vector equals the vector of factor expectations $\boldsymbol{\mu}_f$, and, hence, the corresponding linear beta pricing model implies that $\boldsymbol{\mu}_r = \mathbf{B}\boldsymbol{\mu}_f$.

Defining the residual $\mathbf{u}_t = \mathbf{r}_t - \mathbf{B}\mathbf{f}_t$, we can then write $\mathbf{u}_t = (\mathbf{r}_t - \boldsymbol{\mu}_r) - \mathbf{B}\,(\mathbf{f}_t - \boldsymbol{\mu}_f)$, which implies $\mathrm{E}[\mathbf{u}_t] = \mathbf{0}_N$ and $\mathrm{E}\left[\mathbf{u}_t\mathbf{f}_t'\right] = \mathrm{E}\left[(\mathbf{r}_t - \boldsymbol{\mu}_r)\,\mathbf{f}_t' - \mathbf{B}\,(\mathbf{f}_t - \boldsymbol{\mu}_f)\,\mathbf{f}_t'\right] = \boldsymbol{\Sigma}_{rf} - \mathbf{B}\boldsymbol{\Sigma}_f = \mathbf{0}_{N\times K}$, where $\mathbf{0}_N$ and $\mathbf{0}_{N\times K}$ denote the $N \times 1$ vector and $N \times K$ matrix of zeros, respectively. Hence, we obtain the following time-series regression representation:

$$\mathbf{r}_t = \mathbf{B}\mathbf{f}_t + \mathbf{u}_t, \ \text{with} \quad \mathrm{E}[\mathbf{u}_t] = \mathbf{0}_N, \quad \mathrm{E}[\mathbf{u}_t\mathbf{f}_t'] = \mathbf{0}_{N\times K}. \tag{2}$$

Over the testing period, covering times $t = \tau_1 + 1, \ldots, \tau_1 + \tau_2$, the data generating process in (2) implies that

$$\bar{\mathbf{r}}_2 = \mathbf{1}_N \lambda_0 + \mathbf{B}\boldsymbol{\lambda}_f + \bar{\mathbf{u}}_2 \tag{3}$$

with

$$\lambda_0 = 0, \quad \boldsymbol{\lambda}_f = \bar{\mathbf{f}}_2 = \frac{1}{\tau_2} \sum\nolimits_{t=\tau_1+1}^{\tau_1+\tau_2} \mathbf{f}_t,$$

where

$$\bar{\mathbf{r}}_2 = \frac{1}{\tau_2} \sum\nolimits_{t=\tau_1+1}^{\tau_1+\tau_2} \mathbf{r}_t, \quad \bar{\mathbf{u}}_2 = \frac{1}{\tau_2} \sum\nolimits_{t=\tau_1+1}^{\tau_1+\tau_2} \mathbf{u}_t. \tag{4}$$

Recall that, since T is finite and fixed, our analysis is conducted conditionally on the factor realizations. Hence, following Shanken (1992), among others, we refer to $\boldsymbol{\lambda}_f$ as the ex-post risk premia. When the linear factor model holds, the vector of ex-post risk premia $\boldsymbol{\lambda}_f$ equals the average factor realization over the testing period, namely $\bar{\mathbf{f}}_2$, given that the factors belong to the return space. The object of our inference is the $(K + 1) \times 1$ vector

$$\boldsymbol{\lambda} = [\ \lambda_0 \quad \boldsymbol{\lambda}_f'\ ]'. \tag{5}$$

Defining the $N \times (K + 1)$ matrix $\mathbf{X}$ by

$$\mathbf{X} = [\ \mathbf{1}_N \quad \mathbf{B}\ ], \tag{6}$$

we can rewrite equation (3) as

$$\bar{\mathbf{r}}_2 = \mathbf{1}_N \lambda_0 + \mathbf{B}\boldsymbol{\lambda}_f + \bar{\mathbf{u}}_2 = \mathbf{X}\boldsymbol{\lambda} + \bar{\mathbf{u}}_2. \tag{7}$$

If the true beta matrix $\mathbf{B}$ were known, an N-consistent estimator of $\boldsymbol{\lambda}$ could be obtained by regressing the average excess return vector $\bar{\mathbf{r}}_2$ on a vector of ones and the beta matrix $\mathbf{B}$, under the reasonable assumption of zero limiting cross-sectional correlation between the betas and the shocks. However, the beta matrix $\mathbf{B}$ is not known and has to be estimated using the available data. Natural proxies for $\mathbf{B}$ are the time-series OLS estimators of $\mathbf{B}$ obtained using data from the pretesting period or the testing period. When τ_1 and τ_2 are fixed, as in our framework, the two-pass CSR approach with either proxy yields an inconsistent estimator.

This is a manifestation of the well-known EIV problem as pointed out in Shanken (1992), Jagannathan, Skoulakis, and Wang (2010), and Kim and Skoulakis (2015).

Various approaches have been advanced in the statistics and econometrics literature for dealing with the EIV problem. One such approach, particularly suitable for the case in which multiple proxies of unobserved quantities are available, is the instrumental variable (IV) approach. Starting with the early works of Wald (1940), Reiersøl (1941), and Geary (1943), a long related literature was subsequently developed.[2] Chapter 6 of Carroll, Ruppert, Stefanski, and Crainiceanu (2006) offers a comprehensive account of the IV method, where they state that "One possible source of an instrumental variable is a second, possibly biased, measurement of the (true unobserved) regressor obtained by an independent measuring method."[3] The recent paper by Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2015) also uses an IV approach in the context of asset pricing tests. Our paper differ from the aforementioned paper in a number of important aspects. First, we use beta estimates obtained in the pretesting period and therefore, given that our pretesting and testing periods are non-overlapping and consecutive, our approach is less prone to potential serial correlations in the real data. Second, in addition to past beta estimates, we also use firm characteristics as instruments. As a result, our estimator is an overidentified IV-GMM estimator and not a standard two-stage IV estimator. Third, since we aim to develop asset pricing tests using individual stocks with a focus placed on ex-post risk premia, we exclusively focus on the case of fixed T and large N. Finally, we develop a fully operational asymptotic theory of the IV risk premia estimator. That is, we show its consistency and asymptotic normality as N tends to infinity and, furthermore, construct an estimator of its asymptotic variance-covariance matrix that we finally use to develop novel asset pricing tests.

To develop the IV risk premia estimator, we need to introduce some notation. Define the $N \times \tau_1$ excess return matrix $\mathbf{R}_1$ and the $K \times \tau_1$ factor realization matrix $\mathbf{F}_1$, over the pretesting period, the $N \times \tau_2$ excess return matrix $\mathbf{R}_2$ and the $K \times \tau_2$ factor realization matrix $\mathbf{F}_2$, over the testing period, by

$$\mathbf{R}_1 = [\ \mathbf{r}_1 \quad \cdots \quad \mathbf{r}_{\tau_1}\ ], \quad \mathbf{F}_1 = [\ \mathbf{f}_1 \quad \cdots \quad \mathbf{f}_{\tau_1}\ ] \tag{8}$$

and

$$\mathbf{R}_2 = [\ \mathbf{r}_{\tau_1+1} \quad \cdots \quad \mathbf{r}_{\tau_1+\tau_2}\ ], \quad \mathbf{F}_2 = [\ \mathbf{f}_{\tau_1+1} \quad \cdots \quad \mathbf{f}_{\tau_1+\tau_2}\ ]. \tag{9}$$

[2] Durbin (1954) provides a review of the early EIV literature. Aldrich (1993) offers a historical account of the development of the IV approach to the EIV problem in the 1940s.

[3] In the Online Appendix, we illustrate how the IV method can be used in the context of a linear regression model with regressors subject to the EIV problem.

Then, using the quantities defined in (8) and (9), we express the time-series OLS estimators of the beta matrix $\mathbf{B}$ over the pretesting and testing periods, denoted by $\widehat{\mathbf{B}}_1$ and $\widehat{\mathbf{B}}_2$, respectively, as follows:

$$\widehat{\mathbf{B}}_1 = (\mathbf{R}_1 \mathbf{J}_{\tau_1} \mathbf{F}_1') (\mathbf{F}_1 \mathbf{J}_{\tau_1} \mathbf{F}_1')^{-1}, \quad \widehat{\mathbf{B}}_2 = (\mathbf{R}_2 \mathbf{J}_{\tau_2} \mathbf{F}_2') (\mathbf{F}_2 \mathbf{J}_{\tau_2} \mathbf{F}_2')^{-1}, \tag{10}$$

where $\mathbf{J}_m = \mathbf{I}_m - \frac{1}{m} \mathbf{1}_m \mathbf{1}_m'$, with $\mathbf{I}_m$ and $\mathbf{1}_m$ denoting the $m \times m$ identity matrix and the $m \times 1$ vector of ones, respectively, for any positive integer m.[4]

We introduce the $N \times \tau_1$ idiosyncratic shock matrix $\mathbf{U}_1$, over the pretesting period, and the $N \times \tau_2$ idiosyncratic shock matrix $\mathbf{U}_2$, over the testing period, defined by

$$\mathbf{U}_1 = [\ \mathbf{u}_1 \ \cdots \ \mathbf{u}_{\tau_1}\ ], \quad \mathbf{U}_2 = [\ \mathbf{u}_{\tau_1+1} \ \cdots \ \mathbf{u}_{\tau_1+\tau_2}\ ]. \tag{11}$$

Alternatively, letting $\mathbf{u}_{1,[i]}'$ and $\mathbf{u}_{2,[i]}'$ denote the i-th row of $\mathbf{U}_1$ and $\mathbf{U}_2$, respectively, for $i = 1, \ldots, N$, we can write

$$\mathbf{U}_1 = [\ \mathbf{u}_{1,[1]} \ \cdots \ \mathbf{u}_{1,[N]}\ ]', \quad \mathbf{U}_2 = [\ \mathbf{u}_{2,[1]} \ \cdots \ \mathbf{u}_{2,[N]}\ ]'. \tag{12}$$

Observe that $\overline{\mathbf{u}}_2$, the disturbance term of the equation in (7), and $\mathbf{U}_2$ satisfy the following relationship:

$$\overline{\mathbf{u}}_2 = \frac{1}{\tau_2} \mathbf{U}_2 \mathbf{1}_{\tau_2}. \tag{13}$$

Noting that $\mathbf{R}_1 = \mathbf{B}\mathbf{F}_1 + \mathbf{U}_1$ and $\mathbf{R}_2 = \mathbf{B}\mathbf{F}_2 + \mathbf{U}_2$, we can decompose the beta estimators $\widehat{\mathbf{B}}_1$ and $\widehat{\mathbf{B}}_2$, defined in (10), respectively, into the true beta matrix $\mathbf{B}$ and the corresponding estimation error terms as follows:

$$\widehat{\mathbf{B}}_1 = \mathbf{B} + \mathbf{U}_1 \mathbf{G}_1, \quad \widehat{\mathbf{B}}_2 = \mathbf{B} + \mathbf{U}_2 \mathbf{G}_2, \tag{14}$$

where the $\tau_1 \times K$ matrix $\mathbf{G}_1$ and the $\tau_2 \times K$ matrix $\mathbf{G}_2$ are defined by

$$\mathbf{G}_1 = \mathbf{J}_{\tau_1} \mathbf{F}_1' (\mathbf{F}_1 \mathbf{J}_{\tau_1} \mathbf{F}_1')^{-1}, \quad \mathbf{G}_2 = \mathbf{J}_{\tau_2} \mathbf{F}_2' (\mathbf{F}_2 \mathbf{J}_{\tau_2} \mathbf{F}_2')^{-1}. \tag{15}$$

To illustrate the effect of the beta estimation error in the case that we use $\widehat{\mathbf{B}}_2$ as a beta matrix proxy in the second-pass CSR, we observe that equation (7) is reexpressed as $\overline{\mathbf{r}}_2 =$

[4] Standard matrix algebra shows that $\mathbf{J}_m$ is a symmetric and idempotent matrix, and that $\mathrm{tr}(\mathbf{J}_m) = m - 1$.

$\widehat{\mathbf{X}}_2\boldsymbol{\lambda} + (\mathbf{X} - \widehat{\mathbf{X}}_2)\boldsymbol{\lambda} + \bar{\mathbf{u}}_2$ or

$$\bar{\mathbf{r}}_2 = \widehat{\mathbf{X}}_2\boldsymbol{\lambda} + \boldsymbol{\omega}_2, \tag{16}$$

where

$$\widehat{\mathbf{X}}_2 = [\ \mathbf{1}_N \quad \widehat{\mathbf{B}}_2\ ], \tag{17}$$

$$\boldsymbol{\omega}_2 = (\mathbf{X} - \widehat{\mathbf{X}}_2)\boldsymbol{\lambda} + \bar{\mathbf{u}}_2. \tag{18}$$

Using equations (13) and (14), the disturbance term $\boldsymbol{\omega}_2$ defined in (18) can be expressed as

$$\boldsymbol{\omega}_2 = -(\mathbf{U}_2\mathbf{G}_2)\boldsymbol{\lambda}_f + \bar{\mathbf{u}}_2 = \mathbf{U}_2\left(\frac{1}{\tau_2}\mathbf{1}_{\tau_2} - \mathbf{G}_2\boldsymbol{\lambda}_f\right) = \mathbf{U}_2\mathbf{g}_2, \tag{19}$$

where the $\tau_2 \times 1$ vector of $\mathbf{g}_2$ is given by

$$\mathbf{g}_2 = \frac{1}{\tau_2}\mathbf{1}_{\tau_2} - \mathbf{G}_2\boldsymbol{\lambda}_f = \frac{1}{\tau_2}(\mathbf{I}_{\tau_2} - \mathbf{J}_{\tau_2}\mathbf{F}_2'(\mathbf{F}_2\mathbf{J}_{\tau_2}\mathbf{F}_2')^{-1}\mathbf{F}_2)\mathbf{1}_{\tau_2}. \tag{20}$$

It follows from equations (17) and (18) that the orthogonality condition, necessary for consistency, is violated in the cross-sectional regression (16). In the next subsection, we start our analysis by developing a consistent ex-post risk premia estimator using an instrumental variable approach.

2.2 Estimating ex-post risk premia

It follows from the expressions (14) and (19) that the regressor and disturbance terms in the cross-sectional regression (16) are correlated through the beta estimation error contained in $\widehat{\mathbf{B}}_2$. Hence, ignoring the error-in-variables problem, one would obtain an inconsistent ex-post risk premia estimator. We develop an instrumental variable approach to deal with the error-in-variables problem using past beta estimates and firm characteristics as instruments. Next, we explain that, under mild assumptions, $\widehat{\mathbf{B}}_1$ can serve as an instrumental variable for constructing an N-consistent estimator of $\boldsymbol{\lambda}$ using the cross-sectional regression (16).

Assumption 1 **(i)** *As* $N \to \infty$, $\frac{1}{N}\mathbf{U}'\mathbf{1}_N \xrightarrow{p} \mathbf{0}_T$ *and* $\frac{1}{N}\mathbf{U}'\mathbf{B} \xrightarrow{p} \mathbf{0}_{T\times K}$, *where* $\mathbf{U} = [\ \mathbf{U}_1 \quad \mathbf{U}_2\ ]$. **(ii)** *As* $N \to \infty$, $\frac{1}{N}\mathbf{U}_1'\mathbf{U}_2 \xrightarrow{p} \mathbf{0}_{\tau_1\times\tau_2}$. **(iii)** *As* $N \to \infty$, $\frac{1}{N}\mathbf{B}'\mathbf{1}_N = \frac{1}{N}\sum_{i=1}^N \boldsymbol{\beta}_i \to \boldsymbol{\mu}_\beta$. **(iv)** *As* $N \to \infty$, $\frac{1}{N}\sum_{i=1}^N(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)' \to \mathbf{V}_\beta$, *where* $\mathbf{V}_\beta$ *is a symmetric and positive definite matrix.*

Assumption 1(i) states that, at each time t, the cross-sectional average of the shocks $u_{i,t}$ converges to zero, and the limiting cross-sectional correlation between the shocks $u_{i,t}$ and the betas $\boldsymbol{\beta}_i$ is also zero, as the number of assets N tends to ∞. For $1 \leq t \leq \tau_1$ and $\tau_1 + 1 \leq t' \leq \tau_1 + \tau_2$, Assumption 1(ii) states that the limiting cross-sectional correlation between $u_{i,t}$ and $u_{i,t'}$ vanishes. That is, the pretesting-period and the testing-period shocks are assumed to be cross-sectionally uncorrelated in the limit $N \to \infty$.[5] Assumption 1(iii) states that the limiting cross-sectional average of the betas $\boldsymbol{\beta}_i$ exists while Assumption 1(iv) states that the limiting cross-sectional variance of the betas $\boldsymbol{\beta}_i$ exists and is a symmetric and positive definite matrix.

In light of equation (14), it follows from Assumption 1 that $\widehat{\mathbf{B}}_1$ is correlated with the explanatory variable $\widehat{\mathbf{B}}_2$ in the cross-sectional regression (16), in the sense that $\widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2/N$ converges in probability to $\mathbf{M}_\beta = \mathbf{V}_\beta + \boldsymbol{\mu}_\beta\boldsymbol{\mu}_\beta'$, which is symmetric and positive definite, and hence invertible matrix, as $N \to \infty$. Furthermore, based on equations (14) and (19), Assumptions 1(i) and 1(ii) imply that the proposed instrumental variable $\widehat{\mathbf{B}}_1$ is uncorrelated with the disturbance term $\boldsymbol{\omega}_2$ in the cross-sectional regression (16), in the sense that $\widehat{\mathbf{B}}_1'\boldsymbol{\omega}_2/N \xrightarrow{p} \mathbf{0}_K$, as $N \to \infty$. These properties are formally established in Theorem 1 below, where we establish the N-consistency of the proposed IV-GMM estimator.

In addition to the beta estimates obtained in the pretesting period, we also employ firm characteristics as instrumental variables. We elaborate on the validity of the chosen characteristics as instruments on a model by model basis in the empirical application. For the purposes of the theoretical development, we make the following assumption. Let $\mathbf{C}_1$ denote the $N \times L$ matrix of characteristics observed in the pretesting period, and $\mathbf{c}_{1i}'$ be the i-th row of $\mathbf{C}_1$, $i = 1, \ldots, N$.

Assumption 2 **(i)** *As* $N \to \infty$, $\mathbf{C}_1'\mathbf{U}_1/N \xrightarrow{p} \mathbf{V}_{cu}$, *where* $\mathbf{V}_{cu}$ *is an* $L \times \tau_1$ *matrix.* **(ii)** *As* $N \to \infty$, $\mathbf{C}_1'\mathbf{U}_2/N \xrightarrow{p} \mathbf{0}_{L\times\tau_2}$. **(iii)** *As* $N \to \infty$, $\mathbf{C}_1'\mathbf{1}_N/N = \frac{1}{N}\sum_{i=1}^N \mathbf{c}_{1,i} \to \boldsymbol{\mu}_c$. **(iv)** *As* $N \to \infty$, $\frac{1}{N}\sum_{i=1}^N(\mathbf{c}_{1,i} - \boldsymbol{\mu}_c)(\mathbf{c}_{1,i} - \boldsymbol{\mu}_c)' \to \mathbf{V}_c$, *where* $\mathbf{V}_c$ *is a symmetric and positive definite matrix. (iv) As* $N \to \infty$, $\mathbf{C}_1'\mathbf{B}/N \xrightarrow{p} \mathbf{M}_{c\beta}$, *where* $\mathbf{M}_{c\beta}$ *is an* $L \times K$ *matrix.*

Assumptions 2(i) and 2(ii) state that firm characteristics observed in the pretesting period are potentially correlated with idiosyncratic shocks in the pretesting period but not with those in the testing period. In light of equation (19), Assumption 2(ii) states that $\mathbf{C}_1$ is uncorrelated with the disturbance term $\boldsymbol{\omega}_2$. Assumptions 2(iii) and 2(iv) state that the first two cross-sectional moments of the firm characteristics are well defined. Finally, Assumption 2(v) states that the firm characteristics observed in the pretesting period are correlated with the true betas.

[5] As long as the pretesting and testing periods do not overlap and the shocks over the two periods are cross-sectionally uncorrelated when $N \to \infty$, the IV approach would provide valid inference. In our analysis, we consider the two periods to be consecutive so as to mitigate the effect of potential serial correlation in the real data.

In our empirical applications, the factors are returns on spread portfolios constructed after sorting stocks with respect to a certain firm characteristic, such as size and book-to-market ratio. In this context, it is expected that characteristics and betas with respect to the corresponding spread are highly correlated. We indeed provide evidence that this is the case in Section 4, where we empirically evaluate a number of popular asset pricing models.

Under the aforementioned assumptions, and in particular Assumptions 1 (ii) and 2 (ii), the past beta estimates $\widehat{\mathbf{B}}_1$ and the characteristics $\mathbf{C}_1$ can be used as instruments in the estimation of ex-post risk premia, giving rise to the following overidentified IV-GMM estimator:

$$\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{GMM}} = \left[(\widehat{\mathbf{X}}_2'\widehat{\mathbf{Z}}_1)\widehat{\mathbf{W}}(\widehat{\mathbf{Z}}_1'\widehat{\mathbf{X}}_2) \right]^{-1} (\widehat{\mathbf{X}}_2'\widehat{\mathbf{Z}}_1)\widehat{\mathbf{W}}(\widehat{\mathbf{Z}}_1'\bar{\mathbf{r}}_2), \tag{21}$$

where the $N \times (1 + K + L)$ instrument matrix $\widehat{\mathbf{Z}}_1$ is defined by

$$\widehat{\mathbf{Z}}_1 = [\ \mathbf{1}_N \quad \widehat{\mathbf{B}}_1 \quad \mathbf{C}_1\ ], \tag{22}$$

and $\widehat{\mathbf{W}}$ is a $(1 + K + L) \times (1 + K + L)$ symmetric weighting matrix of full rank which can be computed using the available data. The weighting matrix $\widehat{\mathbf{W}}$ is assumed to converge to a symmetric and positive definite matrix $\mathbf{W}$, as $N \to \infty$. Note that, if we only use the past beta estimates as instruments, i.e., if $\widehat{\mathbf{Z}}_1 = \widehat{\mathbf{X}}_1 = [\ \mathbf{1}_N \quad \widehat{\mathbf{B}}_1\ ]$, then the weighting matrix is irrelevant and the estimator assumes the usual exactly identified IV form: $\widehat{\boldsymbol{\lambda}}_{\text{IV}} = (\widehat{\mathbf{X}}_1'\widehat{\mathbf{X}}_2)^{-1}(\widehat{\mathbf{X}}_1'\bar{\mathbf{r}}_2)$. We will establish the N-consistency and asymptotic normality of the estimator $\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{GMM}}$ for a generic weighting matrix and then show how to obtain the efficient IV-GMM estimator by suitably selecting $\mathbf{W}$ and $\widehat{\mathbf{W}}$. Note that equation (16) implies

$$\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{GMM}} = \boldsymbol{\lambda} + \left(\widehat{\boldsymbol{\Omega}}'\widehat{\mathbf{W}}\widehat{\boldsymbol{\Omega}} \right)^{-1} \widehat{\boldsymbol{\Omega}}'\widehat{\mathbf{W}} \left(\widehat{\mathbf{Z}}_1'\boldsymbol{\omega}_2/N \right), \tag{23}$$

where

$$\widehat{\boldsymbol{\Omega}} = \frac{1}{N}\widehat{\mathbf{Z}}_1'\widehat{\mathbf{X}}_2 \xrightarrow{p} \boldsymbol{\Omega}, \tag{24}$$

and $\boldsymbol{\Omega}$ is a full-rank $(1+K+L)\times(1+K)$ matrix (see equation (77) in the Appendix). Moreover, $\widehat{\mathbf{Z}}_1'\boldsymbol{\omega}_2/N$ converges to a vector of zeros, and hence N-consistency of $\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{GMM}}$ is established. The proof of the following theorem contains the details.

Theorem 1 *Under Assumptions 1 and 2, the IV-GMM ex-post risk premia estimator* $\widehat{\boldsymbol{\lambda}}_{IV}^{GMM}$*, defined in (21), is an* N*-consistent estimator of* $\boldsymbol{\lambda}$*.*

Having established the N-consistency of the proposed IV-GMM estimator $\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{GMM}}$, we proceed to define a measure of aggregate mispricing, that takes into account the time-series alphas of all individual stocks, and describe its limit as the number of stocks increases to infinity.

2.3 A metric of aggregate mispricing

It is common practice to evaluate asset pricing models by examining the corresponding pricing errors. Given the ex-post risk premia estimator $\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{GMM}}$, obtained in the previous subsection, we define the vector of pricing error estimates, traditionally referred to as alphas, as follows

$$\widehat{\boldsymbol{\alpha}} = \bar{\mathbf{r}}_2 - \widehat{\mathbf{X}}_2 \widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{GMM}}. \tag{25}$$

To gauge the magnitude of $\widehat{\boldsymbol{\alpha}}$, we use as metric the following average square pricing error

$$\widehat{\mathcal{Q}} = \frac{1}{N} \widehat{\boldsymbol{\alpha}}' \widehat{\boldsymbol{\alpha}}. \tag{26}$$

Note that $\widehat{\mathcal{Q}}$ is the analogue to the well-known and widely used GRS statistic of Gibbons, Ross, and Shanken (1989) in the small T-large N context. When the asset pricing model is correctly specified, the number of test assets N is small and T tends to infinity, then all individual alphas, as well as $\widehat{\mathcal{Q}}$, vanish in the limit. In contrast, in our context $\widehat{\mathcal{Q}}$ converges to a positive quantity that we denote by $\mathcal{Q}$. To study the sampling properties of $\widehat{\mathcal{Q}}$, we make the following mild assumption on the limiting behavior of the second moment of the disturbances $u_{i,s}$, $i = 1, \ldots, N$ and $s = 1, \ldots, \tau_2$, allowing for time-series heteroscedasticity.

Assumption 3 *As $N \to \infty$, $\frac{1}{N}\mathbf{U}_2'\mathbf{U}_2 \xrightarrow{p} \mathbf{V}_2$, where $\mathbf{V}_2$ is a diagonal matrix with (s,s) element equal to a positive constant $v_{2,s}$, $s = 1, \ldots, \tau_2$.*

The following proposition characterizes the probability limit $\mathcal{Q}$ of the aggregate mispricing metric $\widehat{\mathcal{Q}}$.

Proposition 2 *Under Assumptions 1 and 3, $\widehat{\mathcal{Q}} \xrightarrow{p} \mathcal{Q}$, as $N \to \infty$, where*

$$\mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)' \mathbf{v}_2, \tag{27}$$

the vector $\mathbf{g}_2$ *is defined in (20), and the vector* $\mathbf{v}_2$ *is given by*

$$\mathbf{v}_2 = vec(\mathbf{V}_2). \tag{28}$$

To further study the sampling distribution of the aggregate mispricing metric $\widehat{\mathcal{Q}}$, we need an estimator of $\mathcal{Q}$. Given that $\mathbf{g}_2$ is observed, it suffices to obtain an estimator of the vector $\mathbf{v}_2$. To do so, we introduce the time-series regression residuals over the testing period, $\hat{\mathbf{u}}_s = (\mathbf{r}_s - \bar{\mathbf{r}}_2) - \widehat{\mathbf{B}}(\mathbf{f}_s - \bar{\mathbf{f}}_2)$, $s = \tau_1 + 1, \ldots, \tau_1 + \tau_2$, gathered in the $N \times \tau_2$ matrix

$$\widehat{\mathbf{U}}_2 = \mathbf{R}_2\mathbf{J}_{\tau_2} - \widehat{\mathbf{B}}_2\mathbf{F}_2\mathbf{J}_{\tau_2}. \tag{29}$$

Noting that $\mathbf{R}_2 = \mathbf{B}\mathbf{F}_2 + \mathbf{U}_2$ and using equation (14), we can express $\widehat{\mathbf{U}}_2$ as

$$\widehat{\mathbf{U}}_2 = \mathbf{U}_2\mathbf{H}_2, \tag{30}$$

where the matrix $\mathbf{H}_2$ is defined by[6]

$$\mathbf{H}_2 = \mathbf{J}_{\tau_2} - \mathbf{J}_{\tau_2}\mathbf{F}_2' \left(\mathbf{F}_2\mathbf{J}_{\tau_2}\mathbf{F}_2'\right)^{-1} \mathbf{F}_2\mathbf{J}_{\tau_2}. \tag{31}$$

We assume throughout that the $\tau_2 \times \tau_2$ Hadamard product matrix $\mathbf{H}_2 \odot \mathbf{H}_2$ is of full rank, and hence invertible.[7] The following proposition provides an N-consistent estimator of $\mathbf{v}_2$.

Proposition 3 *Under Assumption 3, the vector*

$$\hat{\mathbf{v}}_2 = \left[\boldsymbol{\mathcal{S}} \left(\mathbf{H}_2 \odot \mathbf{H}_2\right)^{-1} \boldsymbol{\mathcal{S}}'\right] vec\left(\widehat{\mathbf{U}}_2'\widehat{\mathbf{U}}_2/N\right) \tag{32}$$

is an N*-consistent estimator of* $\mathbf{v}_2$*, where* $\widehat{\mathbf{U}}_2$ *is defined in (29),* $\mathbf{H}_2$ *is defined in (31), and* $\boldsymbol{\mathcal{S}}$ *is the* $\tau_2^2 \times \tau_2$ *selection matrix such that the* $(\tau_2 (s-1) + s, s)$ *element of* $\boldsymbol{\mathcal{S}}$ *is 1, for* $s = 1, \ldots, \tau_2$*,*

[6] Standard matrix algebra shows that $\mathbf{H}_2$ is symmetric and idempotent. Moreover, it follows from the properties of the trace operator that $\mathrm{tr}\,(\mathbf{H}_2) = \tau_2 - K - 1$.

[7] It is straightforward to establish that the matrix $\mathbf{H}_2$ is equal to the projection matrix $\mathbf{M}_2 = \mathbf{I}_{\tau_2} - \widetilde{\mathbf{F}}_2(\widetilde{\mathbf{F}}_2'\widetilde{\mathbf{F}}_2)\widetilde{\mathbf{F}}_2'$, where $\widetilde{\mathbf{F}}_2 = [\ \mathbf{1}_{\tau_2} \quad \mathbf{F}_2'\ ]$. The Hadamard product $\mathbf{M}_2 \odot \mathbf{M}_2$, and its invertibility, has come up in early studies of linear models with heteroscedasticity. Hartley, Rao, and Kiefer (1969) and Rao (1970) provide sufficient conditions for the invertibility of $\mathbf{M}_2 \odot \mathbf{M}_2$, while Mallela (1972) provides a necessary and sufficient condition. It follows from the results in Mallela (1972) that a necessary condition for the invertibility of $\mathbf{H}_2 \odot \mathbf{H}_2$ is $\tau_2 \geq 2K + 3$. When this condition is satisfied and the factors are normally distributed, extensive simulation evidence suggests that $\mathbf{H}_2 \odot \mathbf{H}_2$ is indeed invertible. In fact, $\mathbf{H}_2 \odot \mathbf{H}_2$ is always invertible in our empirical applications, where $\tau_2 = 60$ and K takes values up to 5.

and all other elements are zero.

The following theorem follows from Propositions 2 and 3.

Theorem 4 *Let Assumptions 1 and 3 be in effect. Then, as $N \to \infty$,*

$$\widehat{\mathcal{Q}} - (\mathbf{g}_2 \otimes \mathbf{g}_2)' \widehat{\mathbf{v}}_2 \xrightarrow{p} 0. \tag{33}$$

In the previous two subsections, we describe the limits of the ex-post risk premia estimator $\widehat{\boldsymbol{\lambda}}^{\texttt{GMM}}_{\texttt{IV}}$ and the aggregate mispiring metric $\widehat{\mathcal{Q}}$. In the following subsection, we obtain their asymptotic distributions based on which we will devise empirical test statistics.

2.4 Asymptotic distributions

The objective in this subsection is to obtain the asymptotic distributions of the following two statistics: (i) $\sqrt{N}\left(\widehat{\boldsymbol{\lambda}}^{\texttt{GMM}}_{\texttt{IV}} - \boldsymbol{\lambda}\right)$ and (ii) $\sqrt{N}\left(\widehat{\mathcal{Q}} - (\mathbf{g}_2 \otimes \mathbf{g}_2)' \widehat{\mathbf{v}}_2\right)$. To state the results, we need to introduce the following $\mathcal{T}$-dimensional random vector

$$\mathbf{e}_i = \left[\begin{array}{ccccc} \mathbf{u}'_{2,[i]} & \left(\mathbf{u}_{2,[i]} \otimes \mathbf{u}_{1,[i]}\right)' & \left(\mathbf{u}_{2,[i]} \otimes \boldsymbol{\beta}_i\right)' & \left(\mathbf{u}_{2,[i]} \otimes \mathbf{c}_{1,i}\right)' & \operatorname{vec}\left(\mathbf{u}_{2,[i]}\mathbf{u}'_{2,[i]} - \mathbf{V}_2\right)' \end{array}\right]', \tag{34}$$

where $\mathcal{T} = \tau_2(1 + K + L + \tau_1 + \tau_2)$. Assumptions 1(i), 1(ii), 2(ii), and 3 together imply that $\frac{1}{N}\sum_{i=1}^{N} \mathbf{e}_i \xrightarrow{p} \mathbf{0}_{\mathcal{T}}$, as $N \to \infty$. In order to obtain the asymptotic distributions of interest, we make the following mild assumption, where $\xrightarrow{d}$ denotes convergence in distribution, postulating that $\mathbf{e}_i$ satisfies a cross-sectional central limit theorem.

Assumption 4 *As $N \to \infty$, $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \mathbf{e}_i \xrightarrow{d} N(\mathbf{0}_{\mathcal{T}}, \mathbf{V}_e)$, where $\mathbf{V}_e$ is a symmetric and positive definite $\mathcal{T} \times \mathcal{T}$ matrix.*

2.4.1 Asymptotic distribution of the ex-post risk premia estimator

Note that equation (23) yields

$$\sqrt{N}\left(\widehat{\boldsymbol{\lambda}}^{\texttt{GMM}}_{\texttt{IV}} - \boldsymbol{\lambda}\right) = \left(\widehat{\boldsymbol{\Omega}}'\widehat{\mathbf{W}}\widehat{\boldsymbol{\Omega}}\right)^{-1}\widehat{\boldsymbol{\Omega}}'\widehat{\mathbf{W}}\left(\frac{1}{\sqrt{N}}\widehat{\mathbf{Z}}'_1\boldsymbol{\omega}_2\right), \tag{35}$$

where $\widehat{\mathbf{\Omega}} = \frac{1}{N}\widehat{\mathbf{Z}}_1'\widehat{\mathbf{X}}_2$. It is shown in the proof of Theorem 1 that $\widehat{\mathbf{\Omega}} \xrightarrow{p} \mathbf{\Omega}$, where $\mathbf{\Omega}$ is the full-rank $(1+K+L)\times(1+K)$ matrix defined in equation (77) in the Appendix. Letting

$$\widehat{\mathbf{\Psi}}_\lambda = \left(\widehat{\mathbf{\Omega}}'\widehat{\mathbf{W}}\widehat{\mathbf{\Omega}}\right)^{-1}\widehat{\mathbf{\Omega}}'\widehat{\mathbf{W}}, \tag{36}$$

we have

$$\widehat{\mathbf{\Psi}}_\lambda \xrightarrow{p} \mathbf{\Psi}_\lambda, \tag{37}$$

where

$$\mathbf{\Psi}_\lambda = (\mathbf{\Omega}'\mathbf{W}\mathbf{\Omega})^{-1}\mathbf{\Omega}'\mathbf{W}. \tag{38}$$

Hence, to determine the asymptotic distribution of $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}}$, it suffices to determine the asymptotic distribution of $\frac{1}{\sqrt{N}}\widehat{\mathbf{Z}}_1'\boldsymbol{\omega}_2$. It turns out that $\widehat{\mathbf{Z}}_1'\boldsymbol{\omega}_2 = \mathbf{\Pi}_\lambda\sum_{i=1}^N \mathbf{e}_i$, where $\mathbf{\Pi}_\lambda$ is a suitable matrix (see equations (88) and (89) in the Appendix) and $\mathbf{e}_i$ is defined in (34). It follows that $\sqrt{N}\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right) = \mathbf{\Psi}_\lambda\mathbf{\Pi}_\lambda\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{e}_i + o_p(1)$. The proof of the following theorem, which establishes the asymptotic distribution of the estimator $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}}$, contains the details.

Theorem 5 *Under Assumptions 1, 2, and 4, as $N\to\infty$, $\sqrt{N}\left(\widehat{\boldsymbol{\lambda}}_{IV}^{GMM} - \boldsymbol{\lambda}\right) \xrightarrow{d} N\left(\mathbf{0}_{K+1}, \mathbf{V}_\lambda\right)$, where*

$$\mathbf{V}_\lambda = \mathbf{\Psi}_\lambda\mathbf{\Pi}_\lambda\mathbf{V}_e\mathbf{\Pi}_\lambda'\mathbf{\Psi}_\lambda', \tag{39}$$

$\mathbf{\Psi}_\lambda$ is defined in equation (38), $\mathbf{\Pi}_\lambda$ is defined in equation (89) in the Appendix, and $\mathbf{V}_e$ is defined in Assumption 4.

It follows from a standard argument, typically employed in a GMM context, that the optimal (most efficient) IV-GMM estimator is obtained when the weighting matrix is $\mathbf{W}^* = (\mathbf{\Pi}_\lambda\mathbf{V}_e\mathbf{\Pi}_\lambda')^{-1}$, in which case we obtain $\mathbf{V}_\lambda = \left(\mathbf{\Omega}'\left(\mathbf{\Pi}_\lambda\mathbf{V}_e\mathbf{\Pi}_\lambda'\right)^{-1}\mathbf{\Omega}\right)^{-1}$.[8] In the following subsection, we obtain an N-consistent estimator of $\mathbf{\Pi}_\lambda\mathbf{V}_e\mathbf{\Pi}_\lambda'$, based on which an N-consistent estimator of $\mathbf{V}_\lambda$ is readily constructed using equation (39). As expected, the optimal IV-GMM estimator $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}}$ is at least as efficient as the IV estimator $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}$. We formally establish this property in the Online Appendix.

[8]Note that $\mathbf{1}_{\tau_2}'\mathbf{g}_2 = 1$ which implies that $\mathbf{g}_2 \neq \mathbf{0}_{\tau_2}$. Hence, it follows from equation (89) in the Appendix that $\mathbf{\Pi}_\lambda$ has full rank equal to $1+K+L$ and so $\mathbf{\Pi}_\lambda\mathbf{V}_e\mathbf{\Pi}_\lambda'$ is invertible, given that $\mathbf{V}_e$ is positive definite according to Assumption 4.

2.4.2 Asymptotic distribution of the aggregate mispricing metric

Theorem 6 *Let Assumptions 3 and 4 be in effect. Then, the asymptotic distribution of the centered version of* $\widehat{\mathcal{Q}} = \frac{1}{N}\widehat{\boldsymbol{\alpha}}'\widehat{\boldsymbol{\alpha}}$*, as* $N \to \infty$*, is given by*

$$\sqrt{N}\left(\widehat{\mathcal{Q}} - (\mathbf{g}_2 \otimes \mathbf{g}_2)'\,\widehat{\mathbf{v}}_2 - R_{\widehat{\mathcal{Q}}}\right) \xrightarrow{d} N\left(0, v_\alpha\right), \tag{40}$$

where $\widehat{\mathbf{v}}_2$ *is defined in (32),* $R_{\widehat{\mathcal{Q}}}$ *is defined in (80) in the Appendix,*

$$v_\alpha = \boldsymbol{\kappa}_\alpha' \mathbf{V}_e \boldsymbol{\kappa}_\alpha, \tag{41}$$

$\boldsymbol{\kappa}_\alpha$ *is defined in (98) in the Appendix, and* $\mathbf{V}_e$ *is defined in Assumption 4.*

A few comments about the term $R_{\widehat{\mathcal{Q}}}$ are in order. As $N \to \infty$, the difference $\widehat{\mathcal{Q}} - (\mathbf{g}_2 \otimes \mathbf{g}_2)'\,\widehat{\mathbf{v}}_2$ converges to zero in probability and, rescaled by $\sqrt{N}$, converges to the normal distribution $N\left(0, v_\alpha\right)$ given in the above theorem. Hence, one could use $\sqrt{N}\left(\widehat{\mathcal{Q}} - (\mathbf{g}_2 \otimes \mathbf{g}_2)'\,\widehat{\mathbf{v}}_2\right)$ to build a statistic for testing for the magnitude of aggregate mispricing. We incorporate the asymptotically negligible term $R_{\widehat{\mathcal{Q}}}$ in the statistic in (40) as it improves the small-sample properties of the associated tests.

2.4.3 Joint asymptotic distribution of the ex-post risk premia estimator and the aggregate mispricing metric

According to Theorem 5 and 6, the statistics $\sqrt{N}\left(\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{GMM}} - \boldsymbol{\lambda}\right)$ and $\sqrt{N}\left(\widehat{\mathcal{Q}} - (\mathbf{g}_2 \otimes \mathbf{g}_2)'\,\widehat{\mathbf{v}}_2 - R_{\widehat{\mathcal{Q}}}\right)$ follow normal asymptotic distributions. The former statistic can be used to test ex-post risk premia implications while the latter can be used to test mispricing. We next combine the two statistics to obtain a comprehensive statistic that can be used to test both model implications simultaneously. To this end, we define the $(K+2) \times 1$ vector

$$\widehat{\boldsymbol{\delta}} = \begin{bmatrix} \widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{GMM}} - \boldsymbol{\lambda} \\ \widehat{\mathcal{Q}} - (\mathbf{g}_2 \otimes \mathbf{g}_2)'\,\widehat{\mathbf{v}}_2 - R_{\widehat{\mathcal{Q}}} \end{bmatrix}. \tag{42}$$

It follows from the proofs of Theorems 5 and 6 that $\sqrt{N}\left(\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{GMM}} - \boldsymbol{\lambda}\right) = \boldsymbol{\Psi}_\lambda \boldsymbol{\Pi}_\lambda \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{e}_i + o_p(1)$, and $\sqrt{N}\left(\widehat{\mathcal{Q}} - (\mathbf{g}_2 \otimes \mathbf{g}_2)'\,\widehat{\mathbf{v}}_2 - R_{\widehat{\mathcal{Q}}}\right) = \boldsymbol{\kappa}_\alpha' \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{e}_i + o_p(1)$, where $\boldsymbol{\kappa}_\alpha = \boldsymbol{\pi}_\alpha - 2\boldsymbol{\Pi}_\lambda' \boldsymbol{\Psi}_\lambda' \boldsymbol{\rho}_\alpha$. The matrices $\boldsymbol{\Psi}_\lambda$ and $\boldsymbol{\Pi}_\lambda$ are defined in equations (38) and (89), while the vectors $\boldsymbol{\rho}_\alpha$ and $\boldsymbol{\pi}_\alpha$ are defined in equations (83) and (94), respectively. Combining these two results, we obtain that

$\sqrt{N}\widehat{\boldsymbol{\delta}} = \boldsymbol{\Psi}_\delta \boldsymbol{\Pi}_\delta \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{e}_i + o_p(1)$, where the $(K+2) \times (K+2)$ matrix $\boldsymbol{\Psi}_\delta$ and the $(K+2) \times \mathcal{T}$ matrix $\boldsymbol{\Pi}_\delta$ are given by

$$\boldsymbol{\Psi}_\delta = \begin{bmatrix} \boldsymbol{\Psi}_\lambda & \mathbf{0}_{1+K} \\ -2\boldsymbol{\rho}_\alpha' \boldsymbol{\Psi}_\lambda & 1 \end{bmatrix}, \tag{43}$$

and

$$\boldsymbol{\Pi}_\delta = \begin{bmatrix} \boldsymbol{\Pi}_\lambda \\ \boldsymbol{\pi}_\alpha' \end{bmatrix}. \tag{44}$$

The following theorem summarizes the above results and provides the asymptotic distribution of the joint statistic $\widehat{\boldsymbol{\delta}}$.

Theorem 7 *Let Assumptions 1, 2, 3, and 4 be in effect. Then, the asymptotic distribution of* $\widehat{\boldsymbol{\delta}} = \left[\begin{array}{cc} \left(\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{GMM}} - \boldsymbol{\lambda}\right)' & \widehat{\mathcal{Q}} - (\mathbf{g}_2 \otimes \mathbf{g}_2)' \widehat{\mathbf{v}}_2 - R_{\widehat{\mathcal{Q}}} \end{array} \right]'$ *is given by*

$$\sqrt{N}\widehat{\boldsymbol{\delta}} \xrightarrow{d} N(\mathbf{0}_{K+2}, \mathbf{V}_\delta), \tag{45}$$

where

$$\mathbf{V}_\delta = \boldsymbol{\Psi}_\delta \boldsymbol{\Pi}_\delta \mathbf{V}_e \boldsymbol{\Pi}_\delta' \boldsymbol{\Psi}_\delta', \tag{46}$$

$\boldsymbol{\Psi}_\delta$ *and* $\boldsymbol{\Pi}_\delta$ *are defined in equations (43) and (44), and* $\mathbf{V}_e$ *is defined in Assumption 4.*

To make operational the asymptotic distribution obtained in Theorem 7, and obtain feasible test statistics, we need to obtain a consistent estimator of $\mathbf{V}_\delta$. This is the subject of the next subsection.

2.5 Estimation of the asymptotic variance-covariance matrix $\mathbf{V}_\delta$

According to equation (46), the variance-covariance matrix $\mathbf{V}_\delta$ involves the matrix $\mathbf{V}_e$ which, according to Assumption 4, is limiting variance-covariance matrix of $\frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{e}_i$. Hence, the structure of $\mathbf{V}_\delta$ depends on the structure of $\mathbf{V}_e$ which, in turn, depends on potential cross-sectional correlations of the shocks $\mathbf{e}_i$. Note that in the return generating process described by (2), the disturbance vector $\mathbf{u}_t$ could potentially exhibit cross-sectional correlation due to economic links such as industry effects. In that case, the vectors $\mathbf{e}_i$ would be correlated across firms as it follows from definition (34). To incorporate such correlations, we use a clustering approach that we describe next.[9]

We assume that there are M_N clusters and that the m-th cluster consists of N_m stocks, for $m = 1, \ldots, M_N$, so that $\sum_{m=1}^{M_N} N_m = N$. For all N, we assume that the cluster sizes N_m, $m = 1, \ldots, M_N$ are bounded. As $N \to \infty$, the number of clusters, M_N, is assumed to increase so that $\frac{N}{M_N} \to G$, where G is to be interpreted as the limiting average cluster size. For $m = 1, \ldots, M_N$, let I_m be the set of all indices i for which the i-th stock belongs to the m-th cluster, and define the aggregate cluster shocks

$$\boldsymbol{\eta}_m = \sum\nolimits_{i \in I_m} \mathbf{e}_i. \tag{47}$$

In the next assumption, we postulate that the central limit theorem applies to the random sequence $\boldsymbol{\eta}_m$, $m = 1, 2, \ldots$

Assumption 5 *The aggregate cluster shocks $\boldsymbol{\eta}_m$ are independent across clusters and, as $N \to \infty$, $\frac{1}{\sqrt{M_N}} \sum_{m=1}^{M_N} \boldsymbol{\eta}_m \xrightarrow{d} N\left(\mathbf{0}_T, \mathbf{V}_\eta\right)$, where*

$$\mathbf{V}_\eta = p\text{-}\lim \frac{1}{M_N} \sum_{m=1}^{M_N} \boldsymbol{\eta}_m \boldsymbol{\eta}_m'. \tag{48}$$

Utilizing Assumption 5, we obtain

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{e}_i = \sqrt{\frac{M_N}{N}} \frac{1}{\sqrt{M_N}} \sum_{m=1}^{M_N} \boldsymbol{\eta}_m \xrightarrow{d} N\left(\mathbf{0}_T, \mathbf{V}_\eta / G\right),$$

and so it follows that $\mathbf{V}_e = \frac{1}{G}\mathbf{V}_\eta$. Equation (46) then yields $\mathbf{V}_\delta = \frac{1}{G}\boldsymbol{\Psi}_\delta \left(\boldsymbol{\Pi}_\delta \mathbf{V}_\eta \boldsymbol{\Pi}_\delta'\right) \boldsymbol{\Psi}_\delta'$. In light of (43), to estimate $\boldsymbol{\Psi}_\delta$, we need to estimate $\boldsymbol{\Psi}_\lambda$, defined in (43), and $\boldsymbol{\rho}_\alpha$, defined in (83) in the Appendix. Recalling that $\widehat{\boldsymbol{\alpha}} = \boldsymbol{\omega}_2 - \widehat{\mathbf{X}}_2 \left(\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{GMM}} - \boldsymbol{\lambda}\right)$, we have $\frac{\widehat{\mathbf{X}}_2' \widehat{\boldsymbol{\alpha}}}{N} = \frac{\widehat{\mathbf{X}}_2' \boldsymbol{\omega}_2}{N} + \frac{\widehat{\mathbf{X}}_2' \widehat{\mathbf{X}}_2}{N} \left(\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{GMM}} - \boldsymbol{\lambda}\right) \xrightarrow{p} \boldsymbol{\rho}_\alpha$, according to Theorem 1 and the limit in (82) in the Appendix. It follows that

$$\widehat{\boldsymbol{\Psi}}_\delta = \begin{bmatrix} \widehat{\boldsymbol{\Psi}}_\lambda & \mathbf{0}_{1+K} \\ -2\widehat{\boldsymbol{\rho}}_\alpha' \widehat{\boldsymbol{\Psi}}_\lambda & 1 \end{bmatrix} \xrightarrow{p} \boldsymbol{\Psi}_\delta, \tag{49}$$

[9] Our empirical applications, following standard economic intuition, we use an industry classification to determine the clusters. In addition, for robustness purposes, we consider clusters based on firm characteristics such as size and book-to-market ratio.

where $\widehat{\boldsymbol{\Psi}}_\lambda$ is given by (36) and

$$\widehat{\boldsymbol{\rho}}_\alpha = \frac{\widehat{\mathbf{X}}_2'\widehat{\boldsymbol{\alpha}}}{N}. \tag{50}$$

Therefore, to estimate the asymptotic variance-covariance matrix $\mathbf{V}_\delta$, it suffices to obtain an estimator of the $(K+2)\times(K+2)$ matrix $\boldsymbol{\Theta}$ defined by

$$\boldsymbol{\Theta} = \boldsymbol{\Pi}_\delta \mathbf{V}_\eta \boldsymbol{\Pi}_\delta'. \tag{51}$$

To construct such an estimator, we need to introduce some additional notation. Define the cluster selection $M_N \times N$ matrix $\boldsymbol{\mathcal{C}}$ with (m,i) element given by

$$\boldsymbol{\mathcal{C}}(m,i) = \mathbf{1}_{[i\in I_m]}, \quad m=1,\ldots,M_N, \quad i=1,\ldots,N, \tag{52}$$

the $N\times\mathcal{T}$ matrix of $\mathbf{E}$, stacking the firm shocks $\mathbf{e}_i$ in (34), by

$$\mathbf{E} = [\ \mathbf{e}_1 \quad \cdots \quad \mathbf{e}_N\ ]', \tag{53}$$

and the $M_N\times\mathcal{T}$ matrix of $\boldsymbol{\mathcal{H}}$, stacking the cluster aggregate shocks $\boldsymbol{\eta}_m$ in (47), by

$$\boldsymbol{\mathcal{H}} = [\ \boldsymbol{\eta}_1 \quad \cdots \quad \boldsymbol{\eta}_{M_N}\ ]', \tag{54}$$

so that

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{C}}\mathbf{E}. \tag{55}$$

For any $N\times 1$ vector $\mathbf{x} = (x_1,\ldots,x_N)'$, let $\mathbf{D}(\mathbf{x})$ be the $N\times N$ diagonal matrix with (i,i) element equal to x_i, for $i=1,\ldots,N$. Furthermore, let $\widehat{\mathbf{z}}_{1,i}'$ denote the i-th row of $\widehat{\mathbf{Z}}_1$ defined in (22) and $\omega_{2,i}$ denote the i-th element of $\boldsymbol{\omega}_2$ defined in (18). It follows from the proof of Theorem 5 (see equation (88) in the Appendix) that, for $i=1,\ldots,N$, $\widehat{\mathbf{z}}_{1,i}\omega_{2,i}$ equals $\boldsymbol{\Pi}_\lambda\mathbf{e}_i$, which implies that

$$\mathbf{D}(\boldsymbol{\omega}_2)\widehat{\mathbf{Z}}_1 = [\ \widehat{\mathbf{z}}_{1,1}\omega_{2,1} \quad \cdots \quad \widehat{\mathbf{z}}_{1,N}\omega_{2,N}\ ]' = \mathbf{E}\boldsymbol{\Pi}_\lambda'. \tag{56}$$

Furthermore, as shown in the proof of Theorem 6 (see equations (92) and (95) in the Appendix), we have that

$$\nu_i = \omega_{2,i}^2 - (\mathbf{g}_2 \otimes \mathbf{g}_2)' \left[\boldsymbol{\mathcal{S}} \left(\mathbf{H}_2 \odot \mathbf{H}_2\right)^{-1} \boldsymbol{\mathcal{S}}'\right] \operatorname{vec}\left(\widehat{\mathbf{u}}_{2,[i]} \widehat{\mathbf{u}}_{2,[i]}'\right) = \boldsymbol{\pi}_\alpha' \mathbf{e}_i, \tag{57}$$

implying that

$$\boldsymbol{\omega}_2 \odot \boldsymbol{\omega}_2 - \widehat{\boldsymbol{\mathcal{U}}}_2 \left[\boldsymbol{\mathcal{S}} \left(\mathbf{H}_2 \odot \mathbf{H}_2\right)^{-1} \boldsymbol{\mathcal{S}}'\right] (\mathbf{g}_2 \otimes \mathbf{g}_2) = \mathbf{E} \boldsymbol{\pi}_\alpha, \tag{58}$$

where $\widehat{\boldsymbol{\mathcal{U}}}_2$ is $N \times \tau_2^2$ matrix with i-th row equal to $\operatorname{vec}\left(\widehat{\mathbf{u}}_{2,[i]} \widehat{\mathbf{u}}_{2,[i]}'\right)'$. Stacking (56) and (58) together, we have

$$\widetilde{\boldsymbol{\mathcal{E}}} \equiv \left[\begin{array}{cc} \mathbf{D}(\boldsymbol{\omega}_2) \widehat{\mathbf{Z}}_1 & \boldsymbol{\omega}_2 \odot \boldsymbol{\omega}_2 - \widehat{\boldsymbol{\mathcal{U}}}_2 \left[\boldsymbol{\mathcal{S}} \left(\mathbf{H}_2 \odot \mathbf{H}_2\right)^{-1} \boldsymbol{\mathcal{S}}'\right] (\mathbf{g}_2 \otimes \mathbf{g}_2) \end{array}\right] = \mathbf{E} \boldsymbol{\Pi}_\delta'. \tag{59}$$

Let $\widetilde{\boldsymbol{\varepsilon}}_i'$ denote the i-th row of the matrix $\widetilde{\boldsymbol{\mathcal{E}}}$ so that

$$\widetilde{\boldsymbol{\mathcal{E}}} = \left[\begin{array}{ccc} \widetilde{\boldsymbol{\varepsilon}}_1 & \cdots & \widetilde{\boldsymbol{\varepsilon}}_N \end{array}\right]'. \tag{60}$$

For each cluster $m = 1, \ldots, M_N$, we define

$$\widetilde{\boldsymbol{\theta}}_m = \sum_{i \in I_m} \widetilde{\boldsymbol{\varepsilon}}_i, \tag{61}$$

so that $\boldsymbol{\mathcal{C}} \widetilde{\boldsymbol{\mathcal{E}}} = \left[\begin{array}{ccc} \widetilde{\boldsymbol{\theta}}_1 & \cdots & \widetilde{\boldsymbol{\theta}}_{M_N} \end{array}\right]'$. Note that, since $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{C}} \mathbf{E}$, we have

$$\widetilde{\boldsymbol{\theta}}_m = \boldsymbol{\Pi}_\delta \boldsymbol{\eta}_m. \tag{62}$$

It follows that one could consistently estimate $\boldsymbol{\Theta}$ by

$$\widetilde{\boldsymbol{\Theta}} = \frac{1}{M_N} \left(\widetilde{\boldsymbol{\mathcal{E}}}' \boldsymbol{\mathcal{C}}' \boldsymbol{\mathcal{C}} \widetilde{\boldsymbol{\mathcal{E}}}\right). \tag{63}$$

Indeed, as $N \to \infty$,

$$\widetilde{\boldsymbol{\Theta}} = \frac{1}{M_N} \sum_{m=1}^{M_N} \widetilde{\boldsymbol{\theta}}_m \widetilde{\boldsymbol{\theta}}_m' = \boldsymbol{\Pi}_\delta \left(\frac{1}{M_N} \sum_{m=1}^{M_N} \boldsymbol{\eta}_m \boldsymbol{\eta}_m'\right) \boldsymbol{\Pi}_\delta' \xrightarrow{p} \boldsymbol{\Pi}_\delta \mathbf{V}_\eta \boldsymbol{\Pi}_\delta' = \boldsymbol{\Theta},$$

where, in the last step, we make use of definition (48) in Assumption 5. Note, however, that the matrix $\widetilde{\boldsymbol{\mathcal{E}}}$, and hence $\widetilde{\boldsymbol{\Theta}}$, depends on $\boldsymbol{\omega}_2$ which can be thought of as the vector of residuals in the second-pass cross-sectional regression obtained after imposing the null hypothesis. Instead of using $\boldsymbol{\omega}_2$, we proceed in the traditional fashion and define residuals without imposing the null hypothesis as follows

$$\widehat{\boldsymbol{\omega}}_2 = \bar{\mathbf{r}}_2 - \widehat{\mathbf{X}}_2 \widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{GMM}}. \tag{64}$$

Replacing $\boldsymbol{\omega}_2$ by $\widehat{\boldsymbol{\omega}}_2$ in (59) and incorporating the standard degrees-of-freedom adjustment,[10] we propose estimating $\boldsymbol{\Theta}$ by

$$\widehat{\boldsymbol{\Theta}} \;=\; \frac{1}{M_N - K - 1} \widehat{\boldsymbol{\mathcal{E}}}' \boldsymbol{\mathcal{C}}' \boldsymbol{\mathcal{C}} \widehat{\boldsymbol{\mathcal{E}}}, \tag{65}$$

where $\widehat{\boldsymbol{\mathcal{E}}}$ is the $N \times (K+2)$ matrix defined by

$$\widehat{\boldsymbol{\mathcal{E}}} \equiv \left[\begin{array}{cc} \mathbf{D}\left(\widehat{\boldsymbol{\omega}}_2\right) \widehat{\mathbf{Z}}_1 & \widehat{\boldsymbol{\omega}}_2 \odot \widehat{\boldsymbol{\omega}}_2 - \widehat{\boldsymbol{\mathcal{U}}}_2 \left[\boldsymbol{\mathcal{S}} \left(\mathbf{H}_2 \odot \mathbf{H}_2\right)^{-1} \boldsymbol{\mathcal{S}}'\right] \left(\mathbf{g}_2 \otimes \mathbf{g}_2\right) \end{array} \right]. \tag{66}$$

Next, we provide a mild regularity condition under which $\widehat{\boldsymbol{\Theta}}$ is indeed an N-consistent estimator of $\boldsymbol{\Theta}$. The following theorem provides the desired consistent estimator of $\mathbf{V}_\delta$ based on which we can construct feasible test statistics. To this end, for each $i = 1, \ldots, N$, we define the vector $\boldsymbol{\zeta}_i$ by

$$\boldsymbol{\zeta}_i = \left[\begin{array}{cc} \mathbf{e}_i' & \mathbf{d}_i' \end{array} \right]', \tag{67}$$

where $\mathbf{e}_i$ is defined by (34) and

$$\mathbf{d}_i = \left[1 \;\; \mathbf{u}_{1,[i]}' \;\; \boldsymbol{\beta}_i' \;\; \mathbf{c}_{1,i}' \;\; \left(\boldsymbol{\beta}_i \otimes \boldsymbol{\beta}_i\right)' \;\; \left(\boldsymbol{\beta}_i \otimes \mathbf{c}_{1,i}\right)' \;\; \left(\boldsymbol{\beta}_i \otimes \mathbf{u}_{1,[i]}\right)' \;\; \left(\boldsymbol{\beta}_i \otimes \mathbf{u}_{2,[i]}\right)' \;\; \left(\mathbf{u}_{2,[i]} \otimes \mathbf{u}_{2,[i]}\right)'\right]'. \tag{68}$$

Then, for each cluster $m = 1, \ldots, M_N$, we define

$$\boldsymbol{\varphi}_m = \sum\nolimits_{i \in I_m} \boldsymbol{\zeta}_i, \tag{69}$$

and make the following assumption which amounts to the existence of cross-sectional second moments of $\boldsymbol{\varphi}_m$.

Assumption 6 *As $N \to \infty$, $\frac{1}{M_N} \sum_{m=1}^{M_N} \boldsymbol{\varphi}_m \boldsymbol{\varphi}_m'$ converges in probability to a finite matrix.*

[10]The degrees-of-freedom adjustment does not affect the asymptotic properties of $\widehat{\boldsymbol{\Theta}}$ but, following standard econometric practice, we use it to improve the finite sample behavior of the estimator.

The following theorem summarizes the preceding discussion and provides an N-consistent estimator of the asymptotic variance-covariance matrix $\mathbf{V}_\delta$.

Theorem 8 *Under Assumptions 1-6, as $N \to \infty$,*

$$\widehat{\mathbf{V}}_\delta = \frac{M_N}{N} \widehat{\mathbf{\Psi}}_\delta \widehat{\mathbf{\Theta}} \widehat{\mathbf{\Psi}}'_\delta \xrightarrow{p} \mathbf{V}_\delta,$$

where the matrices $\widehat{\mathbf{\Psi}}_\delta$ and $\widehat{\mathbf{\Theta}}$ are defined in (49) and (65), respectively.

In the next subsection, we put together the results obtained in the last two subsections and derive novel test statistics that can be used to test the ex-post risk premia and aggregate mispricing implications of the asset pricing model under consideration.

2.6 Test statistics

Combining Theorems 5, 6, 7, and 8 we can readily obtain statistics for testing the implications of the asset pricing model that form our null hypotheses. The first hypothesis focuses on the ex-post risk premia implications: $H_0^\lambda : [\ \lambda_0 \quad \boldsymbol{\lambda}'_f\ ]' = [\ 0 \quad \overline{\mathbf{f}}'_2\ ]'$. The second hypothesis focuses on the aggregate mispricing metric: $H_0^\alpha : \mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)' \mathbf{v}_2$.

First, the ex-post risk premia implications of the asset pricing model can be tested using the quadratic form $J(\boldsymbol{\lambda}) = N \left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda} \right)' \widehat{\mathbf{V}}_\lambda^{-1} \left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda} \right)$, where $\widehat{\mathbf{V}}_\lambda$ is the $(K+1) \times (K+1)$ upper-left block of $\widehat{\mathbf{V}}_\delta$. To test whether the cross-sectional intercept is equal to zero, which should be the case if the asset pricing model is correctly specified, we can use the t-statistic

$$t(\lambda_0) = \frac{\widehat{\lambda}_0}{\sqrt{\widehat{v}_{\lambda,0}/N}}, \tag{70}$$

where $\widehat{\lambda}_0$ denotes the first element of $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}}$ and $\widehat{v}_{\lambda,0}$ denotes the $(1,1)$ element of $\widehat{\mathbf{V}}_\lambda$. The t-statistic $t(\lambda_0)$ asymptotically follows a standard normal distribution. Similarly, to test whether the k-th factor ex-post risk premium is equal to the factor's average realization, for $k = 1, \ldots, K$, which again should be true under correct model specification, we can use the t-statistic

$$t(\lambda_k) = \frac{\widehat{\lambda}_k - \overline{f}_{k,2}}{\sqrt{\widehat{v}_{\lambda,k}/N}}, \tag{71}$$

where $\widehat{\lambda}_k$ denotes the $(k+1)$-th element of $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}}$ and $\widehat{v}_{\lambda,k}$ denotes the $(k+1, k+1)$ element of $\widehat{\mathbf{V}}_\lambda$.

Second, to test the aggregate mispricing hypothesis $H_0^\alpha : \mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)'\mathbf{v}_2$, we propose using the t-statistic

$$t(\boldsymbol{\alpha}) = \frac{\widehat{\mathcal{Q}} - (\mathbf{g}_2 \otimes \mathbf{g}_2)'\widehat{\mathbf{v}}_2 - R_{\widehat{\mathcal{Q}}}}{\sqrt{\widehat{v}_{\delta,K+2}/N}}, \tag{72}$$

where $\widehat{v}_{\delta,K+2}$ is the $(K+2, K+2)$ element of $\widehat{\mathbf{V}}_\delta$. It follows from Theorems 6 and 8 that the test statistic $t(\boldsymbol{\alpha})$ asymptotically follows a standard normal distribution under the null hypothesis $H_0^\alpha : \mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)'\mathbf{v}_2$.

Both ex-post risk premia and individual stock alpha implications can be tested jointly using the quadratic form $J(\boldsymbol{\delta}) = N\left(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right)' \widehat{\mathbf{V}}_\delta^{-1} \left(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right)$ which asymptotically follows a χ^2 distribution with $K+2$ degrees of freedom under the correct model specification. However, simulation evidence suggests that the joint test statistics $J(\boldsymbol{\delta})$ typically overreject the null hypothesis of correct model specification exhibiting poor performance for empirically relevant finite sample sizes. It appears that the reason is that the variance-covariance matrix estimators $\widehat{\mathbf{V}}_\delta$ can be ill-conditioned in small samples. Motivated by this observation, we propose using the following quadratic form

$$J_d(\boldsymbol{\delta}) = N\left(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right)' \widehat{\mathbf{D}}_\delta^{-1} \left(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right), \tag{73}$$

where $\widehat{\mathbf{D}}_\delta$ is the $(K+2) \times (K+2)$ diagonal matrix consisting of the diagonal elements of $\widehat{\mathbf{V}}_\delta$. While the test statistic $J_d(\boldsymbol{\delta})$ does not asymptotically follow a standard distribution, such as χ^2, under the null hypothesis, one can easily compute p-values associated with $J_d(\boldsymbol{\delta})$ using simulation. Let $\mathbf{Q}_\delta$ be the Cholesky factor of $\mathbf{V}_\delta$ so that $\mathbf{V}_\delta = \mathbf{Q}_\delta\mathbf{Q}_\delta'$. Then, the asymptotic distribution of $J_d(\boldsymbol{\delta})$ is the same as the distribution of the quadratic form $\zeta = \boldsymbol{\zeta}' \left[\mathbf{Q}_\delta'\mathbf{D}_\delta^{-1}\mathbf{Q}_\delta\right] \boldsymbol{\zeta}$, where $\mathbf{D}_\delta$ is the $(K+2) \times (K+2)$ diagonal matrix with (j,j) element equal to the (j,j) element of $\mathbf{V}_\lambda$, for $j = 1, \ldots, K+2$, and $\boldsymbol{\zeta}$ follows a $(K+2)$-dimensional standard normal distribution. The matrix $\mathbf{P}_\delta = \mathbf{Q}_\delta'\mathbf{D}_\delta^{-1}\mathbf{Q}_\delta$ can be N-consistently estimated by $\widehat{\mathbf{P}}_\delta = \widehat{\mathbf{Q}}_\delta'\widehat{\mathbf{D}}_\delta^{-1}\widehat{\mathbf{Q}}_\delta$, where $\widehat{\mathbf{Q}}_\delta$ is the Cholesky factor of $\widehat{\mathbf{V}}_\delta$ so that $\widehat{\mathbf{V}}_\delta = \widehat{\mathbf{Q}}_\delta\widehat{\mathbf{Q}}_\delta'$. Let $\{\boldsymbol{\zeta}_i : i = 1, \ldots, I\}$ be a large sample of simulated draws from $N(\mathbf{0}_{K+2}, \mathbf{I}_{K+2})$ and define $\zeta_i = \boldsymbol{\zeta}_i'\widehat{\mathbf{P}}_\delta\boldsymbol{\zeta}_i$, $i = 1, \ldots, I$. It follows by the Monte Carlo principle that the distribution function of ζ, $F_\zeta(a) = \mathrm{P}[\zeta \leq a]$, can be approximated by $\frac{1}{I}\sum_{i=1}^{I} 1_{[\zeta_i \leq a]}$, with the approximation becoming better as N and I increase. In our simulation exercises and empirical tests, we use $I = 100,000$.

Similarly, instead of using $J(\boldsymbol{\lambda})$ to jointly test the ex-post risk premia implications of the model, we propose using the quadratic form

$$J_d(\boldsymbol{\lambda}) = N\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right)' \widehat{\mathbf{D}}_{\lambda}^{-1}\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right), \tag{74}$$

where $\widehat{\mathbf{D}}_{\lambda}$ is the $(K+1)\times(K+1)$ diagonal matrix consisting of the diagonal elements of $\widehat{\mathbf{V}}_{\lambda}$. We compute p-values for the $J_d(\boldsymbol{\lambda})$ test statistic by simulation, as described above.

When we use both past betas and characteristics as instruments, ex-post risk premia are estimated by the IV-GMM estimator (with $L \geq 1$). In this case, the test statistics described above could be used for inference for any generic weighting matrix $\mathbf{W}$. However, as discussed in Subsection 2.4.1, selecting $\mathbf{W}^* = (\boldsymbol{\Pi}_{\lambda}\mathbf{V}_e\boldsymbol{\Pi}_{\lambda}')^{-1}$ yields the efficient IV-GMM estimator. We follow standard GMM practice and obtain the two-step estimator as follows. First, we set $\widehat{\mathbf{W}} = \mathbf{I}_{1+K+L}$ to obtain an initial N-consistent estimator of $\boldsymbol{\lambda}$, say $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{I}}$. Then, using $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{I}}$, we obtain an N-consistent estimator of $\boldsymbol{\Theta}$ from (65) which then provides an N-consistent estimator of $\mathbf{W}^*$, say $\widehat{\mathbf{W}}^*$. Using $\widehat{\mathbf{W}}^*$ as a weighting matrix in the second step, we obtain the two-step IV-GMM estimator of $\boldsymbol{\lambda}$ which is efficient. In addition, we obtain the iterated IV-GMM estimator by repeating the above process of successively obtaining estimators of the risk premia and the asymptotic variance-covariance matrix, in an alternate fashion, till the sequence of risk premia estimators converges. In practice, we stop the iteration when the L_1-norm of the difference between two successive risk premia estimates becomes less than 10^{-6}. We denote the two-step and iterated IV-GMM estimators by $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$ and $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{IT}}$, respectively, and use both of them in our simulations and empirical applications.

3 Monte Carlo Simulation Evidence

In this section, we investigate the properties of the IV-GMM ex-post risk premia estimator $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}}$ and the associated asset pricing tests for empirically relevant finite sample sizes through a number of Monte Carlo simulation experiments. We illustrate the importance of the EIV correction offered by our IV approach in terms of bias reduction and efficiency enhancement by comparing two versions of the IV-GMM ex-post risk premia estimator with two alternative estimators that do not use an EIV correction. We report the bias as well as the root mean square error of all estimators under consideration. Furthermore, we investigate the finite sample performance of the various test statistics in terms of size and power properties. Finally, we illustrate that the traditional Fama and MacBeth (1973) approach for computing standard errors is not suitable in our small-T context as it leads to severe underrejection. As we explain below, we use three popular asset pricing models in our calibration.

Next, we provide the details of our simulation design. We consider all individual stocks in the CRSP universe, trading between 2005 and 2014, i.e., the last ten years of the sample in our empirical exercise presented in the next section, with price above one dollar and, among those, we select the 1,000 stocks with longest time series histories. We jointly calibrate the betas and the idiosyncratic shock variances of those 1,000 stocks in order to simulate excess return data according to the data generating process (2). Our simulation is based on the following three linear asset pricing models: the single-factor CAPM, the three-factor model of Fama and French (1993) and the four-factor model of Hou, Xue, and Zhang (2015). The factors in the second model, which we refer to as the FF3 model, are the market excess return (MKT), the small size minus big size spread portfolio return (SMB), and the high book-to-market minus low book-to-market spread portfolio return (HML).[11] The factors in the third model, which we refer to as the HXZ4 model, are the market excess return (MKT), the difference between the return on a portfolio of small size stocks and the return on a portfolio of big size stocks (ME), the difference between the return on a portfolio of low investment stocks and the return on a portfolio of high investment stocks (I/A), and the difference between the return on a portfolio of high profitability (return on equity) stocks and the return on a portfolio of low profitability stocks (ROE).[12] In the case of the CAPM, we do not use any characteristics as instruments, and therefore the IV-GMM estimator becomes the standard exactly identified IV estimator, which we denote by $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}$. For the FF3 and HXZ4 models, we use the firm characteristics associated with the factors of each model as instrumental variables. Specifically, for the FF3 model, the instrumental variables are the logarithm of size and the logarithm of the book to market ratio averaged over the 2005-2014 period.[13] For the HXZ4 model, the instrumental variables are the logarithm of size, investment over asset, and the logarithm of return on equity averaged over the 2005-2014 period.[14] We provide empirical evidence illustrating that these characteristics are suitable instruments in Section 4. For both the FF3 and HXZ4 models, we consider two efficient versions of the IV-GMM estimator, namely the two-step estimator $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$ and the iterated estimator $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{IT}}$.

We pay particular attention to the following two aspects of the simulation design: (i) the number of clusters in the stock universe and (ii) the correlation structure among stock returns within clusters. Due to space limitations, we only consider clusters of equal size and assume

[11]The data on the three factors of the Fama and French (1993) model are obtained from Kenneth French's data library.

[12]We thank the authors for providing the data on the four factors of the Hou, Xue, and Zhang (2015) model.

[13]We divide the market capitalization of individual stocks by the contemporaneous aggregate market capitalization for normalization.

[14] I/A is defined as change in inventory, property, plant and equipment (PP&E) over the previous year's total asset. ROE is defined as (IB - DVP + TXDI) over book value of equity where IB is the total earnings before extraordinary items, DVP is the preferred dividends (if available), and TXDI is the deferred taxes (if available).

that correlations within clusters are constant. In the first part of the simulation exercise, that focuses on the bias and the mean square error of the IV-GMM estimators, we set the number of clusters, M_N, equal to 50 and the pairwise correlation ρ, within each cluster, equal to 0.10. In the second part of the simulation exercise, that focuses on the finite sample behavior of the various asset pricing test statistics, we let the number of clusters M_N take the values 50 and 100 and the within-cluster correlation take the values 0, 0.10, and 0.20. Note that in the empirical investigation of Section 4, we consider clustering based on the 49-industry classification of Kenneth French.[15] Following this classification, we estimate an average correlation within industries around 0.10 based on an industry residual model for the shocks, in the spirit of Ang, Liu, and Schwarz (2010) (see their Appendix F.2). Hence, the range of correlation values that we employ in our simulation is empirically relevant.

First, we illustrate the importance of the IV-GMM approach in dealing with the EIV problem by comparing the two IV-GMM estimators, $\widehat{\boldsymbol{\lambda}}^{\texttt{TS}}_{\texttt{IV}}$ and $\widehat{\boldsymbol{\lambda}}^{\texttt{IT}}_{\texttt{IV}}$, with two alternative estimators, $\widehat{\boldsymbol{\lambda}}_1$ and $\widehat{\boldsymbol{\lambda}}_2$, in terms of finite sample bias. The first alternative estimator, denoted by $\widehat{\boldsymbol{\lambda}}_1$, ignores the EIV problem and regresses average excess returns over the testing period on a constant and beta estimates obtained by standard time series regression over the pretesting period, that is $\widehat{\boldsymbol{\lambda}}_1 = (\widehat{\mathbf{X}}_1'\widehat{\mathbf{X}}_1)^{-1}\widehat{\mathbf{X}}_1'\bar{\mathbf{r}}_2$, where $\widehat{\mathbf{X}}_1 = [\ \mathbf{1}_N \quad \widehat{\mathbf{B}}_1\ ]$. Similarly, the second alternative estimator, denoted by $\widehat{\boldsymbol{\lambda}}_2$, also ignores the EIV problem but uses beta estimates from the testing period, that is $\widehat{\boldsymbol{\lambda}}_2 = (\widehat{\mathbf{X}}_2'\widehat{\mathbf{X}}_2)^{-1}\widehat{\mathbf{X}}_2'\bar{\mathbf{r}}_2$. We compute the bias of all four estimators as the average of estimation errors over 30,000 Monte Carlo repetitions. We consider pretesting and testing periods that consist of 60 months and, to provide a comprehensive picture, we repeat the exercise over eight testing periods from 1975-1979 to 2010-2014. In the baseline scenario, the idiosyncratic shocks are assumed to follow a normal distribution. The results, reported in Table 1 in annualized basis points (bps), clearly illustrate the bias reduction gains provided by the IV estimators for all three models. For brevity, we only comment on the results for the HXZ4 model. The average absolute biases of $\widehat{\boldsymbol{\lambda}}_1$ ($\widehat{\boldsymbol{\lambda}}_2$) are 757.6 (726.6), 476.8 (470.1), 435.3 (348.2), 367.0 (305.3), and 484.6 (473.7) annualized bps for λ_0, $\lambda_{\texttt{MKT}}$, $\lambda_{\texttt{ME}}$, $\lambda_{\texttt{I/A}}$, and $\lambda_{\texttt{ROE}}$, respectively. When $\widehat{\boldsymbol{\lambda}}^{\texttt{TS}}_{\texttt{IV}}$ ($\widehat{\boldsymbol{\lambda}}^{\texttt{IT}}_{\texttt{IV}}$) is used, the corresponding values are 39.3 (45.1), 16.8 (20.0), 29.7 (32.8), 20.4 (24.0), and 39.6 (46.1) annualized bps. While this evidence is based on normally distributed disturbances, additional simulation experiments show that these results are robust to the assumption of normality.[16]

Next, we compare the IV-GMM estimators to the alternative estimators in terms of mean

[15]The classification is available at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data-Library/det-49-ind-port.html`.

[16] We repeat the same exercise under the assumption that the idiosyncratic disturbances follow a Student-t distribution with 6 degrees of freedom. The results, reported in Table A1 in the Online Appendix, are almost identical. Hence, our conclusions regarding the superior performance of the IV-GMM estimators in terms of bias reduction is robust to the assumption of normally distributed disturbances.

square error. The purpose of this exercise is to examine whether the gain in bias reduction comes at the cost of a higher variance and perhaps efficiency loss. We compute the root mean square error (RMSE) of each estimator as the square root of the sample mean of squared estimation errors over 30,000 Monte Carlo simulations. The simulation setup is identical to the one used above to examine the finite sample bias. In Table 2, we report, in units of annualized bps, the RMSE of the IV-GMM estimators along with those of the alternative estimators $\widehat{\boldsymbol{\lambda}}_1$ and $\widehat{\boldsymbol{\lambda}}_2$. The results clearly illustrate that the IV-GMM estimators achieve much lower mean square errors in comparison with the alternative estimators for all three models. Again, for brevity, we only comment on the results for the HZX4 model. The average RMSEs of $\widehat{\boldsymbol{\lambda}}_1$ ($\widehat{\boldsymbol{\lambda}}_2$) are 784.9 (754.7), 522.4 (514.3), 463.3 (384.2), 382.2 (321.0), and 499.3 (488.9) annualized bps for λ_0, $\lambda_{\texttt{MKT}}$, $\lambda_{\texttt{ME}}$, $\lambda_{\texttt{I/A}}$, and $\lambda_{\texttt{ROE}}$, respectively. When $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$ ($\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{IT}}$) is used, the corresponding values are 475.6 (477.1), 379.3 (380.1), 338.0 (337.4), 289.7 (290.9), and 467.0 (469.5) annualized bps. As in the case of bias, while this evidence is based on normally distributed disturbances, additional simulation experiments show that the results on mean square error are robust to the assumption of normality.[17] Collectively, the simulation evidence, which is robust across different factor model specifications and distributional assumptions, illustrates that the IV-GMM estimators exhibit superior performance in terms of bias reduction without sacrificing efficiency.

In our next simulation exercise, we investigate the behavior of the asset pricing test statistics based on the IV estimators and the associated variance-covariance matrix estimators. Specifically, we focus on empirical rejection frequencies of (i) the t statistics $t(\lambda_0)$ and $t(\lambda_k)$, $k = 1, \ldots, K$, testing the simple hypotheses $\lambda_0 = 0$ and $\lambda_k = \overline{f}_{k,2}$, for $k = 1, \ldots, K$, given in (70) and (71); (ii) the statistic $J_d(\boldsymbol{\lambda})$ testing the joint ex-post risk premia hypothesis $H_0^{\lambda} : [\ \lambda_0 \quad \boldsymbol{\lambda}_f'\ ]' = [\ 0 \quad \overline{\mathbf{f}}_2'\ ]'$, given in (74); (iii) the t-statistic $t(\boldsymbol{\alpha})$ testing the aggregate mispricing hypothesis $H_0^{\alpha} : \mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)'\mathbf{v}_2$, given in (72); and (iv) the statistic $J_d(\boldsymbol{\delta})$ jointly testing the hypotheses H_0^{λ} and H_0^{α}, given in (73). We fix the factors for the pretesting and testing periods, consisting of $\tau_1 = 60$ and $\tau_2 = 60$ observations, as the historical factors over 2005 to 2009 and 2010 to 2014, respectively. Using the factor realizations and the calibrated pairs of betas and idiosyncratic shock variances, we simulate individual stock returns using the data generating process (2) for $t = 1, \ldots, 120$. Since the asymptotic variance of the IV-GMM estimators crucially depends on the cluster structure, we consider a number of difference scenarios. Specifically, we let the number of clusters M_N take the values 50 and 100 and assume that, within each cluster, pairwise correlations are equal to ρ which takes the following three values: 0, 0.10, and 0.20. We consider three nominal levels of significance, 1%, 5%, and 10%,

[17] We repeat the same exercise under the assumption that the idiosyncratic shocks follow a Student-t distribution with 6 degrees of freedom. Table A2 in the Online Appendix reports the results that are almost identical to the ones obtained under the assumption of normally distributed shocks.

and compute the corresponding empirical rejection frequencies from 30,000 Monte Carlo repetitions. The simulation exercise is first performed for idiosyncratic shocks following a normal distribution and the results are reported in Table 3. We observe that the t statistics $t(\lambda_0)$, $t(\lambda_k)$, $k = 1, \ldots, K$, and $t(\boldsymbol{\alpha})$ as well as the joint test statistics $J_d(\boldsymbol{\lambda})$ and $J_d(\boldsymbol{\delta})$ all yield empirical rejection frequencies that are reasonably close to the corresponding nominal levels of significance. The simulation is repeated for shocks following a Student-t distribution with 6 degrees of freedom and the results are almost identical.[18] The conclusions hold for all three asset pricing factor models considered and under both distributional assumptions.

Next, we illustrate the importance of using the variance-covariance estimator $\widehat{\mathbf{V}}_\lambda$ for obtaining ex-post risk premia tests with good size properties.[19] We do so by examining the empirical rejection frequencies of the various tests when the variance-covariance matrix of the IV ex-post risk premia estimator is estimated using the Fama and MacBeth (1973) (FMB) procedure as suggested by Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2015), for the CAPM, the FF3 model, and the HXZ4 model. It is important to emphasize the difference between our setting and the setting in Jegadeesh, Noh, Pukthuanthong, Roll, and Wang (2015). We perform our simulations and empirical tests over short time intervals, covering 120 months, while they use much longer horizons, ranging from 480 to 720 months. The point of our simulation is to illustrate that using the FMB procedure results in misleading inference about ex-post risk premia in a small T context like ours. Table 4 presents the empirical rejection frequency results for the case of idiosyncratic shocks following a normal distribution under two scenarios. In the first scenario, the pretesting and testing periods are 2000-2004 and 2005-2009, respectively, while in the second scenario, the pretesting and testing periods are 2005-2009 and 2010-2014, respectively. The factor betas and the characteristics are calibrated as before and the factor realizations are kept fixed over simulation repetitions during both pretesting and testing periods. For the majority of the tests, we consistently find that the FMB variance-covariance matrix leads to severe underrejection under both scenarios. The simulation is repeated for shocks following a Student-t distribution with 6 degrees of freedom and the results are almost identical.[20] The conclusion that the FMB procedure leads to severe underrejection holds for all three asset pricing factor models and under both distributional assumptions considered.

Our simulations, so far, have focused on the accuracy of the size of the proposed tests, in finite samples, under the null hypothesis of the asset pricing model is correctly specified and the return generating process over the testing period is given by $\mathbf{R}_2 = \mathbf{B}\mathbf{F}_2 + \mathbf{U}_2$. Another important aspect is the power of the tests, i.e., the frequency with which the null hypothesis is rejected under a different return generating process. To illustrate the power properties of the tests, we

[18]The results are reported in Table A3 in the Online Appendix.

[19]Recall that $\widehat{\mathbf{V}}_\lambda$ is the $(K+1) \times (K+1)$ upper-left block of the estimator $\widehat{\mathbf{V}}_\delta$ provided in Theorem 8.

[20]The results are reported in Table A4 in the Online Appendix.

consider a deviation in the return generating process described by $\mathbf{R}_2 = \boldsymbol{\vartheta}\mathbf{1}'_{\tau_2} + \mathbf{B}\mathbf{F}_2 + \mathbf{U}_2$, where $\boldsymbol{\vartheta}$ is an N-dimensional vector. For illustrative purposes, we restrict attention to the $J_d(\boldsymbol{\delta})$ test statistic defined in (73) and follow the simulation design used for Table 3 while focusing on the case with $\rho = 0.1$ and $M_N = 50$. One alternative hypothesis of interest is the case in which expected asset returns are determined by the so-called characteristics rewards, in addition to factor risk premia. Accordingly, we set the vector $\boldsymbol{\vartheta}$ capturing the deviation from the asset pricing model equal to $\boldsymbol{\vartheta} = \mathbf{c}_{1,\texttt{SIZE}}\vartheta_{\texttt{SIZE}} + \mathbf{c}_{1,\texttt{BTM}}\vartheta_{\texttt{BTM}}$, where $\mathbf{c}_{1,\texttt{SIZE}}$ and $\mathbf{c}_{1,\texttt{BTM}}$ are the N-dimensional vectors of SIZE and BTM characteristics and where $\vartheta_{\texttt{SIZE}}$ and $\vartheta_{\texttt{BTM}}$ are the SIZE and BTM characteristics rewards, respectively. The characteristics are obtained as averages over the pretesting period and are standardized so that they have zero mean and unit variance cross-sectionally. To examine the power of the $J_d(\boldsymbol{\delta})$ test, we let $\vartheta_{\texttt{SIZE}}$ and $\vartheta_{\texttt{BTM}}$ take nine values between -0.4 and 0.4, in percentage, and compute the rejection frequency in each case. Table 5 reports the empirical rejection frequencies of the $J_d(\boldsymbol{\delta})$ test with 5% nominal significance level for the various choices of $\vartheta_{\texttt{SIZE}}$ and $\vartheta_{\texttt{BTM}}$ when shocks are normally distributed. The results indicate that the empirical rejection frequencies quickly increase as $\vartheta_{\texttt{SIZE}}$ or $\vartheta_{\texttt{BTM}}$ starts deviating from zero, illustrating the good power properties of the $J_d(\boldsymbol{\delta})$ test. The simulation is repeated for shocks following a Student-t distribution with 6 degrees of freedom and the results are almost identical.[21]

4 Empirical Evidence

In this section, we use the IV-GMM approach developed in Section 2 to empirically evaluate a number of popular factor models that have been proposed in the asset pricing literature. Specifically, we focus on four models: (i) the standard single-factor CAPM, (ii) the three-factor Fama and French (1993) model (FF3), (iii) the four-factor Hou, Xue, and Zhang (2015) model (HXZ4), and (iv) the five-factor Fama and French (2015) model (FF5). The factors involved in the CAPM, FF3, and FF5 models are obtained from Kenneth French's website. In particular, the market excess return (MKT) is used by all three aforementioned models; the small size minus big size spread portfolio return (SMB) and the high book-to-market minus low book-to-market spread portfolio return (HML) are used by both the FF3 and FF5 models;[22] and, finally, the robust minus weak spread portfolio return (RMW) and conservative minus aggressive spread portfolio return (CMA) are used by FF5. The four factors involved in the HXZ4 q-factor model are the market excess return (MKT), the difference between the return on a portfolio of small

[21] The results are reported in Table A5 in the Online Appendix.

[22] The SMB factor used by the FF3 model is slightly different from the SMB factor used by the FF5 model. Details on how the SMB factor is constructed for each model are provided in Kenneth French's data library website.

size stocks and the return on a portfolio of big size stocks (ME), the difference between the return on a portfolio of low investment stocks and the return on a portfolio of high investment stocks (I/A), and the difference between the return on a portfolio of high profitability (return on equity) stocks and the return on a portfolio of low profitability stocks (ROE).[23] We use individual stock data at the monthly frequency covering the time period between 1970 and 2014 from the CRSP universe and apply the following filters: (i) we require that the share code (SHRCD) is equal to 10 or 11 to keep only ordinary common shares, (ii) we require that the exchange code (EXCHCD) is equal to 1, 2, or 3 to keep only stocks traded at NYSE, AMEX, or NASDAQ, and (iii) we keep a stock in the sample only for the months in which its price (PRC) is at least 1 dollar. When we use clustering based on the 49-industry classification of Kenneth French for estimating the variance-covariance matrix of the IV-GMM estimators, we further we require that stocks have a Standard Industry Classification (SIC) code.[24]

Our pretesting and testing periods consist of five years ($\tau_1 = \tau_2 = 60$ months) resulting in 8 non-overlapping testing periods from 1975 to 2014. The cross section of our test assets consists of all stocks with full histories over both the pretesting and testing periods. Our empirical evidence consists of (i) estimates of $\boldsymbol{\lambda} = \left[\begin{array}{cc} \lambda_0 & \boldsymbol{\lambda}_f' \end{array} \right]'$, where $\boldsymbol{\lambda}_f = \left[\begin{array}{ccc} \lambda_1 & \cdots & \lambda_K \end{array} \right]'$ is the vector of ex-post factor risk premia for the CAPM, FF3, HXZ4, and FF5 models, which they employ $K = 1$, $K = 3$, $K = 4$, and $K = 5$ factors, respectively, and (ii) various test statistics for evaluating the implications of each model and their corresponding p-values.

When estimating ex-post risk premia for the FF3, HXZ4, and FF5 models, in addition to past beta estimates, we employ as instrumental variables the firm characteristics based on which the various factors are constructed. Specifically, for the FF3 model, we use market capitalization (SIZE) and book-to-market ratio (BTM) as firm characteristics. For the HXZ4 model, we use SIZE, investment over asset (I/A), and return on equity (ROE) as firm characteristics. Lastly, for the FF5 model, we use SIZE, BTM, operating profitability (OP) and asset growth (AG) as firm characteristics. Next, we describe how the above characteristics are computed. The SIZE characteristic for month m is defined as the ratio of the market capitalization a given firm at the end of month $m-1$ to the aggregate market capitalization at the end of month $m-1$. The BTM characteristic from July of year $y+1$ till June of year $y+2$ is defined as the ratio of book equity (BE) in the accounting data of fiscal year y to the market capitalization at the end of year y. BE is computed following the method in Kenneth French's database, i.e., BE is defined the book value of stockholders equity (SEQ), plus balance sheet deferred taxes

[23] We are grateful to the authors for providing the data on the factors of the Hou, Xue, and Zhang (2015) model.

[24] SIC codes are obtained from Compustat. If the SIC code does not exist in Compustat, it is obtained from CRSP.

and investment tax credit (TXDITC, if available), minus the book value of preferred stock.[25] The I/A characteristic from July of year $y+1$ till June of year $y+2$ is defined as change in inventory, property, plant and equipment (PP&E) from year $y-1$ to year y over the year $y-1$ total assets. The ROE characteristic from July of year $y+1$ till June of year $y+2$ is defined as the ratio of (IB − DVP + TXDI) for the year y over BE of year y, where IB is the total earnings before extraordinary items, DVP is the preferred dividends (if available), and TXDI is the deferred taxes (if available). The OP characteristic from July of year $y+1$ till June of year $y+2$ is defined as the ratio of (REV − COGS − XINT − XSGA) for year y over BE of year y, where REV is revenue, COGS is cost of goods sold, XINT is interest expense, and XSGA is selling, general and administrative expenses. Finally, the AG characteristic from July of year $y+1$ till June of year $y+2$ is defined as the ratio of change in the total assets from year $y-1$ to year y over the year $y-1$ total assets. With the exception of I/A, the cross-sectional distributions of these characteristics are highly skewed. Hence, for all characteristics except I/A, we follow standard practice and use the logarithm of the average characteristics over the pretesting period as instrumental variables.[26]

The asset pricing models under examination employ factors that are constructed as differences between returns on top and bottom portfolios, or vice versa, after sorting according to a particular firm characteristic. As a result, we expect a firm characteristic to be cross-sectionally correlated with the beta with respect to the corresponding spread factor. This provides a clear rationale for using firm characteristics as instrumental variables, in addition to past beta estimates. In Table 6, we provide evidence supporting this rationale. Specifically, for each characteristic, we consider decile portfolios sorted according to that characteristic and estimate their betas with respect to the related spread factor within the context of each model that we evaluate. For each asset pricing model, the portfolio betas are estimated jointly for all factors using data from 07/1970 to 12/2014. As illustrated in Table 6, there is a clear monotonic pattern in the betas for each spread factor within each model. This evidence justifies our choice of firm characteristics as instruments in the estimation of ex-post risk premia.

We present our empirical results in Tables 7 through 10. We focus on the risk premia t-statistics $t(\lambda_0)$ and $t(\lambda_k)$, $k = 1, \ldots, K$, given in (70) and (71), respectively, the joint risk premia test statistic $J_d(\boldsymbol{\lambda})$, given in (74), the aggregate mispricing t-statistic $t(\boldsymbol{\alpha})$, given in (72), and the joint test statistic $J_d(\boldsymbol{\delta})$, given in (73).[27] In our discussion of the results, we consider both of the conventional 5% and 10% level of significance. The results for the CAPM,

[25] For a more detailed description, the reader is referred to the definition of BE at Kenneth French's website `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/variable_definitions.html`.

[26] For AG, we use the logarithm of one plus average of asset growth.

[27] As mentioned in Section 3, the joint test statistic $J(\boldsymbol{\delta})$ tends to overreject the null hypothesis in small samples. That is why we focus on the $J_d(\boldsymbol{\lambda})$ and $J_d(\boldsymbol{\delta})$ joint test statistics.

based on the IV estimator $\widehat{\boldsymbol{\lambda}}_{\text{IV}}$ and using past beta estimates as instruments, are reported in Table 7. Overall, our evidence points to rejection of the CAPM. The $t(\lambda_0)$ statistic rejects the null hypothesis in five out of eight testing periods at both 5% and 10% levels of significance. Similarly, the $t(\lambda_{\text{MKT}})$ statistic rejects the null hypothesis in five and seven out of eight testing periods at the 5% and 10% level of significance, respectively. The joint test statistics yield similar results. The $J_d(\boldsymbol{\lambda})$ joint statistic rejects the null hypothesis in five and six out of eight testing periods at the 5% or 10% level of significance, respectively. Regarding the joint tests on the magnitude of individual stock alphas, the $t(\boldsymbol{\alpha})$ test rejects the aggregate mispricing null hypothesis $H_0^{\alpha} : \mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)'\mathbf{v}_2$ in all testing periods at both 5% or 10% level of significance. Finally, the same results of uniform rejections at the both significance levels hold for the $J_d(\boldsymbol{\delta})$ statistic used to jointly test the hypotheses $H_0^{\lambda} : [\ \lambda_0 \quad \boldsymbol{\lambda}_f'\ ]' = [\ 0 \quad \overline{\mathbf{f}}_2'\ ]'$ and $H_0^{\alpha} : \mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)'\mathbf{v}_2$. These findings are not very surprising given that the CAPM has been frequently rejected in the literature using portfolios of stocks.

The results for the FF3 model, based on the two-step and iterated IV-GMM estimators, i.e., $\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{TS}}$ and $\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{IT}}$, and using SIZE and BTM as instruments in addition to past beta estimates, are reported in Table 8. The two IV-GMM estimators overall yield similar results. Based on either $\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{TS}}$ or $\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{IT}}$, the $t(\lambda_0)$ statistic rejects the null hypothesis in four out of eight testing periods at the 5% and 10% level of significance. Other t statistics testing the null hypotheses $H_0 : \lambda_k = \overline{f}_{k,2}$, for $k = 1, 2, 3$, exhibit similar behavior except the t statistic corresponding to the HML factor. Consistent with the findings in Kim and Skoulakis (2015), the $t(\lambda_{\text{HML}})$ statistic rejects the null hypothesis only in one and three of eight periods for the 5% and 10% levels of significance, respectively, when the $\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{TS}}$ estimator is used. The results based on the joint test statistics are discussed next. Since the results based on $\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{TS}}$ are similar to those based on $\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{IT}}$, we focus on the results based on the former. The $J_d(\boldsymbol{\lambda})$ joint statistic rejects the null hypothesis in five and six out of eight testing periods at the 5% and 10% level of significance, respectively. The $t(\boldsymbol{\alpha})$ test rejects the aggregate mispricing null hypothesis $H_0^{\alpha} : \mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)'\mathbf{v}_2$ in five out of eight testing periods at both 5% or 10% level of significance. The $J_d(\boldsymbol{\delta})$ statistic, used to jointly test the hypotheses $H_0^{\lambda} : [\ \lambda_0 \quad \boldsymbol{\lambda}_f'\ ]' = [\ 0 \quad \overline{\mathbf{f}}_2'\ ]'$ and $H_0^{\alpha} : \mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)'\mathbf{v}_2$, rejects the null in six out of eight testing periods at the 5% level of significance. Although overall results for the FF3 model are better than those for the CAPM, the null hypothesis is rejected in the majority of the tests.

In Table 9, we report the results for the HZX4 model, based on the two-step and iterated IV-GMM estimators, i.e., $\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{TS}}$ and $\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{IT}}$, and using SIZE, I/A, and ROE as instruments in addition to past beta estimates. Once again, the two IV-GMM estimators yield similar results. Hence, we only mention the results based on $\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{IT}}$. It is worth noting that the $t(\lambda_0)$ statistic rejects the null hypothesis only in two out of eight testing periods at the 10% level of significance.

The additional t statistics point to similar conclusions. The $t(\lambda_{\texttt{MKT}})$, $t(\lambda_{\texttt{ME}})$, $t(\lambda_{\texttt{I/A}})$, and $t(\lambda_{\texttt{ROE}})$ statistics reject the null hypothesis in two, four, two, and two out of eight testing periods, respectively, at the 10% level of significance. Moreover, the $J_d(\boldsymbol{\lambda})$ joint statistic rejects the null hypotheses in three out of eight testing periods at the 10% level of significance. The $t\,(\boldsymbol{\alpha})$ test rejects the aggregate mispricing null hypothesis H_0^{α} : $\mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)'\mathbf{v}_2$ in four out of eight testing periods at the 10% level of significance. As expected, the $J_d(\boldsymbol{\delta})$ statistic, testing jointly all of the hypotheses H_0^{λ} : $[\ \lambda_0 \quad \boldsymbol{\lambda}_f'\ ]' = [\ 0 \quad \overline{\mathbf{f}}_2'\ ]'$ and H_0^{α} : $\mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)'\mathbf{v}_2$, rejects the null hypotheses more frequently, specifically in five out eight testing periods at the 5% level of significance.

The results for the FF5 model, based on the two-step and iterated IV-GMM estimators, i.e., $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$ and $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{IT}}$, and using SIZE, BTM, OP, and AG as instruments in addition to past beta estimates, are reported in Table 10. As before, the results are consistent across the two IV-GMM estimators. Hence, we focus on the results based on $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$. The $t(\lambda_0)$ statistic rejects the null hypothesis in three out of eight testing periods at the 10% level of significance. The $t(\lambda_{\texttt{MKT}})$, $t(\lambda_{\texttt{SMB}})$, $t(\lambda_{\texttt{HML}})$, $t(\lambda_{\texttt{RMW}})$, and $t(\lambda_{\texttt{CMA}})$ statistics reject the null hypothesis in six, four, one, five, and three out of eight testing periods, respectively, at the 10% level of significance. The $J_d(\boldsymbol{\lambda})$ joint statistic rejects the null hypotheses in five out of eight testing periods at the 10% level of significance. However, the performance of FF5 in terms of aggregate mispricing is disappointing. The $t\,(\boldsymbol{\alpha})$ test rejects the aggregate mispricing null hypothesis $H_0^{\alpha} : \mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)'\mathbf{v}_2$ in every testing period at the 10% level of significance. This poor performance in the $t\,(\boldsymbol{\alpha})$ test is transferred to the $J_d(\boldsymbol{\delta})$ statistic jointly testing the hypotheses $H_0^{\lambda} : [\ \lambda_0 \quad \boldsymbol{\lambda}_f'\ ]' = [\ 0 \quad \overline{\mathbf{f}}_2'\ ]'$ and $H_0^{\alpha} : \mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)'\mathbf{v}_2$. The $J_d(\boldsymbol{\delta})$ joint test rejects the null hypotheses seven and eight out of eight testing periods at the 5% and 10% level of significance, respectively.

Collectively, our results show that the CAPM and the FF3 and models are mostly rejected by the IV-GMM test statistics. In contrast, the recent HZX4 FF5 asset pricing models perform better. The overall results are most favorable to the HXZ4 model. In the Online Appendix, we report the test statistics and the corresponding p-values for the four models examined above using alternative clustering schemes and price filters. Tables A6– A9 contain the results based on 49 industry clusters and a \$3 price filter. Tables A10– A13 contain the results based on 49 industry clusters and a \$5 price filter. Finally, tables A14– A17 contain the results based on 30 industry clusters and a \$1 price filter. The results of our tests remain very similar under all alternative scenarios.

5 Conclusion

A linear asset pricing factor model characterizes the average return of an asset as a linear function of its factor betas with the risk premia being the slopes. In theory, such a relationship is supposed to be valid for all individual assets. However, the majority of empirical tests of asset pricing models are based on portfolios. One of the main reasons for this practice is that individual stock beta estimates are plagued by significant sampling error giving rise to the well-known error-in-variables (EIV) problem. When the size of the cross section N is large, while the the time series sample size T is small and fixed, the EIV problem is so severe that it renders the standard two-pass cross-sectional regression (CSR) risk premia estimator inconsistent. To deal with the EIV problem, we develop a modification of the two-pass CSR approach that employs past beta estimates and firm characteristics as instrumental variables and yields an IV-GMM ex-post risk premia estimator. We contribute to the literature by providing a novel method for estimating ex-post risk premia and devising associated tests for evaluating linear factor models using individual stock data over short time horizons. We establish that the ex-post risk premia estimator is N-consistent and asymptotically follows a normal distribution. Furthermore, using a cluster structure for idiosyncratic shock correlations, we provide an estimator of its asymptotic variance-covariance matrix that we then use to construct asset pricing tests focusing on ex-post risk premia.

The good performance of the IV-GMM estimator and the associated variance-covariance matrix estimator for empirically relevant finite sample sizes is illustrated through a number of Monte Carlo simulations. Using three different asset pricing models for calibration, we show that (i) the IV-GMM approach leads to significant bias reduction in the cross-sectional regression intercept and ex-post risk premia estimates without sacrificing efficiency, and (ii) the associated asset pricing test statistics yield empirical rejection frequencies very close to the desired levels of significance. In our empirical investigation, we estimate and evaluate four popular linear asset pricing factor models: the CAPM, the three-factor of model of Fama and French (1993), the q-factor model of Hou, Xue, and Zhang (2015), and the five-factor model of Fama and French (2015). We find that all models are rejected for the majority of our testing periods with the exception of the Hou, Xue, and Zhang (2015) model, for which we find strong support in five out of the eight time periods under all clustering schemes under consideration.

A Proofs

A few facts from matrix algebra are used in the main text and/or in the subsequent proofs. We collect them here for the convenience of the reader. In terms of notation, vec denotes the column-stacking operator and $\otimes$ denotes the Kronecker product.

(F1) For column vectors $\mathbf{x}$ and $\mathbf{y}$, we have $\text{vec}\,(\mathbf{x}\mathbf{y}') = \mathbf{y} \otimes \mathbf{x}$.

(F2) For conformable matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, we have $\text{vec}\,(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\,\text{vec}\,(\mathbf{B})$.

(F3) For conformable matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$, we have $(\mathbf{AC}) \otimes (\mathbf{BD}) = (\mathbf{A} \otimes \mathbf{B})\,(\mathbf{C} \otimes \mathbf{D})$.

(F4) Let $\boldsymbol{\mathcal{S}}$ be the $\tau_2^2 \times \tau_2$ selection matrix such that the $(\tau_2\,(s-1)+s, s)$ element of $\boldsymbol{\mathcal{S}}$ is 1, for $s = 1, \ldots, \tau_2$, and all other elements are zero. Then, $\mathbf{A} \odot \mathbf{B} = \boldsymbol{\mathcal{S}}'(\mathbf{A} \otimes \mathbf{B})\boldsymbol{\mathcal{S}}$, for any $\tau_2 \times \tau_2$ matrices $\mathbf{A}$ and $\mathbf{B}$.

The facts (F1), (F2), (F3), and (F4) follow from Theorem 8.9, Theorem 8.11, Theorem 8.2, and Section 8.5 in Schott (2017), respectively.

Proof of Theorem 1: Recall from equation (23) that

$$\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} = \boldsymbol{\lambda} + \left(\widehat{\boldsymbol{\Omega}}'\widehat{\mathbf{W}}\widehat{\boldsymbol{\Omega}}\right)^{-1}\widehat{\boldsymbol{\Omega}}'\widehat{\mathbf{W}}\left(\widehat{\mathbf{Z}}_1'\boldsymbol{\omega}_2/N\right), \tag{75}$$

where $\widehat{\boldsymbol{\Omega}} = \frac{1}{N}\widehat{\mathbf{Z}}_1'\widehat{\mathbf{X}}_2$. In light of (14), it follows from Assumption 1 that, as $N \to \infty$, $\widehat{\mathbf{B}}_1'\mathbf{1}_N/N = \mathbf{B}'\mathbf{1}_N/N + \mathbf{G}_1'(\mathbf{U}_1'\mathbf{1}_N/N) \xrightarrow{p} \boldsymbol{\mu}_\beta$, $\widehat{\mathbf{B}}_2'\mathbf{1}_N/N = \mathbf{B}'\mathbf{1}_N/N + \mathbf{G}_2'(\mathbf{U}_2'\mathbf{1}_N/N) \xrightarrow{p} \boldsymbol{\mu}_\beta$, and $\widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2/N = \mathbf{B}'\mathbf{B}/N + \mathbf{G}_1'(\mathbf{U}_1'\mathbf{B}/N) + (\mathbf{B}'\mathbf{U}_2/N)\mathbf{G}_2 + \mathbf{G}_1'(\mathbf{U}_1'\mathbf{U}_2/N)\mathbf{G}_2 \xrightarrow{p} \mathbf{M}_\beta$, where $\mathbf{M}_\beta = \mathbf{V}_\beta + \boldsymbol{\mu}_\beta\boldsymbol{\mu}_\beta'$. Moreover, using (14) again, it follows from Assumption 2 that $\mathbf{C}_1'\mathbf{1}_N/N \xrightarrow{p} \boldsymbol{\mu}_c$ and $\mathbf{C}_1'\widehat{\mathbf{B}}_2/N = \mathbf{C}_1'\mathbf{B}/N + (\mathbf{B}'\mathbf{U}_2/N)\mathbf{G}_2 \xrightarrow{p} \mathbf{M}_{c\beta}$. It then follows from definitions (17) and (22) and the above probability limits that

$$\widehat{\boldsymbol{\Omega}} = \widehat{\mathbf{Z}}_1'\widehat{\mathbf{X}}_2/N \xrightarrow{p} \boldsymbol{\Omega}, \tag{76}$$

where

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & \boldsymbol{\mu}_\beta' \\ \boldsymbol{\mu}_\beta & \mathbf{M}_\beta \\ \boldsymbol{\mu}_c & \mathbf{M}_{c\beta} \end{bmatrix}. \tag{77}$$

According to Assumption 1, $\mathbf{V}_\beta$ is positive definite. Since $\mathbf{M}_\beta = \mathbf{V}_\beta + \boldsymbol{\mu}_\beta\boldsymbol{\mu}_\beta'$, it follows that the $(1 + K + L) \times (1 + K)$ matrix $\boldsymbol{\Omega}$ has full rank. Thus, to prove the theorem, it suffices to show that, as $N \to \infty$, $\widehat{\mathbf{Z}}_1'\boldsymbol{\omega}_2/N$ converges to a vector of zeros. Using equations (14), and (19), and invoking Assumption 1(i), Assumption 1(ii), and Assumption 2(ii), we obtain that, as $N \to \infty$, $\mathbf{1}_N'\boldsymbol{\omega}_2/N = (\mathbf{1}_N'\mathbf{U}_2/N)\mathbf{g}_2 \xrightarrow{p} 0$, $\widehat{\mathbf{B}}_1'\boldsymbol{\omega}_2/N = (\mathbf{B}'\mathbf{U}_2/N)\mathbf{g}_2 + \mathbf{G}_1'(\mathbf{U}_1'\mathbf{U}_2/N)\mathbf{g}_2 \xrightarrow{p} \mathbf{0}_K$, and $\mathbf{C}_1'\boldsymbol{\omega}_2/N = (\mathbf{C}_1'\mathbf{U}_2/N)\mathbf{g}_2 \xrightarrow{p} \mathbf{0}_L$. Then, using equation (22), we obtain

$$\widehat{\mathbf{Z}}_1'\boldsymbol{\omega}_2/N \xrightarrow{p} \mathbf{0}_{1+K+L}. \tag{78}$$

Combining (75) with the probability limits in (76) and (78) yields the N-consistency of $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}}$. ■

<u>**Proof of Proposition 2**</u>: First note that $\widehat{\boldsymbol{\alpha}} = \boldsymbol{\omega}_2 - \widehat{\mathbf{X}}_2\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right)$, where $\boldsymbol{\omega}_2$ is defined in (18). Hence,

$$\widehat{\mathcal{Q}} = \boldsymbol{\omega}_2'\boldsymbol{\omega}_2/N - 2\left(\boldsymbol{\omega}_2'\widehat{\mathbf{X}}_2/N\right)\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right) + R_{\widehat{\mathcal{Q}}}. \tag{79}$$

where

$$R_{\widehat{\mathcal{Q}}} = \left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right)'\left(\widehat{\mathbf{X}}_2'\widehat{\mathbf{X}}_2/N\right)\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right). \tag{80}$$

According to equation (19), $\boldsymbol{\omega}_2 = \mathbf{U}_2\mathbf{g}_2$, where $\mathbf{g}_2$ is defined in (20). Assumption 3 then implies that, as $N \to \infty$,

$$\boldsymbol{\omega}_2'\boldsymbol{\omega}_2/N \xrightarrow{p} \mathbf{g}_2'\mathbf{V}_2\mathbf{g}_2 = \left(\mathbf{g}_2' \otimes \mathbf{g}_2'\right)\operatorname{vec}(\mathbf{V}_2) = \left(\mathbf{g}_2 \otimes \mathbf{g}_2\right)'\mathbf{v}_2, \tag{81}$$

where $\mathbf{v}_2$ is defined in (28). Moreover, equation (14) implies $\widehat{\mathbf{X}}_2 = [\ \mathbf{1}_N \quad \mathbf{B}\ ] + [\ \mathbf{0}_N \quad \mathbf{U}_2\mathbf{G}_2\ ]$. It follows from Assumptions 1 and 3 that, as $N \to \infty$,

$$\widehat{\mathbf{X}}_2'\boldsymbol{\omega}_2/N \xrightarrow{p} \boldsymbol{\rho}_\alpha, \tag{82}$$

where

$$\boldsymbol{\rho}_\alpha = [\ 0 \quad \mathbf{g}_2{}'\mathbf{V}_2\mathbf{G}_2\ ]', \tag{83}$$

and $\widehat{\mathbf{X}}_2'\widehat{\mathbf{X}}_2/N$ converges to a finite limit. Finally, it follows from (79), Theorem 1 and the above probability limits that $\widehat{\mathcal{Q}} \xrightarrow{p} (\mathbf{g}_2 \otimes \mathbf{g}_2)' \mathbf{v}_2$. ■

Proof of Proposition 3: Using equation (30), we obtain $\text{vec}\left(\widehat{\mathbf{U}}_2'\widehat{\mathbf{U}}_2/N\right) = \text{vec}\left(\mathbf{H}_2\left(\mathbf{U}_2'\mathbf{U}_2/N\right)\mathbf{H}_2\right) = \left(\mathbf{H}_2 \otimes \mathbf{H}_2\right)\text{vec}\left(\mathbf{U}_2'\mathbf{U}_2/N\right)$ where the last equality follows from fact (F2). It then follows from (32) that

$$\widehat{\mathbf{v}}_2 = \left[\boldsymbol{\mathcal{S}}\left(\mathbf{H}_2 \odot \mathbf{H}_2\right)^{-1}\boldsymbol{\mathcal{S}}'\right]\left(\mathbf{H}_2 \otimes \mathbf{H}_2\right)\text{vec}\left(\mathbf{U}_2'\mathbf{U}_2/N\right). \tag{84}$$

Invoking Assumption 3, we obtain $\widehat{\mathbf{v}}_2 \xrightarrow{p} \left[\boldsymbol{\mathcal{S}}\left(\mathbf{H}_2 \odot \mathbf{H}_2\right)^{-1}\boldsymbol{\mathcal{S}}'\right]\left(\mathbf{H}_2 \otimes \mathbf{H}_2\right)\mathbf{v}_2$. Let $\mathbf{d}_2$ be the $\tau_2 \times 1$ vector with s-th element equal to $v_{2,s}$, $s = 1, \ldots, \tau_2$, so that $\mathbf{v}_2 = \boldsymbol{\mathcal{S}}\mathbf{d}_2$.

It follows that $\left[\boldsymbol{\mathcal{S}}\left(\mathbf{H}_2 \odot \mathbf{H}_2\right)^{-1}\boldsymbol{\mathcal{S}}'\right]\left(\mathbf{H}_2 \otimes \mathbf{H}_2\right)\mathbf{v}_2 = \boldsymbol{\mathcal{S}}\left(\mathbf{H}_2 \odot \mathbf{H}_2\right)^{-1}\left[\boldsymbol{\mathcal{S}}'\left(\mathbf{H}_2 \otimes \mathbf{H}_2\right)\boldsymbol{\mathcal{S}}\right]\mathbf{d}_2$. Finally, fact (F4) yields $\boldsymbol{\mathcal{S}}'\left(\mathbf{H}_2 \otimes \mathbf{H}_2\right)\boldsymbol{\mathcal{S}} = \mathbf{H}_2 \odot \mathbf{H}_2$ and so

$$\left[\boldsymbol{\mathcal{S}}\left(\mathbf{H}_2 \odot \mathbf{H}_2\right)^{-1}\boldsymbol{\mathcal{S}}'\right]\left(\mathbf{H}_2 \otimes \mathbf{H}_2\right)\mathbf{v}_2 = \boldsymbol{\mathcal{S}}\mathbf{d}_2 = \mathbf{v}_2, \tag{85}$$

implying that $\widehat{\mathbf{v}}_2 \xrightarrow{p} \mathbf{v}_2$, as $N \to \infty$. ■

Proof of Theorem 5: Equations (35) and (36) yield

$$\sqrt{N}\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right) = \widehat{\boldsymbol{\Psi}}_\lambda\left(\frac{1}{\sqrt{N}}\widehat{\mathbf{Z}}_1'\boldsymbol{\omega}_2\right). \tag{86}$$

Let $\widehat{\mathbf{z}}_{1,i}'$ denote the i-th row of $\widehat{\mathbf{Z}}_1$ defined in (22), $\widehat{\boldsymbol{\beta}}_{1,i}'$ denote the i-th row of $\widehat{\mathbf{B}}_1$, and $\omega_{2,i}$ denote the i-th element of $\boldsymbol{\omega}_2$ defined in (18). It follows that $\widehat{\mathbf{z}}_{1,i} = \begin{bmatrix} 1 & \widehat{\boldsymbol{\beta}}_{1,i}' & \mathbf{c}_{1,i}' \end{bmatrix}'$, $\widehat{\mathbf{Z}}_1 = \begin{bmatrix} \widehat{\mathbf{z}}_{1,1} & \cdots & \widehat{\mathbf{z}}_{1,N} \end{bmatrix}'$ and $\boldsymbol{\omega}_2 = \begin{bmatrix} \omega_{2,1} & \cdots & \omega_{2,N} \end{bmatrix}'$. It follows that

$$\widehat{\mathbf{Z}}_1'\boldsymbol{\omega}_2 = \sum_{i=1}^N \widehat{\mathbf{z}}_{1,i}\omega_{2,i} = \begin{bmatrix} \sum_{i=1}^N \omega_{2,i} \\ \sum_{i=1}^N \widehat{\boldsymbol{\beta}}_{1,i}\omega_{2,i} \\ \sum_{i=1}^N \mathbf{c}_{1,i}\omega_{2,i} \end{bmatrix}. \tag{87}$$

It follows from equation (19) that $\omega_{2,i} = \mathbf{u}_{2,[i]}'\mathbf{g}_2 = \mathbf{g}_2'\mathbf{u}_{2,[i]}$. Hence, using equations (14) and (19), we obtain $\widehat{\boldsymbol{\beta}}_{1,i}\omega_{2,i} = \boldsymbol{\beta}_i\mathbf{u}_{2,[i]}'\mathbf{g}_2 + \mathbf{G}_1'\mathbf{u}_{1,[i]}\mathbf{u}_{2,[i]}'\mathbf{g}_2 = (\mathbf{g}_2' \otimes \mathbf{I}_K)(\mathbf{u}_{2,[i]} \otimes \boldsymbol{\beta}_i) + (\mathbf{g}_2' \otimes \mathbf{G}_1')(\mathbf{u}_{2,[i]} \otimes \mathbf{u}_{1,[i]})$, and $\mathbf{c}_{1,i}\omega_{2,i} = \mathbf{c}_{1,i}\mathbf{u}_{2,[i]}'\mathbf{g}_2 = (\mathbf{g}_2' \otimes \mathbf{I}_L)(\mathbf{u}_{2,[i]} \otimes \mathbf{c}_{1,i})$, where we use facts (F1), (F2), and (F3). Combining the last three equations, we obtain that

$$\widehat{\mathbf{z}}_{1,i}\omega_{2,i} = \mathbf{\Pi}_\lambda \mathbf{e}_i, \tag{88}$$

where $\mathbf{e}_i$ is defined by (34) and $\mathbf{\Pi}_\lambda$ is the $(1+K+L)\times\mathcal{T}$ matrix, with $\mathcal{T} = (1+\tau_1+K+L+\tau_2)\tau_2$, defined by

$$\mathbf{\Pi}_\lambda = \begin{bmatrix} \mathbf{g}_2' & \mathbf{0}_{\tau_1\tau_2}' & \mathbf{0}_{K\tau_2}' & \mathbf{0}_{L\tau_2}' & \mathbf{0}_{\tau_2^2}' \\ \mathbf{0}_{K\times\tau_2} & \mathbf{g}_2'\otimes\mathbf{G}_1' & \mathbf{g}_2'\otimes\mathbf{I}_K & \mathbf{0}_{K\times(L\tau_2)} & \mathbf{0}_{K\times\tau_2^2} \\ \mathbf{0}_{L\times\tau_2} & \mathbf{0}_{L\times(\tau_1\tau_2)} & \mathbf{0}_{L\times(K\tau_2)} & \mathbf{g}_2'\otimes\mathbf{I}_L & \mathbf{0}_{L\times\tau_2^2} \end{bmatrix}. \tag{89}$$

It follows from (88) that

$$\widehat{\mathbf{Z}}_1'\boldsymbol{\omega}_2 = \mathbf{\Pi}_\lambda \sum_{i=1}^N \mathbf{e}_i. \tag{90}$$

Combining (86), (90), and (37), we obtain that $\sqrt{N}\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right) = \mathbf{\Psi}_\lambda\mathbf{\Pi}_\lambda\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{e}_i + o_p(1)$. Finally, invoking Assumption 4 yields the desired result. ■

Proof of Theorem 6: It follows from equation (79) in the proof of Proposition 2 that

$$\sqrt{N}\left(\widehat{\mathcal{Q}} - (\mathbf{g}_2\otimes\mathbf{g}_2)'\,\widehat{\mathbf{v}}_2 - R_{\widehat{\mathcal{Q}}}\right) = \sqrt{N}\left(\boldsymbol{\omega}_2'\boldsymbol{\omega}_2/N - (\mathbf{g}_2\otimes\mathbf{g}_2)'\,\widehat{\mathbf{v}}_2\right) - 2\left(\boldsymbol{\omega}_2'\widehat{\mathbf{X}}_2/N\right)\sqrt{N}\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right).$$

Invoking equation (82) and Theorem 5 then yields

$$\begin{aligned} &\sqrt{N}\left(\widehat{\mathcal{Q}} - (\mathbf{g}_2\otimes\mathbf{g}_2)'\,\widehat{\mathbf{v}}_2 - R_{\widehat{\mathcal{Q}}}\right) \\ =&\sqrt{N}\left(\boldsymbol{\omega}_2'\boldsymbol{\omega}_2/N - (\mathbf{g}_2\otimes\mathbf{g}_2)'\,\widehat{\mathbf{v}}_2\right) - 2\boldsymbol{\rho}_\alpha'\sqrt{N}\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right) + o_p(1), \end{aligned} \tag{91}$$

where $\boldsymbol{\omega}_2$, $\mathbf{g}_2$, and $\boldsymbol{\rho}_\alpha$ are defined in equations (18), (20), and (83), respectively. It follows from equation (32) that $N\widehat{\mathbf{v}}_2 = \sum_{i=1}^N \left[\boldsymbol{\mathcal{S}}\left(\mathbf{H}_2\odot\mathbf{H}_2\right)^{-1}\boldsymbol{\mathcal{S}}'\right]\text{vec}\left(\widehat{\mathbf{u}}_{2,[i]}\widehat{\mathbf{u}}_{2,[i]}'\right)$. Hence, $\boldsymbol{\omega}_2'\boldsymbol{\omega}_2 - N\left(\mathbf{g}_2\otimes\mathbf{g}_2\right)'\widehat{\mathbf{v}}_2 = \sum_{i=1}^N \nu_i$, where

$$\nu_i = \omega_{2,i}^2 - (\mathbf{g}_2\otimes\mathbf{g}_2)'\left[\boldsymbol{\mathcal{S}}\left(\mathbf{H}_2\odot\mathbf{H}_2\right)^{-1}\boldsymbol{\mathcal{S}}'\right]\text{vec}\left(\widehat{\mathbf{u}}_{2,[i]}\widehat{\mathbf{u}}_{2,[i]}'\right). \tag{92}$$

Equation (19) yields $\omega_{2,i} = \mathbf{u}_{2,[i]}'\mathbf{g}_2$. Hence, $\omega_{2,i}^2 = \mathbf{g}_2'\mathbf{u}_{2,[i]}\mathbf{u}_{2,[i]}'\mathbf{g}_2 = (\mathbf{g}_2\otimes\mathbf{g}_2)'\,\text{vec}\left(\mathbf{u}_{2,[i]}\mathbf{u}_{2,[i]}'\right)$, where the last equality follows from fact (F2). Equation (30) implies $\widehat{\mathbf{u}}_{2,[i]} = \mathbf{H}_2 \mathbf{u}_{2,[i]}$ and so, using fact (F2) again, we obtain $\text{vec}\left(\widehat{\mathbf{u}}_{2,[i]} \widehat{\mathbf{u}}'_{2,[i]}\right) = (\mathbf{H}_2 \otimes \mathbf{H}_2)\,\text{vec}\left(\mathbf{u}_{2,[i]} \mathbf{u}'_{2,[i]}\right)$.

Collecting terms, we obtain $\nu_i = (\mathbf{g}_2 \otimes \mathbf{g}_2)' \,\boldsymbol{\Xi}_\alpha \text{vec}\left(\mathbf{u}_{2,[i]} \mathbf{u}'_{2,[i]}\right)$, where the $\tau_2^2 \times \tau_2^2$ matrix $\boldsymbol{\Xi}_\alpha$ is defined by $\boldsymbol{\Xi}_\alpha = \mathbf{I}_{\tau_2^2} - \left[\boldsymbol{\mathcal{S}} \left(\mathbf{H}_2 \odot \mathbf{H}_2\right)^{-1} \boldsymbol{\mathcal{S}}'\right] (\mathbf{H}_2 \otimes \mathbf{H}_2)$. It follows from equation (85) that $\boldsymbol{\Xi}_\alpha \mathbf{v}_2 = \mathbf{v}_2 - \left[\boldsymbol{\mathcal{S}} \left(\mathbf{H}_2 \odot \mathbf{H}_2\right)^{-1} \boldsymbol{\mathcal{S}}'\right] (\mathbf{H}_2 \otimes \mathbf{H}_2)\,\mathbf{v}_2 = \mathbf{0}_{\tau_2^2}$. Therefore, letting $\boldsymbol{\xi}_\alpha = \boldsymbol{\Xi}'_\alpha (\mathbf{g}_2 \otimes \mathbf{g}_2)$, we have $\nu_i = \boldsymbol{\xi}'_\alpha \left(\text{vec}\left(\mathbf{u}_{2,[i]} \mathbf{u}'_{2,[i]}\right) - \mathbf{v}_2\right)$, where $\boldsymbol{\xi}_\alpha$ is the $\tau_2^2 \times 1$ vector given by

$$\boldsymbol{\xi}_\alpha = (\mathbf{g}_2 \otimes \mathbf{g}_2) - (\mathbf{H}_2 \otimes \mathbf{H}_2) \left[\boldsymbol{\mathcal{S}} \left(\mathbf{H}_2 \odot \mathbf{H}_2\right)^{-1} \boldsymbol{\mathcal{S}}'\right] (\mathbf{g}_2 \otimes \mathbf{g}_2). \tag{93}$$

Defining the $\mathcal{T} \times 1$ vector $\boldsymbol{\pi}_\alpha$ by

$$\boldsymbol{\pi}_\alpha = \begin{bmatrix} \mathbf{0}_{\tau_2^2 \times \tau_2(1+K+L+\tau_1)} & \mathbf{I}_{\tau_2^2} \end{bmatrix}' \boldsymbol{\xi}_\alpha, \tag{94}$$

we have

$$\nu_i = \boldsymbol{\pi}'_\alpha \mathbf{e}_i, \tag{95}$$

where $\mathbf{e}_i$ is defined in (34), and so

$$\sqrt{N} \left(\boldsymbol{\omega}'_2 \boldsymbol{\omega}_2 / N - (\mathbf{g}_2 \otimes \mathbf{g}_2)' \,\widehat{\mathbf{v}}_2\right) = \boldsymbol{\pi}'_\alpha \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{e}_i. \tag{96}$$

It follows from the proof of Theorem 1 that $\sqrt{N}\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right) = \boldsymbol{\Psi}_\lambda \boldsymbol{\Pi}_\lambda \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{e}_i + o_p(1)$, where $\boldsymbol{\Psi}_\lambda$ and $\boldsymbol{\Pi}_\lambda$ are defined in (38) and (89), respectively. Combining this fact with (91) and (96) yields

$$\sqrt{N} \left(\widehat{\mathcal{Q}} - (\mathbf{g}_2 \otimes \mathbf{g}_2)' \,\widehat{\mathbf{v}}_2 - R_{\widehat{\mathcal{Q}}}\right) = \boldsymbol{\kappa}'_\alpha \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{e}_i + o_p(1). \tag{97}$$

where the $\mathcal{T} \times 1$ vector $\boldsymbol{\kappa}_\alpha$ is given by

$$\boldsymbol{\kappa}_\alpha = \boldsymbol{\pi}_\alpha - 2\boldsymbol{\Pi}'_\lambda \boldsymbol{\Psi}'_\lambda \boldsymbol{\rho}_\alpha. \tag{98}$$

Finally, invoking Assumption 4 completes the proof of the theorem. ■

Proof of Theorem 8: It follows from definitions (18) and (64) that

$$\widehat{\boldsymbol{\omega}}_2 = \boldsymbol{\omega}_2 - \widehat{\mathbf{X}}_2 \left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda} \right). \tag{99}$$

Let $\widehat{\boldsymbol{\varepsilon}}_i'$ denote the i-th row of the matrix $\widehat{\boldsymbol{\mathcal{E}}}$, defined by (66), so that $\widehat{\boldsymbol{\mathcal{E}}} = \left[\begin{array}{ccc} \widehat{\varepsilon}_1 & \cdots & \widehat{\varepsilon}_N \end{array} \right]'$. For each cluster $m = 1, \ldots, M_N$, we define $\widehat{\boldsymbol{\theta}}_m = \sum_{i \in I_m} \widehat{\varepsilon}_i$, so that $\boldsymbol{\mathcal{C}}\widehat{\boldsymbol{\mathcal{E}}} = \left[\begin{array}{ccc} \widehat{\boldsymbol{\theta}}_1 & \cdots & \widehat{\boldsymbol{\theta}}_{M_N} \end{array} \right]'$. Definition (65) then yields

$$\widehat{\boldsymbol{\Theta}} = \frac{1}{M_N - K - 1} \sum_{m=1}^{M_N} \widehat{\boldsymbol{\theta}}_m \widehat{\boldsymbol{\theta}}_m'. \tag{100}$$

Moreover, it follows from definitions (59) and (66) that $\widehat{\boldsymbol{\mathcal{E}}} = \widetilde{\boldsymbol{\mathcal{E}}} + \left[\begin{array}{cc} \mathbf{D}\left(\widehat{\boldsymbol{\omega}}_2 - \boldsymbol{\omega}_2\right) \widehat{\mathbf{Z}}_1 & \widehat{\boldsymbol{\omega}}_2 \odot \widehat{\boldsymbol{\omega}}_2 - \boldsymbol{\omega}_2 \odot \boldsymbol{\omega}_2 \end{array} \right]$ which, in light of equation (60), implies that, for $i = 1, \ldots, N$, we have

$$\widehat{\boldsymbol{\varepsilon}}_i' = \widetilde{\boldsymbol{\varepsilon}}_i' + \left[\begin{array}{cc} \left(\widehat{\omega}_{2,i} - \omega_{2,i}\right) \widehat{\mathbf{z}}_{1,i}' & \widehat{\omega}_{2,i}^2 - \omega_{2,i}^2 \end{array} \right]. \tag{101}$$

Equation (99) yields

$$\widehat{\omega}_{2,i} = \omega_{2,i} - \widehat{\mathbf{x}}_{2,i}' (\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}), \tag{102}$$

and so $\widehat{\mathbf{z}}_{1,i}\left(\widehat{\omega}_{2,i} - \omega_{2,i}\right) = -\widehat{\mathbf{z}}_{1,i} \widehat{\mathbf{x}}_{2,i}' \left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda} \right)$. Using fact (F2) and then fact (F1), we obtain

$$\widehat{\mathbf{z}}_{1,i} \widehat{\mathbf{x}}_{2,i}' \left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda} \right) = \left(\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda} \right) \otimes \mathbf{I}_{1+K+L} \right)' \left(\widehat{\mathbf{x}}_{2,i} \otimes \widehat{\mathbf{z}}_{1,i} \right). \tag{103}$$

It follows from (14) that $\widehat{\mathbf{X}}_2 = \left[\begin{array}{cc} \mathbf{1}_N & \mathbf{B} + \mathbf{U}_2 \mathbf{G}_2 \end{array} \right]$ and $\widehat{\mathbf{Z}}_1 = \left[\begin{array}{ccc} \mathbf{1}_N & \mathbf{B} + \mathbf{U}_1 \mathbf{G}_1 & \mathbf{C}_1 \end{array} \right]$ and so

$$\widehat{\mathbf{x}}_{2,i} = \left[\begin{array}{c} 1 \\ \boldsymbol{\beta}_i + \mathbf{G}_2' \mathbf{u}_{2,[i]} \end{array} \right], \quad \widehat{\mathbf{z}}_{1,i} = \left[\begin{array}{c} 1 \\ \boldsymbol{\beta}_i + \mathbf{G}_1' \mathbf{u}_{1,[i]} \\ \mathbf{c}_{1,i} \end{array} \right].$$

Hence,

$$\widehat{\mathbf{x}}_{2,i} \otimes \widehat{\mathbf{z}}_{1,i} = \boldsymbol{\mathcal{K}}_{xz} \begin{bmatrix} 1 \\ \boldsymbol{\beta}_i + \mathbf{G}_1' \mathbf{u}_{1,[i]} \\ \mathbf{c}_{1,i} \\ \boldsymbol{\beta}_i + \mathbf{G}_2' \mathbf{u}_{2,[i]} \\ (\boldsymbol{\beta}_i + \mathbf{G}_2' \mathbf{u}_{2,[i]}) \otimes (\boldsymbol{\beta}_i + \mathbf{G}_1' \mathbf{u}_{1,[i]}) \\ (\boldsymbol{\beta}_i + \mathbf{G}_2' \mathbf{u}_{2,[i]}) \otimes \mathbf{c}_{1,i} \end{bmatrix},$$

where $\boldsymbol{\mathcal{K}}_{xz}$ is a suitable $(K+1)(K+L+1) \times (K+1)(K+L+1)$ matrix with elements equal to 0 or 1 (see Theorem 8.26(e) in Schott (2017)). Fact (F3) implies $(\boldsymbol{\beta}_i + \mathbf{G}_2'\mathbf{u}_{2,[i]}) \otimes (\boldsymbol{\beta}_i + \mathbf{G}_1'\mathbf{u}_{1,[i]}) = \boldsymbol{\beta}_i \otimes \boldsymbol{\beta}_i$ $+ (\mathbf{G}_2' \otimes \mathbf{I}_K)(\mathbf{u}_{2,[i]} \otimes \boldsymbol{\beta}_i) + (\mathbf{I}_K \otimes \mathbf{G}_1')(\boldsymbol{\beta}_i \otimes \mathbf{u}_{1,[i]}) + (\mathbf{G}_2' \otimes \mathbf{G}_1')(\mathbf{u}_{2,[i]} \otimes \mathbf{u}_{1,[i]})$ and $(\boldsymbol{\beta}_i + \mathbf{G}_2'\mathbf{u}_{2,[i]}) \otimes \mathbf{c}_{1,i} =$ $\boldsymbol{\beta}_i \otimes \mathbf{c}_{1,i} + (\mathbf{G}_2' \otimes \mathbf{I}_L)(\mathbf{u}_{2,[i]} \otimes \mathbf{c}_{1,i})$. It follows from the last three equations that

$$\widehat{\mathbf{x}}_{2,i} \otimes \widehat{\mathbf{z}}_{1,i} = \boldsymbol{\Phi}_{xz} \boldsymbol{\zeta}_i, \tag{104}$$

where $\boldsymbol{\zeta}_i$ is defined in (67) and $\boldsymbol{\Phi}_{xz}$ is a suitable matrix that depends on $\mathbf{G}_1$ and $\mathbf{G}_2$. Combining equations (103) and (104) then yields

$$\widehat{\mathbf{z}}_{1,i} \left(\widehat{\omega}_{2,i} - \omega_{2,i} \right) = -\widehat{\mathbf{z}}_{1,i} \widehat{\mathbf{x}}_{2,i}' \left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda} \right) = - \left(\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda} \right) \otimes \mathbf{I}_{1+K+L} \right)' \boldsymbol{\Phi}_{xz} \boldsymbol{\zeta}_i. \tag{105}$$

Next, note that equation (102) implies

$$\widehat{\omega}_{2,i}^2 - \omega_{2,i}^2 = -2\omega_{2,i} \widehat{\mathbf{x}}_{2,i}' \left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda} \right) + \left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda} \right)' \widehat{\mathbf{x}}_{2,i} \widehat{\mathbf{x}}_{2,i}' \left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda} \right). \tag{106}$$

The first term in (106) depends on

$$\omega_{2,i} \widehat{\mathbf{x}}_{2,i} = \left(\mathbf{g}_2' \mathbf{u}_{2,[i]} \right) \begin{bmatrix} 1 \\ \boldsymbol{\beta}_i + \mathbf{G}_2' \mathbf{u}_{2,[i]} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_2' \mathbf{u}_{2,[i]} \\ \boldsymbol{\beta}_i \mathbf{u}_{2,[i]}' \mathbf{g}_2 + \mathbf{G}_2' \mathbf{u}_{2,[i]} \mathbf{u}_{2,[i]}' \mathbf{g}_2 \end{bmatrix}.$$

Note that $\boldsymbol{\beta}_i \mathbf{u}_{2,[i]}' \mathbf{g}_2 = (\mathbf{g}_2' \otimes \mathbf{I}_K) \left(\mathbf{u}_{2,[i]} \otimes \boldsymbol{\beta}_i \right)$ and $\mathbf{G}_2' \mathbf{u}_{2,[i]} \mathbf{u}_{2,[i]}' \mathbf{g}_2 = (\mathbf{g}_2' \otimes \mathbf{G}_2') \left(\mathbf{u}_{2,[i]} \otimes \mathbf{u}_{2,[i]} \right)$, as it follows from facts (F1) and (F2). Hence,

$$\omega_{2,i} \widehat{\mathbf{x}}_{2,i} = \boldsymbol{\Phi}_{\omega x} \boldsymbol{\zeta}_i, \tag{107}$$

where $\boldsymbol{\zeta}_i$ is defined in (67) and $\boldsymbol{\Phi}_{\omega x}$ is a suitable matrix that depends on $\mathbf{g}_2$ and $\mathbf{G}_2$. Using fact (F2) and then fact (F1), we can express the second term in (106) as

$$\left(\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{GMM}} - \boldsymbol{\lambda}\right)' \widehat{\mathbf{x}}_{2,i}\widehat{\mathbf{x}}_{2,i}' \left(\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{GMM}} - \boldsymbol{\lambda}\right) = \left(\left(\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{GMM}} - \boldsymbol{\lambda}\right) \otimes \left(\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{GMM}} - \boldsymbol{\lambda}\right)\right)' \left(\widehat{\mathbf{x}}_{2,i} \otimes \widehat{\mathbf{x}}_{2,i}\right). \tag{108}$$

Since

$$\widehat{\mathbf{x}}_{2,i} = \begin{bmatrix} 1 \\ \boldsymbol{\beta}_i + \mathbf{G}_2'\mathbf{u}_{2,[i]} \end{bmatrix}, \tag{109}$$

we obtain

$$\widehat{\mathbf{x}}_{2,i} \otimes \widehat{\mathbf{x}}_{2,i} = \boldsymbol{\mathcal{K}}_{xx} \begin{bmatrix} 1 \\ \boldsymbol{\beta}_i + \mathbf{G}_2'\mathbf{u}_{2,[i]} \\ \boldsymbol{\beta}_i + \mathbf{G}_2'\mathbf{u}_{2,[i]} \\ \left(\boldsymbol{\beta}_i + \mathbf{G}_2'\mathbf{u}_{2,[i]}\right) \otimes \left(\boldsymbol{\beta}_i + \mathbf{G}_2'\mathbf{u}_{2,[i]}\right) \end{bmatrix},$$

where $\boldsymbol{\mathcal{K}}_{xx}$ is a suitable $(K+1)^2 \times (K+1)^2$ matrix with elements equal to 0 or 1 (see Theorem 8.26(e) in Schott (2017)). Moreover, using fact (F3), we obtain $(\boldsymbol{\beta}_i + \mathbf{G}_2'\mathbf{u}_{2,[i]}) \otimes (\boldsymbol{\beta}_i + \mathbf{G}_2'\mathbf{u}_{2,[i]}) = \boldsymbol{\beta}_i \otimes \boldsymbol{\beta}_i + (\mathbf{G}_2' \otimes \mathbf{I}_K)(\mathbf{u}_{2,[i]} \otimes \boldsymbol{\beta}_i) + (\mathbf{I}_K \otimes \mathbf{G}_2')(\boldsymbol{\beta}_i \otimes \mathbf{u}_{2,[i]}) + (\mathbf{G}_2' \otimes \mathbf{G}_2')(\mathbf{u}_{2,[i]} \otimes \mathbf{u}_{2,[i]})$. It follows that

$$\widehat{\mathbf{x}}_{2,i} \otimes \widehat{\mathbf{x}}_{2,i} = \boldsymbol{\Phi}_{xx}\boldsymbol{\zeta}_i, \tag{110}$$

where $\boldsymbol{\zeta}_i$ is defined in (67) and $\boldsymbol{\Phi}_{xx}$ is a suitable matrix that depends on $\mathbf{G}_2$. In light of equations (107), (108), and (110), it follows from equation (106) that

$$\widehat{\omega}_{2,i}^2 - \omega_{2,i}^2 = -2\left(\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{GMM}} - \boldsymbol{\lambda}\right)' \boldsymbol{\Phi}_{\omega x}\boldsymbol{\zeta}_i + \left(\left(\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{GMM}} - \boldsymbol{\lambda}\right) \otimes \left(\widehat{\boldsymbol{\lambda}}_{\text{IV}}^{\text{GMM}} - \boldsymbol{\lambda}\right)\right)' \boldsymbol{\Phi}_{xx}\boldsymbol{\zeta}_i. \tag{111}$$

Combining equations (101), (105), and (111) yields

$$\widehat{\boldsymbol{\varepsilon}}_i = \widetilde{\boldsymbol{\varepsilon}}_i + \begin{bmatrix} \widehat{\mathbf{z}}_{1,i}\left(\widehat{\omega}_{2,i} - \omega_{2,i}\right) \\ \widehat{\omega}_{2,i}^2 - \omega_{2,i}^2 \end{bmatrix} = \widetilde{\boldsymbol{\varepsilon}}_i - \boldsymbol{\Upsilon}\boldsymbol{\zeta}_i, \tag{112}$$

where

$$\boldsymbol{\Upsilon} = \left[\begin{array}{c} \left(\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right) \otimes \mathbf{I}_{1+K+L}\right)' \boldsymbol{\Phi}_{xz} \\ 2\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right)' \boldsymbol{\Phi}_{\omega x} - \left(\left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right) \otimes \left(\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} - \boldsymbol{\lambda}\right)\right)' \boldsymbol{\Phi}_{xx} \end{array} \right]. \tag{113}$$

In light of equations (61), (62) and (69), equation (112) yields that $\widehat{\boldsymbol{\theta}}_m = \sum_{i \in I_m} \widehat{\boldsymbol{\varepsilon}}_i = \boldsymbol{\Pi}_\delta \boldsymbol{\eta}_m - \boldsymbol{\Upsilon} \boldsymbol{\varphi}_m$. It then follows from equation (100) that

$$\begin{aligned} \frac{M_N - K - 1}{M_N} \widehat{\boldsymbol{\Theta}} =& \boldsymbol{\Pi}_\delta \left(\frac{1}{M_N} \sum\nolimits_{m=1}^{M_N} \boldsymbol{\eta}_m \boldsymbol{\eta}_m' \right) \boldsymbol{\Pi}_\delta' + \boldsymbol{\Upsilon} \left(\frac{1}{M_N} \sum\nolimits_{m=1}^{M_N} \boldsymbol{\varphi}_m \boldsymbol{\varphi}_m' \right) \boldsymbol{\Upsilon}' \\ & - \boldsymbol{\Upsilon} \left(\frac{1}{M_N} \sum\nolimits_{m=1}^{M_N} \boldsymbol{\varphi}_m \boldsymbol{\eta}_m' \right) \boldsymbol{\Pi}_\delta' - \boldsymbol{\Pi}_\delta \left(\frac{1}{M_N} \sum\nolimits_{m=1}^{M_N} \boldsymbol{\varphi}_m \boldsymbol{\eta}_m' \right)' \boldsymbol{\Upsilon}'. \end{aligned} \tag{114}$$

Inspection of definitions (34), (47), (67), (68), and (69) reveals that $\boldsymbol{\eta}_m$ is a subvector of $\boldsymbol{\varphi}_m$, and so Assumption 6 implies that, as $N \to \infty$, both $\frac{1}{M_N} \sum_{m=1}^{M_N} \boldsymbol{\varphi}_m \boldsymbol{\varphi}_m'$ and $\frac{1}{M_N} \sum_{m=1}^{M_N} \boldsymbol{\varphi}_m \boldsymbol{\eta}_m'$ converge in probability to some finite matrices. Moreover, according to Theorem 1, $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{GMM}} \xrightarrow{p} \boldsymbol{\lambda}$ and so it follows from definition (113) that $\boldsymbol{\Upsilon}$ converges in probability to a matrix of zeros. Hence, in light of equation (48), it follows from equation (114) that $\widehat{\boldsymbol{\Theta}} \xrightarrow{p} \boldsymbol{\Pi}_\delta \mathbf{V}_\eta \boldsymbol{\Pi}_\delta' = \boldsymbol{\Theta}$ and thus the proof of the theorem is complete. ■

References

Aldrich, J., 1993, "Reiersøl, Geary and the Idea of Instrumental Variables," *The Economic and Social Review*, 24, 247–273.

Ang, A., and D. Kristensen, 2012, "Testing Conditional Factor Models," *Journal of Financial Economics*, 106, 132–156.

Ang, A., J. Liu, and K. Schwarz, 2010, "Using Stocks or Portfolios in Tests of Factor Models," Working Paper, Columbia University.

Berk, J. B., 1995, "A Critique of Size-Related Anomalies," *Review of Financial Studies*, 8, 275–286.

Black, F., M. C. Jensen, and M. Scholes, 1972, "The Capital Asset Pricing Model: Some Empirical Tests," in *Studies in the Theory of Capital Markets*, ed. by M. C. Jensen. Praeger, New York, NY, pp. 79–121.

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu, 2006, *Measurement Error in Nonlinear Models: A Modern Perspective*. CRC, 2nd edn.

Chen, R., and R. Kan, 2004, "Finite Sample Analysis of Two-Pass Cross-Sectional Regressions," Working Paper, University of Toronto.

Chordia, T., A. Goyal, and J. Shanken, 2015, "Cross-sectional Asset Pricing with Individual Stocks: Betas versus Characteristics," Working Paper, Emory University.

Connor, G., and R. A. Korajczyk, 1988, "Risk and Return in an Equilibrium APT : Application of a New Test Methodology," *Journal of Financial Economics*, 21, 255–289.

——— , 1993, "A Test for the Number of Factors in an Approximate Factor Model," *Journal of Financial Economics*, 48, 1263–1291.

Daniel, K., and S. Titman, 2012, "Testing Factor-Model Explanations of Market Anomalies," *Critical Finance Review*, 1, 103–139.

Durbin, J., 1954, "Errors in Variables," *Review of the International Statistical Institute*, 22, 22–32.

Fama, E. F., and K. R. French, 1992, "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47, 427–465.

——— , 1993, "Common Risk Factors in the Returns on Bonds and Stocks," *Journal of Financial Economics*, 33, 3–56.

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

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

Gagliardini, P., E. Ossola, and O. Scaillet, 2012, "Time-Varying Risk Premium in Large Cross-Sectional Equity Datasets," Working Paper, University of Lugano.

Geary, R., 1943, "Relations between Statistics: The General and the Sampling Problem When the Samples Are Large," *Proceedings of the Royal Irish Academy. Section A: Mathematical and Physical Sciences*, 49, 177–96.

Ghysels, E., 1998, "On Stable Factor Structures in the Pricing of Risk: Do Time-Varying Betas Help or Hurt?," *Journal of Finance*, 53, 549–573.

Gibbons, M., S. Ross, and J. Shanken, 1989, "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57(5), 1121–1152.

Gibbons, M. R., 1982, "Multivariate Tests of Financial Models: A New Approach," *Journal of Financial Economics*, 10, 3–27.

Hansen, L., and R. Jagannathan, 1997, "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, 52(2), 557–590.

Hartley, H. O., J. N. K. Rao, and G. Kiefer, 1969, "Variance Estimation with one Unit per Stratum," *Journal of the American Statistical Association*, 64(327), 841–851.

Harvey, C., 1989, "Time-Varying Conditional Covariances in Tests of Asset Pricing Models," *Journal of Financial Economics*, 24(2), 289–317.

Hou, K., and R. Kimmel, 2010, "On Estimation of Risk Premia in Linear Factor Models," Working Paper, Ohio State University.

Hou, K., C. Xue, and L. Zhang, 2015, "Digesting Anomalies: An Investment Approach," *Review of Financial Studies*, 3(28), 650–705.

Jagannathan, R., G. Skoulakis, and Z. Wang, 2010, "The Analysis of the Cross Section of Security Returns," in *Handbook of Financial Econometrics 2*, ed. by Y. Ait-Sahalia, and L. P. Hansen. Elsevier, UK, pp. 73–134.

Jagannathan, R., and Z. Wang, 1998, "An Asymptotic Theory for Estimating Beta-Pricing Models using Cross-Sectional Regression," *Journal of Finance*, 53, 1285–1309.

——— , 2002, "Empirical Evaluation of Asset-Pricing Models: A Comparison of the SDF and Beta Methods," *Journal of Finance*, 57(5), 2337–2367.

Jegadeesh, N., J. Noh, K. Pukthuanthong, R. Roll, and J. Wang, 2015, "Empirical Tests of Asset Pricing Models with Individual Assets: Resolving the Errors-in-Variables Bias in Risk Premium Estimation," Working Paper, Emory University.

Kan, R., N. Gospodinov, and C. Robotti, 2013, "Chi-Squared Tests for Evaluation and Comparison of Asset Pricing Models," *Journal of Econometrics*, 173, 108–125.

Kan, R., and C. Robotti, 2009, "Model Comparison Using the Hansen-Jagannathan Distance," *Review of Financial Studies*, 22, 3449–3490.

Kan, R., C. Robotti, and J. Shanken, 2013, "Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology," *Journal of Finance*, 68, 2617–2649.

Kan, R., and G. Zhou, 1999, "A Critique of the Stochastic Discount Factor Methodology," *Journal of Finance*, 54, 1021–1048.

Kim, S., and G. Skoulakis, 2015, "Ex-post Risk Premia: Estimation and Inference using Large Cross Sections," Working Paper, Georgia Institute of Technology and University of British Columbia.

Kogan, L., and M. Tian, 2015, "Firm Characteristics and Empirical Factor Models: a Model-Mining Experiment," Working Paper, MIT and Federal Reserve Board.

Lehmann, B. N., and D. M. Modest, 1988, "The Empirical Foundations of the Arbitrage Pricing Theory," *Journal of Financial Economics*, 21, 213–254.

Lewellen, J., and S. Nagel, 2006, "The Conditional CAPM Does Not Explain Asset Pricing Anomalies," *Journal of Financial Economics*, 79, 289–314.

Lewellen, J., S. Nagel, and J. Shanken, 2010, "A Skeptical Appraisal of Asset Pricing Tests," *Journal of Financial Economics*, 96, 175–194.

Lintner, J., 1965, "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance*, 20, 587–615.

Litzenberger, R. H., and K. Ramaswamy, 1979, "The Effect of Personal Taxes and Dividends of Capital Asset Prices: The Theory and Evidence," *Journal of Financial Economics*, 7, 163–196.

Lo, A. W., and A. C. MacKinlay, 1990, "Data-Snooping Biases in Tests of Financial Asset Pricing Models," *Review of Financial Studies*, 3, 431–467.

Mallela, P., 1972, "Necessary and Sufficient Conditions for MINQU-Estimation of Heteroskedastic Variances in Linear Models," *Journal of the American Statistical Association*, 67(338), 486–487.

Mossin, J., 1966, "Equilibrium in a Capital Asset Market," *Econometrica*, 34, 768–783.

Nagel, S., and K. Singleton, 2011, "Estimation and Evaluation of Conditional Asset Pricing Models," *Journal of Finance*, 66, 873–909.

Puntanen, S., G. P. H. Styan, and J. Isotalo, 2013, *Formulas Useful for Linear Regression Analysis and Related Matrix Theory*. Springer.

Rao, C. R., 1970, "Estimation of Heteroscedastic Variances in Linear Models," *Journal of the American Statistical Association*, 65(329), 161–172.

Reiersøl, O., 1941, "Confluence Analysis by Means of Lag Moments and Other Methods of Confluence Analysis," *Econometrica*, 9, 1–24.

Roll, R., 1977, "A Critique of the Asset Pricing Theory's Tests Part I: On Past and Potential Testability of the Theory," *Journal of Financial Economics*, 4, 129–176.

Ross, S. A., 1976, "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360.

Schott, J., 2017, *Matrix Analysis for Statistics*. John Wiley, New York, NY, 3rd edn.

Shanken, J., 1985, "Multivariate Tests of the Zero-Beta CAPM," *Journal of Financial Economics*, 14, 327–348.

——— , 1992, "On the Estimation of Beta-Pricing Models," *Review of Financial Studies*, 5, 1–33.

Shanken, J., and G. Zhou, 2007, "Estimating and Testing Beta Pricing Models: Alternative Methods and Their Performance in Simulations," *Journal of Financial Economics*, 84(1), 40–86.

Sharpe, W. F., 1964, "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425–442.

Wald, A., 1940, "The Fitting of Straight Lines if Both Variables are Subject to Error," *Annals of Mathematical Statistics*, 11, 284–300.

Zhou, G., 1991, "Small Sample Tests of Portfolio Efficiency," *Journal of Financial Economics*, 30, 165–191.

——— , 1993, "Asset Pricing Tests under Alternative Distributions," *Journal of Finance*, 48, 1927–1942.

——— , 1994, "Analytical GMM Tests: Asset Pricing with Time-Varying Risk Premiums," *Review of Financial Studies*, 7, 687–709.

Table 1: **Bias in the estimation of $\boldsymbol{\lambda}$ with normally distributed shocks: the role of the EIV correction through the IV-GMM approach.** This table presents simulation results on the absolute bias, in annualized basis points, in the estimation of $\boldsymbol{\lambda} = [\ \lambda_0 \quad \boldsymbol{\lambda}_f'\ ]'$, where $\boldsymbol{\lambda}_f = [\ \lambda_1 \quad \cdots \quad \lambda_K\ ]'$ is the vector of ex-post risk premia and K is the number of factors. The shocks u_{it} are assumed to follow a normal distribution and the factor realizations are kept fixed throughout. The number of individual stocks, N, is equal to 1,000 and the number of clusters, M_N, is set equal to 50. The pairwise correlation of shocks, assumed to be constant within each cluster, is set equal to 0.10. The simulation is calibrated to the following three linear asset pricing models: the single-factor CAPM, the three-factor Fama and French (1993) model (FF3), and the four-factor Hou, Xue, and Zhang (2015) model (HXZ4). For the CAPM, $K = 1$ and $\lambda_{\texttt{MKT}}$ is the ex-post risk premia of MKT. For FF3, $K = 3$ and $\lambda_{\texttt{MKT}}$, $\lambda_{\texttt{SMB}}$, and $\lambda_{\texttt{HML}}$ are the ex-post risk premia of MKT, SMB, and HML, respectively. For HXZ4, $K = 4$ and $\lambda_{\texttt{MKT}}$, $\lambda_{\texttt{ME}}$, $\lambda_{\texttt{I/A}}$, and $\lambda_{\texttt{ROE}}$ are the ex-post risk premia of MKT, ME, I/A, and ROE, respectively. For the CAPM, we consider the IV estimator $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}$, while for the FF3 and HXZ4 models, we consider both the two-step and iterated IV-GMM estimators, i.e., $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$ and $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{IT}}$. In addition, we consider two alternative estimators $\widehat{\boldsymbol{\lambda}}_1 = (\widehat{\mathbf{X}}_1'\widehat{\mathbf{X}}_1)^{-1}\widehat{\mathbf{X}}_1'\bar{\mathbf{r}}_2$ and $\widehat{\boldsymbol{\lambda}}_2 = (\widehat{\mathbf{X}}_2'\widehat{\mathbf{X}}_2)^{-1}\widehat{\mathbf{X}}_2'\bar{\mathbf{r}}_2$ that ignore the EIV problem. The results are based on 30,000 Monte Carlo repetitions.

CAPM								
Pretesting Period	70-74	75-79	80-84	85-89	90-94	95-99	00-04	05-09
Testing Period	75-79	80-84	85-89	90-94	95-99	00-04	05-09	10-14
Average Factor Realizations								
$\bar{f}_{\texttt{MKT}}$	1118.2	453.0	1248.8	483.0	2069.4	-310.0	-43.4	1552.4
$\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}$								
λ_0	1.8	3.1	2.5	0.8	5.7	1.2	0.3	6.6
$\lambda_{\texttt{MKT}}$	1.3	1.9	2.2	0.6	7.0	0.9	1.5	6.7
$\widehat{\boldsymbol{\lambda}}_1$								
λ_0	358.6	176.9	465.3	161.8	1000.5	131.4	15.3	570.1
$\lambda_{\texttt{MKT}}$	364.1	180.6	472.0	163.9	1013.2	133.4	14.3	578.1
$\widehat{\boldsymbol{\lambda}}_2$								
λ_0	441.4	168.0	414.5	233.9	879.7	108.3	16.4	710.2
$\lambda_{\texttt{MKT}}$	448.0	171.5	420.5	237.0	890.7	110.0	15.4	720.1

Continued on next page

FF3 Model								
Pretesting Period	70-74	75-79	80-84	85-89	90-94	95-99	00-04	05-09
Testing Period	75-79	80-84	85-89	90-94	95-99	00-04	05-09	10-14
Average Factor Realizations								
$\overline{f}_{\texttt{MKT}}$	1118.2	453.0	1248.8	483.0	2069.4	-310.0	-43.4	1552.4
$\overline{f}_{\texttt{SMB}}$	1576.0	470.2	-451.6	158.0	-369.4	1002.4	131.4	149.2
$\overline{f}_{\texttt{HML}}$	651.0	931.4	251.8	269.4	-549.4	1510.8	210.8	-88.0
$\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$								
λ_0	1.2	5.4	6.4	2.3	7.9	8.6	1.9	5.0
$\lambda_{\texttt{MKT}}$	0.1	1.6	5.8	2.7	7.6	5.5	3.1	7.1
$\lambda_{\texttt{SMB}}$	2.5	3.5	2.3	0.2	1.7	2.8	1.9	0.1
$\lambda_{\texttt{HML}}$	0.8	4.2	1.7	0.6	1.6	6.6	1.4	2.2
$\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{IT}}$								
λ_0	1.2	5.6	6.9	2.4	8.3	9.4	1.9	4.6
$\lambda_{\texttt{MKT}}$	0.2	1.7	6.0	2.7	7.6	6.0	3.1	6.5
$\lambda_{\texttt{SMB}}$	3.0	3.5	2.3	0.3	1.5	2.8	1.9	0.4
$\lambda_{\texttt{HML}}$	0.6	4.9	2.2	0.9	2.0	7.4	1.5	2.4
$\widehat{\boldsymbol{\lambda}}_1$								
λ_0	779.3	389.4	632.8	370.0	830.8	754.8	81.0	579.8
$\lambda_{\texttt{MKT}}$	299.1	120.2	741.5	230.3	1136.2	76.8	4.3	689.5
$\lambda_{\texttt{SMB}}$	595.0	148.7	340.7	108.4	374.1	657.8	77.8	49.1
$\lambda_{\texttt{HML}}$	498.2	510.2	206.7	239.7	228.6	913.7	99.4	168.2
$\widehat{\boldsymbol{\lambda}}_2$								
λ_0	603.2	684.7	554.6	379.5	791.5	595.1	12.0	692.7
$\lambda_{\texttt{MKT}}$	159.4	332.8	623.8	275.5	1015.6	36.0	75.4	797.8
$\lambda_{\texttt{SMB}}$	683.8	306.3	294.1	76.5	308.5	568.4	84.5	74.5
$\lambda_{\texttt{HML}}$	281.6	538.1	240.4	191.8	111.4	719.0	113.0	115.8

Continued on next page

HXZ4 Model								
Pretesting Period	70-74	75-79	80-84	85-89	90-94	95-99	00-04	05-09
Testing Period	75-79	80-84	85-89	90-94	95-99	00-04	05-09	10-14
	Average Factor Realizations							
$\overline{f}_{\mathtt{MKT}}$	1115.2	404.8	1212.8	455.0	2014.4	-244.8	37.6	1420.2
$\overline{f}_{\mathtt{ME}}$	1726.6	420.2	-391.4	225.4	-436.2	1435.6	258.2	188.4
$\overline{f}_{\mathtt{I/A}}$	302.8	875.4	657.4	364.4	75.4	1113.4	-56.6	362.6
$\overline{f}_{\mathtt{ROE}}$	359.8	1022.8	921.0	1116.4	731.8	593.2	348.0	186.4
	$\widehat{\boldsymbol{\lambda}}^{\mathtt{TS}}_{\mathtt{IV}}$							
λ_0	29.0	32.0	47.2	55.4	51.2	39.0	18.2	42.0
$\lambda_{\mathtt{MKT}}$	5.5	8.6	16.7	10.9	29.8	23.1	12.4	27.7
$\lambda_{\mathtt{ME}}$	36.7	25.4	34.5	60.9	30.2	21.5	10.2	17.8
$\lambda_{\mathtt{I/A}}$	11.7	38.1	43.8	29.0	15.9	11.3	3.3	10.2
$\lambda_{\mathtt{ROE}}$	43.6	20.3	38.7	95.5	45.6	21.1	18.1	34.1
	$\widehat{\boldsymbol{\lambda}}^{\mathtt{IT}}_{\mathtt{IV}}$							
λ_0	33.8	36.9	55.4	62.7	59.5	44.2	20.9	47.6
$\lambda_{\mathtt{MKT}}$	7.5	10.4	21.2	14.1	34.9	26.5	14.4	30.8
$\lambda_{\mathtt{ME}}$	41.1	27.9	37.2	65.8	33.8	24.1	11.7	20.5
$\lambda_{\mathtt{I/A}}$	13.8	43.9	51.1	33.7	19.9	13.0	3.9	12.6
$\lambda_{\mathtt{ROE}}$	51.3	25.4	44.8	108.4	53.2	24.9	21.4	39.5
	$\widehat{\boldsymbol{\lambda}}_1$							
λ_0	1066.1	579.8	721.4	519.1	1143.6	947.3	173.4	910.0
$\lambda_{\mathtt{MKT}}$	461.8	129.5	668.3	235.9	1250.3	175.8	104.0	788.7
$\lambda_{\mathtt{ME}}$	947.2	466.9	163.8	323.2	277.3	1087.7	141.0	75.4
$\lambda_{\mathtt{I/A}}$	325.4	696.3	489.1	305.2	108.9	676.6	56.3	278.5
$\lambda_{\mathtt{ROE}}$	446.0	617.0	481.9	850.6	458.8	476.9	216.7	329.2
	$\widehat{\boldsymbol{\lambda}}_2$							
λ_0	855.8	661.2	766.7	552.9	1271.8	696.7	170.8	837.0
$\lambda_{\mathtt{MKT}}$	268.6	250.6	656.2	290.5	1272.4	196.8	59.6	766.0
$\lambda_{\mathtt{ME}}$	958.5	429.7	86.5	300.2	74.6	707.5	207.1	21.2
$\lambda_{\mathtt{I/A}}$	234.2	617.0	529.0	272.1	89.3	449.2	38.9	212.8
$\lambda_{\mathtt{ROE}}$	438.4	641.0	642.6	854.4	548.8	194.0	269.1	201.0

Table 2: **Root mean square error in the estimation of $\boldsymbol{\lambda}$ with normally distributed shocks: the role of the EIV correction through the IV-GMM approach.** This table presents simulation results on the root mean square error (RMSE), in annualized basis points, in the estimation of $\boldsymbol{\lambda} = [\ \lambda_0 \quad \boldsymbol{\lambda}_f'\ ]'$, where $\boldsymbol{\lambda}_f = [\ \lambda_1 \quad \cdots \quad \lambda_K\ ]'$ is the vector of ex-post risk premia and K is the number of factors. The shocks u_{it} are assumed to follow a normal distribution and the factor realizations are kept fixed throughout. The number of individual stocks, N, is equal to 1,000 and the number of clusters, M_N, is set equal to 50. The pairwise correlation of shocks, assumed to be constant within each cluster, is set equal to 0.10. The simulation is calibrated to the following three linear asset pricing models: the single-factor CAPM, the three-factor Fama and French (1993) model (FF3), and the four-factor Hou, Xue, and Zhang (2015) model (HXZ4). For the CAPM, $K = 1$ and $\lambda_{\texttt{MKT}}$ is the ex-post risk premia of MKT. For FF3, $K = 3$ and $\lambda_{\texttt{MKT}}$, $\lambda_{\texttt{SMB}}$, and $\lambda_{\texttt{HML}}$ are the ex-post risk premia of MKT, SMB, and HML, respectively. For HXZ4, $K = 4$ and $\lambda_{\texttt{MKT}}$, $\lambda_{\texttt{ME}}$, $\lambda_{\texttt{I/A}}$, and $\lambda_{\texttt{ROE}}$ are the ex-post risk premia of MKT, ME, I/A, and ROE, respectively. For the CAPM, we consider the IV estimator $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}$, while for the FF3 and HXZ4 models, we consider the two-step and iterated IV-GMM estimators, i.e., $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$ and $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{IT}}$. In addition, we consider two alternative estimators $\widehat{\boldsymbol{\lambda}}_1 = (\widehat{\mathbf{X}}_1'\widehat{\mathbf{X}}_1)^{-1}\widehat{\mathbf{X}}_1'\bar{\mathbf{r}}_2$ and $\widehat{\boldsymbol{\lambda}}_2 = (\widehat{\mathbf{X}}_2'\widehat{\mathbf{X}}_2)^{-1}\widehat{\mathbf{X}}_2'\bar{\mathbf{r}}_2$ that ignore the EIV problem. The results are based on 30,000 Monte Carlo repetitions.

CAPM								
Pretesting Period	70-74	75-79	80-84	85-89	90-94	95-99	00-04	05-09
Testing Period	75-79	80-84	85-89	90-94	95-99	00-04	05-09	10-14
Average Factor Realizations								
$\bar{f}_{\texttt{MKT}}$	1118.2	453.0	1248.8	483.0	2069.4	-310.0	-43.4	1552.4
$\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}$								
λ_0	217.8	227.9	226.2	215.2	274.7	236.9	217.3	233.9
$\lambda_{\texttt{MKT}}$	226.7	236.5	235.2	224.1	283.5	246.3	226.3	242.2
$\widehat{\boldsymbol{\lambda}}_1$								
λ_0	389.3	228.8	488.1	219.7	1011.2	194.9	148.7	589.3
$\lambda_{\texttt{MKT}}$	393.7	229.1	493.5	220.2	1022.3	192.9	146.9	596.2
$\widehat{\boldsymbol{\lambda}}_2$								
λ_0	465.3	222.7	440.8	273.2	892.0	183.4	146.4	725.1
$\lambda_{\texttt{MKT}}$	470.2	223.5	445.7	272.1	901.9	182.9	143.5	733.2

Continued on next page

FF3 Model								
Pretesting Period	70-74	75-79	80-84	85-89	90-94	95-99	00-04	05-09
Testing Period	75-79	80-84	85-89	90-94	95-99	00-04	05-09	10-14
Average Factor Realizations								
$\overline{f}_{\mathtt{MKT}}$	1118.2	453.0	1248.8	483.0	2069.4	-310.0	-43.4	1552.4
$\overline{f}_{\mathtt{SMB}}$	1576.0	470.2	-451.6	158.0	-369.4	1002.4	131.4	149.2
$\overline{f}_{\mathtt{HML}}$	651.0	931.4	251.8	269.4	-549.4	1510.8	210.8	-88.0
$\widehat{\boldsymbol{\lambda}}^{\mathtt{TS}}_{\mathtt{IV}}$								
λ_0	308.5	280.0	337.3	325.3	356.9	397.6	304.2	268.7
$\lambda_{\mathtt{MKT}}$	273.2	265.4	291.6	270.1	315.0	335.9	265.6	264.8
$\lambda_{\mathtt{SMB}}$	189.7	190.4	194.5	190.6	202.2	202.0	171.2	195.1
$\lambda_{\mathtt{HML}}$	224.9	230.5	218.8	239.0	236.3	246.9	195.3	218.8
$\widehat{\boldsymbol{\lambda}}^{\mathtt{IT}}_{\mathtt{IV}}$								
λ_0	309.3	280.4	337.9	325.4	357.6	398.5	304.8	268.6
$\lambda_{\mathtt{MKT}}$	274.1	265.5	292.3	270.6	315.7	336.8	266.1	264.8
$\lambda_{\mathtt{SMB}}$	190.0	189.9	194.1	190.0	202.1	201.9	171.6	194.6
$\lambda_{\mathtt{HML}}$	225.6	231.0	219.3	239.1	236.8	247.4	195.9	219.0
$\widehat{\boldsymbol{\lambda}}_1$								
λ_0	798.1	424.3	653.4	404.0	846.7	771.8	181.9	605.4
$\lambda_{\mathtt{MKT}}$	342.9	195.9	758.2	282.6	1146.3	177.4	160.7	706.7
$\lambda_{\mathtt{SMB}}$	608.5	184.4	356.7	150.2	389.8	670.1	154.1	114.5
$\lambda_{\mathtt{HML}}$	512.7	521.3	238.5	258.9	253.8	921.1	162.2	202.2
$\widehat{\boldsymbol{\lambda}}_2$								
λ_0	627.9	703.6	578.4	411.3	806.8	618.8	172.8	713.2
$\lambda_{\mathtt{MKT}}$	225.2	368.2	645.2	312.6	1027.4	171.1	171.0	811.7
$\lambda_{\mathtt{SMB}}$	693.0	324.1	312.3	131.1	333.1	585.0	132.5	121.8
$\lambda_{\mathtt{HML}}$	302.9	551.1	260.3	219.9	160.1	730.8	158.3	148.1

Continued on next page

HXZ4 Model								
Pretesting Period	70-74	75-79	80-84	85-89	90-94	95-99	00-04	05-09
Testing Period	75-79	80-84	85-89	90-94	95-99	00-04	05-09	10-14
Average Factor Realizations								
$\overline{f}_{\texttt{MKT}}$	1115.2	404.8	1212.8	455.0	2014.4	-244.8	37.6	1420.2
$\overline{f}_{\texttt{ME}}$	1726.6	420.2	-391.4	225.4	-436.2	1435.6	258.2	188.4
$\overline{f}_{\texttt{I/A}}$	302.8	875.4	657.4	364.4	75.4	1113.4	-56.6	362.6
$\overline{f}_{\texttt{ROE}}$	359.8	1022.8	921.0	1116.4	731.8	593.2	348.0	186.4
$\widehat{\boldsymbol{\lambda}}^{\texttt{TS}}_{\texttt{IV}}$								
λ_0	444.2	414.9	480.1	485.2	542.4	524.3	431.8	481.8
$\lambda_{\texttt{MKT}}$	349.0	359.7	396.1	369.3	445.2	406.6	351.2	357.1
$\lambda_{\texttt{ME}}$	324.8	351.1	357.7	377.0	359.4	332.9	255.4	345.7
$\lambda_{\texttt{I/A}}$	250.8	379.2	322.3	328.2	316.2	248.4	188.3	284.1
$\lambda_{\texttt{ROE}}$	428.8	509.9	486.3	520.5	531.0	484.6	322.1	452.9
$\widehat{\boldsymbol{\lambda}}^{\texttt{IT}}_{\texttt{IV}}$								
λ_0	446.8	415.0	482.1	485.4	544.5	526.7	432.9	482.9
$\lambda_{\texttt{MKT}}$	350.4	359.2	397.1	369.3	446.4	408.4	352.0	357.9
$\lambda_{\texttt{ME}}$	325.8	349.5	356.7	375.5	357.9	333.5	255.8	344.7
$\lambda_{\texttt{I/A}}$	252.7	380.5	324.1	329.0	317.3	249.7	189.2	284.8
$\lambda_{\texttt{ROE}}$	432.3	511.2	488.5	524.8	533.9	487.3	324.0	454.3
$\widehat{\boldsymbol{\lambda}}_1$								
λ_0	1082.6	611.7	746.9	552.1	1157.3	962.9	239.1	926.8
$\lambda_{\texttt{MKT}}$	494.6	211.8	690.9	294.0	1260.0	238.7	184.6	804.7
$\lambda_{\texttt{ME}}$	956.4	482.5	200.1	339.3	299.0	1095.4	200.0	133.7
$\lambda_{\texttt{I/A}}$	336.7	699.1	495.4	313.2	131.6	682.2	111.7	287.8
$\lambda_{\texttt{ROE}}$	461.4	627.1	495.9	856.3	469.1	486.7	252.1	345.6
$\widehat{\boldsymbol{\lambda}}_2$								
λ_0	877.4	689.1	790.5	580.2	1283.1	717.6	242.4	857.1
$\lambda_{\texttt{MKT}}$	315.6	306.0	680.3	329.2	1282.3	252.7	167.2	781.0
$\lambda_{\texttt{ME}}$	965.9	444.7	136.6	320.1	148.0	722.6	233.6	102.2
$\lambda_{\texttt{I/A}}$	242.0	621.9	534.0	281.7	124.9	459.9	80.4	223.3
$\lambda_{\texttt{ROE}}$	452.4	651.3	650.8	859.9	557.2	236.1	288.2	215.7

Table 3: **Empirical rejection frequencies of IV-GMM asset pricing tests with normally distributed shocks.** This table presents simulation results on the finite-sample performance of the asset pricing tests based on the IV-GMM estimators of $\boldsymbol{\lambda} = [\ \lambda_0 \ \ \boldsymbol{\lambda}_f'\ ]'$, where $\boldsymbol{\lambda}_f = [\ \lambda_1 \ \cdots \ \lambda_K\ ]'$ is the vector of ex-post risk premia and K is the number of factors. For the CAPM, we consider the IV estimator $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}$, while for the FF3 and HXZ4 models, we consider both the two-step and iterated IV-GMM estimators, i.e., $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$ and $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{IT}}$. Reported are empirical rejection frequencies, in percentages, of (i) the t statistics $t(\lambda_0)$ and $t(\lambda_k)$, $k = 1, \ldots, K$, testing the simple hypotheses $\lambda_0 = 0$ and $\lambda_k = \overline{f}_{k,2}$, for $k = 1, \ldots, K$, given in (70) and (71); (ii) the statistic $J_d(\boldsymbol{\lambda})$ testing the joint ex-post risk premia hypothesis $H_0^{\lambda} : [\ \lambda_0 \ \ \boldsymbol{\lambda}_f'\ ]' = [\ 0 \ \ \overline{\mathbf{f}}_2'\ ]'$, given in (74); (iii) the t-statistic $t(\boldsymbol{\alpha})$ testing the aggregate mispricing hypothesis $H_0^{\alpha} : \mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)'\mathbf{v}_2$, given in (72); and (iv) the statistic $J_d(\boldsymbol{\delta})$ jointly testing the hypotheses H_0^{λ} and H_0^{α}, given in (73). The shocks u_{it} are assumed to follow a normal distribution and the factor realizations are kept fixed throughout. We set the number of individual stocks, N, equal to 1,000 and consider two possible values for the number of clusters, M_N, namely 50 and 100. The pairwise correlation of shocks, assumed to be constant within each cluster, is denoted by ρ and takes three values, namely 0, 0.10, and 0.20. Three nominal levels of significance are considered: 1%, 5%, and 10%. The simulation is calibrated to the following three linear asset pricing models: the single-factor CAPM, the three-factor Fama and French (1993) model (FF3), and the four-factor Hou, Xue, and Zhang (2015) model (HXZ4). For the CAPM, $K = 1$ and $\lambda_{\texttt{MKT}}$ is the ex-post risk premia of MKT. For the FF3 model, $K = 3$ and $\lambda_{\texttt{MKT}}$, $\lambda_{\texttt{SMB}}$, and $\lambda_{\texttt{HML}}$ are the ex-post risk premia of MKT, SMB, and HML, respectively. For the HXZ4 model, $K = 4$ and $\lambda_{\texttt{MKT}}$, $\lambda_{\texttt{ME}}$, $\lambda_{\texttt{I/A}}$, and $\lambda_{\texttt{ROE}}$ are the ex-post risk premia of MKT, ME, I/A, and ROE, respectively. The pretesting and testing periods, both consisting of 60 months, are from 2005 to 2009 and from 2010 to 2014, respectively. The results are based on 30,000 Monte Carlo repetitions.

	CAPM																	
ρ	0						0.10						0.20					
M_N	50			100			50			100			50			100		
Nominal Size (%)	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
	$\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}$																	
$t(\lambda_0)$	1.2	5.4	10.5	1.0	5.2	10.0	1.2	5.4	10.6	1.1	5.2	10.3	1.2	5.4	10.4	1.1	5.2	10.3
$t(\lambda_{\texttt{MKT}})$	1.2	5.6	10.6	1.0	5.1	10.2	1.3	5.5	10.7	1.2	5.2	10.5	1.4	5.5	10.8	1.2	5.3	10.3
$J_d(\boldsymbol{\lambda})$	1.2	5.5	10.7	1.0	5.2	10.1	1.3	5.5	10.7	1.2	5.2	10.5	1.3	5.6	10.6	1.1	5.2	10.5
$t(\boldsymbol{\alpha})$	1.7	5.6	10.3	1.7	6.0	10.8	1.8	6.3	11.0	2.1	6.3	11.0	2.5	7.3	12.2	2.2	6.5	11.3
$J_d(\boldsymbol{\delta})$	1.4	5.7	10.8	1.3	5.4	10.4	1.6	5.9	11.0	1.6	5.8	10.8	1.8	6.4	11.7	1.6	5.7	11.0

Continued on next page

Table 3 – continued from previous page

	FF3 Model																	
ρ	0						0.10						0.20					
M_N	50			100			50			100			50			100		
Nominal Size (%)	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
	$\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{TS}}$																	
$t\,(\lambda_0)$	1.4	5.7	10.9	1.1	5.4	10.6	1.3	5.7	10.8	1.2	5.5	10.7	1.4	5.8	11.1	1.2	5.4	10.6
$t\,(\lambda_{\mathtt{MKT}})$	1.4	5.7	11.0	1.1	5.3	10.4	1.4	5.8	11.0	1.3	5.3	10.7	1.5	5.9	11.2	1.1	5.6	10.8
$t\,(\lambda_{\mathtt{SMB}})$	1.3	5.8	10.9	1.1	5.1	10.3	1.4	5.8	11.1	1.0	5.1	10.3	1.6	6.2	11.5	1.1	5.3	10.6
$t\,(\lambda_{\mathtt{HML}})$	1.3	5.9	11.3	1.2	5.7	11.1	1.4	5.8	11.2	1.3	5.6	11.0	1.5	6.1	11.2	1.2	5.5	11.1
$J_d(\boldsymbol{\lambda})$	1.3	5.7	11.4	0.9	5.2	10.6	1.3	5.9	11.5	1.0	5.6	10.8	1.3	6.2	11.9	1.1	5.3	10.8
$t\,(\boldsymbol{\alpha})$	2.7	7.6	12.0	2.9	7.4	11.7	3.0	7.9	12.2	3.1	7.9	12.3	3.3	8.3	12.7	3.1	8.0	12.5
$J_d(\boldsymbol{\delta})$	1.3	5.7	11.4	0.9	5.2	10.6	1.3	5.9	11.5	1.0	5.6	10.8	1.3	6.2	11.9	1.1	5.3	10.8
	$\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{IT}}$																	
$t\,(\lambda_0)$	1.4	5.8	10.9	1.2	5.4	10.5	1.4	5.8	10.9	1.2	5.6	10.7	1.4	5.8	11.2	1.2	5.4	10.6
$t\,(\lambda_{\mathtt{MKT}})$	1.4	5.8	11.0	1.1	5.3	10.4	1.5	5.8	11.0	1.3	5.3	10.6	1.4	5.9	11.3	1.1	5.5	10.7
$t\,(\lambda_{\mathtt{SMB}})$	1.2	5.7	10.9	1.0	5.1	10.2	1.4	5.7	11.1	1.0	5.1	10.4	1.5	6.2	11.6	1.1	5.3	10.5
$t\,(\lambda_{\mathtt{HML}})$	1.4	6.0	11.4	1.2	5.7	11.2	1.4	5.9	11.3	1.3	5.6	11.1	1.5	6.2	11.3	1.2	5.6	11.1
$J_d(\boldsymbol{\lambda})$	1.4	5.8	11.6	0.9	5.2	10.7	1.4	6.0	11.5	1.0	5.6	10.8	1.3	6.3	12.1	1.1	5.3	10.8
$t\,(\boldsymbol{\alpha})$	2.7	7.6	12.0	2.9	7.4	11.7	2.9	7.9	12.2	3.1	7.9	12.3	3.3	8.2	12.6	3.0	7.9	12.4
$J_d(\boldsymbol{\delta})$	1.4	5.8	11.6	0.9	5.2	10.7	1.4	6.0	11.5	1.0	5.6	10.8	1.3	6.3	12.1	1.1	5.3	10.8

Continued on next page

Table 3 – continued from previous page

	HXZ4 Model																	
ρ	0						0.10						0.20					
M_N	50			100			50			100			50			100		
Nominal Size (%)	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
									$\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$									
$t\,(\lambda_0)$	1.2	5.0	10.0	1.0	4.6	9.2	1.1	4.9	9.9	1.0	4.7	9.7	1.1	4.7	9.5	1.0	4.4	9.1
$t\,(\lambda_{\texttt{MKT}})$	1.3	5.2	10.1	1.0	4.8	9.4	1.2	5.0	9.9	1.0	4.8	9.6	1.1	5.0	9.8	0.9	4.4	9.1
$t\,(\lambda_{\texttt{ME}})$	0.7	4.3	9.2	0.5	3.7	8.5	0.8	4.2	9.0	0.5	3.8	8.4	0.7	3.9	8.4	0.5	3.6	8.3
$t\,(\lambda_{\texttt{I/A}})$	1.0	5.1	10.3	0.8	4.3	9.3	1.0	4.8	9.7	0.8	4.3	9.2	1.0	4.7	9.6	0.8	4.3	9.0
$t\,(\lambda_{\texttt{ROE}})$	0.8	4.1	8.9	0.7	4.2	9.1	0.9	4.5	9.4	0.7	4.0	8.9	0.8	4.0	8.5	0.7	3.8	8.4
$J_d(\boldsymbol{\lambda})$	0.9	4.4	9.2	0.7	3.9	8.4	0.9	4.4	9.3	0.7	3.9	8.4	0.8	4.1	8.5	0.6	3.7	8.1
$t\,(\boldsymbol{\alpha})$	2.7	6.6	9.7	3.0	6.6	9.9	2.9	6.7	9.8	3.0	6.5	10.0	2.8	6.6	9.7	2.9	6.7	9.8
$J_d(\boldsymbol{\delta})$	1.2	4.8	9.2	1.1	4.5	8.6	1.2	4.8	9.4	1.2	4.5	8.7	1.2	4.5	8.6	1.1	4.4	8.3
									$\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{IT}}$									
$t\,(\lambda_0)$	1.2	5.2	10.3	1.0	4.7	9.4	1.2	5.2	10.2	1.0	4.8	9.9	1.2	5.0	9.9	1.0	4.5	9.4
$t\,(\lambda_{\texttt{MKT}})$	1.3	5.5	10.5	1.0	4.9	9.5	1.3	5.4	10.3	1.1	4.9	9.9	1.2	5.2	10.2	0.9	4.6	9.2
$t\,(\lambda_{\texttt{ME}})$	0.7	4.4	9.4	0.5	3.8	8.6	0.7	4.3	9.2	0.5	3.7	8.4	0.7	4.0	8.6	0.5	3.6	8.2
$t\,(\lambda_{\texttt{I/A}})$	1.1	5.4	10.7	0.8	4.4	9.5	1.1	5.1	10.1	0.8	4.4	9.2	1.0	5.0	9.9	0.9	4.5	9.2
$t\,(\lambda_{\texttt{ROE}})$	0.9	4.4	9.4	0.7	4.3	9.2	1.0	4.8	10.0	0.7	4.2	9.1	0.9	4.3	8.9	0.7	4.0	8.6
$J_d(\boldsymbol{\lambda})$	1.0	4.6	9.7	0.8	4.1	8.6	1.0	4.8	9.8	0.7	4.0	8.6	0.9	4.4	9.0	0.7	3.8	8.4
$t\,(\boldsymbol{\alpha})$	2.7	6.6	9.9	3.0	6.7	10.1	3.0	6.8	10.1	3.0	6.7	10.1	3.0	6.6	10.0	2.9	6.6	10.0
$J_d(\boldsymbol{\delta})$	1.3	5.1	9.7	1.2	4.7	8.9	1.3	5.2	10.0	1.2	4.6	8.9	1.2	4.8	9.1	1.1	4.5	8.5

Table 4: **Empirical rejection frequencies of IV asset pricing tests with normally distributed shocks based on the Fama-MacBeth variance-covariance estimator.** This table presents simulation results on the finite-sample performance of the asset pricing tests based on the IV-GMM estimators of $\boldsymbol{\lambda} = [\ \lambda_0 \ \ \boldsymbol{\lambda}_f' \ ]'$, where $\boldsymbol{\lambda}_f = [\ \lambda_1 \ \ \cdots \ \ \lambda_K \ ]'$ is the vector of ex-post risk premia and K is the number of factors. The weighting matrix used is the identity matrix ($\mathbf{W} = \mathbf{I}_{1+K+L}$) and the variance-covariance of the ex-post risk premia estimator is estimated using the approach of Fama and MacBeth (1973). Reported are empirical rejection frequencies, in percentages, of the t statistics testing the simple hypotheses $\lambda_0 = 0$ and $\lambda_k = \overline{f}_{k,2}$, for $k = 1, \ldots, K$ (see (70) and (71)), as well as the $J_d(\boldsymbol{\lambda})$ statistic testing the joint ex-post risk premia hypothesis $H_0^{\lambda} : [\ \lambda_0 \ \ \boldsymbol{\lambda}_f' \ ]' = [\ 0 \ \ \overline{\mathbf{f}}_2' \ ]'$ (see (74)). The shocks u_{it} are assumed to follow a normal distribution and the factor realizations are kept fixed throughout. We set the number of individual stocks, N, equal to 1,000 and consider two possible values for the number of clusters, M_N, namely 50 and 100. The pairwise correlation of shocks, assumed to be constant within each cluster, is denoted by ρ and takes three values, namely 0, 0.10, and 0.20. Three nominal levels of significance are considered: 1%, 5%, and 10%. The simulation is calibrated to the following three linear asset pricing models: the single-factor CAPM, the three-factor Fama and French (1993) model (FF3), and the four-factor Hou, Xue, and Zhang (2015) model (HXZ4). For the CAPM, $K = 1$ and $\lambda_{\texttt{MKT}}$ is the ex-post risk premia of MKT. For the FF3 model, $K = 3$ and $\lambda_{\texttt{MKT}}$, $\lambda_{\texttt{SMB}}$, and $\lambda_{\texttt{HML}}$ are the ex-post risk premia of MKT, SMB, and HML, respectively. For the HXZ4 model, $K = 4$ and $\lambda_{\texttt{MKT}}$, $\lambda_{\texttt{ME}}$, $\lambda_{\texttt{I/A}}$, and $\lambda_{\texttt{ROE}}$ are the ex-post risk premia of MKT, ME, I/A, and ROE, respectively. The pretesting and testing periods, both consisting of 60 months, are from from 2000 to 2004 and from 2005 to 2009, respectively, in Panels A-1, B-1, and C-1 and from 2005 to 2009 and from 2010 to 2014, respectively, in Panels A-2, B-2, and C-2. The results are based on 30,000 Monte Carlo repetitions.

ρ	0						0.10						0.20					
M_N	50			100			50			100			50			100		
Nominal Size (%)	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
Panel A: CAPM																		
A-1: Pretesting Period of 2000-2004 and Testing Period of 2005-2009																		
$t\,(\lambda_0)$	1.2	5.6	10.9	1.3	5.6	10.8	1.4	5.9	11.2	1.3	5.6	10.9	1.3	5.6	10.6	1.3	5.9	11.1
$t\,(\lambda_{\texttt{MKT}})$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$J_d(\boldsymbol{\lambda})$	0.4	2.0	4.2	0.4	2.1	4.3	0.4	2.2	4.5	0.4	2.1	4.3	0.4	2.1	4.4	0.4	2.2	4.5
A-2: Pretesting Period of 2005-2009 and Testing Period of 2010-2014																		
$t\,(\lambda_0)$	1.8	6.9	12.8	1.9	6.9	12.7	2.0	7.2	12.9	1.8	7.1	12.8	1.9	7.2	12.7	1.9	6.9	12.5
$t\,(\lambda_{\texttt{MKT}})$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$J_d(\boldsymbol{\lambda})$	0.7	3.1	5.8	0.7	3.0	5.8	0.7	3.1	6.1	0.7	3.1	6.0	0.7	3.2	6.2	0.7	3.1	5.9

Continued on next page

Table 4 – continued from previous page

Panel B: FF3 Model																		
B-1: Pretesting Period of 2000-2004 and Testing Period of 2005-2009																		
$t(\lambda_0)$	1.6	6.2	11.8	1.6	6.3	11.6	1.5	6.0	11.3	1.6	6.2	11.7	1.5	6.2	11.4	1.5	6.3	11.9
$t(\lambda_{\texttt{MKT}})$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$t(\lambda_{\texttt{SMB}})$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
$t(\lambda_{\texttt{HML}})$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$J_d(\boldsymbol{\lambda})$	0.1	0.7	1.6	0.1	0.7	1.5	0.1	0.6	1.5	0.1	0.7	1.5	0.1	0.7	1.6	0.1	0.6	1.6
B-2: Pretesting Period of 2005-2009 and Testing Period of 2010-2014																		
$t(\lambda_0)$	2.2	8.0	14.1	2.2	7.8	13.7	2.1	7.6	13.7	2.2	7.8	13.9	2.3	8.1	13.8	2.2	8.0	13.7
$t(\lambda_{\texttt{MKT}})$	0.0	0.1	0.4	0.0	0.1	0.4	0.0	0.1	0.4	0.0	0.1	0.4	0.0	0.2	0.7	0.0	0.2	0.5
$t(\lambda_{\texttt{SMB}})$	0.1	0.7	2.1	0.1	0.7	2.1	0.1	0.8	2.3	0.1	0.7	2.1	0.1	1.1	3.0	0.1	0.9	2.4
$t(\lambda_{\texttt{HML}})$	0.0	0.5	1.6	0.0	0.5	1.5	0.0	0.5	1.7	0.0	0.4	1.5	0.0	0.6	2.2	0.0	0.5	1.7
$J_d(\boldsymbol{\lambda})$	0.4	1.9	3.8	0.4	1.9	3.8	0.4	1.9	3.9	0.4	1.8	3.8	0.5	2.3	4.4	0.5	2.1	4.2
Panel C: HXZ4 Model																		
C-1: Pretesting Period of 2000-2004 and Testing Period of 2005-2009																		
$t(\lambda_0)$	1.7	6.8	12.5	2.0	7.0	12.8	1.9	6.9	12.5	1.8	6.8	12.3	1.8	6.8	12.5	1.6	6.6	12.3
$t(\lambda_{\texttt{MKT}})$	0.0	0.0	0.2	0.0	0.0	0.2	0.0	0.0	0.1	0.0	0.1	0.2	0.0	0.1	0.3	0.0	0.1	0.2
$t(\lambda_{\texttt{ME}})$	0.0	0.0	0.2	0.0	0.1	0.3	0.0	0.1	0.4	0.0	0.1	0.3	0.0	0.1	0.5	0.0	0.1	0.5
$t(\lambda_{\texttt{I/A}})$	0.0	0.4	1.5	0.0	0.5	1.5	0.0	0.4	1.5	0.0	0.5	1.6	0.0	0.6	2.0	0.1	0.5	1.6
$t(\lambda_{\texttt{ROE}})$	0.0	0.3	1.2	0.0	0.4	1.1	0.0	0.4	1.2	0.0	0.3	1.2	0.0	0.5	1.5	0.0	0.4	1.3
$J_d(\boldsymbol{\lambda})$	0.2	1.0	2.2	0.2	1.1	2.4	0.2	1.1	2.4	0.2	1.0	2.3	0.3	1.2	2.7	0.2	1.0	2.2
C-2: Pretesting Period of 2005-2009 and Testing Period of 2010-2014																		
$t(\lambda_0)$	2.7	8.6	14.5	2.5	8.5	14.5	2.4	8.6	14.8	2.5	8.6	14.7	2.6	8.9	15.1	2.6	8.9	14.9
$t(\lambda_{\texttt{MKT}})$	0.0	0.3	0.8	0.0	0.2	0.8	0.0	0.2	0.7	0.0	0.2	0.8	0.0	0.4	1.2	0.0	0.3	1.0
$t(\lambda_{\texttt{ME}})$	0.3	1.9	4.5	0.3	1.9	4.6	0.3	2.1	5.0	0.3	2.0	4.7	0.4	2.4	5.6	0.3	2.2	5.0
$t(\lambda_{\texttt{I/A}})$	0.4	2.4	5.9	0.4	2.6	6.0	0.5	2.9	6.5	0.4	2.6	5.9	0.6	3.4	7.1	0.5	3.0	6.5
$t(\lambda_{\texttt{ROE}})$	0.9	4.3	8.9	0.9	4.3	8.8	0.9	4.6	9.0	0.9	4.5	9.2	1.1	4.9	9.5	1.0	4.9	9.4
$J_d(\boldsymbol{\lambda})$	1.0	3.9	7.2	1.1	3.9	7.1	1.0	3.9	7.4	1.1	4.0	7.4	1.2	4.4	8.3	1.2	4.4	7.9

Table 5: **Power of the $J_d(\boldsymbol{\delta})$ test with normally distributed shocks.** This table presents empirical rejection frequencies (in percentage) of asset pricing tests based on the IV-GMM estimators of $\boldsymbol{\lambda} = [\ \lambda_0 \ \ \boldsymbol{\lambda}_f'\ ]'$, when, under the true data generating process, the asset pricing model does not hold. We calibrate the simulation to the FF3 model and focus on the $J_d(\boldsymbol{\delta})$ test statistic, defined in (74), that jointly tests the ex-post risk premia hypothesis $H_0^{\lambda} : [\ \lambda_0 \ \ \boldsymbol{\lambda}_f'\ ]' = [\ 0 \ \ \overline{\mathbf{f}}_2'\ ]'$ and the aggregate mispricing hypothesis $H_0^{\alpha} : \mathcal{Q} = (\mathbf{g}_2 \otimes \mathbf{g}_2)'\mathbf{v}_2$. The return generating process over the testing period is assumed to be $\overline{\mathbf{r}}_2 = \mathbf{c}_{1,\mathtt{SIZE}}\vartheta_{\mathtt{SIZE}} + \mathbf{c}_{1,\mathtt{BTM}}\vartheta_{\mathtt{BTM}} + \mathbf{B}\overline{\mathbf{f}}_2 + \overline{\mathbf{u}}_2$, where $\mathbf{c}_{1,\mathtt{SIZE}}$ and $\mathbf{c}_{1,\mathtt{BTM}}$ are the N-dimensional vectors of the SIZE and BTM characteristics (obtained as averages over the pretesting period and standardized to have zero mean and unit variance cross-sectionally), and $\vartheta_{\mathtt{SIZE}}$ and $\vartheta_{\mathtt{BTM}}$ are the SIZE and BTM characteristics rewards. When $\vartheta_{\mathtt{SIZE}} = \vartheta_{\mathtt{BTM}} = 0$, the data generating process conforms with the null hypothesis of the FF3 model being correctly specified. To examine the power of the joint $J_d(\boldsymbol{\delta})$ test, we let $\vartheta_{\mathtt{SIZE}}$ and $\vartheta_{\mathtt{BTM}}$ take nine values between -0.4 and 0.4, in percentage, and compute the rejection frequency in each case. We consider both the two-step and iterated IV-GMM estimators, i.e., $\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{TS}}$ and $\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{IT}}$. The shocks u_{it} are assumed to follow a normal distribution and the factor realizations are kept fixed throughout. We set the number of individual stocks, N, equal to 1,000, the number of clusters, M_N, equal to 50, and the pairwise correlation of shocks, assumed to be constant within each cluster, equal to 0.10. The nominal level of significance is assumed to be 5%. The factor realization of the pretesting and testing periods, both consisting of 60 months, are from 2005 to 2009 and from 2010 to 2014, respectively. The results are based on 10,000 Monte Carlo repetitions.

	$\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{TS}}$									
	$\vartheta_{\mathtt{SIZE}}$									
$\vartheta_{\mathtt{BTM}}$		-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4
	-0.4	94.69	92.22	88.42	83.31	79.37	74.85	72.45	69.72	69.02
	-0.3	85.24	79.13	70.36	61.59	53.57	48.42	44.91	46.86	49.12
	-0.2	66.80	54.13	42.65	32.22	23.56	21.23	21.39	25.57	31.49
	-0.1	44.67	31.69	21.58	12.52	8.16	6.76	9.34	14.92	21.80
	0	34.37	24.40	16.03	10.59	7.71	7.13	10.47	16.67	27.01
	0.1	39.03	34.57	28.08	23.63	21.68	22.17	26.85	34.63	44.83
	0.2	55.12	53.18	50.78	48.80	47.26	48.00	52.67	60.13	68.69
	0.3	72.58	72.83	72.76	71.97	71.86	73.73	78.25	82.47	86.88
	0.4	85.84	86.42	86.13	87.22	88.39	89.75	91.44	94.14	95.91
	$\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{IT}}$									
	$\vartheta_{\mathtt{SIZE}}$									
$\vartheta_{\mathtt{BTM}}$		-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4
	-0.4	91.69	89.15	86.23	83.34	80.84	79.26	78.98	79.10	79.75
	-0.3	82.67	77.83	70.99	64.09	57.77	54.62	53.40	57.15	61.81
	-0.2	69.17	57.35	45.52	35.01	26.04	24.63	26.26	33.99	42.57
	-0.1	50.74	36.21	23.58	12.88	8.31	7.44	11.75	19.99	30.91
	0	41.27	28.76	17.93	11.09	7.98	7.56	12.10	21.82	35.96
	0.1	47.34	40.12	31.46	25.56	23.46	24.98	31.61	41.34	52.55
	0.2	64.01	60.48	57.24	54.97	52.46	53.75	58.23	64.94	73.14
	0.3	81.16	80.44	79.80	79.14	77.88	78.49	80.94	83.75	86.22
	0.4	90.98	91.67	91.10	91.05	90.38	90.45	91.02	91.90	93.66

Table 6: **Betas of decile portfolios sorted by characteristics.** In this table, we consider decile portfolios sorted on a characteristic and present the beta estimates of these portfolios with respect to the corresponding spread factor within the context of the three asset pricing models we empirically examine: the FF3 model, the HXZ4 model, and the FF5 model. For each asset pricing model, the decile portfolio betas are estimated jointly for all factors using data from 07/1970 to 12/2014.

			Decile Portfolios Sorted by Characteristic									
			LOW									HIGH
Model	Factor	Characteristic	1	2	3	4	5	6	7	8	9	10
FF3	SMB	SIZE	1.19	1.10	0.92	0.81	0.69	0.49	0.38	0.28	0.07	-0.29
	HML	BTM	-0.50	-0.10	0.04	0.28	0.33	0.36	0.52	0.69	0.69	0.96
HXZ4	ME	SIZE	1.04	1.00	0.86	0.79	0.66	0.48	0.37	0.26	0.08	-0.27
	I/A	I/A	0.37	0.45	0.29	0.14	0.11	-0.03	-0.12	-0.30	-0.56	-0.42
	ROE	ROE	-0.33	-0.25	0.01	0.01	-0.15	0.06	0.09	0.10	0.23	0.32
FF5	SMB	SIZE	1.12	1.06	0.92	0.83	0.69	0.50	0.38	0.25	0.06	-0.28
	HML	BTM	-0.43	-0.19	-0.06	0.22	0.24	0.35	0.44	0.70	0.69	0.97
	RMW	OP	-0.90	-0.41	-0.27	-0.16	0.01	-0.05	0.07	0.21	0.35	0.43
	CMA	AG	0.63	0.67	0.68	0.27	0.22	0.10	0.04	-0.14	-0.62	-0.53

Table 7: **Testing the CAPM: empirical results for 5-year testing periods from 1975 to 2014 using 49 industry clusters.** This table presents the point estimates of $\boldsymbol{\lambda} = [\ \lambda_0 \ \ \lambda_{\texttt{MKT}}\ ]'$ and the various statistics along with the corresponding p-values for testing the implications of the CAPM. We consider 8 non-overlapping 5-year testing periods from 1975 to 2014. For each testing period, the pretesting period consists of the preceding five years. We report point estimates, in annualized percentages, based on the IV estimator $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}$. The null hypothesis implied by the CAPM is $\lambda_0 = 0$ and $\lambda_{\texttt{MKT}} = \overline{f}_{\texttt{MKT}}$. We report the risk premia t-statistics $t(\lambda_0)$ and $t(\lambda_{\texttt{MKT}})$, given in (70) and (71), respectively, the joint risk premia test statistic $J_d(\boldsymbol{\lambda})$, given in (74), the aggregate mispricing t-statistic $t(\boldsymbol{\alpha})$, given in (72), and the joint test statistic $J_d(\boldsymbol{\delta})$, given in (73). The corresponding p-values are reported in square brackets below the test statistics.

Pretesting Period	70-74	75-79	80-84	85-89	90-94	95-99	00-04	05-09
Testing Period	75-79	80-84	85-89	90-94	95-99	00-04	05-09	10-14
				Number of Test Assets				
N	1204	1736	1706	1942	1967	2098	1999	1926
				Average Factor Realizations				
$\overline{f}_{\texttt{MKT}}$	11.18	4.53	12.49	4.83	20.69	-3.10	-0.43	15.52
				Estimates of $\boldsymbol{\lambda}$: $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}$				
λ_0	-3.22	4.85	25.10	0.24	4.02	18.65	-8.49	13.44
$\lambda_{\texttt{MKT}}$	20.33	4.72	-14.57	9.16	12.13	0.07	11.10	4.10
				Test Statistics				
$t(\lambda_0)$	-1.48	2.20	4.82	0.15	2.11	10.90	-0.96	9.59
p-value	[0.139]	[0.028]	[0.000]	[0.878]	[0.035]	[0.000]	[0.337]	[0.000]
$t(\lambda_{\texttt{MKT}})$	4.85	0.10	-5.35	2.45	-3.63	1.81	1.86	-8.57
p-value	[0.000]	[0.918]	[0.000]	[0.014]	[0.000]	[0.071]	[0.063]	[0.000]
$J_d(\boldsymbol{\lambda})$	25.68	4.84	51.92	6.03	17.61	121.99	4.38	165.34
p-value	[0.000]	[0.109]	[0.000]	[0.077]	[0.002]	[0.000]	[0.140]	[0.000]
$t(\boldsymbol{\alpha})$	-11.36	-3.23	-2.68	-5.95	-5.28	-14.29	-3.78	-5.37
p-value	[0.000]	[0.001]	[0.007]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
$J_d(\boldsymbol{\delta})$	154.83	15.28	59.09	41.49	45.54	326.29	18.65	194.13
p-value	[0.000]	[0.009]	[0.000]	[0.000]	[0.000]	[0.000]	[0.011]	[0.000]

Table 8: **Testing the FF3 model: empirical results for 5-year testing periods from 1975 to 2014 using 49 industry clusters.** This table presents the point estimates of $\boldsymbol{\lambda} = [\ \lambda_0 \quad \lambda_{\texttt{MKT}} \quad \lambda_{\texttt{SMB}} \quad \lambda_{\texttt{HML}}\ ]'$ and the various statistics along with the corresponding p-values for testing the implications of the FF3 model. We consider 8 non-overlapping 5-year testing periods from 1975 to 2014. For each testing period, the pretesting period consists of the preceding five years. We report point estimates, in annualized percentages, based on the two-step and iterated IV-GMM estimators, i.e., $\widehat{\boldsymbol{\lambda}}^{\texttt{TS}}_{\texttt{IV}}$ and $\widehat{\boldsymbol{\lambda}}^{\texttt{IT}}_{\texttt{IV}}$. The null hypothesis implied by the FF3 model is $\lambda_0 = 0$, $\lambda_{\texttt{MKT}} = \overline{f}_{\texttt{MKT}}$, $\lambda_{\texttt{SMB}} = \overline{f}_{\texttt{SMB}}$, and $\lambda_{\texttt{HML}} = \overline{f}_{\texttt{HML}}$. We report the risk premia t-statistics $t(\lambda_0)$ and $t(\lambda_{\texttt{MKT}})$, $t(\lambda_{\texttt{SMB}})$, and $t(\lambda_{\texttt{HML}})$, given in (70) and (71), respectively, the joint risk premia test statistic $J_d(\boldsymbol{\lambda})$, given in (74), the aggregate mispricing t-statistic $t(\boldsymbol{\alpha})$, given in (72), and the joint test statistic $J_d(\boldsymbol{\delta})$, given in (73). The corresponding p-values are reported in square brackets below the test statistics.

Pretesting Period	70-74	75-79	80-84	85-89	90-94	95-99	00-04	05-09
Testing Period	75-79	80-84	85-89	90-94	95-99	00-04	05-09	10-14
Number of Test Assets								
N	1204	1733	1704	1941	1959	2090	1994	1920
Average Factor Realizations								
$\overline{f}_{\texttt{MKT}}$	11.18	4.53	12.49	4.83	20.69	-3.10	-0.43	15.52
$\overline{f}_{\texttt{SMB}}$	15.76	4.70	-4.52	1.58	-3.69	10.02	1.31	1.49
$\overline{f}_{\texttt{HML}}$	6.51	9.31	2.52	2.69	-5.49	15.11	2.11	-0.88
Estimates of $\boldsymbol{\lambda}$: $\widehat{\boldsymbol{\lambda}}^{\texttt{TS}}_{\texttt{IV}}$								
λ_0	6.13	-1.02	14.31	2.31	6.14	14.09	-9.57	11.34
$\lambda_{\texttt{MKT}}$	6.26	4.31	-3.54	3.15	8.90	-9.33	6.02	2.67
$\lambda_{\texttt{SMB}}$	15.43	8.06	-1.91	5.15	4.66	22.61	10.39	5.84
$\lambda_{\texttt{HML}}$	1.97	12.10	6.35	-1.97	-6.26	10.05	7.24	1.94
Estimates of $\boldsymbol{\lambda}$: $\widehat{\boldsymbol{\lambda}}^{\texttt{IT}}_{\texttt{IV}}$								
λ_0	5.73	0.42	18.51	2.32	6.16	17.64	-8.32	11.62
$\lambda_{\texttt{MKT}}$	6.66	2.99	-2.46	2.61	8.88	-12.69	4.82	2.46
$\lambda_{\texttt{SMB}}$	15.39	8.58	-3.75	4.96	4.68	23.99	10.75	5.80
$\lambda_{\texttt{HML}}$	1.96	11.46	-6.86	-0.59	-6.28	9.19	7.87	2.11

Continued on next page

Table 8 – continued from previous page

Pretesting Period	70-74	75-79	80-84	85-89	90-94	95-99	00-04	05-09
Testing Period	75-79	80-84	85-89	90-94	95-99	00-04	05-09	10-14
				Test Statistics: $\widehat{\boldsymbol{\lambda}}^{\mathtt{TS}}_{\mathtt{IV}}$				
$t\,(\lambda_0)$	2.17	-0.22	2.50	1.38	1.51	4.86	-1.43	9.78
p-value	[0.030]	[0.826]	[0.012]	[0.169]	[0.131]	[0.000]	[0.152]	[0.000]
$t\,(\lambda_{\mathtt{MKT}})$	-1.83	-0.05	-3.12	-1.12	-3.86	-2.97	0.93	-8.71
p-value	[0.068]	[0.957]	[0.002]	[0.263]	[0.000]	[0.003]	[0.350]	[0.000]
$t\,(\lambda_{\mathtt{SMB}})$	-0.29	2.92	1.65	4.27	5.66	5.81	1.99	2.53
p-value	[0.774]	[0.003]	[0.099]	[0.000]	[0.000]	[0.000]	[0.047]	[0.011]
$t\,(\lambda_{\mathtt{HML}})$	-1.87	0.60	0.96	-2.16	-0.32	-1.93	0.98	1.16
p-value	[0.061]	[0.547]	[0.338]	[0.030]	[0.748]	[0.053]	[0.329]	[0.245]
$J_d(\boldsymbol{\lambda})$	11.63	8.96	19.65	26.03	49.29	69.85	7.82	179.22
p-value	[0.060]	[0.104]	[0.009]	[0.001]	[0.000]	[0.000]	[0.133]	[0.000]
$t\,(\boldsymbol{\alpha})$	-3.32	0.06	-0.44	-5.87	-5.19	-5.47	-0.60	-5.54
p-value	[0.001]	[0.950]	[0.656]	[0.000]	[0.000]	[0.000]	[0.547]	[0.000]
$J_d(\boldsymbol{\delta})$	22.68	8.97	19.85	60.54	76.25	99.82	8.18	209.87
p-value	[0.013]	[0.157]	[0.021]	[0.000]	[0.000]	[0.000]	[0.184]	[0.000]
				Test Statistics: $\widehat{\boldsymbol{\lambda}}^{\mathtt{IT}}_{\mathtt{IV}}$				
$t\,(\lambda_0)$	2.04	0.09	3.82	1.40	1.52	5.35	-1.17	10.70
p-value	[0.042]	[0.928]	[0.000]	[0.162]	[0.130]	[0.000]	[0.242]	[0.000]
$t\,(\lambda_{\mathtt{MKT}})$	-1.69	-0.38	-3.38	-1.54	-3.86	-4.02	0.72	-9.24
p-value	[0.090]	[0.701]	[0.001]	[0.123]	[0.000]	[0.000]	[0.473]	[0.000]
$t\,(\lambda_{\mathtt{SMB}})$	-0.32	3.39	0.81	4.14	5.67	5.85	2.02	2.52
p-value	[0.749]	[0.001]	[0.417]	[0.000]	[0.000]	[0.000]	[0.043]	[0.012]
$t\,(\lambda_{\mathtt{HML}})$	-1.88	0.47	-2.06	-1.61	-0.33	-2.09	1.06	1.22
p-value	[0.060]	[0.641]	[0.039]	[0.108]	[0.743]	[0.037]	[0.289]	[0.221]
$J_d(\boldsymbol{\lambda})$	10.65	11.87	30.88	24.09	49.42	83.32	7.09	207.80
p-value	[0.072]	[0.057]	[0.001]	[0.002]	[0.000]	[0.000]	[0.159]	[0.000]
$t\,(\boldsymbol{\alpha})$	-3.42	-0.11	-2.63	-5.28	-5.22	-5.70	-0.48	-5.44
p-value	[0.001]	[0.914]	[0.008]	[0.000]	[0.000]	[0.000]	[0.635]	[0.000]
$J_d(\boldsymbol{\delta})$	22.37	11.88	37.82	51.98	76.68	115.83	7.32	237.41
p-value	[0.012]	[0.094]	[0.001]	[0.000]	[0.000]	[0.000]	[0.218]	[0.000]

Table 9: **Testing the HXZ4 model: empirical results for 5-year testing periods from 1975 to 2014 using 49 industry clusters.** This table presents the point estimates of $\boldsymbol{\lambda} = [\ \lambda_0\ \ \lambda_{\texttt{MKT}}\ \ \lambda_{\texttt{ME}}\ \ \lambda_{\texttt{I/A}}\ \ \lambda_{\texttt{ROE}}\ ]'$ and the various statistics along with the corresponding p-values for testing the implications of the HXZ4 model. We consider 8 non-overlapping 5-year testing periods from 1975 to 2014. For each testing period, the pretesting period consists of the preceding five years. We report point estimates, in annualized percentages, based on the two-step and iterated IV-GMM estimators, i.e., $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$ and $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{IT}}$. The null hypothesis implied by the HXZ4 model is $\lambda_0 = 0$, $\lambda_{\texttt{MKT}} = \overline{f}_{\texttt{MKT}}$, $\lambda_{\texttt{ME}} = \overline{f}_{\texttt{ME}}$, $\lambda_{\texttt{I/A}} = \overline{f}_{\texttt{I/A}}$, and $\lambda_{\texttt{ROE}} = \overline{f}_{\texttt{ROE}}$. We report the risk premia t-statistics $t(\lambda_0)$ and $t(\lambda_{\texttt{MKT}})$, $t(\lambda_{\texttt{ME}})$, $t(\lambda_{\texttt{I/A}})$, and $t(\lambda_{\texttt{ROE}})$, given in (70) and (71), respectively, the joint risk premia test statistic $J_d(\boldsymbol{\lambda})$, given in (74), the aggregate mispricing t-statistic $t(\boldsymbol{\alpha})$, given in (72), and the joint test statistic $J_d(\boldsymbol{\delta})$, given in (73). The corresponding p-values are reported in square brackets below the test statistics.

Pretesting Period	70-74	75-79	80-84	85-89	90-94	95-99	00-04	05-09
Testing Period	75-79	80-84	85-89	90-94	95-99	00-04	05-09	10-14
				Number of Test Assets				
N	1059	1598	1527	1709	1627	1740	1610	1560
				Average Factor Realizations				
$\overline{f}_{\texttt{MKT}}$	11.15	4.05	12.13	4.55	20.14	-2.45	0.38	14.20
$\overline{f}_{\texttt{ME}}$	17.27	4.20	-3.91	2.25	-4.36	14.36	2.58	1.88
$\overline{f}_{\texttt{I/A}}$	3.03	8.75	6.57	3.64	0.75	11.13	-0.57	3.63
$\overline{f}_{\texttt{ROE}}$	3.60	10.23	9.21	11.16	7.32	5.93	3.48	1.86
				Estimates of $\boldsymbol{\lambda}$: $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$				
λ_0	1.55	3.19	6.85	2.77	16.50	17.98	-0.20	13.72
$\lambda_{\texttt{MKT}}$	11.26	0.80	6.83	2.21	0.17	-5.10	2.97	2.50
$\lambda_{\texttt{ME}}$	17.48	7.01	-0.66	6.29	1.30	27.86	8.05	0.96
$\lambda_{\texttt{I/A}}$	0.26	11.20	3.12	-3.49	-8.82	0.37	1.56	14.81
$\lambda_{\texttt{ROE}}$	2.22	-5.49	9.91	-0.41	-5.24	11.35	2.97	0.92
				Estimates of $\boldsymbol{\lambda}$: $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{IT}}$				
λ_0	2.52	5.51	4.19	6.84	6.98	16.25	1.12	13.92
$\lambda_{\texttt{MKT}}$	10.16	-1.87	9.69	-3.00	7.06	-4.68	-1.12	2.35
$\lambda_{\texttt{ME}}$	18.12	8.30	-0.53	3.11	4.00	38.52	11.14	0.77
$\lambda_{\texttt{I/A}}$	-0.09	8.42	2.68	-5.25	-4.83	0.72	1.16	14.79
$\lambda_{\texttt{ROE}}$	2.53	-5.06	11.87	-6.83	-5.84	16.85	-1.23	0.42

Continued on next page

Pretesting Period	70-74	75-79	80-84	85-89	90-94	95-99	00-04	05-09
Testing Period	75-79	80-84	85-89	90-94	95-99	00-04	05-09	10-14
				Test Statistics: $\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{TS}}$				
$t\,(\lambda_0)$	0.40	0.34	0.71	0.55	1.63	4.13	-0.04	4.97
p-value	[0.687]	[0.733]	[0.476]	[0.585]	[0.104]	[0.000]	[0.968]	[0.000]
$t\,(\lambda_{\mathtt{MKT}})$	0.03	-0.27	-0.56	-0.40	-2.28	-0.84	0.44	-6.32
p-value	[0.979]	[0.786]	[0.577]	[0.689]	[0.022]	[0.401]	[0.658]	[0.000]
$t\,(\lambda_{\mathtt{ME}})$	0.13	1.13	1.58	0.67	2.16	1.69	1.84	-0.30
p-value	[0.896]	[0.257]	[0.115]	[0.501]	[0.030]	[0.091]	[0.065]	[0.765]
$t\,(\lambda_{\mathtt{I/A}})$	-0.81	0.21	-1.42	-1.52	-2.61	-1.80	0.84	2.54
p-value	[0.420]	[0.831]	[0.156]	[0.128]	[0.009]	[0.071]	[0.400]	[0.011]
$t\,(\lambda_{\mathtt{ROE}})$	-0.16	-1.65	0.16	-0.92	-3.79	0.87	-0.12	-0.17
p-value	[0.876]	[0.099]	[0.875]	[0.357]	[0.000]	[0.386]	[0.905]	[0.869]
$J_d(\boldsymbol{\lambda})$	0.85	4.25	5.34	4.07	33.76	24.60	4.32	71.27
p-value	[0.867]	[0.363]	[0.306]	[0.371]	[0.003]	[0.006]	[0.406]	[0.000]
$t\,(\boldsymbol{\alpha})$	-0.91	-1.38	-1.76	-4.69	-4.63	-2.40	-2.08	-0.43
p-value	[0.364]	[0.167]	[0.078]	[0.000]	[0.000]	[0.016]	[0.037]	[0.669]
$J_d(\boldsymbol{\delta})$	1.68	6.16	8.44	26.08	55.24	30.35	8.65	71.45
p-value	[0.752]	[0.311]	[0.223]	[0.032]	[0.001]	[0.006]	[0.216]	[0.000]
				Test Statistics: $\widehat{\boldsymbol{\lambda}}_{\mathtt{IV}}^{\mathtt{IT}}$				
$t\,(\lambda_0)$	0.69	0.67	0.44	1.29	0.78	3.06	0.26	5.03
p-value	[0.492]	[0.505]	[0.661]	[0.197]	[0.434]	[0.002]	[0.793]	[0.000]
$t\,(\lambda_{\mathtt{MKT}})$	-0.26	-0.56	-0.26	-1.32	-1.85	-0.62	-0.30	-6.47
p-value	[0.798]	[0.574]	[0.797]	[0.186]	[0.065]	[0.538]	[0.765]	[0.000]
$t\,(\lambda_{\mathtt{ME}})$	0.55	1.91	1.57	0.15	3.25	2.70	2.57	-0.36
p-value	[0.583]	[0.056]	[0.117]	[0.884]	[0.001]	[0.007]	[0.010]	[0.718]
$t\,(\lambda_{\mathtt{I/A}})$	-0.99	-0.03	-1.63	-1.82	-1.66	-1.28	0.62	2.56
p-value	[0.322]	[0.974]	[0.102]	[0.069]	[0.097]	[0.202]	[0.538]	[0.010]
$t\,(\lambda_{\mathtt{ROE}})$	-0.13	-1.82	0.56	-1.46	-4.11	1.61	-1.14	-0.25
p-value	[0.896]	[0.069]	[0.574]	[0.143]	[0.000]	[0.108]	[0.252]	[0.799]
$J_d(\boldsymbol{\lambda})$	1.84	7.72	5.70	8.88	34.23	21.21	8.44	73.93
p-value	[0.691]	[0.206]	[0.291]	[0.175]	[0.003]	[0.011]	[0.167]	[0.000]
$t\,(\boldsymbol{\alpha})$	-0.89	-3.82	-1.31	-2.16	-9.82	-1.30	-3.30	-0.43
p-value	[0.374]	[0.000]	[0.192]	[0.031]	[0.000]	[0.194]	[0.001]	[0.668]
$J_d(\boldsymbol{\delta})$	2.63	22.29	7.40	13.53	130.58	22.89	19.33	74.11
p-value	[0.625]	[0.041]	[0.264]	[0.125]	[0.000]	[0.017]	[0.033]	[0.000]

Table 10: **Testing the FF5 model: empirical results for 5-year testing periods from 1975 to 2014 using 49 industry clusters.** This table presents the point estimates of $\boldsymbol{\lambda} = [\ \lambda_0\ \ \lambda_{\texttt{MKT}}\ \ \lambda_{\texttt{SMB}}\ \ \lambda_{\texttt{HML}}\ \ \lambda_{\texttt{RMW}}\ \ \lambda_{\texttt{CMA}}\ ]'$ and the various statistics along with the corresponding p-values for testing the implications of the FF5 model. We consider 8 non-overlapping 5-year testing periods from 1975 to 2014. For each testing period, the pretesting period consists of the preceding five years. We report point estimates, in annualized percentages, based on the two-step and iterated IV-GMM estimators, i.e., $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$ and $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{IT}}$. The null hypothesis implied by the FF5 model is $\lambda_0 = 0$, $\lambda_{\texttt{MKT}} = \overline{f}_{\texttt{MKT}}$, $\lambda_{\texttt{SMB}} = \overline{f}_{\texttt{SMB}}$, $\lambda_{\texttt{HML}} = \overline{f}_{\texttt{HML}}$, $\lambda_{\texttt{RMW}} = \overline{f}_{\texttt{RMW}}$, and $\lambda_{\texttt{CMA}} = \overline{f}_{\texttt{CMA}}$. We report the risk premia t-statistics $t(\lambda_0)$ and $t(\lambda_{\texttt{MKT}})$, $t(\lambda_{\texttt{SMB}})$, $t(\lambda_{\texttt{HML}})$, $t(\lambda_{\texttt{RMW}})$, and $t(\lambda_{\texttt{CMA}})$, given in (70) and (71), respectively, the joint risk premia test statistic $J_d(\boldsymbol{\lambda})$, given in (74), the aggregate mispricing t-statistic $t(\boldsymbol{\alpha})$, given in (72), and the joint test statistic $J_d(\boldsymbol{\delta})$, given in (73). The corresponding p-values are reported in square brackets below the test statistics.

Pretesting Period	70-74	75-79	80-84	85-89	90-94	95-99	00-04	05-09
Testing Period	75-79	80-84	85-89	90-94	95-99	00-04	05-09	10-14
	Number of Test Assets							
N	1193	1719	1692	1882	1811	2049	1956	1885
	Average Factor Realizations							
$\overline{f}_{\texttt{MKT}}$	11.18	4.53	12.49	4.83	20.69	-3.10	-0.43	15.52
$\overline{f}_{\texttt{SMB}}$	17.33	4.14	-4.82	1.57	-5.30	12.78	1.63	1.54
$\overline{f}_{\texttt{HML}}$	6.53	9.35	2.57	2.70	-5.51	15.00	2.08	-0.93
$\overline{f}_{\texttt{RMW}}$	0.10	4.04	5.51	4.70	0.94	10.23	4.73	0.71
$\overline{f}_{\texttt{CMA}}$	1.72	6.24	5.22	1.74	-1.69	13.93	-0.28	2.90
	Estimates of $\boldsymbol{\lambda}$: $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$							
λ_0	1.28	7.42	20.42	3.41	4.38	10.94	-4.57	8.73
$\lambda_{\texttt{MKT}}$	9.51	-3.67	-6.31	0.16	11.06	-7.54	7.99	6.28
$\lambda_{\texttt{SMB}}$	18.39	5.56	-5.84	7.29	3.44	27.70	1.15	5.87
$\lambda_{\texttt{HML}}$	2.16	4.65	3.76	0.05	-6.44	4.32	8.24	-0.53
$\lambda_{\texttt{RMW}}$	-3.36	-0.09	-4.50	-5.67	-2.48	4.27	-5.95	-6.37
$\lambda_{\texttt{CMA}}$	-2.09	6.72	-2.06	-5.37	-4.47	3.64	-7.43	6.53
	Estimates of $\boldsymbol{\lambda}$: $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{IT}}$							
λ_0	2.04	4.23	31.58	3.91	3.45	10.93	0.12	8.64
$\lambda_{\texttt{MKT}}$	9.26	-2.46	-13.05	-0.89	11.67	-7.38	3.43	6.38
$\lambda_{\texttt{SMB}}$	16.79	8.28	-11.77	6.51	3.74	26.93	3.17	6.85
$\lambda_{\texttt{HML}}$	1.84	-1.73	6.57	1.82	-6.53	4.52	13.55	-1.33
$\lambda_{\texttt{RMW}}$	-2.24	0.83	-8.68	-4.33	-2.18	3.95	-6.53	-8.44
$\lambda_{\texttt{CMA}}$	-1.14	7.84	-6.35	-4.75	-3.80	2.97	-4.37	6.22

Continued on next page

Table 10 – continued from previous page

Pretesting Period	70-74	75-79	80-84	85-89	90-94	95-99	00-04	05-09
Testing Period	75-79	80-84	85-89	90-94	95-99	00-04	05-09	10-14
				Test Statistics: $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{TS}}$				
$t\,(\lambda_0)$	0.24	1.17	1.94	1.58	0.88	4.35	-0.93	6.37
p-value	[0.811]	[0.243]	[0.052]	[0.115]	[0.377]	[0.000]	[0.354]	[0.000]
$t\,(\lambda_{\texttt{MKT}})$	-0.34	-1.22	-1.99	-1.95	-2.50	-2.10	1.73	-5.46
p-value	[0.733]	[0.224]	[0.046]	[0.051]	[0.013]	[0.036]	[0.084]	[0.000]
$t\,(\lambda_{\texttt{SMB}})$	0.72	1.13	-0.31	4.33	4.47	2.82	-0.14	2.75
p-value	[0.470]	[0.259]	[0.760]	[0.000]	[0.000]	[0.005]	[0.889]	[0.006]
$t\,(\lambda_{\texttt{HML}})$	-1.83	-0.99	0.31	-1.08	-0.41	-2.93	1.14	0.28
p-value	[0.067]	[0.322]	[0.755]	[0.279]	[0.683]	[0.003]	[0.254]	[0.779]
$t\,(\lambda_{\texttt{RMW}})$	-0.88	-1.02	-2.60	-4.03	-1.47	-2.21	-2.24	-2.95
p-value	[0.378]	[0.310]	[0.009]	[0.000]	[0.142]	[0.027]	[0.025]	[0.003]
$t\,(\lambda_{\texttt{CMA}})$	-0.96	0.18	-1.62	-4.08	-0.78	-1.01	-1.56	2.19
p-value	[0.338]	[0.861]	[0.105]	[0.000]	[0.438]	[0.312]	[0.118]	[0.029]
$J_d(\boldsymbol{\lambda})$	5.74	6.16	17.31	59.11	29.92	45.86	12.64	91.55
p-value	[0.341]	[0.354]	[0.053]	[0.000]	[0.002]	[0.000]	[0.102]	[0.000]
$t\,(\boldsymbol{\alpha})$	-4.42	-5.76	-2.12	-4.44	-6.22	-3.69	-2.14	-1.88
p-value	[0.000]	[0.000]	[0.034]	[0.000]	[0.000]	[0.000]	[0.033]	[0.060]
$J_d(\boldsymbol{\delta})$	25.29	39.32	21.82	78.86	68.63	59.48	17.20	95.08
p-value	[0.029]	[0.001]	[0.047]	[0.000]	[0.000]	[0.000]	[0.064]	[0.000]
				Test Statistics: $\widehat{\boldsymbol{\lambda}}_{\texttt{IV}}^{\texttt{IT}}$				
$t\,(\lambda_0)$	0.43	0.68	1.66	1.96	0.74	4.36	0.02	5.94
p-value	[0.665]	[0.498]	[0.096]	[0.049]	[0.458]	[0.000]	[0.984]	[0.000]
$t\,(\lambda_{\texttt{MKT}})$	-0.43	-1.07	-1.53	-2.59	-2.47	-2.12	0.74	-5.28
p-value	[0.664]	[0.283]	[0.127]	[0.010]	[0.013]	[0.034]	[0.462]	[0.000]
$t\,(\lambda_{\texttt{SMB}})$	-0.44	3.06	-1.22	3.81	4.59	2.70	0.46	3.27
p-value	[0.663]	[0.002]	[0.222]	[0.000]	[0.000]	[0.007]	[0.645]	[0.001]
$t\,(\lambda_{\texttt{HML}})$	-2.45	-2.12	0.76	-0.38	-0.45	-2.93	2.31	-0.30
p-value	[0.014]	[0.034]	[0.448]	[0.700]	[0.655]	[0.003]	[0.021]	[0.767]
$t\,(\lambda_{\texttt{RMW}})$	-0.71	-0.79	-2.12	-3.75	-1.36	-2.37	-2.32	-3.37
p-value	[0.475]	[0.429]	[0.034]	[0.000]	[0.172]	[0.018]	[0.020]	[0.001]
$t\,(\lambda_{\texttt{CMA}})$	-0.82	0.61	-1.59	-4.16	-0.61	-1.10	-0.93	1.92
p-value	[0.414]	[0.541]	[0.111]	[0.000]	[0.539]	[0.273]	[0.353]	[0.055]
$J_d(\boldsymbol{\lambda})$	7.74	16.46	14.20	56.53	30.12	46.19	12.32	88.97
p-value	[0.248]	[0.048]	[0.097]	[0.000]	[0.002]	[0.000]	[0.111]	[0.000]
$t\,(\boldsymbol{\alpha})$	-6.51	-3.07	-0.83	-4.45	-5.08	-4.34	-1.25	-1.75
p-value	[0.000]	[0.002]	[0.408]	[0.000]	[0.000]	[0.000]	[0.210]	[0.080]
$J_d(\boldsymbol{\delta})$	50.15	25.89	14.89	76.32	55.96	64.99	13.89	92.03
p-value	[0.002]	[0.013]	[0.121]	[0.000]	[0.000]	[0.000]	[0.113]	[0.000]