Journal of Financial Economics 00 (1999) 000–000

GMM Tests of Stochastic Discount Factor Models with Useless Factors[1]

Raymond Kan[a], Chu Zhang[b,c,*]

[a]*Joseph L. Rotman School of Management, University of Toronto, Toronto, Ontario, Canada M5S 3E6*

[b]*Department of Finance, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong*

[c]*Faculty of Business, University of Alberta. Edmonton, Alberta, Canada T6G 2R6*

(Received 23 March 1998; received in revised form 24 September 1998)

Abstract

This paper studies generalized method of moments tests for the stochastic discount factor representation of asset pricing models when one of the proposed factors is in fact useless, defined as being independent of the asset returns. Analytic results on asymptotic distributions and simulation results on finite sample distributions both show that (i) the Wald test tends to overreject the hypothesis of a zero factor premium for a useless factor when the model is misspecified, (ii) with the presence of a useless factor, the power of the over-identifying restriction test in rejecting misspecified models is reduced, and in some cases a misspecified model with a useless factor is more likely to be accepted than the true model.

JEL Classification: G12

Keywords: Stochastic discount factor models; Generalized method of moments; Nonidentifiability; Useless factors; Misspecification

---

*Corresponding author. Tel.: 852/2358-7684; fax: 852/2358-1749; e-mail: czhang@ust.hk.

[1]We would like to thank Burton Hollifield, Guofu Zhou, seminar participants at Hong Kong University and an anonymous referee for their helpful comments and suggestions. All remaining errors are ours. Kan gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada.

1. Introduction

In empirical tests of asset pricing models, macroeconomic variables are often proposed as candidates for systematic factors. The macroeconomic variables are typically motivated by theory or economic intuition, but many have statistically insignificant correlations with the returns on financial assets. Taking a skeptic's point of view, some of these macroeconomic variables might be useless "factors," in the sense that they are independent of all the asset returns. Recently, Kan and Zhang (1999) show that the t-test in the Fama-MacBeth (1973) two-pass methodology overrejects the hypothesis that the risk premium associated with a useless factor is zero, implying that a useless factor tends to be mistaken as an important factor. In this paper, we focus on the stochastic discount factor representation of beta pricing models, and investigate the performance of Hansen's (1982) Generalized Method of Moments (GMM) when the model includes a useless factor.

There are two main reasons that further investigation of the useless factor problem is warranted. To illustrate the first reason, consider the following stochastic discount factor model,

$$E[r_t - r_t g_t \gamma] = 0_N, \tag{1}$$

where r_t is the N-vector of returns in excess of the risk-free rate, g_t is a marketwide variable with zero mean, and γ is the risk premium associated with g_t. To exclude the uninteresting case, suppose $E[r_t]$ is not a zero vector. If g_t is useless, which implies $E[r_t g_t] = 0_N$, then Eq.(1) cannot hold and, theoretically, γ is undefined. Of course, the sample version of $E[r_t g_t]$, $\frac{1}{T}\sum_{t=1}^{T} r_t g_t$, will not be exactly zero, so γ can still be estimated. The easiest way to see the property of such an estimate is when $N = 1$. The estimate of γ is $\frac{1}{T}\sum_{t=1}^{T} r_t / \frac{1}{T}\sum_{t=1}^{T} r_t g_t$, which tends to infinity in absolute value as the sample size T increases. This is basically the problem caused by a useless factor. Now consider the model

$$E[r_t - r_t f_t' \lambda - r_t g_t \gamma] = 0_N, \tag{2}$$

where f_t is the k-vector of all the true factors, λ is the vector of risk premiums associated with f_t, and g_t and γ are defined the same as above. What will happen to the estimate of γ now? Although γ is still undefined, we will show that the useless factor is unlikely to be accepted as an important factor and is less harmful than in Eq.(1). The key difference is that Eq.(2) is a valid equation despite the presence of a useless factor. We establish in this paper that, in a sense, the seriousness of the problem caused by a useless factor is related to the degree of model misspecification.

The second issue is whether the over-identifying restriction (OIR) test in GMM can reliably detect misspecified models that contain useless factors. As we remarked in the last paragraph, a useless factor causes serious problems primarily when the model is misspecified, which is exactly what the OIR test is designed to detect. It is natural to conjecture that the OIR test will effectively reject a misspecified model with a useless factor. But can we really count on the OIR test to detect a misspecified model that contains a useless factor? Surprisingly, the answer is "no." By definition, the presence of a useless factor does not make a misspecified first-moment condition more, or less, incorrect. However, it blows up the estimated second-moment function whose inverse is used as a weighting matrix in estimation and testing. As a result, the OIR test becomes less powerful and a misspecified model with a useless factor may pass as a correct one.

We emphasize that our results do not imply a flaw in the GMM itself. The validity of the GMM requires some regularity conditions that are sometimes overlooked by empirical researchers and sometimes difficult to verify. Many macroeconomic variables used in empirical asset pricing studies have low sample correlations with asset returns. Although the hypothesis that a macroeconomic variable has zero correlations with the returns on a set of assets cannot be rejected, researchers often find themselves reluctant to throw away the variable because of the concern about statistical power. On the other hand, a useless factor could pass the test for zero correlations due to a few outliers, or violations of other joint hypotheses in forming the test of zero correlations. These possibilities complicate direct tests of useless factors. It is important, therefore, to understand the statistical properties of asset pricing tests in the presence of a useless factor. It is important also because a pure useless factor serves as the limiting case of a true factor observed with noise when the amount of noise increases.

Empirical models are unlikely to include all true factors, and the findings in our paper illustrate the dangers of blindly interpreting empirical results without a rigorous diagnosis of the model. In a common situation, a researcher estimates a misspecified asset pricing model that does not include all the true factors and the OIR test rejects the model. Consequently, another variable is added as a factor. The new variable appears to be priced, the p-value of the OIR test of the new model increases, so the research does not reject the new model. Our analysis indicates a potential problem with the interpretations that the added variable is a priced factor, and that the new model is the correctly specified model.

The rest of the paper is organized as follows. Section 2 discusses the asymptotic distributions of GMM parametric tests of risk premiums and OIR tests for models containing a useless factor. Section 3 presents simulation results that illustrate the magnitude of the problem in finite sample. The last section concludes and the Appendix contains proofs of all propositions.

2. Asymptotic distribution of GMM tests with useless factors

Suppose r_t is the vector of returns in excess of the risk-free rate at time t on N risky assets, generated by a factor structure with factors f_t:

$$r_t = \nu + Bf_t + \varepsilon_t, \tag{3}$$

where $E[f_t] = 0_k$, $E[\varepsilon_t|f_t] = 0_N$, $\nu \equiv E[r_t] \neq 0_N$, and $\mathrm{Var}[f_t] = I_k$. Standardizing macroeconomic factors is a nontrivial issue. In practice, the conditional distribution of the factors should be modeled and estimated as part of the moment conditions. We ignore this complications because it would distract from the main issue regarding useless factors, which does not depend on the standardization assumption. We assume f_t and r_t are both stationary and ergodic with finite fourth moments. In addition, since only unexpected shocks matter for unexpected returns, f_t can be modeled as a martingale difference sequence.

If the multibeta asset pricing model holds, the model can be written as a stochastic discount factor model

$$E[r_t(1 - f_t'\lambda^*)] = 0_N, \tag{4}$$

where each component of the k-vector of the true risk premiums λ^* is nonzero. For simplicity, we assume the model is estimated and tested without conditioning on information variables at time $t-1$. Cochrane (1996) define this simple form of the test as "unconditional test of the unconditional model." Let $u_t = r_t(1 - f_t'\lambda^*)$ and $\bar{u}_T = \frac{1}{T}\sum_{t=1}^{T} u_t$, where T is the number of time-series observations. We assume that

$$\sqrt{T}\bar{u}_T \overset{A}{\sim} N(0_N, S) \tag{5}$$

for some positive-definite matrix S, where $\overset{A}{\sim}$ denotes an asymptotic distribution.

Suppose an econometrician builds a stochastic discount factor model using a subvector f_{1t} of f_t, and a useless factor g_t,

$$E[r_t(1 - f_{1t}'\lambda_1 - g_t\gamma)] = 0_N. \tag{6}$$

A variable is defined as useless if it is independent of f_s and ε_s for all t and s (at all leads and lags). Without loss of generality, g_t is assumed to be stationary and ergodic martingale difference sequence with $E[g_t] = 0$ and $\mathrm{Var}[g_t] = 1$. The vector f_{1t} has dimension k_1, with $0 \leq k_1 \leq k$. For $k_1 = 0$, all the true factors drop out of Eq.(6) and only the useless factor g_t is used in the model. For $k_1 = k$, $f_{1t} = f_t$, so the model includes the full set of true factors.

When a useless factor is included in the model, its risk premium is not identifiable because all values of γ give the same moment condition. Consequently, the moment conditions violate the identifiability condition required by GMM (Hansen 1982, Assumption 3.4), implying that the standard asymptotic theory for GMM is not applicable. A model represented by Eq.(6) is said to be misspecified if the moment condition does not hold for any set of parameter values. In the context here, the model is misspecified when $k_1 < k$, while the model is correctly specified when $k_1 = k$. Although γ is nonidentifiable in both cases, we will show that nonidentifiability causes serious problems only in misspecified models.[2]

Without knowing whether the model is misspecified or a factor is useless, the econometrician estimates the model with the usual GMM procedure, by minimizing a quadratic form of the moment function in Eq.(6). Let

$$\bar{m}_T(\lambda_1, \gamma) = \frac{1}{T}\sum_{t=1}^{T} r_t(1 - f_{1t}'\lambda_1 - g_t\gamma), \tag{7}$$

where λ_1 and γ are treated as parameters. In the GMM methodology, the parameters are estimated by

$$\mathrm{argmin}_{\lambda_1,\gamma}\, \bar{m}_T(\lambda_1, \gamma)' W_T \bar{m}_T(\lambda_1, \gamma), \tag{8}$$

where W_T is a (possibly stochastic) positive-definite weighting matrix. Standard GMM chooses the statistically optimal weighting matrix, equal to the inverse of the second moment

[2] Staiger and Stock (1997) and Stock and Wright (1997) discuss a related problem known as weak identification. The weak identification problem occurs when poor instruments are used in conditional tests of asset pricing models. This problem does not affect the unconditional tests we discuss here. In contrast, model misspecification with useless factors affects both conditional and unconditional tests of asset pricing models.

matrix of $\bar{m}_T(\lambda_1, \gamma)$, and typically requires iterations. Nonstandard GMM selects some predetermined weighting matrix. We discuss the two GMM approaches in turn.

2.1. The standard GMM parametric tests

When the optimal weighting matrix depends on unknown parameters, an iterative method is required. In the first round, a positive-definite matrix, such as the identity matrix, is used as the weighting matrix to estimate the parameters. In the second round, the model is re-estimated using the optimal weighting matrix based on the parameter estimates from the first round. When the model is correctly specified and T is large, two rounds of iteration are sufficient because the GMM parameter estimator in the first round is consistent for any positive-definite weighting matrix under regularity conditions given by Hansen (1982). In practice, however, the iterative procedure can be repeated for many rounds to give better estimates, as Ferson and Foerster (1994) suggest.

Let $\bar{r}_T = \frac{1}{T}\sum_{t=1}^{T} r_t$, $\bar{B}_{1T} = \frac{1}{T}\sum_{t=1}^{T} r_t f_{1t}'$, $\bar{D}_T = \frac{1}{T}\sum_{t=1}^{T} r_t g_t$, and let $(\hat{S}_T^{(0)})^{-1}$ be the initial weighting matrix. In the lth round for $l \geq 1$, the parameter estimates and the estimate of the second moment matrix of $\sqrt{T}\bar{m}_T(\lambda_1, \gamma)$ are

$$\begin{pmatrix} \hat{\lambda}_{1T}^{(l)} \\ \hat{\gamma}_T^{(l)} \end{pmatrix} = \left[(\bar{B}_{1T}, \bar{D}_T)' \left(\hat{S}_T^{(l-1)}\right)^{-1} (\bar{B}_{1T}, \bar{D}_T)\right]^{-1} \left[(\bar{B}_{1T}, \bar{D}_T)' \left(\hat{S}_T^{(l-1)}\right)^{-1} \bar{r}_T\right], \tag{9}$$

$$\hat{S}_T^{(l)} = \frac{1}{T}\sum_{t=1}^{T} \left[r_t(1 - f_{1t}'\hat{\lambda}_{1T}^{(l)} - g_t\hat{\gamma}_T^{(l)})\right] \left[r_t(1 - f_{1t}'\hat{\lambda}_{1T}^{(l)} - g_t\hat{\gamma}_T^{(l)})\right]'. \tag{10}$$

In this paper, we are interested in the performance of the Wald test statistic (i.e., the square of the asymptotic z-ratio) of the hypothesis that γ equals 0. Let $H = (0_{k_1}', 1)$ and consider the Wald test based on the weighting matrix obtained *after* the minimization,

$$\xi_{P_a}^{(l)} = \frac{T(\hat{\gamma}_T^{(l)})^2}{H\left[(\bar{B}_{1T}, \bar{D}_T)' \left(\hat{S}_T^{(l)}\right)^{-1} (\bar{B}_{1T}, \bar{D}_T)\right]^{-1} H'}, \qquad l = 2, 3, \cdots, \tag{11}$$

and the Wald test based on the weighting matrix *before* the minimization,

$$\xi_{P_b}^{(l)} = \frac{T(\hat{\gamma}_T^{(l)})^2}{H\left[(\bar{B}_{1T}, \bar{D}_T)' \left(\hat{S}_T^{(l-1)}\right)^{-1} (\bar{B}_{1T}, \bar{D}_T)\right]^{-1} H'}, \qquad l = 2, 3, \cdots. \tag{12}$$

If the model is correctly specified and g_t is a nonpriced factor ($\gamma = 0$) with $E[r_t g_t] \neq 0_N$, the Wald test statistic is asymptotically distributed as χ_1^2. However, standard asymptotic results of GMM do not apply if the model contains a useless factor. The following proposition provides the large sample properties of the GMM Wald tests in this case.

Proposition 1. (A) Suppose the model in Eq.(6) is misspecified with $k_1 < k$. The Wald tests $\xi_{P_a}^{(l)}$ for $l \geq 2$, and $\xi_{P_b}^{(l)}$ for $l \geq 3$ of the hypothesis $\gamma = 0$ are asymptotically distributed as $\chi^2_{N-k_1}$. (B) Suppose the model in Eq.(6) is correctly specified with $k_1 = k$. The limiting distributions of the Wald tests $\xi_{P_a}^{(l)}$ for $l \geq 2$, and $\xi_{P_b}^{(l)}$ for $l \geq 2$ of the hypothesis $\gamma = 0$ are stochastically dominated by the χ^2_1 distribution.

The proposition indicates that there is a major difference between misspecified and correctly specified models with useless factors. Consider the misspecified model first. Since the Wald tests are compared to the χ^2_1 distribution for statistical inference, the fact that the tests are asymptotically $\chi^2_{N-k_1}$ means that the null hypothesis will be overrejected when the model is misspecified.[3] The magnitude of overrejection is especially severe when N is large. We can compare the results for the misspecified model to those of Kan and Zhang (1999), who analyze the Fama and MacBeth (1973) two-pass methodology when the model contains only the useless factor. Kan and Zhang show that for the hypothesis of a zero risk premium, the t-test unadjusted for errors-in-variables (EIV) bias diverges to infinity as T goes to infinity, while the t-test adjusted for EIV bias converges to a finite random variable, both over-rejecting the hypothesis. The limit of the GMM Wald tests in the stochastic discount factor model behave more like the EIV adjusted t-tests because the standard errors in GMM account for estimation errors in the betas. The overrejection rates are reduced when the model contains a partial set of true factors, compared to the case of a useless factor alone. But since N is typically much greater than k_1, the reduction is very limited for the asymptotic distributions.

Part (B) of the proposition shows that a useless factor is less harmful when the model includes all the true factors. The risk premium estimates for the true factors are not affected in the limit, but the standard errors of the estimates are still incorrect due to the useless factor. The risk premium estimate for the useless factor does not explode to infinity, but it also does not converge to zero, or any constant. The limiting distribution is unbounded and

[3]We slightly abuse the expression "overrejection" here. The standard use of the term refers to the case in which the hypothesis $\gamma = 0$ is true, but the risk premium of a useless factor is simply undefined.

centered at zero. The Wald test statistic has a distribution dominated by the χ_1^2 distribution, so asymptotically there is no over-rejection problem at all. Typically, a useless factor will be excluded as a nonpriced factor and, therefore, correctly specified models do not suffer greatly from the presence of useless factors. It is model misspecification that causes the problem.

2.2. The nonstandard GMM parametric tests

Nonstandard GMM employs a weighting matrix, W_T, that is not statistically optimal because it does not minimize the variance matrix of the parameter estimator. Nonstandard GMM is preferable in some situations. For example, Zhou (1994) suggests that analytical GMM tests are possible for some choice of the weighting matrix, and Hansen and Jagannathan (1997) suggest that the sample version of the inverse of the second moment matrix of gross returns can be used as the weighting matrix for the purpose of comparing models that are potentially misspecified.

In nonstandard GMM, the Wald test for $\gamma = 0$ is given by

$$\begin{aligned} \xi_P^W &= T(\hat{\gamma}_T^W)^2 \Big(H \left[(\bar{B}_{1T}, \bar{D}_T)' W_T (\bar{B}_{1T}, \bar{D}_T) \right]^{-1} \left[(\bar{B}_{1T}, \bar{D}_T)' W_T \hat{S}_T^W W_T (\bar{B}_{1T}, \bar{D}_T) \right] \\ &\quad \cdot \left[(\bar{B}_{1T}, \bar{D}_T)' W_T (\bar{B}_{1T}, \bar{D}_T) \right]^{-1} H' \Big)^{-1}, \end{aligned} \tag{13}$$

where

$$\begin{pmatrix} \hat{\lambda}_{1T}^W \\ \hat{\gamma}_T^W \end{pmatrix} = \left[(\bar{B}_{1T}, \bar{D}_T)' W_T (\bar{B}_{1T}, \bar{D}_T) \right]^{-1} \left[(\bar{B}_{1T}, \bar{D}_T)' W_T \bar{r}_T \right], \tag{14}$$

and $\hat{S}_T^W$ is estimated as in Eq.(10) using $\hat{\lambda}_{1T}^W$ and $\hat{\gamma}_T^W$. When the stochastic discount factor model is estimated and tested with nonstandard GMM, useless factors still cause an overrejection problem, but to a lesser extent. We present the results in the following proposition.

Proposition 2. (A) Suppose the model in Eq.(6) is misspecified with $k_1 < k$ and the parameters are estimated using a weighting matrix W_T with positive-definite nonstochastic limit W. Then the limit of the nonstandard Wald test ξ_P^W for the hypothesis $\gamma = 0$ as T tends to infinity is bounded in distribution by $\xi_1 \overset{A}{\leq} \xi_P^W \overset{A}{\leq} \xi_{N-k_1}$, where ξ_1 is distributed as χ_1^2 and ξ_{N-k_1} is distributed as $\chi_{N-k_1}^2$.[4] *(B) Suppose the model in Eq.(6) is correctly specified*

[4]For a sequence of random variables x_T and a random variable y, $x_T \overset{A}{\leq} y$ means the limit of x_T is dominated by y, which in turn implies that the limit of x_T is stochastically dominated by y. A similar definition applies to $y \overset{A}{\leq} x_T$.

with $k_1 = k$. Then the limit of the nonstandard Wald test ξ_P^W for the hypothesis $\gamma = 0$ is stochastically dominated by the χ_1^2 distribution.

The proposition says that ξ_P^W dominates a χ_1^2 variable when the model is misspecified, which indicates that the nonstandard Wald test still overrejects the hypothesis that $\gamma = 0$. The upper bound given in part (A), however, suggests that the overrejection problem of the nonstandard Wald test is not as serious as the problem in standard GMM. The overrejection rate depends on the fixed weighting matrix. If W is proportional to $(E[r_t r_t'])^{-1}$, the overrejection problem for the nonstandard Wald test is just as serious as that of standard GMM. In the proof of Proposition 1, it is shown that $\frac{1}{T}\hat{S}_T^{(l)}$ is asymptotically equivalent to a matrix proportional to $E[r_t r_t']$. Therefore, when W is proportional to $(E[r_t r_t'])^{-1}$, the nonstandard GMM is asymptotically equivalent to standard GMM. When W becomes more and more different from $(E[r_t r_t'])^{-1}$, the rejection rate of the nonstandard Wald test potentially declines. Nevertheless, we do not advocate the use of a weighting matrix that is different from the optimal weighting matrix because doing so also causes the variance matrix of the parameter estimates to be large. Reducing the accuracy of estimates cannot be an effective way of eliminating the problem caused by useless factors.

2.3. The standard GMM over-identifying restriction tests

The GMM over-identifying restriction (OIR) test is designed to detect misspecified models. In the previous subsections, we show that misspecification causes the over-rejection problem of the Wald tests. If the OIR test is effective in detecting misspecified models when a useless factor is present, then we do not need to be concerned with the problem of the Wald tests. In this and the next subsection, however, we show that the OIR tests for both standard and nonstandard GMM cannot reliably detect a misspecified model that contains a useless factor, even asymptotically. Worse yet, in some cases, a misspecified model that contains a useless factor can pass the OIR test even more likely than the true model, making the OIR test completely unreliable in detecting such misspecified models.

In standard GMM, the OIR test statistic can be computed from the estimated parameters after the first round. More specifically, the OIR test statistic of the model in Eq.(6) in the

lth round is given by

$$\xi_O^{(l)} = T\bar{m}_T(\hat{\lambda}_{1T}^{(l)}, \hat{\gamma}_T^{(l)})' \left(\hat{S}_T^{(l-1)}\right)^{-1} \bar{m}_T(\hat{\lambda}_{1T}^{(l)}, \hat{\gamma}_T^{(l)}), \qquad l = 2, 3, \ldots. \tag{15}$$

If the model in Eq.(6) is correctly specified and the factors are not useless, $\xi_O^{(l)}$ is asymptotically distributed as $\chi^2_{N-k_1-1}$. An unusually high value of $\xi_O^{(l)}$ compared with the $\chi^2_{N-k_1-1}$ distribution is taken as evidence that the model in Eq.(6) is misspecified. For a misspecified model that does not contain a useless factor, it is easy to show that $\xi_O^{(l)}$ goes to infinity as T goes to infinity. Therefore, misspecified models that do not contain useless factors can be detected by the OIR test almost surely when T is large.

However, adding a useless factor to a misspecified model complicates the situation. The addition of a useless factor by itself does not make the misspecified moment condition more or less incorrect, but it always increases $\hat{S}_T^{(l-1)}$ because the useless factor adds noise to the estimated risk premium. As a result, the OIR test statistic becomes smaller when a useless factor is added to the model, and the power of the OIR test in detecting the misspecified model declines. In fact, from the proof of Proposition 1, we know that $(\hat{S}_T^{(l-1)})^{-1}$ actually tends to zero when the model is misspecified. Therefore, even though $\sqrt{T}\bar{m}_T(\hat{\lambda}_{1T}^{(l)}, \hat{\gamma}_T^{(l)})$ tends to infinity, $\xi_O^{(l)}$ does not. Instead, $\xi_O^{(l)}$ converges to a finite random variable and the misspecification of the model cannot be reliably detected. The following proposition summarizes the asymptotic properties of $\xi_O^{(l)}$ for models that contain a useless factor.

Proposition 3. (A) Suppose the model in Eq.(6) is misspecified with $k_1 < k$. Then for all $l \geq 2$, $\xi_O^{(l)}$ converges to a finite random variable and the asymptotic probability of rejecting Eq.(6) is less than one. For $l \geq 3$, $\xi_O^{(l)} \overset{A}{\sim} \left(1 + \frac{\xi_{N-k_1-1}}{\xi_1}\right)\xi_{N-k_1-1} > \xi_{N-k_1-1}$, where ξ_1 is distributed as χ^2_1, ξ_{N-k_1-1} is distributed as $\chi^2_{N-k_1-1}$, and ξ_1 and ξ_{N-k_1-1} are independent. (B) Suppose the model in Eq.(6) is correctly specified with $k_1 = k$. Then the limits of standard OIR tests $\xi_O^{(l)}$ for $l \geq 2$ are stochastically dominated by the χ^2_{N-k-1} distribution.

Proposition 3 suggests that adding a useless factor will always increase the chance for a model to pass the OIR test, whether or not the model is correctly specified. For a misspecified model, adding a useless factor drives the asymptotic rejection rates to less than one, so the misspecification cannot be reliably detected. The asymptotic probability of rejection depends on whether we use the OIR test from the second round or from a subsequent round.

Proposition 3 advocates the use of the OIR test from the third or a subsequent round, instead of the more popular test from the second round. The asymptotic rejection rates of $\xi_O^{(l)}$ for $l \geq 3$ are guaranteed to be higher than the size of the test, whereas $\xi_O^{(2)}$ does not always satisfy this modest requirement of a good test. Note that in part (A), we do not present the limiting distribution of $\xi_O^{(2)}$ because it depends on the initial weighting matrix. In general, there is no guarantee that $\xi_O^{(2)}$ asymptotically dominates the $\chi^2_{N-k_1-1}$ distribution, so it is possible that the test $\xi_O^{(2)}$ accepts the misspecified model more likely than the true model. In the proof of Proposition 3, we give such an example.

In part (B) the model is correctly specified but includes a useless factor. The fact that the OIR test statistic is stochastically dominated by χ^2_{N-k-1} is not a serious problem because it only makes the correctly specified model more likely to be accepted. This result parallels the case of the Wald test about risk premiums in Subsection 2.1.

2.4. The nonstandard GMM over-identifying restriction tests

For nonstandard GMM, Jagannathan and Wang (1996) give one version of the OIR test and Zhou (1994) discusses another version. For the sake of brevity, we present the analysis only for the first version here. If the model in Eq.(6) is correctly specified and the factors are not useless, we have

$$\xi_O^W = T\,\bar{m}_T(\hat{\lambda}_{1T}^W, \hat{\gamma}_T^W)' W_T \bar{m}_T(\hat{\lambda}_{1T}^W, \hat{\gamma}_T^W) \overset{A}{\sim} \sum_{i=1}^{N-k_1-1} \rho_i z_i^2, \tag{16}$$

where z_i's are independent standard normal variates and ρ_is are the nonzero eigenvalues of the limit of

$$\hat{S}_T^W \left[W_T - W_T(\bar{B}_{1T}, \bar{D}_T) \left[(\bar{B}_{1T}, \bar{D}_T)' W_T (\bar{B}_{1T}, \bar{D}_T) \right]^{-1} (\bar{B}_{1T}, \bar{D}_T)' W_T \right], \tag{17}$$

which must be estimated. One of the advantages of using a fixed weighting matrix W_T that does not depend on parameter estimates is robustness of the OIR test. When a model is misspecified, large residuals cause a large $\hat{S}_T^{(l-1)}$, and hence a small $(\hat{S}_T^{(l-1)})^{-1}$ where $(\hat{S}_T^{(l-1)})^{-1}$ is the weighting matrix in Eq.(15). As a result, the "optimal" weighting matrix in fact rewards misspecification. A fixed weighting matrix does not depend on residuals, and therefore, avoids this undesired result. In the last subsection, we show that the standard GMM OIR test does not reliably reject a misspecified model with a useless factor because the estimated variance matrix of the sample moment conditions, $\hat{S}_T^{(l-1)}$, goes to infinity. The nonstandard GMM OIR test avoids using $(\hat{S}_T^{(l)-1})^{-1}$, and the test goes to infinity as T increases for a misspecified model. Thus, one may think that a misspecified model with a useless factor should be detected reliably by nonstandard GMM OIR tests. The following proposition, however, shows that this intuition is false.

Proposition 4. (A) Suppose the model in Eq.(6) is misspecified with $k_1 < k$. Then the asymptotic probability of rejection remains less than one for the OIR test ξ_O^W, compared to the distribution of $\sum_{i=1}^{N-k_1-1} \rho_i z_i^2$, where ρ_is are the nonzero eigenvalues of the matrix in Eq.(17). (B) Suppose the model in Eq.(6) is correctly specified with $k_1 = k$. Then the asymptotic probability of rejection is less than the size of the test for the OIR test ξ_O^W, compared to the distribution of $\sum_{i=1}^{N-k-1} \rho_i z_i^2$, where ρ_is are the nonzero eigenvalues of the matrix in Eq.(17).

To understand why the predetermined weighting matrix does not help detect a misspecified model in part (A), note that $\hat{S}_T^W$ is used to calculate the matrix in Eq.(17), even though it is not used as the weighting matrix in Eq.(15). The OIR test ξ_O^W goes to infinity, but the estimated ρ_is, and hence the distribution that ξ_O^W is compared with in making the acceptance/rejection decision, also tend to infinity at the same rate. Therefore, the nonstandard GMM OIR test does not escape the curse of a useless factor. Unlike the usual cases of misspecification without a useless factor, the adoption of a fixed weighting matrix does not solve the problem. In comparing a misspecified model without a useless factor and the same model with an added useless factor, one will still be allured to conclude that the added useless factor saves the model. Note that in part (A), we do not specify the asymptotic probability of rejection because it depends on the weighting matrix W_T. In general, there is no guarantee that the asymptotic probability of rejection is higher than the size of the test. Similar to the standard GMM OIR test based on the second round, it is possible that the nonstandard OIR test accepts a misspecified model which contains a useless factor more likely than the true model. Part (B) says that the nonstandard OIR test, like the standard OIR test, tends to accept a correctly specified model more likely when a useless factor is added to the model.

3. Simulation Results

In this section, we present simulation evidence to evaluate the magnitude of the problems in finite sample. In our simulation experiment, we generate returns on ten assets, corresponding to the popular use of NYSE/AMEX size decile portfolios. The returns are generated from a 2-factor model

$$r_t = \nu + \beta_1 f_{1t} + \beta_2 f_{2t} + \varepsilon_t, \tag{18}$$

where $f_{1t} \sim N(0, \sigma_m^2)$, $f_{2t} \sim N(0,1)$, $\varepsilon_t \sim N(0_N, \Sigma_\varepsilon)$, and f_{1t}, f_{2t} and ε_t are independent of each other and over t. The parameter ν is chosen as the sample mean of actual excess returns on size decile portfolios during the period July 1963 through December 1991. β_1 is set equal to the sample beta of actual returns with respect to the actual market portfolios (the value-weighted NYSE/AMEX portfolio) and σ_m^2 is set equal to the sample variance of the market returns. We set $\beta_2 = \nu - \beta_1 \lambda_1^*$, where $\lambda_1^* = \operatorname{argmin}_{\lambda_1} (\nu - \beta_1 \lambda_1)'(\nu - \beta_1 \lambda_1)$, and $\Sigma_\varepsilon = V - \beta_1 \beta_1' \sigma_m^2 - \beta_2 \beta_2'$, where V is the sample variance matrix of actual excess returns. As long as β_2 is not a zero vector, f_{1t} is a partial set of true factors in the simulation. By design, (f_{1t}, f_{2t}) will be the full set of priced true factors. The choice of Σ_ε makes the theoretical variance of the simulated returns equal to the sample variance of the actual returns. We verify numerically that Σ_ε is positive-definite. A useless factor g_t is generated as an $N(0,1)$ variable independent of f_{1t}, f_{2t} and ε_t.

In the first panel of Table 1, we report the rejection rate of the hypothesis $\gamma = 0$ for the useless factor using the standard GMM Wald test. We test three different models. The first model contains only the useless factor. The table shows that for finite samples, the overrejection rate is not overwhelming for small T, but it is still not negligible. As T increases, the rejection rate increases steadily. The fact that the overrejection problem exacerbates as T increases places GMM in an awkward position, because GMM relies on a large sample size for inferences. The second model contains one of the two true factors together with the useless factor. Compared with the rejection rates of the first model, we can see that the rejection rates become lower when one of the true factors is included. Nevertheless, the rejection rates are still much higher than the level of significance even though only one true factor is missing. In the third model, we include all the true factors as well as the useless factor. The Wald tests on the useless factor actually underreject. In addition, the rejection rate does not increase with T as in the two misspecified models.

Table 1 here

In the second panel of Table 1, we present the corresponding results using nonstandard GMM. In general, the seriousness of the over-rejection problem depends on the choice of the weighting matrix and the degree of misspecification. Our simulation results are based on the weighting matrix suggested by Hansen and Jagannathan (1997).[5] The results show that overrejection still occurs in nonstandard GMM, but the rates are reduced compared with those in standard GMM.

Our asymptotic results suggest that a useless factor also causes problems with the OIR test. Table 2 reports the finite sample simulation results of the OIR tests. In the first panel of Table 2, we report simulation results of the standard GMM OIR test from the second round, using the identity matrix as the initial weighting matrix. The first model includes only the useless factor, the second model includes only one of the two true factors, and the third model includes one of the two true factors and an added useless factor. Since all three models are misspecified, a higher rejection rate implies a more powerful OIR test.

Table 2 here

It is very surprising that the rejection rates are lowest for the model with the useless factor alone. While this model is the most unreasonable one, it is the one that is most likely to pass the second-round OIR test. In some cases (for example, $\alpha = 0.1$), the asymptotic probability of rejecting the model with the useless factor alone is even lower than the size of the test, which means that the useless factor could pass the OIR test more likely than the true model! The second model, while it contains one of the two true factors, is rejected with the highest probability. Consistent with Proposition 3, the rejection rates are reduced for this misspecified model when we add a useless factor, and the degree of reduction is often quite substantial.

[5]This is simply a convenient choice of a fixed weighting matrix, not an exercise in calculating the Hansen-Jagannathan distance, because the moment conditions in Eq.(6) are for excess returns rather than gross returns.

The results in Table 2 highlight three common situations in which OIR test results may be misinterpreted. In the first situation, suppose a researcher estimates the incumbent model $E[r_t(1 - f_{1t}'\lambda_1)] = 0_N$ and rejects the model based on $\xi_O^{(l)}$. Then the researcher adds another factor, finds that it is priced, and finds that the p-value of the OIR test becomes much larger. The common interpretation is that the newly added factor "salvages" the model. However, the analytical results in Propositions 3 and 4, coupled with the simulation results in Table 2, indicate the dangers of concluding that the added factor is a true factor and the augmented model is a true model. It could well be the case that the added factor is a useless factor and the model with the added factor is still misspecified. The GMM OIR test cannot rule out this possibility. In the second situation, suppose a researcher proposes a factor, which happens to be a useless factor g_t, as an alternative to an imperfect model $E[r_t(1 - f_{1t}'\lambda_1)] = 0_N$. By comparing the OIR test p-values of the two models, the model with a useless factor is more likely to win the contest. Finally, consider a situation in which the researcher likes to reduce, for the sake of parsimony, the number of factors from the model $E[r_t(1 - f_{1t}'\lambda_1 - g_t\gamma)] = 0_N$. More often than not, some of the true factors will be dropped rather than the useless factor g_t, and the final model might end up having g_t alone.

On a positive note, Proposition 3 advocates using the standard OIR test statistics from the third or subsequent rounds, instead of the popular one from the second round. The finite-sample simulation results in the second panel of Table 2 convey the same message. The third round OIR test has the same power as the second round test in rejecting misspecified models that do not contain a useless factor (i.e., the second model). However, the third round OIR test is much more powerful than the second round OIR test in rejecting misspecified models that *do* contain a useless factor (i.e., the first and third models). In addition, unlike the second round OIR test, misspecified models do not pass the third round OIR test more often than the true model. In view of these results, we suggest that researchers should perform the OIR test using parameters estimated from the third or subsequent rounds.

The third panel of Table 2 presents the rejection rates of the nonstandard GMM OIR tests for the three misspecified models. For the Wald tests in the second panel of Table 1, the power of tests is reduced because of the predetermined weighting matrix. In contrast, the rejection rates for the nonstandard OIR test are higher in some cases than the rates for the "optimal" weighting matrix. Similar to our findings for standard GMM, the model with just one useless factor passes the nonstandard GMM OIR test much more likely than the model with a partial set of true factors. In addition, adding a useless factor to a misspecified model can substantially reduce the power of the nonstandard OIR test in rejecting a misspecified model, especially when T is small.

4. Conclusion

Like the two-pass methodology, the GMM methodology of estimating stochastic discount factor models suffers in the presence of useless factors. Specifically, the Wald tests tend to overreject the hypothesis of a zero factor premium for the useless factor when the model is misspecified. Increasing the sample size or the number of assets in the model increases the severity of the overrejection problem. The degree of misspecification of the asset pricing model also affects the Wald test of the risk premiums. The higher the degree of misspecification, in terms of the number of true factors missing from the set of true factors, the more likely a useless factor will be mistaken as a priced factor. For a correctly specified model, adding a useless factor does not cause serious problems. The problems created by useless factors are a combination of nonidentifiability and misspecification of the asset pricing model.

A unique contribution of the paper is that we provide an analysis of the asymptotic properties of the GMM OIR tests for models with useless factors. Unlike usual misspecified models, which can be detected with an asymptotic probability of one, we show that OIR tests do not reject a misspecified model that contains a useless factor with an asymptotic probability of one. More surprisingly, the standard GMM OIR test from the second round can actually reject a misspecified model with a useless factor less likely than the size of the test, which renders the specification test completely incapable of detecting misspecified models with useless factors. These findings are not easily anticipated from the results of Kan and Zhang (1999). In short, a useless factor can be mistaken as a priced factor, and it can also cause wrong inferences about the model.

While GMM has many advantages in terms of the robustness of its distributional assumptions, it remains vulnerable to the presence of useless factors when the model is misspecified. Kan and Zhang (1999) suggest several diagnostic procedures for the two-pass methodology in detecting useless factors, including the use of subsample joint tests and goodness-of-fit measures. These procedures can be similarly applied to the stochastic discount factor model with GMM methodology.

Appendix. Proof of propositions

Proof of Proposition 1: (A) Let $D_t = r_t g_t$. Since r_t and g_t are independent, D_t is also a stationary and ergodic martingale difference sequence with finite variance. A central limit theorem (see, for example, Davidson, 1994 p.385) implies that

$$\sqrt{T}\bar{D}_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} D_t \overset{D}{\to} N(0_N, U), \tag{A.1}$$

where $U = E[D_t D_t'] = E[g_t^2]E[r_t r_t'] = E[r_t r_t']$. Eq.(A.1) shows that $\bar{D}_T$ is $O_p(T^{-\frac{1}{2}})$, or has a stochastic order of $T^{-\frac{1}{2}}$ (see Davidson, 1994 p.187). Let $\tilde{B}_1 = (\hat{S}_T^{(l-1)})^{-\frac{1}{2}}\bar{B}_{1T}$, $\tilde{D} = (\hat{S}_T^{(l-1)})^{-\frac{1}{2}}\bar{D}_T$, $\tilde{r} = (\hat{S}_T^{(l-1)})^{-\frac{1}{2}}\bar{r}_T$, and $M_{B_1} = I_N - \tilde{B}_1(\tilde{B}_1'\tilde{B}_1)^{-1}\tilde{B}_1'$. Using the partitioned matrix inverse formula, it can be shown that the GMM estimate of γ is

$$\hat{\gamma}_T^{(l)} = (\tilde{D}'M_{B_1}\tilde{D})^{-1}(\tilde{D}'M_{B_1}\tilde{r}). \tag{A.2}$$

By assumption there does not exist a λ_1 such that $\nu = B_1\lambda_1$, where $B_1 = E[r_t f_{1t}']$, so $M_{B_1}\tilde{r}$ converges to a nonzero vector. As a result, $\hat{\gamma}_T^{(l)} = O_p(T^{\frac{1}{2}})$. The variance matrix of $\sqrt{T}\bar{m}_T$ calculated from the estimates $\hat{\lambda}_{1T}^{(l)}$ and $\hat{\gamma}_T^{(l)}$ is

$$\begin{aligned}
\hat{S}_T^{(l)} &= \frac{1}{T}\sum_{t=1}^{T} r_t r_t'(1 - f_{1t}'\hat{\lambda}_{1T}^{(l)} - g_t\hat{\gamma}_T^{(l)})^2 \\
&= \frac{1}{T}\sum_{t=1}^{T} r_t r_t'\left[(1 - f_{1t}'\hat{\lambda}_{1T}^{(l)})^2 - 2(1 - f_{1t}'\hat{\lambda}_{1T}^{(l)})g_t\hat{\gamma}_T^{(l)} + (g_t\hat{\gamma}_T^{(l)})^2\right].
\end{aligned} \tag{A.3}$$

All the terms are $O_p(1)$, except for the last term which is $O_p(T)$. From the Cramer-Slutsky theorem,

$$\frac{1}{T}\hat{S}_T^{(l)} \overset{LD}{=} \frac{(\hat{\gamma}_T^{(l)})^2}{T}E[r_t r_t' g_t^2] = \frac{(\hat{\gamma}_T^{(l)})^2}{T}U, \tag{A.4}$$

where $\overset{LD}{=}$ means equal in limiting distribution. Therefore, the Wald test, $\xi_{P_a}^{(l)}$ for $l \geq 2$, is

$$\xi_{P_a}^{(l)} = \frac{T(\hat{\gamma}_T^{(l)})^2}{H[(\bar{B}_{1T}, \bar{D}_T)'(\hat{S}_T^{(l)})^{-1}(\bar{B}_{1T}, \bar{D}_T)]^{-1}H'} \overset{LD}{=} T\left[H[(B_1, \bar{D}_T)'U^{-1}(B_1, \bar{D}_T)]^{-1}H'\right]^{-1}, \tag{A.5}$$

where $H = [0_{k_1}', 1]$. Using the partitioned matrix inverse formula, it can be shown that

$$\begin{aligned}
\xi_{P_a}^{(l)} &\overset{LD}{=} \sqrt{T}\bar{D}_T'[U^{-1} - U^{-1}B_1(B_1'U^{-1}B_1)^{-1}B_1'U^{-1}]\sqrt{T}\bar{D}_T \\
&= (\sqrt{T}U^{-\frac{1}{2}}\bar{D}_T)'[I_N - U^{-\frac{1}{2}}B_1(B_1'U^{-1}B_1)^{-1}B_1'U^{-\frac{1}{2}}](\sqrt{T}U^{-\frac{1}{2}}\bar{D}_T) \overset{A}{\sim} \chi^2_{N-k_1},
\end{aligned} \tag{A.6}$$

because $\sqrt{T}U^{-\frac{1}{2}}\bar{D}_T \overset{A}{\sim} N(0_N, I_N)$ and $I_N - U^{-\frac{1}{2}}B_1(B_1'U^{-1}B_1)^{-1}B_1'U^{-\frac{1}{2}}$ is symmetric and idempotent with trace $N - k_1$. The case for $\xi_{P_b}^{(l)}$, $l \geq 3$ is similar. From Eq.(A.4), it is clear that the limit of $\frac{1}{T}\hat{S}_T^{(l)}$ for all l can be different by, at most, a scalar due to different $(\hat{\gamma}_T^{(l)})^2$. On the other hand, from Eq.(A.2) and the definition of M_{B_1}, $\hat{\gamma}_T^{(l)}$ is not affected by a scalar multiplier of $\hat{S}_T^{(l-1)}$. It follows that the limiting distributions of $\frac{1}{\sqrt{T}}\hat{\gamma}_T^{(l)}$ and $\frac{1}{T}\hat{S}_T^{(l)}$ for all $l \geq 2$ are the same. Therefore, although $\xi_{P_b}^{(l)}$ uses $\hat{S}_T^{(l-1)}$ instead of $\hat{S}_T^{(l)}$, it still has the same limiting distribution as $\xi_{P_a}^{(l)}$ when $l \geq 3$.

(B) Following the proof of part (A), let $\tilde{B} = (\hat{S}_T^{(l-1)})^{-\frac{1}{2}}\bar{B}_T$, $\tilde{D} = (\hat{S}_T^{(l-1)})^{-\frac{1}{2}}\bar{D}_T$, $\tilde{r} = (\hat{S}_T^{(l-1)})^{-\frac{1}{2}}\bar{r}_T$, $\tilde{u} = (\hat{S}_T^{(l-1)})^{-\frac{1}{2}}\bar{u}_T$, and $M_B = I_N - \tilde{B}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'$. Then,

$$\hat{\gamma}_T^{(l)} = (\tilde{D}'M_B\tilde{D})^{-1}(\tilde{D}'M_B\tilde{r}). \tag{A.7}$$

$M_B\tilde{r} = M_B\tilde{u}$ is $O_p(T^{-\frac{1}{2}})$ with zero mean and the order of $\hat{S}_T^{(l-1)}$ cancels out, so it follows that $\hat{\gamma}_T^{(l)}$ is $O_p(1)$.[6] The calculated variance matrix of $\sqrt{T}\bar{m}_T$ is

$$\hat{S}_T^{(l)} = \frac{1}{T}\sum_{t=1}^{T} r_t r_t' \left[(1 - f_t'\hat{\lambda}_T^{(l)})^2 - 2(1 - f_t'\hat{\lambda}_T^{(l)})g_t\hat{\gamma}_T^{(l)} + (g_t\hat{\gamma}_T^{(l)})^2\right]. \tag{A.8}$$

The first term of $\hat{S}_T^{(l)}$ converges to S because $\hat{\lambda}_T^{(l)}$ converges to λ^*. The middle term converges to a zero matrix because g_t is independent of f_t and r_t with $E[g_t] = 0$. The last term is asymptotically equivalent to $(\hat{\gamma}_T^{(l)})^2U$, as we show in the proof of part (A). It can be shown that the second moment matrix S for the true model can be expressed as

$$S = (1 + \lambda^{*\prime}\lambda^*)V + BCB', \tag{A.9}$$

for some $k \times k$ matrix C. Therefore, for $l \geq 2$,

$$\begin{aligned}
\hat{S}_T^{(l-1)} &\overset{LD}{=} S + (\hat{\gamma}_T^{(l-1)})^2U = (1 + \lambda^{*\prime}\lambda^*)V + BCB' + (\hat{\gamma}_T^{(l-1)})^2(V + \nu\nu') \\
&= (1 + \lambda^{*\prime}\lambda^* + (\hat{\gamma}_T^{(l-1)})^2)V + B[C + (\hat{\gamma}_T^{(l-1)})^2\lambda^*\lambda^{*\prime}]B' \\
&\equiv \tau V + BC_1B',
\end{aligned} \tag{A.10}$$

where $\tau = 1 + \lambda^{*\prime}\lambda^* + (\hat{\gamma}_T^{(l-1)})^2$ and $C_1 = C + (\hat{\gamma}_T^{(l-1)})^2\lambda^*\lambda^{*\prime}$. Algebra manipulation yields

$$(\hat{S}_T^{(l-1)})^{-\frac{1}{2}}M_B(\hat{S}_T^{(l-1)})^{-\frac{1}{2}} \overset{LD}{=} (\hat{S}_T^{(l-1)})^{-1} - (\hat{S}_T^{(l-1)})^{-1}B[B'(\hat{S}_T^{(l-1)})^{-1}B]^{-1}B'(\hat{S}_T^{(l-1)})^{-1}$$

[6] It can be proved that the limiting distribution of $\hat{\gamma}_T^{(l)}$ for $l \geq 2$ is $t_{N-k}/(N-k)^{\frac{1}{2}}$ where t_{N-k} is a central-t distribution with $N - k$ degrees of freedom. The proof is available upon request.

$$
\begin{aligned}
&= \frac{1}{\tau}[V^{-1} - V^{-1}B(B'V^{-1}B)^{-1}B'V^{-1}] \\
&\equiv \frac{1}{\tau}V^{-\frac{1}{2}}MV^{-\frac{1}{2}}, \qquad (A.11)
\end{aligned}
$$

where $M = I_N - V^{-\frac{1}{2}}B(B'V^{-1}B)^{-1}B'V^{-\frac{1}{2}}$. From the partitioned matrix inverse formula, the Wald test $\xi_{P_b}^{(l)}$ for $l \geq 2$ is

$$
\xi_{P_b}^{(l)} = T(\hat{\gamma}_T^{(l)})^2(\tilde{D}'M_B\tilde{D}) = \frac{T(\tilde{D}'M_B\tilde{r})^2}{\tilde{D}'M_B\tilde{D}}. \qquad (A.12)
$$

Using Eq.(A.11) and noting $\tau = 1 + \lambda^{*\prime}\lambda^* + (\hat{\gamma}_T^{(l-1)})^2 > 1$, we have

$$
\xi_{P_b}^{(l)} \overset{LD}{=} \frac{T(\bar{D}'V^{-\frac{1}{2}}MV^{-\frac{1}{2}}\bar{r}_T)^2}{\tau(\bar{D}'V^{-\frac{1}{2}}MV^{-\frac{1}{2}}\bar{D})} < \frac{T(\bar{D}'V^{-\frac{1}{2}}MV^{-\frac{1}{2}}\bar{r}_T)^2}{\bar{D}'V^{-\frac{1}{2}}MV^{-\frac{1}{2}}\bar{D}} \equiv (x'y)^2, \qquad (A.13)
$$

where $x = P'V^{-\frac{1}{2}}\bar{D}/(\bar{D}'V^{-\frac{1}{2}}MV^{-\frac{1}{2}}\bar{D})^{\frac{1}{2}}$, $y = \sqrt{T}P'V^{-\frac{1}{2}}\bar{r}_T$, and P is an $N \times (N-k)$ orthonormal matrix such that $PP' = M$. It is easy to verify that $P'B = O_{N-k,k}$ and hence $y \overset{A}{\sim} N(0_{N-k}, I_{N-k})$. Since y is asymptotically independent of x, $x'y$ is asymptotically conditional normal with conditional mean $x'E[y|x] = x'E[y] = 0$ and conditional variance $x'\text{Var}[y|x]x = x'\text{Var}[y]x = x'x = 1$, independent of x. Therefore, unconditionally, $x'y \overset{A}{\sim} N(0,1)$, and $(x'y)^2 \overset{A}{\sim} \chi_1^2$. The proof for $\xi_{P_a}^{(l)}$, $l \geq 2$ is similar. *Q.E.D.*

Proof of Proposition 2: (A) Let $X \equiv (\tilde{B}_1, \tilde{D}) \equiv W^{\frac{1}{2}}(B_1, \bar{D}_T)$ and $\tilde{U} \equiv W^{\frac{1}{2}}UW^{\frac{1}{2}}$. Since $\frac{1}{T}\hat{S}_T^W \overset{LD}{=} \frac{1}{T}(\hat{\gamma}_T^W)^2U$, which can be shown as in the proof of Proposition 1, the nonstandard Wald test of $\gamma = 0$ can be written as

$$
\xi_P^W \overset{LD}{=} T\left[H\left((X'X)(X'\tilde{U}X)^{-1}(X'X)\right)^{-1}H'\right]^{-1}. \qquad (A.14)
$$

From the partitioned matrix inverse formula, it can be shown that

$$
\xi_P^W \overset{LD}{=} \frac{T(\tilde{D}'M_{B_1}\tilde{D})^2}{\tilde{D}'M_{B_1}\tilde{U}M_{B_1}\tilde{D}}, \qquad (A.15)
$$

where $M_{B_1} = I_N - \tilde{B}_1(\tilde{B}_1'\tilde{B}_1)^{-1}\tilde{B}_1'$. Let $M_{B_1}\tilde{U}M_{B_1} = P\Theta P'$ be the spectral decomposition of $M_{B_1}\tilde{U}M_{B_1}$, where $\Theta = \text{Diag}(\theta_1, \cdots, \theta_{N-k_1})$ is the matrix of eigenvalues listed in descending order, and let $y = \sqrt{T}P'\tilde{D} \overset{A}{\sim} N(0_{N-k_1}, \Theta)$. Since $M_{B_1} = PP'$ and

$$
y'\Theta y = \sum_{i=1}^{N-k_1}\theta_i y_i^2 \leq \theta_1\sum_{i=1}^{N-k_1}y_i^2 = \theta_1(y'y), \qquad (A.16)
$$

it follows that,

$$\xi_P^W = \frac{T(\tilde{D}'M_{B_1}\tilde{D})^2}{\tilde{D}'M_{B_1}\tilde{U}M_{B_1}\tilde{D}} = \frac{(y'y)^2}{y'\Theta y} > \frac{y_1^2}{\theta_1} \equiv \xi_1 \overset{A}{\sim} \chi_1^2. \tag{A.17}$$

Using the Cauchy-Schwarz inequality, we can establish the upper bound of ξ_P^W as

$$\xi_P^W = (y'y)^2(y'\Theta y)^{-1} \leq y'\Theta^{-1}y \equiv \xi_{N-k_1} \overset{A}{\sim} \chi^2_{N-k_1}. \tag{A.18}$$

(B) Let $X \equiv (\tilde{B}, \tilde{D}) \equiv W^{\frac{1}{2}}(B, \bar{D}_T)$, $\tilde{r} = W^{\frac{1}{2}}\bar{r}_T$, and $M_B = I_N - \tilde{B}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'$. The estimate of γ is $\hat{\gamma}_T^W \overset{LD}{=} (\tilde{D}'M_B\tilde{r})/(\tilde{D}'M_B\tilde{D})$. Following the proof in Proposition 1 part (B), we can show that $\hat{\gamma}_T^W$ converges in distribution to a random variable with zero mean. Let $\hat{S}_T^W$ be the estimate of S with $\hat{S}_T^W \overset{LD}{=} S + (\hat{\gamma}_T^W)^2 U$. Using Eq.(A.9), we have $\hat{S}_T^W \overset{LD}{=} \tau V + BC_1B'$ where $\tau = 1 + \lambda^{*\prime}\lambda^* + (\hat{\gamma}_T^W)^2$ and $C_1 = C + (\hat{\gamma}_T^W)^2\lambda^*\lambda^{*\prime}$. From the partitioned matrix inverse formula, the nonstandard Wald test is

$$\begin{aligned}
\xi_P^W &\overset{LD}{=} T(\hat{\gamma}_T^W)^2\left[H\left((X'X)(X'W^{\frac{1}{2}}\hat{S}_T^W W^{\frac{1}{2}}X)^{-1}(X'X)\right)^{-1}H'\right]^{-1} \\
&= T(\hat{\gamma}_T^W)^2\frac{(\tilde{D}'M_B\tilde{D})^2}{\tilde{D}'M_BW^{\frac{1}{2}}\hat{S}_T^W W^{\frac{1}{2}}M_B\tilde{D}} \overset{LD}{=} \frac{T(\tilde{D}'M_B\tilde{r})^2}{\tau(\tilde{D}'M_BW^{\frac{1}{2}}VW^{\frac{1}{2}}M_B\tilde{D})}
\end{aligned} \tag{A.19}$$

because $M_BW^{\frac{1}{2}}BC_1B'W^{\frac{1}{2}}M_B = O_{N,N}$ as $M_BW^{\frac{1}{2}}B = M_B\tilde{B} = O_{N,k}$. Since $\tau > 1$, we have

$$\xi_P^W \overset{LD}{<} \frac{T(\tilde{D}'M_B\tilde{r})^2}{\tilde{D}'M_BW^{\frac{1}{2}}VW^{\frac{1}{2}}M_B\tilde{D}} \equiv (x'y)^2, \tag{A.20}$$

where $y = \sqrt{T}\Theta^{-\frac{1}{2}}P'\tilde{r}$ and $x = \Theta^{\frac{1}{2}}P'\tilde{D}/(\tilde{D}'M_BW^{\frac{1}{2}}VW^{\frac{1}{2}}M_B\tilde{D})^{\frac{1}{2}}$, with $P\Theta P'$ being the spectral decomposition of $M_BW^{\frac{1}{2}}VW^{\frac{1}{2}}M_B$ and $\Theta = \text{Diag}(\theta_1, \cdots, \theta_{N-k})$ being the matrix of nonzero eigenvalues listed in descending order. It is easy to verify that $PP' = M_B$ and $P'\tilde{B} = O_{N-k,k}$. Therefore, using the same argument as in the proof of Proposition 1 part (B), $(x'y)^2$ is asymptotically a χ_1^2 variable. This completes the proof. *Q.E.D.*

Proof of Proposition 3: (A) $\bar{D}_T$ is $O_p(T^{-\frac{1}{2}})$ and we prove in Proposition 1 part (A) that $\hat{\gamma}_T^{(l)} = O_p(T^{\frac{1}{2}})$. Therefore, $\bar{r}_T - \bar{B}_{1T}\hat{\lambda}_{1T}^{(l)} - \bar{D}_T\hat{\gamma}_T^{(l)}$ is $O_p(1)$ and $\hat{S}_T^{(l-1)}$ is $O_p(T)$. It follows that for $l \geq 2$,

$$\xi_O^{(l)} = T\left[\bar{r}_T - \bar{B}_{1T}\hat{\lambda}_{1T}^{(l)} - \bar{D}_T\hat{\gamma}_T^{(l)}\right]'\left(\hat{S}_T^{(l-1)}\right)^{-1}\left[\bar{r}_T - \bar{B}_{1T}\hat{\lambda}_{1T}^{(l)} - \bar{D}_T\hat{\gamma}_T^{(l)}\right] = O_p(1). \tag{A.21}$$

Substituting from Eq.(9) into Eq.(A.21) and rearranging gives

$$\xi_O^{(l)} \overset{LD}{=} T\tilde{r}'\left[M_{B_1} - M_{B_1}\tilde{D}(\tilde{D}'M_{B_1}\tilde{D})^{-1}\tilde{D}'M_{B_1}\right]\tilde{r}. \tag{A.22}$$

Eq.(A.22) implies that, conditioned on $\hat{S}_T^{(l-1)}/T$, $\xi_O^{(l)}$ is a noncentral quadratic form of the returns whose support is $(0,\infty)$. So is the support of the unconditional distribution of $\xi_O^{(l)}$. Therefore, for any percentile value, c, of the $\chi^2_{N-k_1-1}$ distribution, $P[\xi_O^{(l)} > c]$ remains less than one.

To derive the asymptotic distribution of $\xi_O^{(l)}$ for $l \geq 3$, we define $Z = \sqrt{T}P'U^{-\frac{1}{2}}\bar{D}_T \overset{A}{\sim} N(0_{N-k_1}, I_{N-k_1})$, $\eta = P'U^{-\frac{1}{2}}\nu$ where PP' is the spectral decomposition of $M_1 = I_N - U^{-\frac{1}{2}}B_1(B_1'U^{-1}B_1)^{-1}B_1'U^{-\frac{1}{2}}$. Using Eq.(A.4), we have for $l \geq 2$,

$$\frac{\hat{\gamma}_T^{(l)}}{\sqrt{T}} = \frac{\bar{D}_T'(\hat{S}_T^{(l-1)})^{-\frac{1}{2}}M_{B_1}(\hat{S}_T^{(l-1)})^{-\frac{1}{2}}\bar{r}_T}{\sqrt{T}(\bar{D}_T'(\hat{S}_T^{(l-1)})^{-\frac{1}{2}}M_{B_1}(\hat{S}_T^{(l-1)})^{-\frac{1}{2}}\bar{D}_T)} \overset{LD}{=} \frac{\bar{D}_T'U^{-\frac{1}{2}}M_1U^{-\frac{1}{2}}\bar{r}_T}{\sqrt{T}(\bar{D}_T'U^{-\frac{1}{2}}M_1U^{-\frac{1}{2}}\bar{D}_T)} \overset{LD}{=} \frac{Z'\eta}{Z'Z}. \qquad (A.23)$$

For $l \geq 3$,

$$\begin{aligned}
\xi_O^{(l)} &= T\left[\bar{r}_T - \bar{D}_T\hat{\gamma}_T^{(l)}\right]'(\hat{S}_T^{(l-1)})^{-\frac{1}{2}}M_{B_1}(\hat{S}_T^{(l-1)})^{-\frac{1}{2}}\left[\bar{r}_T - \bar{D}_T\hat{\gamma}_T^{(l)}\right] \\
&\overset{LD}{=} \frac{T}{(\hat{\gamma}_T^{(l-1)})^2}\left[\bar{r}_T - \bar{D}_T\hat{\gamma}_T^{(l)}\right]'U^{-\frac{1}{2}}M_1U^{-\frac{1}{2}}\left[\bar{r}_T - \bar{D}_T\hat{\gamma}_T^{(l)}\right] \\
&\overset{LD}{=} \frac{(Z'Z)^2}{(Z'\eta)^2}\left[\eta'\eta - \frac{(\eta'Z)^2}{Z'Z}\right] \\
&= \frac{(Z'Z)(\eta'\eta)}{(Z'\eta)^2}Z'\left[I_{N-k_1} - \frac{\eta\eta'}{\eta'\eta}\right]Z = \left(\frac{\xi_1 + \xi_{N-k_1-1}}{\xi_1}\right)\xi_{N-k_1-1}, \qquad (A.24)
\end{aligned}$$

by writing $\xi_1 = \frac{Z'\eta\eta'Z}{\eta'\eta} \overset{A}{\sim} \chi^2_1$ and $\xi_{N-k_1-1} = Z'\left[I_{N-k_1} - \frac{\eta\eta'}{\eta'\eta}\right]Z \overset{A}{\sim} \chi^2_{N-k_1-1}$. It is easy to verify that ξ_1 and ξ_{N-k_1-1} are independent of each other, and $\xi_1 + \xi_{N-k_1-1} = Z'Z$.

Using a similar proof, the asymptotic distribution of $\xi_O^{(2)}$ is given by

$$\xi_O^{(2)} \overset{LD}{=} \frac{(Z'\Theta Z)^2}{(Z'\Theta\eta)^2}\left[\eta'\eta - \frac{(\eta'Z)^2}{Z'Z}\right], \qquad (A.25)$$

where Θ is defined in the proof of Proposition 2 (A) with W being the limit of the initial weighting matrix. Note that there is no guarantee that $\xi_O^{(2)}$ stochastically dominates $\chi^2_{N-k_1-1}$. To give a counterexample, take $N - k_1 = 2$, $\eta = [1,\ 0]'$ and $\Theta = \mathrm{Diag}(\theta_1,\ 1)$. We have

$$\xi_O^{(2)} \overset{LD}{=} \frac{(Z'\Theta Z)^2}{(Z'\Theta\eta)^2}\left[\eta'\eta - \frac{(\eta'Z)^2}{Z'Z}\right] = \frac{(\theta_1 Z_1^2 + Z_2^2)^2}{\theta_1^2 Z_1^2}\left(\frac{Z_2^2}{Z_1^2 + Z_2^2}\right). \qquad (A.26)$$

When $\theta_1 \to \infty$,

$$\xi_O^{(2)} \overset{LD}{=} Z_1^2\left(\frac{Z_2^2}{Z_1^2 + Z_2^2}\right) < Z_1^2 \overset{A}{\sim} \chi^2_1. \qquad (A.27)$$

Therefore, when θ_1 is large enough, it is possible that $\xi_O^{(2)}$ will accept the misspecified model with a useless factor more often than the true model.

(B) Let $X = \sqrt{T}P'V^{-\frac{1}{2}}\bar{D}_T \overset{A}{\sim} N(0_{N-k}, I_{N-k})$ and $Y = \sqrt{T}P'V^{-\frac{1}{2}}\bar{r}_T \overset{A}{\sim} N(0_{N-k}, I_{N-k})$, where PP' is the spectral decomposition of $M = I_N - V^{-\frac{1}{2}}B(B'V^{-1}B)^{-1}B'V^{-\frac{1}{2}}$. For $l \geq 2$, using Eq.(A.7) and Eq.(A.11), we have

$$\hat{\gamma}_T^{(l)} \overset{LD}{=} \frac{\bar{D}_T'V^{-\frac{1}{2}}MV^{-\frac{1}{2}}\bar{r}_T}{\bar{D}_T'V^{-\frac{1}{2}}MV^{-\frac{1}{2}}\bar{D}_T} = \frac{X'Y}{X'X}. \tag{A.28}$$

For $l \geq 2$, using Eq.(A.11),

$$\begin{aligned}
\xi_O^{(l)} &= T\left[\bar{r}_T - \bar{D}_T\hat{\gamma}_T^{(l)}\right]'(\hat{S}_T^{(l-1)})^{-\frac{1}{2}}M_B(\hat{S}_T^{(l-1)})^{-\frac{1}{2}}\left[\bar{r}_T - \bar{D}_T\hat{\gamma}_T^{(l)}\right] \\
&\overset{LD}{=} T\left[\bar{r}_T - \bar{D}_T\hat{\gamma}_T^{(l)}\right]'\frac{V^{-\frac{1}{2}}MV^{-\frac{1}{2}}}{\tau}\left[\bar{r}_T - \bar{D}_T\hat{\gamma}_T^{(l)}\right] \\
&< (Y - X\hat{\gamma}_T^{(l)})'(Y - X\hat{\gamma}_T^{(l)}) = Y'\left[I_{N-k} - \frac{XX'}{X'X}\right]Y \overset{A}{\sim} \chi^2_{N-k-1},
\end{aligned} \tag{A.29}$$

where $\tau = 1 + \lambda^{*\prime}\lambda^* + (\hat{\gamma}_T^{(l-1)})^2 > 1$. The χ^2_{N-k-1} distribution is obtained by noting that X and Y are independent, so conditioned on any X, the distribution is χ^2_{N-k-1}, which is independent of X. This completes the proof. *Q.E.D.*

Proof of Proposition 4: (A) The result that ξ_O^W tends to infinity is obvious. To show that $\rho_i \to \infty$ as $T \to \infty$, let $\eta_i(P)$ be the ith largest eigenvalue of a matrix P. It is well known that for two symmetric $N \times N$ matrices, P and Q, $\eta_i(PQ) \geq \eta_i(P)\eta_N(Q)$. Similar to the proof of Proposition 1, it can be shown that $\hat{\gamma}_T^W$ is $O_p(T^{\frac{1}{2}})$ and $\hat{S}_T^W/T \overset{LD}{=} (\hat{\gamma}_T^W)^2U/T$. Therefore, for $i = 1, \cdots, N - k_1 - 1$,

$$\begin{aligned}
\frac{\rho_i}{T} &= \frac{1}{T}\eta_i\left((W_T^{\frac{1}{2}}\hat{S}_T^W W_T^{\frac{1}{2}})\left[I_N - W_T^{\frac{1}{2}}(\bar{B}_{1T}, \bar{D}_T)[(\bar{B}_{1T}, \bar{D}_T)'W_T(\bar{B}_{1T}, \bar{D}_T)]^{-1}(\bar{B}_{1T}, \bar{D}_T)'W_T^{\frac{1}{2}}\right]\right) \\
&\geq \frac{1}{T}\eta_N(W_T^{\frac{1}{2}}\hat{S}_T^W W_T^{\frac{1}{2}}) \overset{LD}{=} \frac{(\hat{\gamma}_T^W)^2}{T}\eta_N(W_T^{\frac{1}{2}}UW_T^{\frac{1}{2}}) = O_p(1).
\end{aligned} \tag{A.30}$$

The inequality follows from the fact that the first $N - k_1 - 1$ nonzero eigenvalues of a symmetric and idempotent matrix are one. The last equality follows because $(\hat{\gamma}_T^W)^2/T$ is $O_p(1)$ and $\eta_N(W_T^{\frac{1}{2}}UW_T^{\frac{1}{2}}) \to \eta_N(W^{\frac{1}{2}}UW^{\frac{1}{2}}) = \eta_N(WU)$ which is a positive constant.

Since the rejection decision is based on the percentile value of $\sum_{i=1}^{N-k_1-1}\rho_i z_i^2$ where z_is are independently drawn $N(0,1)$, the p-value of rejection is

$$P\left[\xi_O^W > \sum_{i=1}^{N-k_1-1} \rho_i z_i^2\right] \leq P\left[\frac{\xi_O^W}{\rho_{N-k_1-1}} > \sum_{i=1}^{N-k_1-1} z_i^2\right]. \tag{A.31}$$

Let $x = \sum_{i=1}^{N-k_1-1} z_i^2$ be the $\chi^2_{N-k_1-1}$ variable. It suffices to show that the probability on the right hand side is less than one, or $P[\xi_O^W/\rho_{N-k_1-1} < x] > 0$. This is guaranteed because x has a support on $(0,\infty)$ and is independent of $\xi_O^W/\rho_{N-k_1-1} = O_p(1)$.

(B) Denote $M_B = I_N - W^{\frac{1}{2}}B(B'WB)^{-1}B'W^{\frac{1}{2}}$, $Z = \sqrt{T}P'W^{\frac{1}{2}}\bar{D}_T \overset{A}{\sim} N(0_{N-k},\Theta)$, $Y_1 = \sqrt{T}P'W^{\frac{1}{2}}\bar{r}_T \overset{A}{\sim} N(0_{N-k},\Theta)$ where P and Θ are defined in the proof of Proposition 2 part (B). We have

$$\hat{\gamma}_T^W \overset{LD}{=} \frac{\bar{D}_T'W^{\frac{1}{2}}M_B W^{\frac{1}{2}}\bar{r}_T}{\bar{D}_T'W^{\frac{1}{2}}M_B W^{\frac{1}{2}}\bar{D}_T} \overset{LD}{=} \frac{Z'Y_1}{Z'Z}. \tag{A.32}$$

Define $M = I_{N-k} - Z(Z'Z)^{-1}Z'$. The nonstandard OIR test statistic can be written as

$$\xi_O^W \overset{LD}{=} T(\bar{r}_T - \bar{D}_T\hat{\gamma}_T^W)'W^{\frac{1}{2}}M_B W^{\frac{1}{2}}(\bar{r}_T - \bar{D}_T\hat{\gamma}_T^W) \overset{LD}{=} Y_1'MY_1. \tag{A.33}$$

Define $X = W^{\frac{1}{2}}(B,\ \bar{D}_T)$. The OIR test, ξ_O^W, is compared with the distribution of

$$\begin{aligned}
\xi_C &= y'(\hat{S}_T^W)^{\frac{1}{2}}[W_T - W_T(\bar{B}_T,\bar{D}_T)[(\bar{B}_T,\bar{D}_T)'W_T(\bar{B}_T,\bar{D}_T)]^{-1}(\bar{B}_T,\bar{D}_T)'W_T](\hat{S}_T^W)^{\frac{1}{2}}y \\
&\overset{LD}{=} y'(\hat{S}_T^W)^{\frac{1}{2}}W^{\frac{1}{2}}\left[I_N - X(X'X)^{-1}X'\right]W^{\frac{1}{2}}(\hat{S}_T^W)^{\frac{1}{2}}y \\
&= y'(\hat{S}_T^W)^{\frac{1}{2}}W^{\frac{1}{2}}\left[M_B - M_B W^{\frac{1}{2}}\bar{D}_T(\bar{D}_T'W^{\frac{1}{2}}M_B W^{\frac{1}{2}}\bar{D}_T)^{-1}\bar{D}_T'W^{\frac{1}{2}}M_B\right]W^{\frac{1}{2}}(\hat{S}_T^W)^{\frac{1}{2}}y \\
&= y'(\hat{S}_T^W)^{\frac{1}{2}}W^{\frac{1}{2}}P\left[I_{N-k} - P'W^{\frac{1}{2}}\bar{D}_T(\bar{D}_T'W^{\frac{1}{2}}M_B W^{\frac{1}{2}}\bar{D}_T)^{-1}\bar{D}_T'W^{\frac{1}{2}}P\right]P'W^{\frac{1}{2}}(\hat{S}_T^W)^{\frac{1}{2}}y,
\end{aligned} \tag{A.34}$$

where $y \sim N(0_N, I_N)$. Let $Y_2 \equiv P'W^{\frac{1}{2}}(\hat{S}_T^W)^{\frac{1}{2}}y \overset{A}{\sim} N(0_{N-k},\tau\Theta)$, where $\tau = 1+\lambda^{*\prime}\lambda^{*}+(\hat{\gamma}_T^W)^2 > 1$. Then,

$$\xi_C \overset{LD}{=} Y_2'MY_2 \overset{LD}{=} \tau Y_1'MY_1 > Y_1'MY_1 = \xi_O^W. \tag{A.35}$$

Therefore, the asymptotic probability of rejection of the nonstandard OIR test is less than the size of the test. This completes the proof. *Q.E.D.*

References

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

Davidson, J., 1994. Stochastic Limit Theory: An Introduction to Econometricians, Oxford University Press, New York.

Fama, E. F., MacBeth, J. D., 1973. Risk, return and equilibrium: empirical tests. Journal of Political Economy 81, 607–636.

Ferson, W., Foerster, S., 1994. Finite sample properties of the generalized method of moments in tests of conditional asset pricing models. Journal of Financial Economics 36, 29–55.

Hansen, L. P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hansen, L. P., Jagannathan, R., 1997. Assessing specification errors in stochastic discount factor models. Journal of Finance 52, 557–590.

Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. Journal of Finance 51, 3–53.

Kan, R., Zhang, C., 1999. Two-pass tests of asset pricing models with useless factors. Journal of Finance 54, 203–235.

Staiger, D., Stock, J. H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.

Stock, J. H., Wright, J., 1997. GMM with weak identification, Unpublished working paper, Harvard University, Cambridge, MA.

Zhou, G., 1994. Analytical GMM tests: asset pricing with time-varying risk premiums. Review of Financial Studies 7, 687–710.

Table 1
Rejection Rates of the Standard and Nonstandard GMM Parametric Tests of $\gamma = 0$

This table presents the probability of rejecting the hypothesis $\gamma = 0$ in three different models where γ is the risk premium of a useless factor. A finite T is the number of time-series observations, and the rejection rates for a finite T are based on simulation. For $T = \infty$, the rejection rates are obtained from theoretical asymptotic distributions. In the simulation, the excess returns on the N assets, r_t, are generated by $r_t = \nu + \beta_1 f_{1t} + \beta_2 f_{2t} + \varepsilon_t$, where $f_{1t} \sim N(0, \sigma_m^2)$, $f_{2t} \sim N(0, 1)$, $\varepsilon_t \sim N(0_N, \Sigma_\varepsilon)$, and f_{1t}, f_{2t} and ε_t are independent of each other and across t. The parameter ν is set equal to the sample mean of actual excess returns on 10 size portfolios from NYSE and AMEX. β_1 is set equal to the sample betas of the 10 size portfolios with respect to the value-weighted NYSE/AMEX market returns. σ_m^2 is the sample variance of the market returns. β_2 is determined by $\beta_2 = \nu - \beta_1 \lambda_1^*$ where $\lambda_1^* = \text{argmin}_{\lambda_1} (\nu - \beta_1 \lambda_1)'(\nu - \beta_1 \lambda_1)$, and $\Sigma_\varepsilon = V - \beta_1 \beta_1' \sigma_m^2 - \beta_2 \beta_2'$ where V is the sample variance matrix of the excess returns. The useless factor g_t is $N(0, 1)$, independently and identically distributed across t, and is independent of f_{1t}, f_{2t} and ε_t. For standard GMM, the initial weighting matrix is the identity matrix and the rejection rates reported in the table are based on the Wald test using the estimated γ of the second round ($l = 2$), with the variance estimated at the end of the second round. For nonstandard GMM, the weighting matrix W_T is the inverse of the sample version of $E[(1 + r_f + r_t)(1 + r_f + r_t)']$ where r_f is the average T-bill rate. $\chi_1^2(1 - \alpha)$ is the $100(1 - \alpha)$ percentile of the χ_1^2 distribution. The simulation has 10000 replications.

Panel A: Rejection rates of $\gamma = 0$ of standard GMM, $P[\xi_{P_a}^{(2)} > \chi_1^2(1 - \alpha)]$

		Significance level α		
Model	T	0.10	0.05	0.01
$E[r_t(1 - g_t\gamma)] = 0_N$	250	0.226	0.125	0.021
	500	0.312	0.183	0.038
	1000	0.434	0.289	0.079
	∞	0.988	0.954	0.759
$E[r_t(1 - f_{1t}\lambda_1 - g_t\gamma)] = 0_N$	250	0.209	0.107	0.018
	500	0.288	0.164	0.031
	1000	0.414	0.270	0.066
	∞	0.975	0.922	0.675
$E[r_t(1 - f_{1t}\lambda_1 - f_{2t}\lambda_2 - g_t\gamma)] = 0_N$	250	0.082	0.028	0.002
	500	0.060	0.020	0.001
	1000	0.051	0.014	0.000
	∞	0.004	0.000	0.000

Panel B: Rejection rates of $\gamma = 0$ of nonstandard GMM, $P[\xi_P^W > \chi_1^2(1 - \alpha)]$

		Significance level α		
Model	T	0.10	0.05	0.01
$E[r_t(1 - g_t\gamma)] = 0_N$	250	0.187	0.092	0.011
	500	0.282	0.154	0.024
	1000	0.400	0.251	0.060
	∞	0.975	0.922	0.676
$E[r_t(1 - f_{1t}\lambda_1 - g_t\gamma)] = 0_N$	250	0.184	0.087	0.010
	500	0.276	0.145	0.021
	1000	0.391	0.242	0.052
	∞	0.952	0.872	0.578
$E[r_t(1 - f_{1t}\lambda_1 - f_{2t}\lambda_2 - g_t\gamma)] = 0_N$	250	0.058	0.016	0.001
	500	0.047	0.014	0.000
	1000	0.043	0.011	0.000
	∞	0.003	0.000	0.000

Table 2
Rejection Rates of the Standard and Nonstandard GMM Over-identifying Restriction Tests

This table presents the probability of rejecting the standard and nonstandard GMM over-identifying restriction test in three different models. A finite T is the number of time-series observations, the rejection rates for a finite T are based on simulation. For $T = \infty$, the rejection rates are obtained from theoretical asymptotic distributions. In the simulation, the excess returns on the N assets, r_t, are generated by $r_t = \nu + \beta_1 f_{1t} + \beta_2 f_{2t} + \varepsilon_t$, where $f_{1t} \sim N(0, \sigma_m^2)$, $f_{2t} \sim N(0, 1)$, $\varepsilon_t \sim N(0_N, \Sigma_\varepsilon)$, and f_{1t}, f_{2t} and ε_t are independent of each other and across t. The parameter ν is set equal to the sample mean of actual excess returns on 10 size portfolios of NYSE and AMEX. β_1 is set equal to the sample betas of the 10 size portfolios with respect to the value-weighted NYSE/AMEX market returns. σ_m^2 is the sample variance of the market returns. β_2 is determined by $\beta_2 = \nu - \beta_1 \lambda_1^*$, where $\lambda_1^* = \mathrm{argmin}_{\lambda_1} (\nu - \beta_1 \lambda_1)'(\nu - \beta_1 \lambda_1)$, and $\Sigma_\varepsilon = V - \beta_1 \beta_1' \sigma_m^2 - \beta_2 \beta_2'$, where V is the sample variance matrix of the excess returns. The useless factor g_t is $N(0, 1)$, independently and identically distributed across t, and is independent of f_{1t}, f_{2t} and ε_t. For standard GMM, the initial weighting matrix is the identity matrix and the rejection rates reported in the first two panels of the table are based on the over-identifying restriction test of the second round ($l = 2$) and the third round ($l = 3$), respectively. $\chi_m^2(1 - \alpha)$ is the $100(1 - \alpha)$ percentile of the χ_m^2 distribution. For nonstandard GMM, the weighting matrix W_T is the inverse of the sample version of $E[(1 + r_f + r_t)(1 + r_f + r_t)']$ where r_f is the average T-bill rate. c is the $100(1 - \alpha)$ percentile of the distribution $\sum_{i=1}^m \rho_i z_i^2$, where z_is are independent standard normal variates. The ρ_is are the nonzero eigenvalues of $\hat{S}_T^W [W_T - W_T X (X' W_T X)^{-1} X' W_T]$, where X is the matrix of estimated betas and $\hat{S}_T^W$ is the estimated variance matrix of the sample moments. The simulation has 10000 replications.

Panel A: Rejection rates of the second round standard GMM OIR test, $P[\xi_O^{(2)} > \chi_m^2(1 - \alpha)]$

			Significance level α		
Model	m =DF	T	0.10	0.05	0.01
$E[r_t(1 - g_t\gamma)] = 0_N$	9	250	0.111	0.074	0.029
	9	500	0.129	0.099	0.057
	9	1000	0.113	0.091	0.059
	9	∞	0.094	0.074	0.049
$E[r_t(1 - f_{1t}\lambda_1)] = 0_N$	9	250	0.715	0.586	0.328
	9	500	0.962	0.926	0.797
	9	1000	1.000	0.999	0.995
	9	∞	1.000	1.000	1.000
$E[r_t(1 - f_{1t}\lambda_1 - g_t\gamma)] = 0_N$	8	250	0.243	0.169	0.066
	8	500	0.380	0.301	0.178
	8	1000	0.431	0.367	0.262
	8	∞	0.404	0.340	0.247

Table 2 continued
Rejection Rates of the Standard and Nonstandard GMM Over-identifying Restriction Tests

Panel B: Rejection rates of the third round standard GMM OIR test, $P[\xi_O^{(3)} > \chi_m^2(1-\alpha)]$

Model	m =DF	T	Significance level α		
			0.10	0.05	0.01
$E[r_t(1-g_t\gamma)] = 0_N$	9	250	0.567	0.444	0.224
	9	500	0.805	0.737	0.575
	9	1000	0.889	0.859	0.796
	9	∞	0.944	0.931	0.902
$E[r_t(1-f_{1t}\lambda_1)] = 0_N$	9	250	0.718	0.590	0.332
	9	500	0.963	0.927	0.800
	9	1000	1.000	1.000	0.996
	9	∞	1.000	1.000	1.000
$E[r_t(1-f_{1t}\lambda_1-g_t\gamma)] = 0_N$	8	250	0.517	0.395	0.195
	8	500	0.757	0.683	0.499
	8	1000	0.864	0.829	0.743
	8	∞	0.926	0.910	0.875

Panel C: Rejection rates of the nonstandard GMM OIR test, $P[\xi_O^W > c]$

Model	m =DF	T	Significance level α		
			0.10	0.05	0.01
$E[r_t(1-g_t\gamma)] = 0_N$	9	250	0.538	0.421	0.222
	9	500	0.759	0.680	0.501
	9	1000	0.862	0.824	0.741
	9	∞	0.924	0.907	0.871
$E[r_t(1-f_{1t}\lambda_1)] = 0_N$	9	250	0.735	0.622	0.380
	9	500	0.967	0.933	0.812
	9	1000	1.000	0.999	0.996
	9	∞	1.000	1.000	1.000
$E[r_t(1-f_{1t}\lambda_1-g_t\gamma)] = 0_N$	8	250	0.506	0.390	0.204
	8	500	0.728	0.652	0.478
	8	1000	0.832	0.795	0.702
	8	∞	0.900	0.878	0.835