FEDERAL RESERVE BANK *of* ATLANTA

Further Results on the Limiting Distribution of GMM Sample Moment Conditions

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti

Working Paper 2010-11
July 2010

WORKING PAPER SERIES

FEDERAL RESERVE BANK *of* ATLANTA WORKING PAPER SERIES

Further Results on the Limiting Distribution of GMM Sample Moment Conditions

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti

Working Paper 2010-11
July 2010

Abstract: In this paper, we extend the results in Hansen (1982) regarding the asymptotic distribution of generalized method of moments (GMM) sample moment conditions. In particular, we show that the part of the scaled sample moment conditions that gives rise to degeneracy in the asymptotic normal distribution is T-consistent and has a nonstandard limiting distribution. We derive the asymptotic distribution for a given linear combination of the sample moment conditions and show how to conduct statistical inference. We demonstrate the finite-sample properties of the proposed asymptotic approximation using simulation.

JEL classification: C13, C32, G12

Key words: GMM

The authors thank Jonathan Wright and Chu Zhang for helpful comments and suggestions. Gospodinov gratefully acknowledges financial support from Fonds de Recherche sur la Société et la Culture (FQRSC), Institut de Finance Mathématique de Montréal (IFM2), and the Social Sciences and Humanities Research Council of Canada. Kan gratefully acknowledges financial support from the National Bank Financial of Canada, the Social Sciences and Humanities Research Council of Canada, and the Center for Financial Innovation and Stability at the Federal Reserve Bank of Atlanta. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Nikolay Gospodinov (corresponding author), Concordia University and CIREQ, 1455 de Maisonneuve Boulevard West, Montreal, Quebec, Canada H3G 1M8, 514-848-2424, ext. 3935, nikolay.gospodinov@concordia.ca; Raymond Kan, Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6, kan@chass.utoronto.ca; or Cesare Robotti, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street, N.E., Atlanta, GA 30309-4470, 404-498-8543, cesare.robotti@atl.frb.org.

Federal Reserve Bank of Atlanta working papers, including revised versions, are available on the Atlanta Fed's Web site at frbatlanta.org/pubs/WP/. Use the WebScriber Service at frbatlanta.org to receive e-mail notifications about new papers.

FURTHER RESULTS ON THE LIMITING DISTRIBUTION OF GMM SAMPLE MOMENT CONDITIONS

1. INTRODUCTION

Over the past thirty years, the generalized method of moments (GMM) has established itself as arguably the most popular method for estimating economic models defined by a set of moment conditions. In his seminal paper, Hansen (1982) developed the asymptotic distributions of the GMM estimator, sample moment conditions, and test of over-identifying restrictions for possibly nonlinear models with sufficiently general dependence structure. This large sample theory proved to cover a large class of models and estimators that are of interest to researchers in economics and finance.

There are cases, however, in which the root-T convergence and asymptotic normality of the GMM sample moment conditions and estimators based on these moment conditions do not accurately characterize their limiting behavior. For example, Gospodinov, Kan, and Robotti (2010) demonstrate that some GMM estimators, which are functions of the sample moment conditions, are proportional to the GMM objective function and, hence, cannot be root-T consistent and asymptotically normally distributed for correctly specified models. This situation is directly related to the results in Lemma 4.1 and its subsequent discussion in Hansen (1982) which correctly point out that the covariance matrix of the sample moment conditions is singular.

In this paper, we study the case that gives rise to degeneracy in the asymptotic approximation in Lemma 4.1 of Hansen (1982) and establish the appropriate limiting theory. Interestingly, we show that in this case, the scaled sample moment conditions evaluated at the GMM estimator are characterized by a non-standard asymptotic behavior. In particular, we demonstrate that the estimated GMM moment conditions converge to zero (the value implied by the population moment conditions) at rate T and are asymptotically distributed as a product of jointly normally distributed random vectors.

The rest of this paper is organized as follows. Section 2 introduces the general framework and notation and discusses some motivating examples that illustrate the discontinuity in the asymptotic approximation of the sample moment conditions. This section also provides the main theoretical results on the limiting behavior of linear combinations of sample moment conditions and presents

an easy-to-implement rank test that determines which asymptotic approximation should be used. Section 3 reports simulation results based on a problem in empirical asset pricing and Section 4 concludes.

2. ASYMPTOTICS FOR GMM SAMPLE MOMENT CONDITIONS

2.1. NOTATION AND ANALYTICAL FRAMEWORK

Let $\theta \in \Theta$ denote a $p \times 1$ parameter vector of interest with true value θ_0 that lies in the interior of the parameter space Θ and $g_t(\theta)$ be a known function $\{g : \mathbb{R}^p \to \mathbb{R}^m, m > p\}$ of the data and θ that satisfies the set of population orthogonality conditions

$$E[g_t(\theta_0)] = 0_m. \tag{1}$$

The GMM estimator of θ_0 is defined as

$$\hat{\theta} = \text{argmin}_{\theta \in \Theta} \bar{g}_T(\theta)' W_T \bar{g}_T(\theta), \tag{2}$$

where W_T is an $m \times m$ positive-definite weight matrix and

$$\bar{g}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} g_t(\theta). \tag{3}$$

The matrix W_T is allowed to be a fixed matrix that does not depend on the data and θ (identity matrix, for example), a matrix that depends on the data but not on θ, or a matrix that depends on the data and a preliminary consistent estimator of θ_0 as in the two-step and iterated GMM estimation. Given the first-order asymptotic equivalence of the two-step, iterated, and continuously-updated GMM estimators, our results below can be easily modified to accommodate the continuously-updated (one-step) GMM estimator.

Let $D_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial g_t(\theta)}{\partial \theta'}$, $D(\theta) = E\left[\frac{\partial g_t(\theta)}{\partial \theta'}\right]$ and make the following assumptions.

Assumption A: *Assume that*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_t(\theta_0) \xrightarrow{d} N(0_m, V), \tag{4}$$

where $V = \sum_{j=-\infty}^{\infty} E\left[g_t(\theta_0) g_{t+j}(\theta_0)'\right]$ *is a finite positive-definite matrix.*

Assumption B: *Assume that*

(i) *$g_t(\theta)$ is continuous in θ almost surely, $E\left[\sup_{\theta\in\Theta}|g_t(\theta)|\right] < \infty$, and the parameter space Θ is a compact subset of $\mathbb{R}^p$,*

(ii) *there exists a unique $\theta_0 \in \Theta$ such that $E\left[g_t(\theta_0)\right] = 0_m$ and $E\left[g_t(\theta)\right] \neq 0_m$ for all $\theta \neq \theta_0$,*

(iii) *$W_T \xrightarrow{p} W$, where W is a non-stochastic symmetric positive definite matrix,*

(iv) *$D_T(\theta) \xrightarrow{p} D(\theta)$ uniformly in θ on some neighborhood of θ_0 and $D_0 \equiv D(\theta_0)$ is of rank p.*

Assumption A is a high-level assumption that implicitly imposes restrictions on the data and the vector $g_t(\theta)$. The validity of this assumption can either be verified in the particular context or it can be replaced by a set of explicit primitive conditions. Assumption A can be further strengthened in order to allow for more general dependence structure (see, for instance, Stock and Wright, 2000). Assumption B imposes sufficient conditions that ensure $\hat{\theta} \xrightarrow{p} \theta_0$ in the interior of the compact parameter space Θ. The uniform convergence and the full rank condition in Assumption B (iv) are required for establishing the asymptotic distributions of $\hat{\theta}$ and $\bar{g}_T(\hat{\theta})$.

Under Assumptions A and B (Hansen, 1982),

$$\sqrt{T}\bar{g}_T(\hat{\theta}) = [I_m - D_0(D_0'WD_0)^{-1}D_0'W]\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\theta_0) + o_p(1). \tag{5}$$

Hansen (1982, Lemma 4.1) states the asymptotic normality of $\sqrt{T}\bar{g}_T(\hat{\theta})$ with an asymptotic covariance matrix

$$\Omega_0 = [I_m - D_0(D_0'WD_0)^{-1}D_0'W]V[I_m - D_0(D_0'WD_0)^{-1}D_0'W]'. \tag{6}$$

However, Hansen (1982) notes that Ω_0 is singular and that the asymptotic covariance matrix of $\sqrt{T}D_0'W\bar{g}_T(\hat{\theta})$ reduces to a $p \times p$ matrix of zeros. Provided that W_T is a consistent estimator of W, a similar degeneracy occurs for the object $\sqrt{T}D_0'W_T\bar{g}_T(\hat{\theta}) = \sqrt{T}D_0'h_T(\hat{\theta})$, where $h_T(\hat{\theta}) \equiv W_T\bar{g}_T(\hat{\theta})$.

For our analysis, it is more convenient to rewrite the asymptotic normality result in terms of the nonzero parts of the covariance matrices of $\sqrt{T}\bar{g}_T(\hat{\theta})$ and $\sqrt{T}h_T(\hat{\theta})$. Let Q denote an $m \times (m-p)$ orthonormal matrix whose columns are orthogonal to $W^{\frac{1}{2}}D_0$. Then,

$$QQ' = I_m - W^{\frac{1}{2}}D_0(D_0'WD_0)^{-1}D_0'W^{\frac{1}{2}}. \tag{7}$$

Lemma 1: *Under Assumptions A and B,*

$$\sqrt{T}Q'W^{\frac{1}{2}}\bar{g}_T(\hat{\theta}) \xrightarrow{d} N(0_{m-p}, Q'W^{\frac{1}{2}}VW^{\frac{1}{2}}Q) \tag{8}$$

and

$$\sqrt{T}Q'W^{-\frac{1}{2}}h_T(\hat{\theta}) \xrightarrow{d} N(0_{m-p}, Q'W^{\frac{1}{2}}VW^{\frac{1}{2}}Q). \tag{9}$$

Lemma 1 shows that $\sqrt{T}Q'W^{\frac{1}{2}}\bar{g}_T(\hat{\theta})$ and $\sqrt{T}Q'W^{-\frac{1}{2}}h_T(\hat{\theta})$ have a non-degenerate asymptotic normal distribution. This is a well-known result which allows us to easily establish the limiting distribution of the over-identifying restrictions test. However, little is known about the limiting behavior of those linear combinations of $\bar{g}_T(\hat{\theta})$ or $h_T(\hat{\theta})$ that do not have an asymptotic normal distribution. The purpose of this paper is to establish the rate of convergence and asymptotic distributions of $D_0'W\bar{g}_T(\hat{\theta})$ and $D_0'h_T(\hat{\theta})$. While it is desirable to obtain the limiting behavior of these scaled sample moment conditions for completeness, our interest in this issue does not arise only from theoretical considerations. For instance, in asset pricing, some GMM estimators based on the Hansen-Jagannathan (HJ, 1997) distance have a similar structure and deriving the rate of convergence and asymptotic distribution of $D_0'h_T(\hat{\theta})$ has important practical implications for conducting statistical inference and evaluating asset pricing models. Before we present our main result, we first provide two examples to illustrate the discontinuous nature of the asymptotic analysis for linear combinations of $\bar{g}_T(\hat{\theta})$ or $h_T(\hat{\theta})$.

2.2. MOTIVATING EXAMPLES

Example 1

Suppose that we observe for $t = 1, \ldots, T$ two samples $y_{1t} \sim N(\mu_1, \sigma_1^2)$ and $y_{2t} \sim N(\mu_2, \sigma_2^2)$ that are independent of each other and over time with $\mu_1 = \mu_2 = \theta_0$ and $\sigma_1^2 = \sigma_2^2 = 1$. We assume that the econometrician does not know the variance of y_1 and y_2 and is interested in estimating the common mean parameter θ_0. Let $\hat{\mu}_1 = \frac{1}{T}\sum_{t=1}^{T} y_{1t}$, $\hat{\mu}_2 = \frac{1}{T}\sum_{t=1}^{T} y_{2t}$, and $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ denote the corresponding sample variances. The econometrician estimates θ_0 by minimizing $\bar{g}_T(\theta)'W_T\bar{g}_T(\theta)$, where

$$\bar{g}_T(\theta) = \begin{bmatrix} \hat{\mu}_1 - \theta \\ \hat{\mu}_2 - \theta \end{bmatrix}, \qquad W_T = \begin{bmatrix} \frac{1}{\hat{\sigma}_1^2} & 0 \\ 0 & \frac{1}{\hat{\sigma}_2^2} \end{bmatrix}. \tag{10}$$

The resulting GMM estimator of θ_0 has the form

$$\hat{\theta} = \frac{\hat{\sigma}_2^2\hat{\mu}_1 + \hat{\sigma}_1^2\hat{\mu}_2}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2} \tag{11}$$

with sample moments given by

$$\bar{g}_{1T}(\hat{\theta}) = \hat{\mu}_1 - \hat{\theta} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}(\hat{\mu}_1 - \hat{\mu}_2), \tag{12}$$

$$\bar{g}_{2T}(\hat{\theta}) = \hat{\mu}_2 - \hat{\theta} = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}(\hat{\mu}_2 - \hat{\mu}_1). \tag{13}$$

Given that $D_0 = [-1,\ -1]'$ and $W = I_2$, it can be easily shown that the distribution of $D_0' W \bar{g}_T(\hat{\theta})$ is given by

$$T D_0' W \bar{g}_T(\hat{\theta}) \xrightarrow{d} -\sqrt{2} u_1 u_2, \tag{14}$$

where u_1 and u_2 are two independent standard normal random variables. Hence, the distribution is non-normal and $D_0' W \bar{g}_T(\hat{\theta})$ converges to its true value of zero at rate T. This should also be the case for any linear combination of $W\bar{g}_T(\hat{\theta})$ (or $\bar{g}_T(\hat{\theta})$ since $W = I_2$) with a vector of weights $\alpha = (\alpha_1, \alpha_2)'$ with $\alpha_1 = \alpha_2$, i.e., for a vector α that is in the span of the column space of D_0. In contrast, when α is not in the span of D_0 ($\alpha_1 \neq \alpha_2$), then

$$\sqrt{T}\alpha' \bar{g}_T(\hat{\theta}) \xrightarrow{d} N\left(0, \frac{(\alpha_1 - \alpha_2)^2}{2}\right). \tag{15}$$

The degeneracy of this standard asymptotic distribution occurs when $\alpha_1 = \alpha_2$.

Example 2

Let $y_t(\theta)$ be a candidate stochastic discount factor (SDF) at time t, where θ is a p vector of the parameters of the SDF. Suppose we use m test assets to estimate the true SDF parameter vector θ_0 as well as to test if the proposed SDF is correctly specified. Denote by R_t the payoffs of the m test assets at time t and by q the vector of the costs of the m test assets. Let

$$g_t(\theta) = R_t y_t(\theta) - q. \tag{16}$$

If the model is correctly specified, we have $E[g_t(\theta_0)] = 0_m$. A popular method of estimating θ_0 is to choose θ to minimize the sample squared HJ-distance, defined as

$$\delta_T^2 = \min_{\theta} \bar{g}_T(\theta)' W_T \bar{g}_T(\theta), \tag{17}$$

where $W_T = \left(\frac{1}{T}\sum_{t=1}^{T} R_t R_t'\right)^{-1}$.

To determine whether the proposed SDF is correctly specified, we can examine the sample pricing errors of the m test assets, i.e., $\bar{g}_T(\hat{\theta})$, where $\hat{\theta}$ is the vector of estimated parameters chosen

to minimize the sample HJ-distance. Alternatively, we can examine the m vector of estimated Lagrange multipliers

$$\hat{\lambda} = W_T \bar{g}_T(\hat{\theta}), \tag{18}$$

which is a transformation of the sample pricing errors. Hansen and Jagannathan (1997) show that if the proposed SDF does not price the test assets correctly, then it is possible to correct the mispricing of the SDF by subtracting $\lambda' R_t$ from $y_t(\theta)$. As a result, researchers are often interested in testing $H_0 : \lambda_i = 0$, i.e., in determining whether asset i is responsible for the proposed SDF to deviate from the true SDF.

Gospodinov, Kan, and Robotti (2010) show that for a linear SDF, $q'\hat{\lambda} = -\hat{\delta}_T^2$ where $\hat{\delta}_T^2 = \bar{g}_T(\hat{\theta})' W_T \bar{g}_T(\hat{\theta})$ is the squared sample HJ-distance. For the special case of $q = [1,\ 0'_{m-1}]'$ (i.e., the payoff of the first test asset is a gross return and the rest are excess returns), the estimate of the Lagrange multiplier associated with the first test asset, $\hat{\lambda}_1$, is T-consistent and shares the weighted chi-squared distribution of $\hat{\delta}_T^2$ under the assumption of a correctly specified model. This result is of practical importance since applied researchers often resort to testing the statistical significance of individual Lagrange multipliers in evaluating specification errors in asset pricing models (see Hodrick and Zhang, 2001, for example). More generally, as we show below,

$$T D_0' \hat{\lambda} \xrightarrow{d} -(I_p \otimes v_2') v_1, \tag{19}$$

where v_1 and v_2 are jointly normally distributed vectors of random variables. As a result, any linear combinations of $\hat{\lambda}$ with a vector of weights that is in the span of the column space of D_0 is also T-consistent with a non-standard (product of normals) asymptotic distribution.[1]

It is interesting to note that a similar type of discontinuity in the asymptotic approximation and accelerated rate of convergence have been established by Sims, Stock, and Watson (1990) in an AR(p) model, $p > 1$, with a unit root in the AR polynomial. In particular, Sims, Stock, and Watson (1990) show that a linear combination of $W_T \bar{g}_T(\theta_0)$ with a vector of weights $(\alpha_1, ..., \alpha_p)' \neq (\bar{\alpha}, ..., \bar{\alpha})'$ is root-T and asymptotically normally distributed while a linear combination of $W_T \bar{g}_T(\theta_0)$ with a vector of weights $(\alpha_1, ..., \alpha_p)' = (\bar{\alpha}, ..., \bar{\alpha})'$ yields a T-consistent and asymptotically non-normally distributed estimator.

[1] Detailed derivations of the results in Examples 1 and 2 are available from the authors upon request.

2.3. MAIN RESULTS

We now turn to deriving the asymptotic distributions of $D_0'W\bar{g}_T(\hat{\theta})$ and $D_0'h_T(\hat{\theta})$. Due to the similarities in their structure, we first present the results for $D_0'h_T(\hat{\theta})$ and discuss the $D_0'W\bar{g}_T(\hat{\theta})$ case in the next subsection. First, we make an additional assumption on the joint limiting behavior of $\hat{D}_T = D_T(\hat{\theta})$ and $h_T(\hat{\theta})$ that is needed to establish the asymptotic distribution of $D_0'h_T(\hat{\theta})$.

Assumption C: *Assume that*

$$\sqrt{T}\begin{bmatrix} \mathrm{vec}(Q'W^{\frac{1}{2}}\hat{D}_T) \\ Q'W^{-\frac{1}{2}}h_T(\hat{\theta}) \end{bmatrix} \xrightarrow{d} N(0_{(m-p)(p+1)}, \Sigma) \tag{20}$$

for some finite positive semidefinite matrix Σ.

The asymptotic normality of the $m-p$ vector $Q'W^{-\frac{1}{2}}h_T(\hat{\theta})$ follows directly from Lemma 1. The main requirement is on the limiting behavior of the matrix $\hat{D}_T$ which is, however, rather weak and rules out only some trivial cases. It is important to note that we do not need to impose any restriction on the rate of convergence of W_T apart from being a consistent estimator of W (Assumption B (iii)). In contrast, as we argue later, deriving the asymptotic distribution of $D_0'W\bar{g}_T(\hat{\theta})$ requires explicit assumptions on the rate of convergence of W_T that can differ for parametric and nonparametric heteroskedasticity and autocorrelation consistent (HAC) estimators.

We now state our main result in the following theorem.

Theorem 1: *Under Assumptions A, B, and C,*

$$TD_0'h_T(\hat{\theta}) \xrightarrow{d} -(I_p \otimes v_2')v_1, \tag{21}$$

where v_1 *and* v_2 *are* $(m-p)p$ *and* $(m-p)$ *vectors, respectively, and* $(v_1',\ v_2')' \sim N(0_{(m-p)(p+1)}, \Sigma)$.

Proof. See Appendix A.

In order to make the asymptotic approximation derived in Theorem 1 operational for conducting inference, we need an estimate of the covariance matrix Σ. In the following, we provide explicit expressions that can be used for consistent estimation of the covariance matrix Σ in Theorem 1. Let $G_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \partial\mathrm{vec}(\partial g_t(\theta)/\partial\theta')/\partial\theta'$, $G(\theta) = \partial\mathrm{vec}(D(\theta))/\partial\theta'$, and $G_0 = G(\theta_0)$.

Assumption D: *Assume that* $G_T(\theta) \xrightarrow{p} G(\theta)$ *uniformly in* θ *on some neighborhood of* θ_0, *where* $G(\theta)$ *exists, is finite, and is continuous in* $\theta \in \Theta$ *almost surely.*

In the following lemma, we provide the explicit form of the matrix Σ.

LEMMA 2. *Let* $\tilde{G} = (I_p \otimes Q'W^{\frac{1}{2}})G_0$. *Under Assumptions A, B, and D, we have*

$$\Sigma = \sum_{j=-\infty}^{\infty} E[d_t d'_{t+j}], \tag{22}$$

where $d_t = [d'_{1,t},\ d'_{2,t}]'$ *and*

$$d_{1,t} = \tilde{G}(D'_0 W D_0)^{-1} D'_0 W g_t(\theta_0) + \text{vec}\left(Q'W^{\frac{1}{2}} \frac{\partial g_t(\theta_0)}{\partial \theta'}\right), \tag{23}$$

$$d_{2,t} = Q'W^{\frac{1}{2}} g_t(\theta_0). \tag{24}$$

PROOF. See Appendix A.

Consistent estimators of $d_{1,t}$ and $d_{2,t}$ can be obtained by replacing the population quantities (parameters) in Lemma 2 with their sample analogs (estimators). The consistent estimation of the long-run covariance matrix Σ can then proceed by using a HAC estimator (see Andrews, 1991, for example).

2.4. DISCUSSION

The result in Theorem 1 has important implications for the asymptotic distribution of a linear combination of $h_T(\hat{\theta})$ with a weighting vector α that is in the span of the column space of D_0. In particular, if $\alpha = D_0\tilde{c}$ for a constant nonzero p vector $\tilde{c}$, then we have

$$T\alpha' h_T(\hat{\theta}) \xrightarrow{d} -\tilde{v}'_1 v_2, \tag{25}$$

where $(\tilde{v}'_1,\ v'_2)' \sim N(0_{2(m-p)}, \tilde{\Sigma})$ and $\tilde{v}_1$ is the limit of $\sqrt{T}Q'W^{\frac{1}{2}}\hat{D}_T\tilde{c}$.[2] Instead of expressing the asymptotic distribution as the inner product of two normal random vectors, the following lemma shows that we can alternatively express it as a linear combination of independent χ^2_1 random variables.

[2] It is easy to show that

$$\tilde{\Sigma} = \sum_{j=-\infty}^{\infty} E[\tilde{d}_t \tilde{d}'_{t+j}],$$

where $\tilde{d}_t = (\tilde{d}'_{1,t},\ d'_{2,t})'$ with $\tilde{d}_{1,t} = (\tilde{c}' \otimes Q'W^{\frac{1}{2}})G_0(D'_0 W D_0)^{-1} D'_0 W g_t(\theta_0) + Q'W^{\frac{1}{2}} \frac{\partial g_t(\theta_0)}{\partial \theta'}\tilde{c}$ and $d_{2,t} = Q'W^{\frac{1}{2}} g_t(\theta_0)$. When $\tilde{c}$ is unknown, one could plug in a consistent estimator of $\tilde{c}$. For example, a consistent estimator of $\tilde{c}$ can be obtained as

$$\widehat{\tilde{c}} = (\hat{D}'_T W_T \hat{D}_T)^{-1}(\hat{D}'_T W_T \alpha).$$

Lemma 3. *Suppose that $z = [z_1', \ z_2']'$, where z_1 and z_2 are both $n \times 1$ vectors, is multivariate normally distributed*

$$z \sim N(0_{2n}, \Psi), \tag{26}$$

where Ψ is a positive semidefinite matrix with rank $l \leq 2n$. Let $\Psi = S\Upsilon S'$, where Υ is an $l \times l$ diagonal matrix of the nonzero eigenvalues of Ψ and S is a $2n \times l$ matrix of the corresponding eigenvectors. In addition, let

$$\Gamma = \Upsilon^{\frac{1}{2}} S' \begin{bmatrix} 0_{n\times n} & \frac{1}{2}I_n \\ \frac{1}{2}I_n & 0_{n\times n} \end{bmatrix} S\Upsilon^{\frac{1}{2}}. \tag{27}$$

Then,

$$z_1' z_2 \sim \sum_{i=1}^{k} \gamma_i \xi_i, \tag{28}$$

where the γ_i's are the $k \leq l$ nonzero eigenvalues of Γ and the ξ_i's are independent χ_1^2 random variables.

Proof. See Appendix A.

This lemma shows that the inner product of two vectors of normal random variables (with mean zero) can always be written as a linear combination of independent chi-squared random variables. This result proves very useful since it allows us to adopt numerical procedures for obtaining the p-value of a weighted chi-squared test that are already available in the literature.[3] Furthermore, this result helps us to reconcile the form of the asymptotic approximation proposed in Theorem 1 with the weighted chi-squared distribution that arises in some special cases as in Example 2 above.

Extending the result in Theorem 1 to cover the limiting behavior of $A_0'\bar{g}_T(\hat{\theta})$, where $A_0 = WD_0$, requires stronger conditions. Defining $\hat{A}_T = W_T\hat{D}_T$, we need to replace Assumption C by assuming that

$$\sqrt{T} \begin{bmatrix} \text{vec}(Q'W^{-\frac{1}{2}}\hat{A}_T) \\ Q'W^{\frac{1}{2}}\bar{g}_T(\hat{\theta}) \end{bmatrix} \xrightarrow{d} N(0_{(m-p)(p+1)}, \Xi) \tag{29}$$

for some finite positive definite matrix Ξ. The conditions that (29) imposes on the mp vector

[3] See, for example, Imhof (1961), Davies (1980), and Lu and King (2002). A Matlab program for computing the p-value of a weighted chi-squared test is available from the authors upon request.

$\text{vec}(\hat{A}_T - A_0)$ can be best seen using the decomposition

$$\begin{aligned}\sqrt{T}(\hat{A}_T - A_0) &= \sqrt{T}(W_T\hat{D}_T - WD_0) \\ &= \sqrt{T}W(\hat{D}_T - D_0) + \sqrt{T}(W_T - W)D_0 + \sqrt{T}(W_T - W)(\hat{D}_T - D_0) \\ &= \sqrt{T}W(\hat{D}_T - D_0) + \sqrt{T}(W_T - W)D_0 + o_p(1). \end{aligned} \tag{30}$$

While the conditions for the matrix $\hat{D}_T$ are easily satisfied (Assumption C), the requirement of root-T convergence for W_T rules out nonparametric HAC estimators (see Andrews, 1991, for example) but allows for some parametric HAC estimators (West, 1997). In general, this assumption requires that W_T is computed using a martingale difference sequence process or a dependent process for which the form of serial correlation is known. Then, under the assumption in (29), it can be shown, using similar arguments as in the proof of Theorem 1, that

$$TA_0'\bar{g}_T(\hat{\theta}) \xrightarrow{d} -(I_p \otimes u_2')u_1, \tag{31}$$

where $(u_1',\ u_2')' \sim N(0_{(m-p)(p+1)}, \Xi)$.[4]

2.5. RANK RESTRICTION TEST

The result in equation (25) crucially depends on prior knowledge that a given m vector α is in the column span of D_0. This is the case, for instance, in our Examples 1 and 2. If this information is not available, then one needs to resort to pre-testing in order to determine which asymptotic framework should be used for the particular problem at hand. Below we propose a computationally attractive pre-test that determines if α is in the span of the column space of D_0.

Let P_α be an $m \times (m-1)$ orthonormal matrix whose columns are orthogonal to α such that

$$P_\alpha P_\alpha' = I_m - \alpha(\alpha'\alpha)^{-1}\alpha'. \tag{32}$$

Also, let $\Pi = P_\alpha' D_0$. It turns out that determining if α is in the span of the column space of D_0 is equivalent to determining if Π is of reduced rank.

Under the null that Π is of (reduced) rank $p-1$, $H_0 : \text{rank}(\Pi) = p-1$, there exists a nonzero p vector $\tilde{c}$ such that $D_0\tilde{c} = \alpha$, or equivalently (by premultiplying by P_α' and using the properties of

[4]Note that the factor $\sqrt{2}$ in (14) is due to the fact that u_1 and u_2 in this expression are standardized to have variance equal to one.

P_α) $\Pi\tilde{c} = 0_{m-1}$ with the normalization $\tilde{c}'\tilde{c} = 1$. As discussed in Cragg and Donald (1997), if Π has a reduced column rank of $p-1$, we can use an alternative normalization and express one column of this matrix, say π_j, as a linear combination of the others columns, assuming that $\tilde{c}_j \neq 0$. Without any loss of generality, we can order this column first and define the rearranged partitioned matrix $\Pi = [\pi_1,\ \Pi_2]$ such that

$$[\pi_1,\ \ \Pi_2] \begin{bmatrix} -1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} = 0_{m-1} \tag{33}$$

or

$$\Pi_2 c_0 = \pi_1, \tag{34}$$

where $c_0 = (c_2, ..., c_p)'$. This is equivalent to imposing a normalization on $\tilde{c}$ such that its first element is -1. With such a normalization, c_0 is uniquely defined provided that $\text{rank}(\Pi) = p-1$.

Let $\hat{\Pi}_T = P_\alpha' \hat{D}_T$. Using Assumption C and the proof of Lemma 2, it can be shown that

$$\sqrt{T}\text{vec}(\hat{\Pi}_T - \Pi) \xrightarrow{d} N(0_{(m-1)p}, M), \tag{35}$$

where $M = \sum_{j=-\infty}^{\infty} E[m_t m_{t+j}']$ and

$$m_t = (I_p \otimes P_\alpha')G_0(D_0'WD_0)^{-1}D_0'Wg_t(\theta_0) + \text{vec}\left(P_\alpha' \frac{\partial g_t(\theta_0)}{\partial \theta'}\right). \tag{36}$$

Let $l_T(c) = \hat{\Pi}_{2,T}c - \hat{\pi}_{1,T}$. Define the test statistic

$$LM = \min_c T[l_T(c)'\hat{\Lambda}_T(c)^{-1}l_T(c)], \tag{37}$$

where $\Lambda(c) = ((-1, c') \otimes I_{m-1})M((-1, c')' \otimes I_{m-1})$ and $\hat{\Lambda}_T(c)$ denotes its consistent estimator. The following lemma shows that the rank test statistic LM is chi-squared distributed with $m-p$ degrees of freedom under the null hypothesis that Π is of rank $p-1$.

Lemma 4. *Under Assumptions A to D, and* $H_0 : \text{rank}(\Pi) = p-1$,

$$LM \xrightarrow{d} \chi^2_{m-p}. \tag{38}$$

Proof. See Appendix A.

It is important to note that the rank test statistic in equation (37) has the form of the continuously-updated GMM objective function and is invariant to scaling of c. Furthermore, we would like to emphasize that the minimization in (37) is with respect to only a $p-1$ vector c, and the complexity of the minimization problem does not increase with m. Although the LM test statistic in (37) can be shown to be equivalent to the test statistic proposed by Cragg and Donald (1997),[5] it offers substantial computational advantages over the highly dimensional optimization problem in Cragg and Donald's (1997) test. Finally, our simulation experiments show that the test in (38) enjoys excellent size and power properties (see footnote 7 below).

3. MONTE CARLO EXPERIMENT

In this section, we report the results from a small Monte Carlo experiment that assesses the accuracy of the proposed asymptotic approximation in finite samples. In particular, we adopt the setup of Example 2 and evaluate the size of the weighted chi-squared test on the Lagrange multiplier associated with the first asset when $q = [1,\ 0'_{m-1}]'$ (i.e., the payoff of the first asset is a gross return and the payoffs of the other assets are excess returns). We consider two model specifications that are calibrated to monthly data for the period January 1932 – December 2006. The first one is calibrated to the capital asset pricing model (CAPM) with the value-weighted market excess return as risk factor. For the CAPM, the returns on the test assets are the gross return on the risk-free asset and the excess returns on 10 size ranked portfolios. The second specification is calibrated to the three-factor model (FF3) of Fama and French (1993) with risk factors given by the value-weighted market excess return, the return difference between portfolios of small and large stocks, and the return difference between portfolios of high and low book-to-market ratios. For FF3, the returns on the test assets are the gross return on the risk-free asset and the excess returns on 25 size and book-to-market ranked portfolios. All data are obtained from Kenneth French's website. The SDFs of the CAPM and FF3 include an intercept term.

For each model, the factors and the returns on the test assets are drawn from a multivariate normal distribution. The covariance matrix of the factors and returns is chosen based on the covariance matrix estimated from the data. The mean return vector is chosen such that the asset pricing model holds exactly for the test assets. For each simulated set of returns and factors, the unknown parameters θ_0 of the linear SDF $y(\theta_0) = \tilde{f}'\theta_0$, where $\tilde{f} = (1,\ f')'$, are estimated by

[5] The proof of this result is available from the authors upon request.

minimizing the sample HJ-distance, which yields

$$\hat{\theta} = (\hat{D}_T' W_T \hat{D}_T)^{-1} (\hat{D}_T' W_T q), \tag{39}$$

where $\hat{D}_T = \frac{1}{T}\sum_{t=1}^{T} R_t \tilde{f}_t'$, $W_T = \left(\frac{1}{T}\sum_{t=1}^{T} R_t R_t'\right)^{-1}$, and $q = [1,\ 0_{m-1}']'$. The estimated Lagrange multipliers are given by

$$\hat{\lambda} = W_T \left[\frac{1}{T}\sum_{t=1}^{T} R_t y_t(\hat{\theta}) - q\right], \tag{40}$$

and we consider the first element $\hat{\lambda}_1$. From our discussion in Section 2.4, if we set $\tilde{c} = \theta_0$, then $\alpha'\hat{\lambda} = q'\hat{\lambda} = \hat{\lambda}_1$ and

$$T\hat{\lambda}_1 = Tq'\hat{\lambda} \xrightarrow{d} -v_2' v_2. \tag{41}$$

This result shows that $\sqrt{T}\hat{\lambda}_1$ is not asymptotically normally distributed but instead $T\hat{\lambda}_1$ has a weighted chi-squared distribution. Appendix B provides detailed derivations.

In the analysis of the empirical size of our asymptotic approximation, the computed p-values from this weighted chi-squared distribution are compared to the 10%, 5%, and 1% theoretical sizes of the test. For a comparison, we also provide the empirical size of a standard normal test of $H_0 : \lambda_1 = 0$ used, for example, in Hodrick and Zhang (2001). The empirical rejection probabilities are computed based on 100,000 Monte Carlo replications.

Table I about here

For different sample sizes T, we report the simulation results for the two model specifications in Panels A and B of Table I. In Panel A, the weighted chi-squared distribution provides a very accurate approximation to the finite-sample behavior of $\hat{\lambda}_1$. In contrast, the standard normal test leads to severe size distortions and rejects the true null hypothesis about 92% of the time at the 5% significance level.[6] In the case of 25 risky assets (Panel B), our approximation tends to over-reject for small sample sizes. This over-rejection is a well documented fact in empirical finance and occurs when the number of test assets m is large relative to the number of time series observations T (see,

[6]The substantially different behavior of the two tests documented in the simulations is also observed in real data. For example, using data from the sample period January 1932 – December 2006, the standard normal test suggests that the CAPM fails to price the risk-free asset correctly at the 5% nominal level (p-value of 0.035). In contrast, the weighted chi-squared test delivers the opposite conclusion at any conventional significance level (p-value of 0.887).

for instance, Ahn and Gadarowki, 2004). As T increases, the empirical size of the weighted chi-squared approximation approaches its nominal level. In contrast, the standard normal test always rejects the true null hypothesis 100% of the time and does not improve as T increases.[7]

While the incorrect size of the normal test is expected from our theoretical analysis, the severity of these size distortions is somewhat surprising and deserves a few remarks. It can be shown that in our simulation setup, the normal test statistic of $H_0 : \lambda_1 = 0$ is asymptotically distributed as $-\sqrt{\chi^2_{m-p}}$.[8] One important implication of this result is that although $\lambda_1 = 0$, the correct asymptotic distribution of the normal test statistic of $\hat{\lambda}_1$ is miscentered compared to the standard normal approximation and the shift to the left increases with the degree of over-identification. For example, the medians of this limiting distribution for the CAPM (with $m - p = 9$) and FF3 (with $m - p = 22$) are -2.89 and -4.62, respectively. The 5th and 95th percentiles for the CAPM are -4.11 and -1.82 whereas for FF3, the respective percentiles are -5.82 and -3.51. In summary, this experiment clearly illustrates that the standard asymptotic inference can be grossly misleading.

4. CONCLUSION

This paper derives some new results on the asymptotic distribution of linear combinations of GMM sample moment conditions. These results complement Lemma 4.1 of Hansen (1982) with the cases that give rise to singularity of the asymptotic covariance matrix and degeneracy of the asymptotic distribution. Interestingly, we establish that in these cases, the GMM sample moment conditions converge at rate T to their population analogs and obey a non-standard (product of normals) limiting distribution. We also explain how to consistently estimate the nuisance parameters of the proposed limiting distributions. Finally, we propose an easy-to-implement rank test to determine which asymptotic framework should be adopted for the particular problem at hand.

[7] We also examined the statistical properties of the rank test proposed in Section 2.5 and the sequential test (that includes a pre-test of reduced rank) of $H_0 : \lambda_1 = 0$. Our rank test possesses excellent size and power properties. For example, for FF3 with $T = 900$, the empirical size of the rank test at the 10%, 5%, and 1% nominal levels is 10%, 5%, and 0.9%, respectively; the empirical power of the rank test obtained by setting $\alpha = 1_m$ is always 100% at the 10%, 5%, and 1% nominal levels. The results from the sequential test are very similar to those for the weighted chi-squared approximation. Detailed simulation results can be found in a separate appendix on the authors' websites.

[8] The proof of this result (and a generalization of it) is not presented to preserve space but can be found in a separate appendix on the authors' websites.

REFERENCES

[1] Ahn, S. C., and C. Gadarowski (2004): "Small Sample Properties of the Model Specification Test Based on the Hansen-Jagannathan Distance," *Journal of Empirical Finance*, 11, 109–132.

[2] Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

[3] Cragg, J. G., and S. G. Donald (1997): "Inferring the Rank of a Matrix," *Journal of Econometrics*, 76, 223–250.

[4] Davies, R. B. (1980): "Algorithm AS 155: The Distribution of a Linear Combination of χ^2 Random Variables, " *Applied Statistics*, 29, 323–333.

[5] Fama, E. F., and K. R. French (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

[6] Gospodinov, N., R. Kan, and C. Robotti (2010): "On the Hansen-Jagannathan Distance with a No-Arbitrage Constraint," Federal Reserve Bank of Atlanta Working Paper 2010–4, Available at SSRN: `http://ssrn.com/abstract=1571668`.

[7] Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

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

[9] Hodrick, R., and X. Zhang (2001): "Evaluating the Specification Errors of Asset Pricing Models," *Journal of Financial Economics*, 62, 327–376.

[10] Imhof, J. P. (1961): "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426.

[11] Lu, Z. H., and M. L. King (2002): "Improving the Numerical Technique for Computing the Accumulated Distribution of a Quadratic Form in Normal Variables," *Econometric Reviews*, 21, 149–165.

[12] Newey, W. K., and R. J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–255.

[13] Sims, C. A., J. H. Stock, and M. W. Watson (1990): "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113–144.

[14] Stock, J. H., and J. H. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

[15] West, K. D. (1997): "Another Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Journal of Econometrics*, 76, 171–191.

APPENDIX A

Proof of Theorem 1: Using the first order condition $\hat{D}_T' h_T(\hat{\theta}) = 0_p$, we can express $D_0' h_T(\hat{\theta})$ as

$$\begin{aligned} D_0' h_T(\hat{\theta}) &= -(\hat{D}_T - D_0)' h_T(\hat{\theta}) \\ &= -(\hat{D}_T - D_0)' W^{\frac{1}{2}} (QQ' + W^{\frac{1}{2}} D_0 (D_0' W D_0)^{-1} D_0' W^{\frac{1}{2}}) W^{-\frac{1}{2}} h_T(\hat{\theta}). \end{aligned} \tag{A1}$$

Then,

$$T D_0' h_T(\hat{\theta}) = -\sqrt{T}(\hat{D}_T - D_0)' W^{\frac{1}{2}} (QQ' + W^{\frac{1}{2}} D_0 (D_0' W D_0)^{-1} D_0' W^{\frac{1}{2}}) \sqrt{T} W^{-\frac{1}{2}} h_T(\hat{\theta}). \tag{A2}$$

Since $\sqrt{T} D_0' h_T(\hat{\theta}) = o_p(1)$ and $Q' W^{\frac{1}{2}} D_0 = 0_{(m-p)\times p}$, it follows that

$$T D_0' h_T(\hat{\theta}) = -\left[\sqrt{T} \hat{D}_T' W^{\frac{1}{2}} Q\right] \left[\sqrt{T} Q' W^{-\frac{1}{2}} h_T(\hat{\theta})\right] + o_p(1). \tag{A3}$$

Using Assumption C, let $\sqrt{T}\mathrm{vec}(Q' W^{\frac{1}{2}} \hat{D}_T)$ converge to a vector of normal random variables v_1. Similarly, using (9) in Lemma 1, let $\sqrt{T} Q' W^{-\frac{1}{2}} h_T(\hat{\theta})$ converge to a vector of normal random variables v_2 and write the joint distribution of $(v_1',\ v_2')'$ as

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \sim N\left(0_{(m-p)(p+1)}, \Sigma\right). \tag{A4}$$

Thus,

$$T D_0' h_T(\hat{\theta}) = \mathrm{vec}(T h_T(\hat{\theta})' D_0) \xrightarrow{d} -(I_p \otimes v_2') v_1. \tag{A5}$$

This completes the proof of Theorem 1. ■

Proof of Lemma 2: To obtain the asymptotic distribution of $\mathrm{vec}(\hat{D}_T - D_0)$, define $\tilde{D}_T = \frac{1}{T}\sum_{t=1}^{T} \partial g_t(\theta_0)/\partial\theta'$ and write

$$\sqrt{T}\mathrm{vec}(\hat{D}_T - D_0) = \sqrt{T}\mathrm{vec}(\hat{D}_T - \tilde{D}_T) + \sqrt{T}\mathrm{vec}(\tilde{D}_T - D_0). \tag{A6}$$

For the first term, we use the mean-value theorem to obtain

$$\begin{aligned} \sqrt{T}\mathrm{vec}(\hat{D}_T - \tilde{D}_T) &= G_0 \sqrt{T}(\hat{\theta} - \theta_0) + o_p(1) \\ &= G_0 (D_0' W D_0)^{-1} D_0' W \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_t(\theta_0) + o_p(1), \end{aligned} \tag{A7}$$

where the first equality follows from Assumption D and the second equality is ensured by the conditions imposed in Assumption B. For the second term, we have

$$\sqrt{T}\mathrm{vec}(\tilde{D}_T - D_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\mathrm{vec}\left(\frac{\partial g_t(\theta_0)}{\partial\theta'}\right) - \mathrm{vec}(D_0)\right]. \tag{A8}$$

Using expressions (A6), (A7), and (A8), we have

$$\begin{aligned}
& \sqrt{T}\text{vec}(Q'W^{\frac{1}{2}}\hat{D}_T) \\
= \; & \sqrt{T}\text{vec}(Q'W^{\frac{1}{2}}(\hat{D}_T - D_0)) \\
= \; & (I_p \otimes Q'W^{\frac{1}{2}})\sqrt{T}\text{vec}(\hat{D}_T - D_0) \\
= \; & \tilde{G}(D_0'WD_0)^{-1}D_0'W\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\theta_0) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\text{vec}\left(Q'W^{\frac{1}{2}}\frac{\partial g_t(\theta_0)}{\partial \theta'}\right) + o_p(1) \qquad \text{(A9)}
\end{aligned}$$

using that $\tilde{G} = (I_p \otimes Q'W^{\frac{1}{2}})G_0$. Stacking the expression for $\sqrt{T}\text{vec}(\hat{D}_T'W^{\frac{1}{2}}Q)$ with $Q'W^{\frac{1}{2}}g_t(\theta_0)$, we have

$$\Sigma = \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}'], \qquad \text{(A10)}$$

where $d_t = [d_{1,t}', \; d_{2,t}']'$ and

$$\begin{aligned}
d_{1,t} &= \tilde{G}(D_0'WD_0)^{-1}D_0'Wg_t(\theta_0) + \text{vec}\left(Q'W^{\frac{1}{2}}\frac{\partial g_t(\theta_0)}{\partial \theta'}\right), \qquad \text{(A11)} \\
d_{2,t} &= Q'W^{\frac{1}{2}}g_t(\theta_0). \qquad \text{(A12)}
\end{aligned}$$

This completes the proof of Lemma 2. ■

Proof of Lemma 3: Defining $\tilde{z} = S'z \sim N(0_l, \Upsilon)$, we can write

$$z_1'z_2 = z'\begin{bmatrix} 0_{n\times n} & \frac{1}{2}I_n \\ \frac{1}{2}I_n & 0_{n\times n} \end{bmatrix} z = \tilde{z}'S'\begin{bmatrix} 0_{n\times n} & \frac{1}{2}I_n \\ \frac{1}{2}I_n & 0_{n\times n} \end{bmatrix} S\tilde{z}. \qquad \text{(A13)}$$

Let $e = \Upsilon^{-\frac{1}{2}}\tilde{z} \sim N(0_l, I_l)$. Then, we can write

$$z_1'z_2 = e'\Upsilon^{\frac{1}{2}}S'\begin{bmatrix} 0_{n\times n} & \frac{1}{2}I_n \\ \frac{1}{2}I_n & 0_{n\times n} \end{bmatrix} S\Upsilon^{\frac{1}{2}}e = e'\Gamma e. \qquad \text{(A14)}$$

Since e is standard normal, it follows that

$$z_1'z_2 \sim \sum_{i=1}^{k}\gamma_i\xi_i, \qquad \text{(A15)}$$

where the γ_i's are the $k \leq l$ nonzero eigenvalues of Γ and the ξ_i's are independent χ_1^2 random variables. This completes the proof of Lemma 3. ■

Proof of Lemma 4: Combining $l_T(c) = \hat{\Pi}_{2,T}c - \hat{\pi}_{1,T} = \text{vec}(\hat{\Pi}_{2,T}c - \hat{\pi}_{1,T}) = ((-1, \; c') \otimes I_{m-1})\text{vec}(\hat{\Pi}_T)$ and equation (35), we have

$$\sqrt{T}l_T(c_0) \xrightarrow{d} N(0_{m-1}, \Lambda(c_0)), \qquad \text{(A16)}$$

where $\Lambda(c_0) = ((-1,\ c_0') \otimes I_{m-1})M((-1,\ c_0')' \otimes I_{m-1})$. Let

$$\hat{c} = \arg\min_c l_T(c)'\Lambda_T^{-1}(c)l_T(c) \tag{A17}$$

be the estimator of c_0. Noting that $\hat{c}$ is a continuously-updated GMM estimator and using the equivalence between the continuously-updated GMM estimator and the generalized empirical likelihood estimator with a quadratic discrepancy function (Newey and Smith, 2004, for example), the first-order conditions for the minimization problem in (A17) are given by

$$\left[\sum_{t=1}^{T} \frac{1+\hat{\rho}' l_t(\hat{c})}{T}\left(\frac{\partial l_t(\hat{c})}{\partial c'}\right)\right]' \Lambda_T^{-1}(\hat{c})l_T(\hat{c}) = 0_{p-1},$$

where $\hat{\rho} = -\Lambda_T^{-1}(\hat{c})l_T(\hat{c})$. Furthermore, using that $\hat{\rho} \xrightarrow{p} 0$, $\hat{\rho} = O_p(T^{-1/2})$ and $\sqrt{T}\hat{\rho}$ is asymptotically independent of $\sqrt{T}(\hat{c}-c_0)$, we have (Newey and Smith, 2004, p. 240)

$$\sqrt{T}(\hat{c}-c_0) = -[\Pi_2'\Lambda^{-1}\Pi_2]^{-1}\Pi_2'\Lambda^{-1}\sqrt{T}l_T(c_0) + o_p(1), \tag{A18}$$

where $\Lambda \equiv \Lambda(c_0)$.

Then,

$$\begin{aligned}
\sqrt{T}l_T(\hat{c}) &= \sqrt{T}l_T(c_0) + \Pi_2'\sqrt{T}(\hat{c}-c_0) + o_p(1) \\
&= \left[I_{m-1} - \Pi_2(\Pi_2'\Lambda^{-1}\Pi_2)^{-1}\Pi_2'\Lambda^{-1}\right]\sqrt{T}l_T(c_0) + o_p(1) \\
&= \Lambda^{\frac{1}{2}}\left[I_{m-1} - \Lambda^{-\frac{1}{2}}\Pi_2(\Pi_2'\Lambda^{-1}\Pi_2)^{-1}\Pi_2'\Lambda^{-\frac{1}{2}}\right]\Lambda^{-\frac{1}{2}}\sqrt{T}l_T(c_0) + o_p(1) \\
&= \Lambda^{\frac{1}{2}}\left(I_{m-1} - B\right)\Lambda^{-\frac{1}{2}}\sqrt{T}l_T(c_0) + o_p(1),
\end{aligned} \tag{A19}$$

where $B = \Lambda^{-\frac{1}{2}}\Pi_2[\Pi_2'\Lambda^{-1}\Pi_2]^{-1}\Pi_2'\Lambda^{-\frac{1}{2}}$ is an $(m-1)\times(m-1)$ idempotent matrix with $\operatorname{rank}(B) = p-1$. The test statistic $LM(\hat{c})$ can then be expressed as

$$\begin{aligned}
LM(\hat{c}) &= \sqrt{T}l_T(\hat{c})'\hat{\Lambda}_T(\hat{c})^{-1/2}\hat{\Lambda}_T(\hat{c})^{-1/2}\sqrt{T}l_T(\hat{c}) \\
&= \left[\sqrt{T}l_T(c_0)'\Lambda^{-\frac{1}{2}}\left(I_{m-1} - B\right)\Lambda^{\frac{1}{2}}\right]\Lambda^{-1}\left[\Lambda^{\frac{1}{2}}\left(I_{m-1} - B\right)\Lambda^{-\frac{1}{2}}\sqrt{T}l_T(c_0)\right] + o_p(1) \\
&= \sqrt{T}l_T(c_0)'\Lambda^{-\frac{1}{2}}\left(I_{m-1} - B\right)\Lambda^{-\frac{1}{2}}\sqrt{T}l_T(c_0) + o_p(1) \\
&\xrightarrow{d} \xi'\left(I_{m-1} - B\right)\xi,
\end{aligned} \tag{A20}$$

which is χ^2_{m-p} since $\sqrt{T}\Lambda^{-\frac{1}{2}}l_T(c_0) \xrightarrow{d} \xi \sim N\left(0_{m-1}, I_{m-1}\right)$ and $\operatorname{rank}(I_{m-1}-B) = (m-1)-(p-1) = m-p$. This completes the proof of Lemma 4. ■

APPENDIX B

In the case of asset pricing models with a pricing constraint $\bar{g}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} R_t y_t(\theta) - q$ as in Example 2, the expressions for $d_{1,t}$ and $d_{2,t}$ in the covariance matrix $\Sigma = \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}']$ in Lemma 2 specialize to

$$d_{1,t} = \tilde{G}(D_0' W D_0)^{-1} D_0' W (R_t y_t(\theta_0) - q) + (Q' W^{\frac{1}{2}} R_t \otimes I_p) \frac{\partial y_t(\theta_0)}{\partial \theta'}, \tag{B1}$$

$$d_{2,t} = Q' W^{\frac{1}{2}} (R_t y_t(\theta_0) - q), \tag{B2}$$

where

$$D_0 = E\left[R_t \frac{\partial y_t(\theta_0)}{\partial \theta'}\right] \tag{B3}$$

and

$$G_0 = E\left[(R_t \otimes I_p) \frac{\partial^2 y_t(\theta_0)}{\partial \theta \partial \theta'}\right]. \tag{B4}$$

For the special case of a linear SDF that prices the test assets correctly, these expressions can be further simplified and have the form

$$d_{1,t} = Q' W^{\frac{1}{2}} R_t \otimes \tilde{f}_t, \tag{B5}$$

$$d_{2,t} = Q' W^{\frac{1}{2}} R_t \tilde{f}_t' \theta_0 \tag{B6}$$

since G_0 is a null matrix and $Q' W^{\frac{1}{2}} q = Q' W^{\frac{1}{2}} D_0 \theta_0 = 0_{m-p}$ from the definition of Q.

For the linear combination $T\alpha' h_T(\hat{\theta})$, where $\alpha = D_0 \tilde{c}$ and $h_T(\hat{\theta}) = \hat{\lambda}$, we have from the proof of Theorem 1 that

$$T\tilde{c}' D_0' \hat{\lambda} = -\left[\sqrt{T} Q' W^{\frac{1}{2}} \hat{D}_T \tilde{c}\right]' \left[\sqrt{T} Q' W^{-\frac{1}{2}} h_T(\hat{\theta})\right] + o_p(1). \tag{B7}$$

It is straightforward to show using the results above that

$$\sqrt{T} Q' W^{\frac{1}{2}} \hat{D}_T \tilde{c} \xrightarrow{d} N\left(0_{m-p}, \sum_{j=-\infty}^{\infty} E[\tilde{d}_{1,t} \tilde{d}_{1,t+j}']\right), \tag{B8}$$

where $\tilde{d}_{1,t} = Q' W^{\frac{1}{2}} R_t \tilde{f}_t' \tilde{c}$. When $\tilde{c} = \theta_0$, i.e., $\tilde{c}' D_0' \hat{\lambda} = q' \hat{\lambda}$ as in the simulation experiment in Section 3, we have $\tilde{d}_{1t} = d_{2,t}$ and it follows that

$$T q' \hat{\lambda} \xrightarrow{d} -v_2' v_2, \tag{B9}$$

which is a linear combination of $m - p$ independent chi-squared random variables with one degree of freedom.

Table I
Empirical Sizes of $H_0 : \lambda_1 = 0$

Panel A: Capital Asset Pricing Model

	Standard Normal			Mixture of χ^2		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.978	0.929	0.689	0.144	0.082	0.022
300	0.977	0.925	0.682	0.121	0.065	0.015
450	0.976	0.923	0.679	0.115	0.060	0.014
600	0.976	0.924	0.679	0.111	0.057	0.013
750	0.975	0.923	0.679	0.109	0.057	0.012
900	0.976	0.923	0.680	0.107	0.055	0.011

Panel B: Fama-French Three-Factor Model

	Standard Normal			Mixture of χ^2		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	1.000	1.000	1.000	0.284	0.189	0.072
300	1.000	1.000	1.000	0.178	0.105	0.031
450	1.000	1.000	0.999	0.151	0.084	0.022
600	1.000	1.000	0.999	0.138	0.074	0.018
750	1.000	1.000	0.999	0.130	0.070	0.016
900	1.000	1.000	0.999	0.125	0.067	0.015

The table presents the actual probabilities of rejection for the asymptotic tests of $H_0 : \lambda_1 = 0$ with different levels of significance under the null hypothesis of correctly specified models, assuming that the factors and returns are generated from a multivariate normal distribution. We consider two model specifications that are calibrated to monthly data for the period January 1932 – December 2006. The model specification in Panel A is calibrated to the capital asset pricing model. The model specification in Panel B is calibrated to the three-factor model of Fama and French (1993). The results for different values of the number of time series observations (T) are based on 100,000 simulations.

FURTHER RESULTS ON THE LIMITING DISTRIBUTION OF GMM SAMPLE MOMENT CONDITIONS

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti

Supplementary Material

SIMULATION SETUP

This appendix contains some additional simulation and analytical results regarding the properties of the standard normal test, the mixture of χ^2 test, the LM rank test, and the sequential test considered in the paper. In the simulation experiment, the factors (f) and the returns (R) on the test assets for the CAPM (1 factor and 11 test asset returns) and FF3 (3 factors and 26 test asset returns) are drawn from a multivariate normal distribution with a covariance matrix estimated from the data. The mean return vector is chosen such that the asset pricing model holds exactly for the test assets. For each simulated set of returns and factors, the unknown parameters θ_0 of the linear SDF $y(\theta_0) = \tilde{f}'\theta_0$, where $\tilde{f} = (1,\ f')'$, are estimated by minimizing the sample HJ-distance, which yields

$$\hat{\theta} = (\hat{D}_T' W_T \hat{D}_T)^{-1}(\hat{D}_T' W_T q), \tag{1}$$

where $\hat{D}_T = \frac{1}{T}\sum_{t=1}^{T} R_t \tilde{f}_t'$, $W_T = \left(\frac{1}{T}\sum_{t=1}^{T} R_t R_t'\right)^{-1}$, and $q = [1,\ 0_{m-1}']'$. The estimated Lagrange multipliers are given by

$$\hat{\lambda} = W_T \left[\frac{1}{T}\sum_{t=1}^{T} R_t y_t(\hat{\theta}) - q\right], \tag{2}$$

where $y_t(\hat{\theta}) = \tilde{f}_t'\hat{\theta}$.

We consider linear combinations of sample Lagrange multipliers with different choices of an $m \times 1$ nonzero weighting vector α, i.e., $\alpha'\hat{\lambda}$. Let matrix Q_c denote the null space of the p vector $E[\tilde{f}_t\tilde{f}_t']\theta_0$ and Q_c^1 be the first column of Q_c. Also, let $\Pi = P_\alpha' D_0$, where P_α is an $m \times (m-1)$ orthonormal matrix whose columns are orthogonal to α. In Tables I through IV, we analyze the empirical sizes of four tests – (i) standard normal test of $H_0 : \alpha'\lambda = 0$, (ii) mixture of χ^2 test of $H_0 : \alpha'\lambda = 0$, (iii) LM rank test of $H_0 : \text{rank}(\Pi) = p-1$, and (iv) sequential test of $H_0 : \alpha'\lambda = 0$ with a pre-test of $H_0 : \text{rank}(\Pi) = p-1$, using three choices of α:

1. $\alpha = q = [1\ , 0_{m-1}']'$,
2. $\alpha = D_0 1_p$,
3. $\alpha = D_0 Q_c^1$.

We also analyze the statistical properties of the rank and sequential tests when α in not in the span of the column space of D_0. Specifically, in Table V, we analyze the empirical power of the

rank test for $\alpha = 1_m$ and $\alpha = \sqrt{m}q + 1_m$. In Table VI, we report results for the empirical size of the sequential test for $\alpha = 1_m$ and $\alpha = \sqrt{m}q + 1_m$. The empirical rejection probabilities are computed based on 100,000 Monte Carlo replications.

STANDARD NORMAL TEST

Panels A and B of Table I show that the use of the standard normal test leads to severe over-rejections when α is in the span of the column space of D_0. To understand why, we provide a theoretical analysis of the normal test for particular linear combinations of the Lagrange multipliers λ when the underlying asset pricing model is linear (see Appendix B in the paper).

When $\alpha = q$ (Panel A), the normal test statistic is given by

$$z = \frac{\sqrt{T} q' \hat{\lambda}}{\left[\frac{1}{T}\sum_{t=1}^{T} \hat{h}_t^2\right]^{\frac{1}{2}}}, \tag{3}$$

where

$$\begin{aligned} \hat{h}_t &= q'[W_T \hat{D}_T (\hat{D}_T' W_T \hat{D}_T)^{-1} \hat{D}_T' - I_m] W_T (R_t \tilde{f}_t' \hat{\theta} - q) \\ &= (\hat{\theta}' \hat{D}_T' - q') W_T (R_t \tilde{f}_t' \hat{\theta} - q) \\ &= \hat{e}' W_T (R_t \tilde{f}_t' \hat{\theta} - q). \end{aligned} \tag{4}$$

The numerator can be rewritten as

$$\sqrt{T} q' \hat{\lambda} = -\sqrt{T} \hat{e}' W_T \hat{e}. \tag{5}$$

The denominator can be rewritten as

$$\frac{1}{T}\sum_{t=1}^{T} \hat{h}_t^2 = \hat{e}' W_T \hat{S} W_T \hat{e}, \tag{6}$$

where

$$\hat{S} = \frac{1}{T}\sum_{t=1}^{T} (R_t \tilde{f}_t' \hat{\theta} - q)(R_t \tilde{f}_t' \hat{\theta} - q)'. \tag{7}$$

We can then write the normal test statistic as

$$z = -\frac{\sqrt{T} \hat{e}' W_T \hat{e}}{(\hat{e}' W_T \hat{S} W_T \hat{e})^{\frac{1}{2}}}. \tag{8}$$

Let Q be an $m \times (m-p)$ orthonormal matrix with its columns orthogonal to $W^{\frac{1}{2}} D_0$. We have

$$QQ' = I_m - W^{\frac{1}{2}} D_0 (D_0' W D_0)^{-1} D_0' W^{\frac{1}{2}}. \tag{9}$$

When the model is correctly specified, we have

$$u \equiv \sqrt{T} Q' W_T^{\frac{1}{2}} \hat{e} \xrightarrow{d} N(0_{m-p}, Q' W^{\frac{1}{2}} S W^{\frac{1}{2}} Q), \tag{10}$$

where $\xrightarrow{d}$ denotes "convergence in distribution." It follows that

$$\begin{aligned} T\hat{e}' W_T \hat{e} &= u'u + o_p(1), & (11) \\ T\hat{e}' W_T \hat{S} W_T \hat{e} &= u' Q' W^{\frac{1}{2}} S W^{\frac{1}{2}} Q u + o_p(1). & (12) \end{aligned}$$

Let $P\Lambda P'$ be the spectral decomposition of $Q' W^{\frac{1}{2}} S W^{\frac{1}{2}} Q$ and $\tilde{u} = \Lambda^{-\frac{1}{2}} P' u \xrightarrow{d} N(0_{m-p}, I_{m-p})$. We can then write

$$z \xrightarrow{d} -\frac{\tilde{u}' \Lambda \tilde{u}}{(\tilde{u}' \Lambda^2 \tilde{u})^{\frac{1}{2}}}. \tag{13}$$

When $(R_t', \ f_t')'$ are jointly normally distributed, we have $\Lambda = q_3 I_{m-p}$, where $q_3 = \theta_0' E[\tilde{f}_t \tilde{f}_t'] \theta_0$ (see Proposition 3 of Kan and Zhou, 2004). It follows that

$$z \xrightarrow{d} -\frac{\tilde{u}' \tilde{u}}{(\tilde{u}' \tilde{u})^{\frac{1}{2}}} = -(\tilde{u}' \tilde{u})^{\frac{1}{2}}. \tag{14}$$

In particular $z^2 \xrightarrow{d} \tilde{u}' \tilde{u} = \chi^2_{m-p}$, and it is not χ^2_1. This expression shows that we have an over-rejection problem when we use the normal test and the over-rejection rate increases with $m-p$. In addition, the mean of z is negative and is given by

$$E[z] = -E[(\tilde{u}\tilde{u})^{\frac{1}{2}}] = -\frac{\sqrt{2}\Gamma\left(\frac{m-p+1}{2}\right)}{\Gamma\left(\frac{m-p}{2}\right)}. \tag{15}$$

These theoretical findings explain why the standard normal test strongly over-rejects in Panel A of Table I.

When $\alpha = D_0 1_p$ and $\alpha = D_0 Q_c^1$ (Panels B and C of Table I, respectively), we need to consider a (more general) normal test of $H_0 : \alpha' \lambda = 0$, where $\alpha = D_0 \tilde{c}$ and $\tilde{c}$ is a nonzero p vector. Then, the normal test statistic is given by

$$z = \frac{\sqrt{T} \alpha' \hat{\lambda}}{\left[\frac{1}{T} \sum_{t=1}^T \hat{h}_t^2\right]^{\frac{1}{2}}} = \frac{T \alpha' \hat{\lambda}}{\left[\sum_{t=1}^T \hat{h}_t^2\right]^{\frac{1}{2}}}, \tag{16}$$

where

$$\begin{aligned} \hat{h}_t &= \alpha' [W_T \hat{D}_T (\hat{D}_T' W_T \hat{D}_T)^{-1} \hat{D}_T' - I_m] W_T (R_t \tilde{f}_t' \hat{\theta} - q) \\ &= -\tilde{c}' D_0' W_T^{\frac{1}{2}} \hat{Q} \hat{Q}' W_T^{\frac{1}{2}} (R_t \tilde{f}_t' \hat{\theta} - q). & (17) \end{aligned}$$

The numerator can be written as

$$\begin{aligned} T\tilde{c}'D_0'\hat{\lambda} &= T(\tilde{c}'D_0' - \tilde{c}'\hat{D}_T')\hat{\lambda} \\ &= T(\tilde{c}'D_0' - \tilde{c}'\hat{D}_T')W^{\frac{1}{2}}QQ'W^{-\frac{1}{2}}\hat{\lambda} + o_p(1) \\ &= -[\sqrt{T}\tilde{c}'\hat{D}_T'W^{\frac{1}{2}}Q][\sqrt{T}Q'W^{-\frac{1}{2}}\hat{\lambda}] + o_p(1) \\ &\xrightarrow{d} -z_1'z_2, \end{aligned} \tag{18}$$

where z_1 is the limiting distribution of $\sqrt{T}Q'W^{\frac{1}{2}}\hat{D}_T\tilde{c}$ and z_2 is the limiting distribution of $\sqrt{T}Q'W^{-\frac{1}{2}}\hat{\lambda}$.

The term inside the squared root of the denominator can be rewritten as

$$\sum_{t=1}^{T}\hat{h}_t^2 = T\tilde{c}'D_0'W_T^{\frac{1}{2}}\hat{Q}\hat{Q}'W_T^{\frac{1}{2}}\hat{S}W_T^{\frac{1}{2}}\hat{Q}\hat{Q}'W_T^{\frac{1}{2}}D_0\tilde{c}. \tag{19}$$

Since

$$\begin{aligned} \sqrt{T}\hat{Q}'W_T^{\frac{1}{2}}D_0\tilde{c} &= \sqrt{T}\hat{Q}'W_T^{\frac{1}{2}}(D_0 - \hat{D}_T)\tilde{c} \\ &= \sqrt{T}Q'W^{\frac{1}{2}}(D_0 - \hat{D}_T)\tilde{c} + o_p(1) \\ &= -\sqrt{T}Q'W^{\frac{1}{2}}\hat{D}_T\tilde{c} + o_p(1) \\ &\xrightarrow{d} -z_1, \end{aligned} \tag{20}$$

it follows that

$$\sum_{t=1}^{T}\hat{h}_t^2 \xrightarrow{d} z_1'Q'W^{\frac{1}{2}}SW^{\frac{1}{2}}Qz_1. \tag{21}$$

Therefore, we have

$$z \xrightarrow{d} -\frac{z_1'z_2}{[z_1'Q'W^{\frac{1}{2}}SW^{\frac{1}{2}}Qz_1]^{\frac{1}{2}}}. \tag{22}$$

The joint distribution of z_1 and z_2 is given by

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \sim N(0_{2(m-p)}, \Sigma), \tag{23}$$

where

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}'], \tag{24}$$

and $d_t = [d_{1,t}',\ d_{2,t}']'$ is given by

$$d_{1,t} = Q'W^{\frac{1}{2}}R_t\tilde{f}_t'\tilde{c}, \tag{25}$$

$$d_{2,t} = Q'W^{\frac{1}{2}}R_t\tilde{f}_t'\theta_0. \tag{26}$$

For the special case when R_t and f_t are jointly multivariate normally distributed, it can be easily verified that

$$\Sigma_{11} = q_1 I_{m-p}, \qquad \Sigma_{12} = q_2 I_{m-p}, \qquad \Sigma_{22} = q_3 I_{m-p}, \tag{27}$$

with

$$q_1 = \tilde{c}' E[\tilde{f}_t \tilde{f}_t'] \tilde{c}, \qquad q_2 = \tilde{c}' E[\tilde{f}_t \tilde{f}_t'] \theta_0, \qquad q_3 = \theta_0' E[\tilde{f}_t \tilde{f}_t'] \theta_0, \tag{28}$$

and we have $q_1 q_3 \geq q_2^2$.

Conditional on z_1, we have

$$z_2 \sim N(\Sigma_{21}\Sigma_{11}^{-1} z_1, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}). \tag{29}$$

Noting that $Q'W^{\frac{1}{2}}SW^{\frac{1}{2}}Q = \Sigma_{22}$, we have conditional on z_1,

$$z \sim N\left(-\frac{z_1'\Sigma_{21}\Sigma_{11}^{-1}z_1}{(z_1'\Sigma_{22}z_1)^{\frac{1}{2}}}, \frac{z_1'(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})z_1}{z_1'\Sigma_{22}z_1}\right). \tag{30}$$

Letting $u = \Sigma_{11}^{-\frac{1}{2}} z_1 \sim N(0_{m-p}, I_{m-p})$ and $w \sim N(0,1)$ be independent of each other, we can then write

$$z = -\frac{u'\Sigma_{11}^{\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}u}{(u'\Sigma_{11}^{\frac{1}{2}}\Sigma_{22}\Sigma_{11}^{\frac{1}{2}}u)^{\frac{1}{2}}} + \left[\frac{u'(\Sigma_{11}^{\frac{1}{2}}\Sigma_{22}\Sigma_{11}^{\frac{1}{2}} - \Sigma_{11}^{\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{\frac{1}{2}})u}{u'\Sigma_{11}^{\frac{1}{2}}\Sigma_{22}\Sigma_{11}^{\frac{1}{2}}u}\right]^{\frac{1}{2}} w. \tag{31}$$

The unconditional mean of z is therefore given by

$$E[z] = -E\left[\frac{u'\Sigma_{11}^{\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}u}{(u'\Sigma_{11}^{\frac{1}{2}}\Sigma_{22}\Sigma_{11}^{\frac{1}{2}}u)^{\frac{1}{2}}}\right], \tag{32}$$

which is generally nonzero unless $\Sigma_{21} = 0_{(m-p)\times(m-p)}$.

When $[R_t', \ f_t']'$ are jointly normally distributed (as in our simulation setup), the distribution of z can be simplified to

$$z = -\frac{q_2}{\sqrt{q_1 q_3}}\sqrt{u'u} + \left(1 - \frac{q_2^2}{q_1 q_3}\right)^{\frac{1}{2}} w = r\sqrt{u'u} + \sqrt{1-r^2}w, \tag{33}$$

where $r = -q_2/\sqrt{q_1 q_3}$. It follows that

$$E[z] = -\frac{q_2\sqrt{2}\Gamma\left(\frac{m-p+1}{2}\right)}{\sqrt{q_1 q_3}\Gamma\left(\frac{m-p}{2}\right)}, \tag{34}$$

and its sign is determined by q_2. In addition, $E[z^2]$ is given by

$$E[z^2] = r^2(m-p) + (1-r^2) = 1 + r^2(m-p-1), \tag{35}$$

which is greater than or equal to 1 when $m \geq p+1$. The only case in which the normal test is correct is when $r = 0$, or equivalently $q_2 = \tilde{c}'E[\tilde{f}_t\tilde{f}_t']\theta_0 = 0$. The over-rejection rate of the normal test depends on r^2 and $m-p$. $E[z^2]$ is maximized when $r^2 = 1$ and this occurs when z_1 is proportional to z_2 or, equivalently, when $\tilde{c}$ is proportional to θ_0, i.e., α is proportional to q.

These theoretical findings explain why the standard normal test strongly over-rejects in Panel B of Table I. They also explain why the normal test behaves well in Panel C. Since α in Panel C is set such that $q_2 = 0$, the normal test works well in this scenario.[1]

MIXTURE OF χ^2 TEST

In Table II, we report the empirical size of the mixture of χ^2 test. For the CAPM, our asymptotic approximation works very well even for relatively small sample sizes. For FF3, we need a larger T for the asymptotic approximation to work well. This is a well-known problem in empirical asset pricing that arises when the number of test assets m is large relative to T (see, e.g., Ahn and Gadarowski, 2004).

RANK TEST

Tables III and V report the empirical size and power of the rank test. Overall, the test has excellent size and power properties. Some modest under-rejections only occur for FF3 when $T = 150$.

SEQUENTIAL TEST

In Tables IV and VI, we analyze the empirical size of the sequential test (that includes a reduced rank pre-test) of $H_0 : \lambda_1 = 0$ when α is in the span of the column space of D_0 and when α is not. The sequential test we consider has the following structure. If we reject the null of reduced rank, then we use the normal test in the second stage; otherwise, we use the weighted chi-squared test. Acceptance and rejection of $H_0 : \alpha'\lambda = 0$ is based on the outcome of the second test. Let η_1 be the

[1] Note that our conclusions are not affected by the particular choice of the column of Q_c (the matrix described in the simulation setup).

asymptotic size of the rank restriction test and η_2 be the asymptotic size of either the normal test or the weighted chi-squared test used in the second stage.

When α is in the span of the column space of D_0 (Table IV), the rank restriction test will accept the null of reduced rank with probability $1 - \eta_1$ (asymptotically). Therefore, the probability of using the normal test in the second stage is η_1. Unconditionally, the normal test will reject with probability $p_1 \geq \eta_2$ (in our simulation setup) and the mixture of chi-squared test will reject with probability η_2. Therefore, if the two tests are independent, the size of the sequential test is given by

$$\eta_1 p_1 + (1 - \eta_1)\eta_2 \geq \eta_2.$$

In general, the two tests are dependent because both the rank restriction test and the test of $H_0 : \alpha'\lambda = 0$ are specification tests. In this case, we can only establish an upper bound on the probability of rejection of the sequential test, which is given by

$$\eta_1 + \eta_2.$$

When α is not in the span of the column space of D_0 (Table VI), the rank restriction test will reject the null of reduced rank with probability one (asymptotically), so the normal test will be chosen in the second stage. As a result, the asymptotic size of the sequential test is simply η_2.

The results in Tables IV and VI (which are obtained by setting the asymptotic sizes of the first and second tests equal to each other, i.e., $\eta_1 = \eta_2$) show that the proposed sequential test tends to behave well in our simulation setup.

REFERENCES

[1] Ahn, S. C., and C. Gadarowski (2004): "Small Sample Properties of the Model Specification Test Based on the Hansen-Jagannathan Distance," *Journal of Empirical Finance*, 11, 109–132.

[2] Kan, R., and G. Zhou (2004): "Hansen-Jagannathan Distance: Geometry and Exact Distribution," Working Paper, University of Toronto.

Table I
Empirical Size of the Standard Normal Test

Panel A: $\alpha = q = [1\ , 0'_{m-1}]'$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.978	0.929	0.689	1.000	1.000	1.000
300	0.977	0.925	0.682	1.000	1.000	1.000
450	0.976	0.923	0.679	1.000	1.000	0.999
600	0.976	0.924	0.679	1.000	1.000	0.999
750	0.975	0.923	0.679	1.000	1.000	0.999
900	0.976	0.923	0.680	1.000	1.000	0.999

Panel B: $\alpha = D_0 1_p$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.968	0.910	0.661	1.000	1.000	0.998
300	0.965	0.907	0.650	1.000	1.000	0.998
450	0.964	0.904	0.650	1.000	1.000	0.998
600	0.965	0.905	0.647	1.000	1.000	0.998
750	0.966	0.904	0.648	1.000	1.000	0.998
900	0.965	0.904	0.648	1.000	1.000	0.997

Panel C: $\alpha = D_0 Q_c^1$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.129	0.071	0.017	0.187	0.115	0.037
300	0.114	0.059	0.013	0.141	0.079	0.020
450	0.109	0.056	0.012	0.127	0.068	0.017
600	0.107	0.055	0.012	0.120	0.063	0.015
750	0.106	0.053	0.011	0.117	0.062	0.014
900	0.105	0.053	0.011	0.115	0.060	0.013

Table II
Empirical Size of the Mixture of χ^2 Test

Panel A: $\alpha = q = [1\ , 0'_{m-1}]'$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.144	0.082	0.022	0.284	0.189	0.072
300	0.121	0.065	0.015	0.178	0.105	0.031
450	0.115	0.060	0.014	0.151	0.084	0.022
600	0.111	0.057	0.013	0.138	0.074	0.018
750	0.109	0.057	0.012	0.130	0.070	0.016
900	0.107	0.055	0.011	0.125	0.067	0.015

Panel B: $\alpha = D_0 1_p$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.124	0.068	0.018	0.209	0.137	0.052
300	0.111	0.058	0.013	0.136	0.077	0.021
450	0.109	0.057	0.012	0.123	0.066	0.015
600	0.106	0.054	0.012	0.115	0.061	0.014
750	0.105	0.054	0.011	0.112	0.058	0.013
900	0.104	0.054	0.012	0.112	0.058	0.012

Panel C: $\alpha = D_0 Q_c^1$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.132	0.072	0.018	0.185	0.111	0.034
300	0.116	0.061	0.013	0.138	0.076	0.019
450	0.109	0.056	0.012	0.124	0.067	0.016
600	0.108	0.055	0.012	0.119	0.062	0.014
750	0.108	0.054	0.011	0.115	0.060	0.013
900	0.105	0.053	0.010	0.111	0.059	0.013

Table III
Empirical Size of the Rank Test

Panel A: $\alpha = q = [1\ , 0'_{m-1}]'$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.095	0.044	0.007	0.069	0.024	0.001
300	0.098	0.048	0.009	0.093	0.044	0.007
450	0.099	0.050	0.009	0.098	0.047	0.008
600	0.099	0.049	0.010	0.099	0.047	0.009
750	0.100	0.050	0.010	0.100	0.049	0.009
900	0.099	0.050	0.010	0.100	0.050	0.009

Panel B: $\alpha = D_0 1_p$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.096	0.045	0.007	0.072	0.026	0.001
300	0.099	0.047	0.009	0.093	0.043	0.007
450	0.100	0.050	0.010	0.098	0.046	0.008
600	0.100	0.050	0.010	0.098	0.048	0.008
750	0.101	0.050	0.010	0.100	0.048	0.009
900	0.101	0.050	0.010	0.100	0.050	0.009

Panel C: $\alpha = D_0 Q_c^1$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.084	0.036	0.004	0.048	0.015	0.001
300	0.093	0.044	0.007	0.079	0.033	0.004
450	0.097	0.046	0.008	0.088	0.039	0.006
600	0.097	0.046	0.008	0.091	0.043	0.007
750	0.097	0.047	0.008	0.094	0.044	0.008
900	0.097	0.048	0.009	0.095	0.045	0.008

Table IV
Empirical Size of the Sequential Test
When α is in the Span of the Column Space of D_0

Panel A: $\alpha = q = [1\ , 0'_{m-1}]'$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.145	0.082	0.022	0.284	0.189	0.072
300	0.121	0.065	0.015	0.178	0.105	0.031
450	0.115	0.060	0.014	0.151	0.085	0.022
600	0.111	0.058	0.013	0.138	0.074	0.018
750	0.109	0.057	0.012	0.130	0.070	0.016
900	0.107	0.055	0.011	0.125	0.067	0.015

Panel B: $\alpha = D_0 1_p$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.141	0.072	0.018	0.210	0.137	0.052
300	0.146	0.072	0.014	0.145	0.080	0.021
450	0.149	0.075	0.015	0.143	0.073	0.016
600	0.149	0.074	0.015	0.142	0.072	0.015
750	0.149	0.074	0.015	0.145	0.072	0.015
900	0.149	0.075	0.015	0.147	0.074	0.015

Panel C: $\alpha = D_0 Q_c^1$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.119	0.067	0.017	0.180	0.110	0.034
300	0.103	0.055	0.012	0.130	0.073	0.019
450	0.095	0.050	0.012	0.116	0.063	0.015
600	0.094	0.049	0.011	0.110	0.058	0.014
750	0.093	0.048	0.010	0.106	0.056	0.013
900	0.091	0.047	0.010	0.102	0.055	0.012

Table V
Empirical Power of the Rank Test

Panel A: $\alpha = 1_m$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.999	0.997	0.965	0.977	0.913	0.531
300	1.000	1.000	1.000	1.000	1.000	1.000
450	1.000	1.000	1.000	1.000	1.000	1.000
600	1.000	1.000	1.000	1.000	1.000	1.000
750	1.000	1.000	1.000	1.000	1.000	1.000
900	1.000	1.000	1.000	1.000	1.000	1.000

Panel B: $\alpha = \sqrt{m}q + 1_m$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.999	0.997	0.965	0.974	0.904	0.508
300	1.000	1.000	1.000	1.000	1.000	1.000
450	1.000	1.000	1.000	1.000	1.000	1.000
600	1.000	1.000	1.000	1.000	1.000	1.000
750	1.000	1.000	1.000	1.000	1.000	1.000
900	1.000	1.000	1.000	1.000	1.000	1.000

Table VI
Empirical Size of the Sequential Test
When α is not in the Span of the Column Space of D_0

Panel A: $\alpha = 1_m$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.123	0.067	0.022	0.177	0.124	0.130
300	0.110	0.057	0.012	0.132	0.072	0.018
450	0.106	0.054	0.011	0.121	0.065	0.014
600	0.104	0.053	0.011	0.116	0.061	0.013
750	0.104	0.052	0.011	0.112	0.059	0.013
900	0.103	0.051	0.011	0.110	0.057	0.012

Panel B: $\alpha = \sqrt{m}q + 1_m$

	CAPM			FF3		
	Level of Significance			Level of Significance		
T	10%	5%	1%	10%	5%	1%
150	0.124	0.065	0.022	0.211	0.151	0.142
300	0.110	0.057	0.012	0.151	0.086	0.023
450	0.108	0.054	0.011	0.134	0.073	0.018
600	0.106	0.053	0.011	0.126	0.067	0.016
750	0.105	0.052	0.011	0.119	0.063	0.015
900	0.104	0.052	0.010	0.116	0.061	0.013