Econometrica, Vol. 85, No. 5 (September, 2017), 1641–1656

SPURIOUS INFERENCE IN REDUCED-RANK ASSET-PRICING MODELS

NIKOLAY GOSPODINOV
Research Department, Federal Reserve Bank of Atlanta

RAYMOND KAN
Joseph L. Rotman School of Management, University of Toronto

CESARE ROBOTTI
Terry College of Business, University of Georgia

This note studies some seemingly anomalous results that arise in possibly misspecified, reduced-rank linear asset-pricing models estimated by the continuously updated generalized method of moments. When a spurious factor (that is, a factor that is uncorrelated with the returns on the test assets) is present, the test for correct model specification has asymptotic power that is equal to the nominal size. In other words, applied researchers will erroneously conclude that the model is correctly specified even when the degree of misspecification is arbitrarily large. The rejection probability of the test for overidentifying restrictions typically decreases further in underidentified models where the dimension of the null space is larger than 1.

KEYWORDS: Asset pricing, spurious risk factors, reduced-rank models, model misspecification, continuously updated GMM, rank test, test for overidentifying restrictions.

1. INTRODUCTION

THIS NOTE CHARACTERIZES the limiting behavior of the specification test based on the continuously updated generalized method of moments (CU-GMM) estimator in linear asset-pricing models when the derivative matrix of the moment conditions is rank deficient. For example, this could arise when the model includes spurious factors; that is, factors that are uncorrelated with the returns on the test assets.

In a recent paper, Gospodinov, Kan, and Robotti (2014) analyzed the detrimental effects of lack of identification on estimation, testing, and evaluation of asset-pricing mod-

Nikolay Gospodinov: Nikolay.Gospodinov@atl.frb.org
Raymond Kan: kan@chass.utoronto.ca
Cesare Robotti: robotti@uga.edu

We would like to thank a co-editor and three anonymous referees for numerous insightful suggestions that led to a substantially improved presentation. For helpful discussions and comments, we also thank Seung Ahn, Alex Horenstein, Lei Jiang, Robert Kimmel, Frank Kleibergen, Francisco Peñaranda, Enrique Sentana, Chu Zhang, and seminar participants at EDHEC, ESSEC, Federal Reserve Bank of New York, Boston University, National University of Singapore, Queen Mary University of London, University of California at San Diego, University of Cantabria, University of Exeter, University of Geneva, University of Reading, University of Rome Tor Vergata, University of Southampton, Vanderbilt University, and Western University, as well as conference participants at the 2013 All-Georgia Finance Conference, the 2013 Metro-Atlanta Econometric Study Group Meeting, the 2013 Seventh International Conference on Computational and Financial Econometrics, the 2014 China International Conference in Finance, the 2014 NBER-NSF Time-Series Conference, the 2014 Northern Finance Association Conference, the 2014 SoFiE Conference, the 2014 Tsinghua Finance Workshop, the 2015 Brunel Workshop in Empirical Finance, the 2015 Toulouse Financial Econometrics Conference, and the 2015 York Conference on Macroeconomic, Financial, and International Linkages. Kan gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

 DOI: 10.3982/ECTA13750

els using the Hansen and Jagannathan (1997) distance. Concerns about the reliability of goodness-of-fit measures for noninvariant estimators in beta-pricing models and stochastic discount factor (SDF) models with excess returns have also been raised by Kan and Zhang (1999), Kleibergen and Zhan (2015), and Burnside (2016). In this study, we show that the use of CU-GMM[1] does not alleviate these inference problems and, somewhat surprisingly, makes them substantially worse when spurious factors are included in the model.

In particular, we demonstrate that in the presence of a spurious factor, the power of the specification test is equal to its size. As a consequence, an applied researcher would conclude with high probability that the model is correctly specified, and proceed with constructing standard errors and test statistics that assume correct model specification. Since these statistics would not take into account the extra uncertainty arising from potential model misspecification, the inference on the model parameters would be distorted and would manifest itself in highly significant estimates for factors that do not contribute to improved pricing. In addition, we derive explicitly the limiting distribution of the specification test when the dimension of the null space of the model exceeds 1. In this case, the power of the test under the alternative of a misspecified model is below the size of the test that uses critical values from the standard chi-squared approximation.

The rest of this note is organized as follows. Section 2 derives the limiting distribution of the model specification test in correctly specified and misspecified models with possible underidentification. This section also presents Monte Carlo simulation results. Section 3 concludes.

We adopt the following notation throughout the note: $E[y_t]$ and $\text{Var}[y_t]$ denote the expected value and the variance of a random variable y_t, respectively; 1_N is an $N \times 1$ vector of ones; 0_N is an $N \times 1$ vector of zeros; I_N is the identity matrix of dimension $N \times N$; $\text{rank}(A)$ denotes the column rank of a matrix A; $\text{vec}(A)$ signifies column vectorization of a matrix A; $\otimes$ denotes the Kronecker product; $\stackrel{p}{\to}$ and $\stackrel{d}{\to}$ stand for *convergence in probability* and *convergence in distribution*, respectively; $\sim$ stands for *distributed as*; $\mathcal{N}(\cdot)$ denotes the normal distribution and χ^2_m denotes the chi-squared distribution with m degrees of freedom.

2. LIMITING BEHAVIOR UNDER RANK DEFICIENCY

2.1. *Model and Assumptions*

Let $x_t'\lambda$ be a candidate SDF at time t, where $x_t = [1, f_t']'$, f_t is a $(K-1)$ vector of systematic risk factors, and $\lambda = [\lambda_0, \lambda_1']'$ is a K vector of SDF parameters.[2] Also, let R_t denote the gross returns on N $(N > K)$ test assets and let $e_t(\lambda) = D_t\lambda - 1_N$, where $D_t =$

[1]The CU-GMM estimator is invariant to data scaling, reparameterizations and normalizations, curvature-altering and stationarity-inducing transformations, etc. (Hall (2005)). Peñaranda and Sentana (2015) demonstrate convincingly the appeal of the CU-GMM estimator by showing the numerical equality of prices of risk, overidentifying restrictions tests, and pricing errors in alternative representations of asset-pricing models estimated by CU-GMM. Also, note that the invariance property is not special to CU-GMM and is a feature of the class of generalized empirical likelihood estimators.

[2]The SDF is typically defined in terms of conditional expectations (Hansen and Richard (1987)). Our specification of the SDF remains valid if conditioning information is incorporated through scaled factors and returns (see, for example, Section 8.1 in Cochrane (2005)). We should note that, given the widely documented weak predictive ability of conditioning variables for future returns, this approach could further exacerbate the spurious factor problem discussed in this note.

$R_t x_t'$.[3] The CU-GMM estimator of λ is defined as the solution to (Hansen, Heaton, and Yaron (1996))

$$\mathcal{J} = T \min_{\lambda} \bar{e}(\lambda)' \hat{V}_e(\lambda)^{-1} \bar{e}(\lambda), \tag{1}$$

where $\bar{e}(\lambda) = T^{-1} \sum_{t=1}^{T} e_t(\lambda)$ and $\hat{V}_e(\lambda)$ is a consistent estimator of the long-run variance matrix of the sample pricing errors $V_e(\lambda) = \sum_{j=-\infty}^{\infty} E[(e_t(\lambda) - \bar{e}(\lambda))(e_{t+j}(\lambda) - \bar{e}(\lambda))']$.[4] This is the $\mathcal{J}$ test of the validity of the asset-pricing model restriction $D\lambda = 1_N$, where $D = E[D_t]$. The model is said to be misspecified if $D\lambda \neq 1_N$ for all λ.

Let $Y_t = [f_t', R_t']'$ with $E[Y_t] \equiv \left[\begin{smallmatrix} \mu_f \\ \mu_R \end{smallmatrix}\right]$ and $\mathrm{Var}[Y_t] \equiv V = \left[\begin{smallmatrix} V_f & V_{fR} \\ V_{Rf} & V_R \end{smallmatrix}\right]$. We now formally define a spurious factor.

DEFINITION—Spurious Factor: A spurious factor $f_{t,i}$ is defined such that $E[R_t f_{t,i}] = \mu_R \mu_{f,i}$, where $\mu_{f,i} \equiv E[f_{t,i}]$.

It follows from this definition that the presence of a spurious factor renders the D matrix rank deficient. For example, by writing D more explicitly as

$$D = \big[E[R_t], E[R_t f_{t,1}], \dots, E[R_t f_{t,i}], \dots, E[R_t f_{t,K-1}]\big], \tag{2}$$

it is easy to see that $E[R_t f_{t,i}] = \mu_R \mu_{f,i}$ and $E[R_t] = \mu_R$ are collinear, and D has one degree of rank deficiency.

To gain some intuition for the results to follow, consider the $N \times (K+1)$ matrix $H \equiv [1_N, D]$ and note that the asset-pricing model restriction $D\lambda = 1_N$ can be rewritten as $Hv = 0_N$, where $v = [1, -\lambda']'$. Hence, the specification test is essentially testing that matrix H is of reduced rank since, if the asset-pricing model holds, the vector 1_N is in the column space of the matrix D. This implies that there exists a nonzero vector v that solves $Hv = 0_N$. When the model is correctly specified ($D\lambda = 1_N$) and well identified (matrix D is of full column rank), $v = [1, -\lambda']'$ and there is a member of the null space of H for which the first entry of v is equal to, or can be normalized to, 1. However, there are also cases where H is of reduced rank but the first entry of v is 0. For instance, when the model is misspecified ($D\lambda \neq 1_N$) with a spurious factor (D is rank deficient of degree 1), the vector v that solves $Hv = 0_N$ is proportional to $[0, -\mu_{f,K-1}, 0_{K-2}', 1]'$, where, for convenience, the spurious factor is ordered last. Any attempt to normalize the first element of v to 1 in a given sample when its population value is actually 0 will be approximately offset by making the rest of the coefficients large in magnitude. Despite these mathematical differences, a test that is insensitive to the scaling of a vector in the null space will fail to make a meaningful distinction in these two cases. In summary, the invariant $\mathcal{J}$ test cannot distinguish whether the reduced rank of H arises because the vector 1_N lies in the column space of D (correctly specified model) or because the vector 1_N is not in the column space of D but D is of reduced column rank $K-1$ (for example, a misspecified model with a spurious factor).

[3]When R_t is a vector of payoffs with initial cost $q \neq 0_N$, we just need to replace 1_N with q. In addition, the analysis in the note can be adapted to handle the case of excess returns with $q = 0_N$.

[4]In the case of independent and identically distributed (i.i.d.) data, Newey and Smith (2004, footnote 2) and Antoine, Bonnal, and Renault (2007) establish the equivalence of this CU-GMM estimator and the CU-GMM estimator based on the uncentered optimal weighting matrix. This equivalency continues to hold for time series data when the weighting matrix is of general form.

From this discussion, it proves useful to rewrite the objective function $\mathcal{J}(\lambda) \equiv T\bar{e}(\lambda)'\hat{V}_e(\lambda)^{-1}\bar{e}(\lambda)$ in (1) in a slightly different form. Let $H_t = [1_N, D_t]$, and let $\hat{V}_d$ and $\hat{V}_h$ denote consistent estimators of the long-run variance matrices $V_d = \lim_{T\to\infty} \mathrm{Var}[T^{-1/2} \times \sum_{t=1}^T \mathrm{vec}(D_t)]$ and $V_h = \lim_{T\to\infty} \mathrm{Var}[T^{-1/2} \sum_{t=1}^T \mathrm{vec}(H_t)]$, respectively. Then we have

$$e_t(\lambda) = (-v' \otimes I_N)\,\mathrm{vec}(H_t), \tag{3}$$

$$\hat{V}_e(\lambda) = (-v' \otimes I_N)\hat{V}_h(-v' \otimes I_N)' = (\lambda' \otimes I_N)\hat{V}_d(\lambda \otimes I_N), \tag{4}$$

and

$$\mathcal{J}(\lambda) = T(\hat{D}\lambda - 1_N)'\big[(\lambda' \otimes I_N)\hat{V}_d(\lambda \otimes I_N)\big]^{-1}(\hat{D}\lambda - 1_N), \tag{5}$$

where $\hat{D} = \frac{1}{T}\sum_{t=1}^T D_t$. The expression (5) for the CU-GMM objective function is convenient because it requires a consistent estimator of V_d that is only a function of the data. We show below that this form of the CU-GMM objective function is directly related to the objective function for testing the reduced rank of a matrix.

For our main results, we impose the following assumption.

ASSUMPTION 1: *Assume that Y_t is a jointly stationary and ergodic process with a finite fourth moment and that V is a positive-definite matrix. In addition, assume that $\hat{V}_d \xrightarrow{p} V_d$, where V_d is a positive-definite matrix.*

Assumption 1 provides primitive conditions for the central limit theorem approximation of the product of returns and factors. It allows for general heteroskedasticity and serial correlation in the covariance matrix V_d. This assumption is sufficient for bounding the asymptotic distribution of the $\mathcal{J}$ test under possible underidentification. But to obtain the explicit limiting distribution of the $\mathcal{J}$ test when the null space of the model is more than one dimensional, we impose some further restrictions on the data and the model.

2.2. *Asymptotic Distribution of the Specification Test*

As mentioned above, the asset-pricing model restriction can be expressed as $Hv = 0_N$, where $H = [1_N, D]$ is of dimension $N \times (K+1)$. In this section, we study the limiting behavior of the model specification test when H is rank deficient of degree r ($r = 1, 2, \ldots, K$), where the rank deficiency arises either because the asset-pricing restrictions are satisfied or D itself is rank deficient, or both. In that sense, the limiting distribution of the specification test will depend only on the degree of rank deficiency r and not on whether the model is correctly specified or misspecified. More specifically, suppose that the matrix H has a column rank $K + 1 - r$ ($r = 1, 2, \ldots, K$), that is, there exist r distinct linear combinations of the columns of H that are equal to zero vectors. Also, let P_1 be an $N \times (N-1)$ orthonormal matrix whose columns are orthogonal to 1_N such that $P_1'P_1 = I_{N-1}$ and

$$P_1P_1' = I_N - 1_N(1_N'1_N)^{-1}1_N'. \tag{6}$$

Note that premultiplying by P_1' removes the column of 1s from the matrix H. Thus, performing a rank test on the $(N-1) \times K$ matrix $P_1'D$ provides a convenient way of testing for rank deficiency of H. Under the null that $P_1'D$ is of reduced rank $K-1$, there exists a nonzero K vector c such that $P_1'Dc = 0_{N-1}$ with the normalization $c'c = 1$.[5] As a result, for the purpose of testing whether the asset-pricing model is correctly specified, one could use the Cragg and Donald (1997) test[6] of $H_0 : \text{rank}(P_1'D) = K - 1$, which can be rewritten as an invariant test of the form

$$\mathcal{CD} = T \min_{c:c'c=1} \left(P_1'\hat{D}c\right)' \left[\left(c' \otimes P_1'\right)\hat{V}_d(c \otimes P_1)\right]^{-1} \left(P_1'\hat{D}c\right). \tag{7}$$

Let r denote the dimension of the null space of H. We first establish the limiting behavior of the $\mathcal{J}$ test for $r \geq 1$ under the general conditions in Assumption 1. More specifically, the limiting behavior of the $\mathcal{J}$ test is obtained under a general structure of the V_d matrix for which a consistent, possibly heteroskedasticity and autocorrelation consistent (HAC), estimator is available.

THEOREM 1: *Suppose that the matrix H has a column rank $K + 1 - r$ for an integer $r \geq 1$ and that Assumption 1 holds. Then the limiting behavior of the $\mathcal{J}$ test for correct model specification can be characterized as follows*: (a) *when* $r = 1$, $\mathcal{J} \xrightarrow{d} \chi^2_{N-K}$, *and* (b) *when* $r \geq 2$, $\lim_{T\to\infty} \Pr[\mathcal{J} \leq a] \geq \Pr[x_{N-1} \leq a]$, *where* $x_{N-1} \sim \chi^2_{N-1}$.

The proofs of Theorems 1 and 2 are given in the Appendix.

The result in Theorem 1 shows that the limiting behavior of the $\mathcal{J}$ test is determined entirely by the dimension of the null space of H and not by whether the model is correctly specified or misspecified. This observation is key in understanding the lack of power of the $\mathcal{J}$ test when the model is misspecified. While the rank reduction in H in well identified models arises only when the asset-pricing restrictions are satisfied, the presence of spurious factors leads to a rank reduction in H through a rank deficiency in D even when the asset-pricing restrictions do not hold. In the latter case, the test will have difficulties rejecting the null of correct model specification even if the degree of model misspecification is arbitrarily large. The proof of this result explores the numerical equality between the $\mathcal{J}$ and $\mathcal{CD}$ tests, which has important implications for testing the validity of the asset-pricing model (see also Kleibergen and Mavroeidis (2009), and Arellano, Hansen, and Sentana (2012)). Part (a) of Theorem 1 is concerned with the situation when the null space of H is one dimensional ($r = 1$) and the parameter vector c is uniquely identified, up to a sign. Part (a) embeds two cases: (i) the model is correctly specified and identified, and (ii) the model is misspecified with a spurious factor. In the first case, the asset-pricing restrictions hold ($D\lambda = 1_N$) and the vector 1_N lies in the column space of D so that matrix $P_1'D$ is of reduced column rank with a one-dimensional null space. The asymptotic result for this case, $\mathcal{J} = \mathcal{CD} \xrightarrow{d} \chi^2_{N-K}$, is standard (Hansen (1982), Hansen, Heaton, and Yaron (1996), Cragg and Donald (1997)).

In the second case, 1_N is not in the column space of D (that is, $D\lambda \neq 1_N$), but D is of reduced column rank. Hence, $P_1'D$ is of reduced rank and the null space is again one dimensional. Because D is rank deficient due to the presence of a spurious factor, we have $\text{rank}(P_1'D) = K - 1$ and $\mathcal{CD} \xrightarrow{d} \chi^2_{N-K}$. From the numerical equality of the $\mathcal{J}$ and $\mathcal{CD}$

[5]Solving $P_1'Dc = 0_{N-1}$ requires some normalization since this condition only determines the direction of the vector c (up to a sign) but not its length (Hillier (1990)). While various normalizations are possible, here we employ the normalization $c'c = 1$. The unit norm also imposes compactness (see Hansen (2012)). The relationship between c and λ is made explicit in the proof of Theorem 1 in the Appendix.

[6]See Cragg and Donald (1997), Robin and Smith (2000), and Kleibergen and Paap (2006) for a detailed analysis of rank restriction tests.

tests, we have that $\mathcal{J} \xrightarrow{d} \chi^2_{N-K}$ even when the model is misspecified and the $\mathcal{J}$ test does not exhibit power in rejecting the null hypothesis of correct model specification.[7]

When the null space of $P_1'D$ is more than one dimensional, there is a multiplicity of solutions to $P_1'Dc = 0_{N-1}$ and the model is underidentified (see Arellano, Hansen, and Sentana (2012), Manresa, Peñaranda, and Sentana (2016)). Part (b) in Theorem 1 establishes that the $\mathcal{J}$ test is asymptotically bounded by the χ^2_{N-1} distribution. Since the set of solutions of c for $r > 1$ is multidimensional, this underidentification implies that there are more degrees of freedom in selecting a vector c for solving $P_1'Dc = 0_{N-1}$. In this situation, the specification test will lack power when the model is indeed misspecified and the rejection rates of the test will be bounded by the nominal size of the χ^2_{N-1} distribution. In the context of asset-pricing models for equity returns, N is typically much larger than K and the χ^2_{N-1} asymptotic bound is not substantially more conservative than the asymptotic distribution in part (a) of Theorem 1.

Theorem 1 characterizes the limiting distribution of the $\mathcal{J}$ test under very general conditions but provides only a conservative upper bound when $r \geq 2$. Theorem 2 below presents the explicit asymptotic distribution for this underidentified case, where the dimension of the null space of H is $r \geq 2$, at the cost of some more restrictive assumptions. To introduce these assumptions, let $\Sigma = V_R - V_{Rf}V_f^{-1}V_{fR}$, $\beta = V_{Rf}V_f^{-1}$, $\alpha = \mu_R - \beta\mu_f$, and $B = [\alpha, \beta]$, and let $\hat{B}$ be the sample counterpart of B.

THEOREM 2: *Suppose that the matrix H has a column rank $K + 1 - r$ for an integer $r \geq 1$ and Assumption 1 holds. In addition, assume that $\sqrt{T}\,\mathrm{vec}(\hat{B} - B) \xrightarrow{d} \mathcal{N}(0_{NK}, E[x_t x_t']^{-1} \otimes \Sigma)$. Then we have*

$$\mathcal{J} \xrightarrow{d} w_r, \tag{8}$$

where w_r is the smallest eigenvalue of $W_r \sim \mathcal{W}_r(N-K-1+r, I_r)$, and $\mathcal{W}_r(N-K-1+r, I_r)$ denotes the Wishart distribution with $N-K-1+r$ degrees of freedom and a scaling matrix I_r. Furthermore, $\Pr[w_r \leq a] \geq \Pr[x_{N-K} \leq a]$, *where $x_{N-K} \sim \chi^2_{N-K}$.*

Sufficient conditions for the high level assumption on $\mathrm{vec}(\hat{B})$ in the beta representation of the model are contemporaneous conditional homoskedasticity and a martingale difference sequence requirement for the projection error of returns on factors. Note that these conditions are only sufficient and the result in Theorem 2 may continue to hold when more general features of the data are allowed. For example, under the assumption that $[f_t', R_t']'$ are jointly elliptically distributed, the returns R_t can exhibit conditional heteroskedasticity but Theorem 2 still holds (see the Supplemental Material (Gospodinov, Kan, and Robotti (2017))). Also, Theorem 2 imposes lack of serial correlation on the projection errors but not on the data. In a more general context of a multiperiod conditional linear factor model (Hansen and Richard (1987)), the implied errors will have a zero conditional expectation under correct model specification but not necessarily under model misspecification with spurious factors. Even in this case, it may be better practice to impose the martingale structure (or approximate martingale structure as in Hansen (1985)) on the error term. While this would render the weighting matrix and the asymptotic limit

[7]There are other cases when D is rank deficient with similar consequences on the power of the $\mathcal{J}$ test. For example, the full-rank condition on D may also be violated when the model includes two factors that are noisy versions of the same underlying factor.

in Theorem 2 invalid, our main point that model misspecification will be difficult to detect in the presence of spurious factors remains intact. For full generality, however, one should resort to the asymptotic bound result in Theorem 1.

The more restrictive conditions in Theorem 2 are imposed to ensure that the weighting matrix in (7) has an approximate Kronecker structure.[8] This allows us to express the objective function in (7) as a ratio of quadratic forms and an eigenvalue problem, which gives rise to the Wishart limiting distribution. As shown in part (a) of Theorem 1, the case $r = 1$, where the Wishart distribution specializes to the χ^2_{N-K} distribution, holds under general conditions but it is included in Theorem 2 for completeness. Theorem 2 also shows that the $\mathcal{J}$ test is asymptotically bounded by the χ^2_{N-K} distribution, which is a sharper bound than the asymptotic bound in part (b) of Theorem 1. Figure 1 plots the limiting distribution of the $\mathcal{J}$ test for $r = 1$, 2, and 3 when $N - K = 7$.

The lack of power of the specification tests in underidentified models suggests that the decision regarding the model specification should be augmented with additional diagnostics. For instance, the tests developed by Arellano, Hansen, and Sentana (2012), Peñaranda and Sentana (2015), and Manresa, Peñaranda, and Sentana (2016) can detect whether the lack of rejection of the model specification tests is genuine or is due to the presence of a spurious factor.

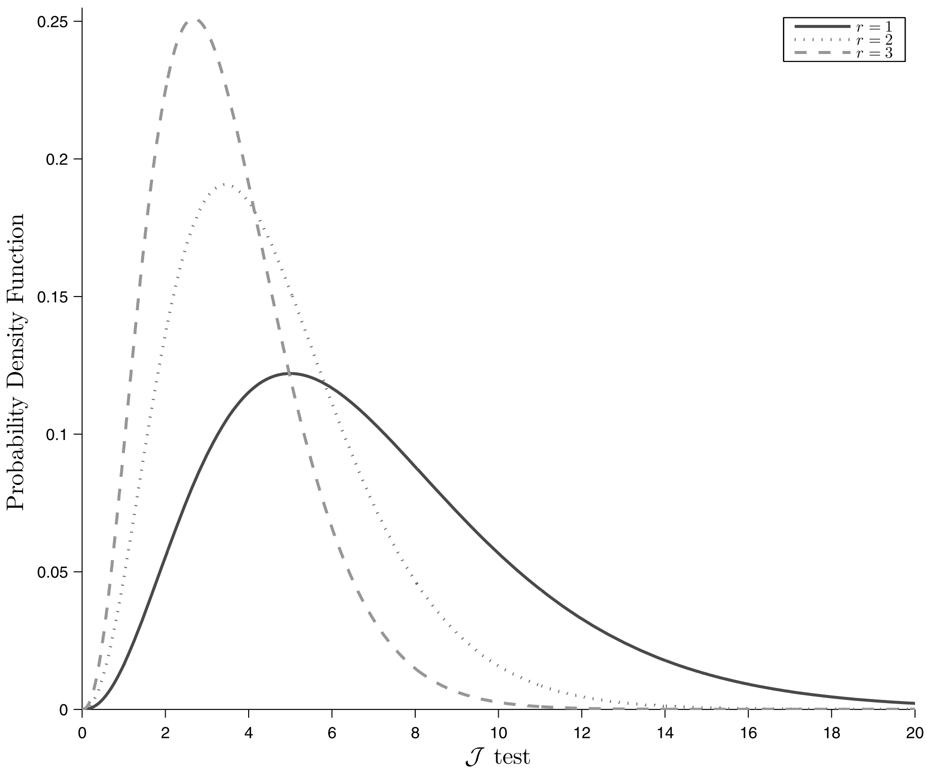

FIGURE 1.—Limiting distribution of the specification test $\mathcal{J}$. The figure plots the asymptotic distributions of $\mathcal{J}$ presented in Theorem 2 for $r = 1$, 2, and 3 (for $N - K = 7$).

[8]See, for instance, Guggenberger, Kleibergen, Mavroeidis, and Chen (2012, p. 2655) for a discussion on the importance of the Kronecker product assumption for this type of result.

2.3. *Simulation Results*

In this section, we undertake a small Monte Carlo simulation experiment to evaluate the empirical rejection rates of the specification test for the CU-GMM estimator. We consider four linear models: (i) a model with a constant term and a useful factor, (ii) a model with a constant term and a spurious factor, (iii) a model with a constant term, a useful factor, and a spurious factor, and (iv) a model with a constant term, three useful factors, and two spurious factors. For each of the four specifications, we consider separately the case of a correctly specified and a misspecified model. This allows us to assess the properties of the specification test when the null space of H is of dimension $r = 0$, 1, 2, and 3.

The returns on the test assets and the useful factors are drawn from a multivariate normal distribution. In all simulation designs, the covariance matrix of the simulated test asset returns is set equal to the sample covariance matrix from the 1959:2–2012:12 sample of monthly gross returns on the 25 Fama–French size and book-to-market ranked portfolios. For misspecified models, the means of the simulated returns are set equal to the means of the actual returns. For correctly specified models, the means of the simulated returns are set such that the asset-pricing model restrictions are satisfied. The means and variances of the simulated useful factors are calibrated to the sample means and variances of the three Fama–French factors (see Fama and French (1993)).[9] The covariances between the useful factors and the returns are chosen based on the sample covariances estimated from the data. The spurious factors are generated as standard normal random variables that are independent of the returns and the useful factors. The time series sample size is $T = 200$, 600, and 1000, and all results are based on 100,000 Monte Carlo replications. We also report the limiting rejection probabilities (denoted by $T = \infty$) for the specification test based on our asymptotic results in Theorem 2 since our simulation setup satisfies the assumptions of Theorem 2.

Table I presents the probabilities of rejection of the model specification test at the 10%, 5%, and 1% nominal levels.

When the model contains only a useful factor (Panel A), the $\mathcal{J}$ test is correctly sized and consistent under the alternative as $T \to \infty$. Some size distortions occur for small T, but this is a well documented finding and is mainly due to the relatively large number of test assets used in our simulations.

Consistent with our theoretical results, the empirical rejection probabilities of the specification test are less than the nominal size when the model is correctly specified but it contains one or more spurious factors. In addition, the specification test does not exhibit any power in the presence of a spurious factor and the empirical rejection probabilities approach the nominal size under the alternative of a misspecified model (last three columns of Panels B and C). As a result, when a spurious factor is included in the model, the researcher will erroneously conclude (with probability 1 minus the nominal size of the test) that the model is correctly specified even when the misspecification of the model is arbitrarily large. Finally, the last three columns of Panel D show that when $r = 2$, the power of the CU-GMM test under the alternative of a misspecified model is below the size of the test that uses critical values from the standard chi-squared approximation. These spurious results should serve as a warning signal in applied work where many macroeconomic factors are only weakly correlated with the returns on the test assets.

[9]When the model contains only one useful factor, the mean and variance of the simulated useful factor is calibrated to the sample mean and variance of the value-weighted market excess return from Fama and French (1993).

TABLE I

REJECTION RATES OF THE SPECIFICATION TEST[a]

	Size			Power		
T	10%	5%	1%	10%	5%	1%
			Panel A: Model With 1 Useful Factor Only			
200	0.211	0.128	0.040	0.900	0.831	0.636
600	0.134	0.073	0.018	1.000	1.000	0.999
1000	0.121	0.065	0.014	1.000	1.000	1.000
∞	0.100	0.050	0.010	1.000	1.000	1.000
			Panel B: Model With 1 Spurious Factor Only			
200	0.022	0.007	0.000	0.127	0.060	0.010
600	0.008	0.002	0.000	0.114	0.057	0.011
1000	0.007	0.002	0.000	0.108	0.054	0.011
∞	0.005	0.001	0.000	0.100	0.050	0.010
			Panel C: Model With 1 Useful and 1 Spurious Factor			
200	0.017	0.005	0.000	0.102	0.044	0.006
600	0.008	0.002	0.000	0.109	0.055	0.010
1000	0.007	0.002	0.000	0.108	0.054	0.010
∞	0.005	0.001	0.000	0.100	0.050	0.010
			Panel D: Model With 3 Useful and 2 Spurious Factors			
200	0.000	0.000	0.000	0.003	0.000	0.000
600	0.000	0.000	0.000	0.005	0.001	0.000
1000	0.000	0.000	0.000	0.006	0.001	0.000
∞	0.000	0.000	0.000	0.006	0.001	0.000

[a]The table presents the rejection rates of Hansen, Heaton, and Yaron's (1996) test for overidentifying restrictions ($\mathcal{J}$) under correctly specified and misspecified models. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama–French portfolio returns for the period 1959:2–2012:12. The $\mathcal{J}$ test statistic is compared with the critical values from a χ^2_{N-K} distribution. The rejection rates for the limiting case ($T = \infty$) in Panels B, C, and D are based on the results in Theorem 2.

3. CONCLUDING REMARKS

In this note, we establish the limiting properties of the CU-GMM specification test of asset-pricing models and show that the inference based on this test can be misleading when spurious factors are present. It is important to stress that this is not an isolated problem limited to a particular sample, test assets, and asset-pricing models.

While the results in this note are developed in the context of linear factor models, we conjecture that similar results characterize the limiting behavior of specification tests in a more general setup. For example, Cragg and Donald (1996) establish the inconsistency of the Anderson–Rubin test for overidentifying restrictions in underidentified linear instrumental variable models while Dovonon and Renault (2013) derive the asymptotic distribution of the specification test under lack of first-order identification. Furthermore, Guggenberger et al. (2012) characterize the asymptotic behavior of the upper bound of the subset Anderson–Rubin statistic in linear instrumental variables regression models with potentially weak identification. Extending the results to the class of generalized empirical likelihood estimators Newey and Smith (2004) is also a promising direction for future research.

APPENDIX: PROOFS

A.1. *Proof of Theorem 1*

We start by showing the numerical equality between the $\mathcal{J}$ and $\mathcal{CD}$ tests. Note that this is an algebraic equality and does not depend on statistical assumptions. Let $P = [1_N/\sqrt{N}, P_1]$, where P_1 is the orthonormal matrix defined in the text. Then we can write

$$\begin{aligned}\mathcal{J}(\lambda) &= T(\hat{D}\lambda - 1_N)'\big[(\lambda'\otimes I_N)\hat{V}_d(\lambda\otimes I_N)\big]^{-1}(\hat{D}\lambda - 1_N)\\ &= T(\hat{D}\lambda - 1_N)'P\big(P'\hat{V}_e(\lambda)P\big)^{-1}P'(\hat{D}\lambda - 1_N) \qquad \text{(A.1)}\\ &= T\begin{bmatrix}\dfrac{1_N'(\hat{D}\lambda - 1_N)}{\sqrt{N}}\\ P_1'\hat{D}\lambda\end{bmatrix}'\begin{bmatrix}\dfrac{1_N'\hat{V}_e(\lambda)1_N}{N} & \dfrac{1_N'\hat{V}_e(\lambda)P_1}{\sqrt{N}}\\ \dfrac{P_1'\hat{V}_e(\lambda)1_N}{\sqrt{N}} & P_1'\hat{V}_e(\lambda)P_1\end{bmatrix}^{-1}\begin{bmatrix}\dfrac{1_N'(\hat{D}\lambda - 1_N)}{\sqrt{N}}\\ P_1'\hat{D}\lambda\end{bmatrix}.\end{aligned}$$

Denote the matrix in the middle as

$$A \equiv \begin{bmatrix}\dfrac{1_N'\hat{V}_e(\lambda)1_N}{N} & \dfrac{1_N'\hat{V}_e(\lambda)P_1}{\sqrt{N}}\\ \dfrac{P_1'\hat{V}_e(\lambda)1_N}{\sqrt{N}} & P_1'\hat{V}_e(\lambda)P_1\end{bmatrix} = \begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}. \qquad \text{(A.2)}$$

Using the following formula for the inverse of a partitioned matrix,

$$A^{-1} = \begin{bmatrix}0 & 0_{N-1}'\\ 0_{N-1} & A_{22}^{-1}\end{bmatrix} + \frac{1}{A_{11} - A_{12}A_{22}^{-1}A_{21}}\begin{bmatrix}-1\\ A_{22}^{-1}A_{21}\end{bmatrix}\begin{bmatrix}-1\\ A_{22}^{-1}A_{21}\end{bmatrix}', \qquad \text{(A.3)}$$

we obtain

$$\begin{aligned}\mathcal{J}(\lambda) &= \mathcal{CD}(\lambda) + \frac{T}{N\big(A_{11} - A_{12}A_{22}^{-1}A_{21}\big)}\\ &\quad \times \big[1_N'\hat{V}_e(\lambda)P_1\big(P_1'\hat{V}_e(\lambda)P_1\big)^{-1}P_1'\hat{D}\lambda - 1_N'(\hat{D}\lambda - 1_N)\big]^2,\end{aligned} \qquad \text{(A.4)}$$

where $\mathcal{CD}(\lambda) = T\lambda'\hat{D}'P_1[(\lambda'\otimes P_1')\hat{V}_d(\lambda\otimes P_1)]^{-1}P_1'\hat{D}\lambda$.

Note first that

$$\begin{aligned}N\big(A_{11} - A_{12}A_{22}^{-1}A_{21}\big) &= 1_N'\hat{V}_e(\lambda)1_N - 1_N'\hat{V}_e(\lambda)P_1\big(P_1'\hat{V}_e(\lambda)P_1\big)^{-1}P_1'\hat{V}_e(\lambda)1_N\\ &= 1_N'\hat{V}_e(\lambda)\big[1_N - P_1\big(P_1'\hat{V}_e(\lambda)P_1\big)^{-1}P_1'\hat{V}_e(\lambda)1_N\big]\\ &= 1_N'\hat{V}_e(\lambda)\hat{V}_e(\lambda)^{-1/2}\\ &\quad \times \big[I_N - \hat{V}_e(\lambda)^{1/2}P_1\big(P_1'\hat{V}_e(\lambda)P_1\big)^{-1}P_1'\hat{V}_e(\lambda)^{1/2}\big]\hat{V}_e(\lambda)^{1/2}1_N \qquad \text{(A.5)}\\ &= 1_N'\hat{V}_e(\lambda)\hat{V}_e(\lambda)^{-1/2}\\ &\quad \times \big[\hat{V}_e(\lambda)^{-1/2}1_N\big(1_N'\hat{V}_e(\lambda)^{-1}1_N\big)^{-1}1_N'\hat{V}_e(\lambda)^{-1/2}\big]\hat{V}_e(\lambda)^{1/2}1_N\\ &= \frac{N^2}{1_N'\hat{V}_e(\lambda)^{-1}1_N}.\end{aligned}$$

Similarly, rearranging the term in the square brackets gives

$$\begin{aligned}
&1_N'\hat{V}_e(\lambda)P_1\big(P_1'\hat{V}_e(\lambda)P_1\big)^{-1}P_1'\hat{D}\lambda - 1_N'\hat{D}\lambda + N \\
&\quad = -\big[1_N - P_1\big(P_1'\hat{V}_e(\hat{c})P_1\big)^{-1}P_1'\hat{V}_e(\hat{c})1_N\big]'\hat{D}\lambda + N \\
&\quad = -\frac{N1_N'\hat{V}_e(\lambda)^{-1}(\hat{D}\lambda - 1_N)}{1_N'\hat{V}_e(\lambda)^{-1}1_N}.
\end{aligned} \tag{A.6}$$

Thus,

$$\begin{aligned}
&\frac{T}{N\big(A_{11} - A_{12}A_{22}^{-1}A_{21}\big)}\big[1_N'\hat{V}_e(\lambda)P_1\big(P_1'\hat{V}_e(\lambda)P_1\big)^{-1}P_1'\hat{D}\lambda - 1_N'(\hat{D}\lambda - 1_N)\big]^2 \\
&\quad = \frac{T\big(1_N'\hat{V}_e(\lambda)^{-1}(\hat{D}\lambda - 1_N)\big)^2}{1_N'\hat{V}_e(\lambda)^{-1}1_N}.
\end{aligned} \tag{A.7}$$

For the numerical equality of the $\mathcal{J}$ and $\mathcal{CD}$ tests, we need to show that $1_N'\hat{V}_e(\lambda)^{-1} \times (\hat{D}\lambda - 1_N) = 0$ when evaluated at the minimizer of $\mathcal{J}(\lambda)$, $\hat{\lambda}$. The first-order conditions for the CU-GMM estimator are given by (see Antoine, Bonnal, and Renault (2007))[10]

$$\left[\sum_{t=1}^{T}\pi_t\left(\frac{\partial e_t(\hat{\lambda})}{\partial \lambda'}\right)'\right]\hat{V}_e(\hat{\lambda})^{-1}\bar{e}(\hat{\lambda}) \equiv \hat{D}_\pi'\hat{V}_e(\hat{\lambda})^{-1}(\hat{D}\hat{\lambda} - 1_N) = 0_K, \tag{A.8}$$

where the weights $\pi_t = (1 - \bar{e}(\hat{\lambda})'\hat{V}_e(\hat{\lambda})^{-1}[e_t(\hat{\lambda}) - \bar{e}(\hat{\lambda})])/T$ induce the moment conditions to be exactly satisfied, $\sum_{t=1}^{T}\pi_t e_t(\hat{\lambda}) = 0_N$. Since this forces the vector 1_N to be in the column span of $\hat{D}_\pi$, it follows that $1_N'\hat{V}_e(\hat{\lambda})^{-1}(\hat{D}\hat{\lambda} - 1_N) = 0$. Note that the $\mathcal{CD}$ test is invariant to scaling and normalizations of the parameter vector so that $\min_\lambda \mathcal{CD}(\lambda) = \min_{c:c'c=1}\mathcal{CD}(c) \equiv \mathcal{CD}$ and $\min_\lambda \mathcal{J}(\lambda) \equiv \mathcal{J} = \mathcal{CD}$.[11] Since under the null $H_0 : \text{rank}(P_1'D) = K - 1$ we have $\mathcal{CD} \xrightarrow{d} \chi^2_{N-K}$ (Cragg and Donald (1997)), it immediately follows by $\mathcal{J} = \mathcal{CD}$ that $\mathcal{J} \xrightarrow{d} \chi^2_{N-K}$. This completes the proof of part (a).

For part (b), consider the more general case when the true dimension of the null space of H is $r \geq 1$. Let c_* denote the parameter vector under the null $H_0 : \text{rank}(P_1'D) = K - 1$ that solves $P_1'Dc_* = 0_{N-1}$. Note that this restriction can hold for both correctly specified and misspecified models when $r \geq 1$. Furthermore, a nonzero vector c_* that solves $P_1'Dc_* = 0_{N-1}$ always exists although the set of solutions for c_* is r dimensional. This is not the case for the parameter vector λ for which λ_* that solves $D\lambda_* - 1_N = 0_N$ cannot be defined when the model is misspecified and it contains spurious factors. Since $\hat{c}$ is the minimizer of $\min_{c:c'c=1}\mathcal{CD}(c) \equiv \mathcal{CD}$, we have that $\mathcal{CD}(c_*) \geq \mathcal{CD}$. Also, from Stock and Wright (2000), we have $\mathcal{CD}(c_*) \xrightarrow{d} \chi^2_{N-1}$. Then it follows that the test $\mathcal{J} = \mathcal{CD}$ is asymptotically bounded by the χ^2_{N-1} distribution when $r \geq 2$. This completes the proof of part (b).

[10]Note that $\hat{V}_e(\lambda)^{-1}(\hat{D}\lambda - 1_N)$ is, up to a sign, the vector of Lagrange multipliers associated with the N moment conditions (Antoine, Bonnal, and Renault (2007)).

[11]If we let $\hat{\lambda} = \hat{a}\hat{c}$, the constant $\hat{a}$ that solves $1_N'[\hat{a}^2\hat{V}_e(\hat{c})]^{-1}(\hat{D}\hat{a}\hat{c} - 1_N) = 0$ is given by $\hat{a} = \frac{1_N'\hat{V}_e(\hat{c})^{-1}1_N}{1_N'\hat{V}_e(\hat{c})^{-1}\hat{D}\hat{c}}$.

A.2. *Proof of Theorem 2*

First we rewrite the $\mathcal{CD}$ test in an asymptotically equivalent but simpler form. Let X be a $T \times K$ matrix with a typical row x_t' and note that

$$\hat{B} = \hat{D}\left(\frac{X'X}{T}\right)^{-1}. \tag{A.9}$$

Then, using that

$$\sqrt{T}\,\mathrm{vec}(\hat{B} - B) \xrightarrow{d} \mathcal{N}\left(0_{NK}, E\left[x_t x_t'\right]^{-1} \otimes \Sigma\right) \tag{A.10}$$

and the delta method, we have

$$\sqrt{T}\,\mathrm{vec}(\hat{D} - D) \xrightarrow{d} \mathcal{N}\left(0_{NK}, E\left[x_t x_t'\right] \otimes \Sigma + (I_K \otimes B)V_x\left(I_K \otimes B'\right)\right), \tag{A.11}$$

where V_x is the asymptotic variance of $\sqrt{T}\,\mathrm{vec}((X'X)/T - E[x_t x_t'])$.[12] Therefore, for any nonzero vector c, we have

$$\sqrt{T}\left(P_1'\hat{D}c - P_1'Dc\right) \xrightarrow{d} \mathcal{N}\left(0_{N-1}, c'E\left[x_t x_t'\right]cP_1'\Sigma P_1 + \left(c' \otimes P_1'B\right)V_x\left(c \otimes B'P_1\right)\right). \tag{A.12}$$

Hence under the assumptions of Theorem 2, we can consistently estimate $(c' \otimes P_1')V_d \times (c \otimes P_1)$ using $\mathcal{A}(c) = \mathcal{A}_1(c) + \mathcal{A}_2(c)$, where $\mathcal{A}_1(c) = c'(X'X/T)cP_1'\hat{\Sigma}P_1$ and $\mathcal{A}_2(c) = (c' \otimes P_1'\hat{B})\hat{V}_x(c \otimes \hat{B}'P_1)$, with $\hat{\Sigma}$ and $\hat{V}_x$ being consistent estimators of Σ and V_x, respectively. When $P_1'D$ has a reduced rank, we have

$$\mathcal{J} = \mathcal{CD} = T \min_{c:c'c=1} \left(P_1'\hat{D}c\right)'\left[\left(c' \otimes P_1'\right)\hat{V}_d(c \otimes P_1)\right]^{-1}\left(P_1'\hat{D}c\right) = \mathcal{J}_A + o_p(1), \tag{A.13}$$

where

$$\mathcal{J}_A = T \min_{c:c'c=1} \left(P_1'\hat{D}c\right)'\mathcal{A}(c)^{-1}\left(P_1'\hat{D}c\right). \tag{A.14}$$

Next, we show that

$$\mathcal{J}_A = T \min_{c:c'c=1} \left(P_1'\hat{D}c\right)'\mathcal{A}_1(c)^{-1}\left(P_1'\hat{D}c\right) + o_p(1). \tag{A.15}$$

Let $\hat{c}$ be the optimal c in (A.14) and note that $\hat{c}$ is $O_p(1)$ since $\hat{c}'\hat{c} = 1$ by the adopted normalization. Since $\mathcal{A}_2(c)$ is a positive-definite matrix, it follows that $\mathcal{A}(c)^{-1} \preceq \mathcal{A}_1(c)^{-1}$ and

$$\begin{aligned} T\left(P_1'\hat{D}\hat{c}\right)'\mathcal{A}(\hat{c})^{-1}\left(P_1'\hat{D}\hat{c}\right) &\le T \min_{c:c'c=1} \left(P_1'\hat{D}c\right)'\mathcal{A}_1(c)^{-1}\left(P_1'\hat{D}c\right) \\ &\le T\left(P_1'\hat{D}\hat{c}\right)'\mathcal{A}_1(\hat{c})^{-1}\left(P_1'\hat{D}\hat{c}\right). \end{aligned} \tag{A.16}$$

Then, so as to establish the result in (A.15), it is sufficient to show that

$$T\left(P_1'\hat{D}\hat{c}\right)'\mathcal{A}(\hat{c})^{-1}\left(P_1'\hat{D}\hat{c}\right) = T\left(P_1'\hat{D}\hat{c}\right)'\mathcal{A}_1(\hat{c})^{-1}\left(P_1'\hat{D}\hat{c}\right) + o_p(1). \tag{A.17}$$

[12]Note that V_x is singular because x_t has 1 as its first element.

Using that

$$\mathcal{A}_1(\hat{c})^{-1} = \mathcal{A}(\hat{c})^{-1} + \mathcal{A}(\hat{c})^{-1}\mathcal{A}_2(\hat{c})\big[\mathcal{A}_2(\hat{c}) - \mathcal{A}_2(\hat{c})\mathcal{A}(\hat{c})^{-1}\mathcal{A}_2(\hat{c})\big]^{-1}\mathcal{A}_2(\hat{c})\mathcal{A}(\hat{c})^{-1}, \quad \text{(A.18)}$$

where $[\mathcal{A}_2(\hat{c}) - \mathcal{A}_2(\hat{c})\mathcal{A}(\hat{c})^{-1}\mathcal{A}_2(\hat{c})]^{-1}$ is a positive-definite matrix, we obtain

$$\begin{aligned} T\big(P_1'\hat{D}\hat{c}\big)'\mathcal{A}_1(\hat{c})^{-1}\big(P_1'\hat{D}\hat{c}\big) &= T\big(P_1'\hat{D}\hat{c}\big)'\mathcal{A}(\hat{c})^{-1}\big(P_1'\hat{D}\hat{c}\big) \\ &\quad + \sqrt{T}\big(P_1'\hat{D}\hat{c}\big)'\mathcal{A}(\hat{c})^{-1}\mathcal{A}_2(\hat{c}) \\ &\quad \times \big[\mathcal{A}_2(\hat{c}) - \mathcal{A}_2(\hat{c})\mathcal{A}(\hat{c})^{-1}\mathcal{A}_2(\hat{c})\big]^{-1}\mathcal{A}_2(\hat{c})\mathcal{A}(\hat{c})^{-1}\sqrt{T}\big(P_1'\hat{D}\hat{c}\big). \end{aligned} \quad \text{(A.19)}$$

Since $\mathcal{A}(\hat{c})$, $\mathcal{A}_2(\hat{c})$, and $[\mathcal{A}_2(\hat{c}) - \mathcal{A}_2(\hat{c})\mathcal{A}(\hat{c})^{-1}\mathcal{A}_2(\hat{c})]^{-1}$ are $O_p(1)$, it suffices to show that

$$\mathcal{A}_2(\hat{c})\mathcal{A}(\hat{c})^{-1}\sqrt{T}P_1'\hat{D}\hat{c} = o_p(1). \quad \text{(A.20)}$$

From the first-order conditions of (A.14), we have

$$\begin{aligned} &\sqrt{T}\big(P_1'\hat{D}\hat{c}\big)'\mathcal{A}(\hat{c})^{-1}P_1'\hat{D} \\ &\quad - \frac{1}{\sqrt{T}}\big[\sqrt{T}\big(P_1'\hat{D}\hat{c}\big)'\mathcal{A}(\hat{c})^{-1}\mathcal{B}(\hat{c}) \otimes \sqrt{T}\big(P_1'\hat{D}\hat{c}\big)'\mathcal{A}(\hat{c})^{-1}\big]\big[I_K \otimes \operatorname{vec}(I_{N-1})\big] = 0_K', \end{aligned} \quad \text{(A.21)}$$

where $\mathcal{B}(\hat{c}) = (\hat{c}' \otimes I_{N-1})[(X'X/T) \otimes P_1'\hat{\Sigma}P_1 + (I_K \otimes P_1'\hat{B})\hat{V}_x(I_K \otimes \hat{B}'P_1)]$. In deriving the limiting distribution below, we show that the middle term in (A.16), $T\min_{c:c'c=1}(P_1'\hat{D}c)' \times \mathcal{A}_1(c)^{-1}(P_1'\hat{D}c)$, is $O_p(1)$. The first inequality in (A.16) then implies that a quadratic form in $\sqrt{T}(P_1'\hat{D}\hat{c})'\mathcal{A}(\hat{c})^{-1/2}$ is bounded by an $O_p(1)$ random variable. Since $\mathcal{B}(\hat{c})$ and $\mathcal{A}(\hat{c})$ are $O_p(1)$, then

$$\frac{1}{\sqrt{T}}\big[\sqrt{T}\big(P_1'\hat{D}\hat{c}\big)'\mathcal{A}(\hat{c})^{-1}\mathcal{B}(\hat{c}) \otimes \sqrt{T}\big(P_1'\hat{D}\hat{c}\big)'\mathcal{A}(\hat{c})^{-1}\big]\big[I_K \otimes \operatorname{vec}(I_{N-1})\big] = o_p(1) \quad \text{(A.22)}$$

and

$$\hat{D}'P_1\mathcal{A}(\hat{c})^{-1}\sqrt{T}P_1'\hat{D}\hat{c} = o_p(1). \quad \text{(A.23)}$$

Furthermore, substituting for $\hat{D} = \hat{B}(X'X/T)$ and premultiplying both sides by the $O_p(1)$ matrix $(X'X/T)^{-1}$, we obtain

$$\hat{B}'P_1\mathcal{A}(\hat{c})^{-1}\sqrt{T}P_1'\hat{D}\hat{c} = o_p(1). \quad \text{(A.24)}$$

Finally, substituting for $\mathcal{A}_2(\hat{c})$, we have

$$\mathcal{A}_2(\hat{c})\mathcal{A}(\hat{c})^{-1}\sqrt{T}P_1'\hat{D}\hat{c} = \big(\hat{c}' \otimes P_1'\hat{B}\big)\hat{V}_x\big(\hat{c} \otimes \hat{B}'P_1\mathcal{A}(\hat{c})^{-1}\sqrt{T}P_1'\hat{D}\hat{c}\big) = o_p(1), \quad \text{(A.25)}$$

where the last equality follows from (A.24). Thus,

$$T\big(P_1'\hat{D}\hat{c}\big)'\mathcal{A}(\hat{c})^{-1}\big(P_1'\hat{D}\hat{c}\big) = T\big(P_1'\hat{D}\hat{c}\big)'\mathcal{A}_1(\hat{c})^{-1}\big(P_1'\hat{D}\hat{c}\big) + o_p(1) \quad \text{(A.26)}$$

and

$$\begin{aligned}\mathcal{J} &= T \min_{c:c'c=1} \frac{c'\hat{D}'P_1\left(P_1'\hat{\Sigma}P_1\right)^{-1}P_1'\hat{D}c}{c'\left(X'X/T\right)c} + o_p(1) \\ &= T \min_{\tilde{c}:\tilde{c}'\tilde{c}=1} \frac{\tilde{c}'\hat{B}'P_1\left(P_1'\hat{\Sigma}P_1\right)^{-1}P_1'\hat{B}\tilde{c}}{\tilde{c}'\left[\left(X'X/T\right)^{-1}\right]\tilde{c}} + o_p(1),\end{aligned} \tag{A.27}$$

where $\tilde{c}$ is proportional to $(X'X/T)c$. Using (A.27) and the invariance property of the estimator, it then follows that the $\mathcal{J}$ test is asymptotically distributed as T times the smallest eigenvalue of (Anderson (1951), Sargan (1958))

$$\tilde{\Omega} = \left(X'X/T\right)\hat{B}'P_1\left(P_1'\hat{\Sigma}P_1\right)^{-1}P_1'\hat{B}. \tag{A.28}$$

Let L_f be a lower triangular matrix such that $L_f L_f' = V_f$ and define

$$L = \begin{bmatrix} 1 & 0_{K-1}' \\ \mu_f & L_f \end{bmatrix}. \tag{A.29}$$

Using that $(X'X)/T \xrightarrow{p} LL'$ and $\hat{\Sigma} \xrightarrow{p} \Sigma$, the $\mathcal{J}$ test has the same distribution as the smallest eigenvalue of

$$W = TL'\hat{B}'P_1\left(P_1'\Sigma P_1\right)^{-1}P_1'\hat{B}L. \tag{A.30}$$

Define $Z = (P_1'\Sigma P_1)^{-1/2}P_1'\hat{B}L$ and $M = (P_1'\Sigma P_1)^{-1/2}P_1'BL$. We have

$$\sqrt{T}\operatorname{vec}(Z - M) \xrightarrow{d} \mathcal{N}(0_{(N-1)K}, I_{(N-1)K}). \tag{A.31}$$

Since $P_1'BL$ has rank $K - r$, there exists a $K \times r$ orthonormal matrix C_1 such that $MC_1 = 0_{(N-1)\times r}$. Let $C = [C_1, C_2]$ be a $K \times K$ orthonormal matrix, and define $\tilde{Z} = [\tilde{Z}_1, \tilde{Z}_2] = [ZC_1, ZC_2]$. Using (A.10) and $MC_1 = 0_{(N-1)\times r}$, we have $MC_1 = 0_{(N-1)\times r}$, we have

$$\begin{aligned}&\sqrt{T}\begin{bmatrix} \operatorname{vec}(\tilde{Z}_1) \\ \operatorname{vec}(\tilde{Z}_2 - M_2) \end{bmatrix} \\ &\quad \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0_{(N-1)r} \\ 0_{(N-1)(K-r)} \end{bmatrix}, \begin{bmatrix} I_{(N-1)r} & 0_{(N-1)r\times(N-1)(K-r)} \\ 0_{(N-1)(K-r)\times(N-1)r} & I_{(N-1)(K-r)} \end{bmatrix}\right),\end{aligned} \tag{A.32}$$

where $M_2 = MC_2$ and all columns of M_2 are nonzero vectors. From the fact that $W = T(Z'Z)$ and $\tilde{W} = T(\tilde{Z}'\tilde{Z})$ share the same eigenvalues, it is sufficient to obtain the limiting distribution of the smallest eigenvalue of $\tilde{W}$, which is equal to the reciprocal of the largest eigenvalue of

$$\tilde{W}^{-1} = \begin{bmatrix} \tilde{W}^{11} & \tilde{W}^{12} \\ \tilde{W}^{21} & \tilde{W}^{22} \end{bmatrix}. \tag{A.33}$$

Using the formula for the inverse of a partitioned matrix, we have

$$\tilde{W}^{11} = \left(\sqrt{T}\tilde{Z}_1'\left[I_{N-1} - \tilde{Z}_2\left(\tilde{Z}_2'\tilde{Z}_2\right)^{-1}\tilde{Z}_2'\right]\sqrt{T}\tilde{Z}_1\right)^{-1} \xrightarrow{d} \mathcal{W}_r(N-K-1+r, I_r)^{-1}, \tag{A.34}$$

$$\tilde{W}^{12} = -\tilde{W}^{11}\tilde{Z}_1'\tilde{Z}_2(\tilde{Z}_2'\tilde{Z}_2)^{-1} = O_p(T^{-1/2}), \tag{A.35}$$

$$\tilde{W}^{22} = (T\tilde{Z}_2'\tilde{Z}_2)^{-1} + (\tilde{Z}_2'\tilde{Z}_2)^{-1}(\tilde{Z}_2'\tilde{Z}_1)\tilde{W}^{11}(\tilde{Z}_1'\tilde{Z}_2)(\tilde{Z}_2'\tilde{Z}_2)^{-1} = O_p(T^{-1}), \tag{A.36}$$

where $\mathcal{W}_r(N-K-1+r, I_r)$ denotes the Wishart distribution with $N-K-1+r$ degrees of freedom and a scaling matrix I_r. Therefore, the limiting distribution of the largest eigenvalue of $\tilde{W}^{-1}$ is the same as the limiting distribution of the largest eigenvalue of $\tilde{W}^{11}$. Equivalently, the smallest eigenvalue of $T\tilde{\Omega}$ has the same limiting distribution as w_r, the smallest eigenvalue of $W_r \sim \mathcal{W}_r(N-K-1+r, I_r)$, where W_r denotes the limit of the inverse of $\tilde{W}^{11}$.

We now show that $\Pr[w_r \le a] \ge \Pr[x_{N-K} \le a]$, where $x_{N-K} \sim \chi^2_{N-K}$. When $r = 1$, $\mathcal{J} \xrightarrow{d} w_1 \sim \chi^2_{N-K}$. When $r \ge 2$, by the Bartlett decomposition of a Wishart matrix, we can write

$$W_r = \begin{bmatrix} W_{r-1} & W_{r-1}^{1/2} z \\ z' W_{r-1}^{1/2} & x_{N-K} + z'z \end{bmatrix}, \tag{A.37}$$

where $W_{r-1} \sim \mathcal{W}_{r-1}(N-K-2+r, I_{r-1})$, $z \sim \mathcal{N}(0_{r-1}, I_{r-1})$, and they are independent of each other and x_{N-K}. Using the fact that the eigenvalues of W_r are the same as the reciprocal of the eigenvalues of W_r^{-1}, it follows that

$$\begin{aligned} w_r &= \min_{\omega:\omega'\omega=1} \omega' W_r \omega = \Big(\max_{\omega:\omega'\omega=1} \omega' W_r^{-1}\omega\Big)^{-1} \\ &\le \big([0'_{r-1}, 1] W_r^{-1} [0'_{r-1}, 1]'\big)^{-1} = x_{N-K} \sim \chi^2_{N-K}. \end{aligned} \tag{A.38}$$

This completes the proof.

REFERENCES

ANDERSON, T. W. (1951): "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *The Annals of Mathematical Statistics*, 22, 327–351. [1654]

ANTOINE, B., H. BONNAL, AND E. RENAULT (2007): "On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood," *Journal of Econometrics*, 138, 461–487. [1643,1651]

ARELLANO, M., L. P. HANSEN, AND E. SENTANA (2012): "Underidentification?" *Journal of Econometrics*, 170, 256–280. [1645-1647]

BURNSIDE, C. A. (2016): "Identification and Inference in Linear Stochastic Discount Factor Models With Excess Returns," *Journal of Financial Econometrics*, 14, 295–330. [1642]

COCHRANE, J. H. (2005): *Asset Pricing*. Princeton: Princeton University Press. [1642]

CRAGG, J. G., AND S. G. DONALD (1996): "Testing Overidentifying Restrictions in Unidentified Models," Working Paper. [1649]

——— (1997): "Inferring the Rank of a Matrix," *Journal of Econometrics*, 76, 223–250. [1645,1651]

DOVONON, P., AND E. RENAULT (2013): "Testing for Common Conditionally Heteroskedastic Factors," *Econometrica*, 81, 2561–2586. [1649]

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. [1648]

GOSPODINOV, N., R. KAN, AND C. ROBOTTI (2014): "Misspecification-Robust Inference in Linear Asset-Pricing Models With Irrelevant Risk Factors," *Review of Financial Studies*, 27, 2139–2170. [1641]

——— (2017): "Supplement to 'Spurious Inference in Reduced-Rank Asset-Pricing Models'," *Econometrica Supplemental Material*, 85, http://dx.doi.org/10.3982/ECTA13750. [1646]

GUGGENBERGER, P., F. KLEIBERGEN, S. MAVROEIDIS, AND L. CHEN (2012): "On the Asymptotic Sizes of Subset Anderson–Rubin and Lagrange Multiplier Tests in Linear Instrumental Variables Regression," *Econometrica*, 80, 2649–2666. [1647,1649]

HALL, A. R. (2005): *Generalized Method of Moments*. Oxford: Oxford University Press. [1642]

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054. [1645]

——— (1985): "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators," *Journal of Econometrics*, 30, 203–238. [1646]

——— (2012): "Proofs for Large Sample Properties of Generalized Method of Moments Estimators," *Journal of Econometrics*, 170, 325–330. [1645]

HANSEN, L. P., AND R. JAGANNATHAN (1997): "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, 52, 557–590. [1642]

HANSEN, L. P., AND S. F. RICHARD (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica*, 50, 587–613. [1642,1646]

HANSEN, L. P., J. HEATON, AND A. YARON (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262–280. [1643,1645,1649]

HILLIER, G. H. (1990): "On the Normalization of Structural Equations: Properties of Direction Estimators," *Econometrica*, 58, 1181–1194. [1645]

KAN, R., AND C. ZHANG (1999): "Two-Pass Tests of Asset Pricing Models With Useless Factors," *Journal of Finance*, 54, 203–235. [1642]

KLEIBERGEN, F., AND S. MAVROEIDIS (2009): "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve," *Journal of Business and Economic Statistics*, 27, 293–311. [1645]

KLEIBERGEN, F., AND R. PAAP (2006): "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126. [1645]

KLEIBERGEN, F., AND Z. ZHAN (2015): "Unexplained Factors and Their Effects on Second Pass R-Squared's," *Journal of Econometrics*, 189, 101–116. [1642]

MANRESA, E., F. PEÑARANDA, AND E. SENTANA (2016): "Empirical Evaluation of Overspecified Asset Pricing Models," Working Paper. [1646,1647]

NEWEY, W. K., AND R. J. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–255. [1643,1649]

PEÑARANDA, F., AND E. SENTANA (2015): "A Unifying Approach to the Empirical Evaluation of Asset Pricing Models," *Review of Economics and Statistics*, 97, 412–435. [1642,1647]

ROBIN, J.-M., AND R. J. SMITH (2000): "Tests of Rank," *Econometric Theory*, 16, 151–175. [1645]

SARGAN, J. D. (1958): "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415. [1654]

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM With Weak Identification," *Econometrica*, 68, 1055–1096. [1651]

Co-editor Lars Peter Hansen handled this manuscript.

Manuscript received 31 August, 2015; final version accepted 31 May, 2017; available online 8 June, 2017.