Size Matters: Covariance Matrix Estimation Under the Alternative

Jason Allen*

Department of Economics, Queen's University at Kingston

Department of Monetary and Financial Analysis, The Bank of Canada

September 29, 2006

ABSTRACT

The purpose of this paper is to investigate, using Monte Carlo methods, whether Hall's (2000) centered test of overidentifying restrictions for parameters estimated by Generalized Method of Moments (GMM) is more powerful, once the test is size-adjusted, than the standard test introduced by Hansen (1982). The Monte Carlo evidence shows that very little size-adjusted power is gained over the standard uncentered calculation.

Keywords: Size, Power, Generalized Method of Moments, Overidentifying Restrictions.

JEL *classification:* C15, C52, G12

1. Introduction

Generalized Method of Moments (GMM) (Hansen, 1982) is widely used in applied economics to estimate and test macroeconomic models. In terms of testing for model misspecification, the most popular test is Hansen's (1982) *J*-test for overidentifying restrictions. While the test has widespread use, Altonji and Segal (1996), and Hall and Horowitz (1996), among others, have documented that it frequently over-rejects in small samples. Most recent research efforts have been directed at improving the size performance of the *J*-test. On the other hand, Hall (2000) proposes a centered version of the popular *J*-test with the aim of gaining power. Smith (1999) provides Monte Carlo evidence that the standard *J*-test can have low size-adjusted power in a nonlinear asset pricing model. Hall (2000)'s Monte Carlo evidence shows a modest power gain of 10 per cent assuming a linear data generating process.

Hall observes that the standard estimator of the long-run covariance matrix with unknown heteroscedasticity and serial correlation (HAC) used in Hansen's test statistics is consistent only under the null. If the null hypothesis is false, power could fall from using an inconsistent estimate. While power is not usually a problem in most practical situations, it is worthwhile to pursue the performance of the statistic under the alternative. Hall uses a two-step procedure to construct a new HAC matrix. The procedure amounts to subtracting the sample mean of the moment conditions from the second step weighting matrix. This gives a consistent estimator both under the null and alternative hypotheses. Hall's Monte Carlo experiments show a considerable size distortion of the centered calculation over the uncentered calculation as well as a 10 per cent increase in the number of rejections when the null is false.

This paper presents Monte Carlo evidence on the properties of the two test statistics to show that the size-adjusted power gain of Hall's test is not significantly greater than the standard *J*-test. Two experimental designs are considered. The first is identical to Hall's and involves independent data. The second experiment augments the data generating process with serially correlated data, which better mimics most applications. Unfortunately, the Monte Carlo findings indicate that the centered *J*-test once adjusted for its size distortion is only a marginal improvement over the uncentered counterpart.

2. Generalized Method of Moments Estimation

GMM estimates the parameters of a model, matching the moments of the theoretical model to those of the data as closely as possible. A weighting matrix determines the relative importance of matching each moment.

Let $\mathbf{X} = (x_1, \ldots, x_t)$, where $x_i \in \mathbb{R}^k$ is a $k \times 1$ random variable, and $t = 1, \ldots T$, be a set of observables from a stationary sequence. Suppose for some true parameter-value θ_0 ($k \times 1$) the following moment conditions (m equations) hold and $m \geq k$:

$$E\left[g(x_t, \theta_0)\right] = 0. \tag{1}$$

This is, of course, the usual set-up for GMM and leads to the estimator:

$$\hat{\theta}_T = \arg\min_{\theta \in \Theta} \left(T^{-1} \sum_{t=1}^{T} g(x_t, \theta)\right)' W_T \left(T^{-1} \sum_{t=1}^{T} g(x_t, \theta)\right), \tag{2}$$

where the positive semi-definite weighting matrix, W_T converges to a positive definite matrix of constants. The GMM estimator $\hat{\theta}$ is consistent for any arbitrary weighting matrix, subject to some regularity conditions. Hansen shows that the optimal weighting matrix converges to S_T^{-1} where $S_T = \sum_{j=-\infty}^{\infty} E g_t(\theta_0) g_{t-j}(\theta_0)' \equiv \Gamma_0 + \sum_{j=0}^{\infty} (\Gamma_j + \Gamma_j')$.

The estimation is usually done in two steps. The initial weighting matrix uses the instruments and the final estimate uses the optimal weighting matrix. The centered and uncentered estimators differ in the second stage. The centered estimator subtracts the sample mean from the moment condition which is non-zero in expectation under the alternative. The centered HAC estimator is consistent under the null and alternative, whereas the standard estimator is consistent only under the null.

2.1. Hypothesis Testing

A primary objective of this paper is to compare the size-adjusted power of Hall's centered J-test with that of the standard J-test introduced by Hansen. The standard way of testing is to take the second step estimate of the parameters, $\hat{\theta}_T$, and construct a test statistic that is distributed χ^2_{m-k}:

$$\hat{J}_T = T g_T(\hat{\theta}_T)' \hat{S}_T^{-1} g_T(\hat{\theta}_T). \tag{3}$$

The centered test statistic is essentially the same, except $g_T(\hat{\theta}_T)$ is demeaned for the covariance calculation. However, Hall demonstrates that the centered J-test is more powerful than the standard J-test in large samples. This is because the HAC matrix continues to be consistent under the alternative. I wish to examine the finite sample power issue when the two tests are size-adjusted.

3. Monte Carlo Experiment

Hall demonstrates an asymptotic gain in the divergence of the centered J-test over the uncentered test for a non-local alternative. He also presents Monte Carlo results that suggest this test is 10 per cent more powerful that Hansen's original test. An unrealistic feature of Hall's Monte Carlo design is the independence of the data so that there is no need to calculate a weight matrix based on autocovariances. Hall's comparisons are based on asymptotic critical values for the centered test, even though it has considerably more size distortion than does the standard test. I replicate Hall's experiment below and find that the power gain is almost completely lost once there is an adjustment for size. I also consider simple simulations using a model with dependent data, that nests Hall's experiments:

$$
\begin{aligned}
y_t &= x_t + \gamma z_{1,t} + \mu_t, \\
x_t &= z_{1,t} + z_{2,t} + \varepsilon_t, \\
z_{1,t} &= \rho_1 z_{1,t-1} + \omega_{1,t}, \\
z_{2,t} &= \rho_2 z_{2,t-1} + \omega_{2,t},
\end{aligned}
\tag{4}
$$

with $t = 1, 2, ..., T$, and $(z_{1,t}, z_{2,t}, \mu_t, \varepsilon_t)' \sim N(0, \Sigma)$ with Σ having elements $\sigma_{ii} = 1$, $i = 1, 2, 3, 4$ and $\sigma_{12} = \sigma_{34} = 0.5$. For a model with independent data ρ_1 and ρ_2 are set equal to zero. For dependent data $\rho_1 = \rho_2 = 0.9$. The following moment condition is tested:

$$E[z_t(y_t - x_t\theta)] = 0. \tag{5}$$

The model is estimated by GMM where $\hat{S}_T = \Omega_0 + \sum_{j=1}^{N} \hat{k}_j(\hat{\Omega}_j + \hat{\Omega}_j')$ is a consistent estimator of S_T. The kernel weights are determined using the Newey and West (1994) automatic selection method with $N = c \times int[(T/100)^{2/9}]$ and $c = 4, 12$. The moment condition holds only if $\theta = 1$ and $\gamma = 0$. I vary γ from 0.00 to 10.00 to measure the power of the J-test with 10,000 replications and sample size, $T = 300$. Results are presented in Tables 1 and 2. The median test statistic and the power of each test are reported in Table 1 for $i.i.d$ data and Table 2 for serially correlated data. The median of the automatic lag length criteria of Newey and West using the Bartlett kernel is also reported for the standard and centered tests (b, b_c). For both $i.i.d$ and serially correlated data, the power of the centered J-test (J_c) is greater than the power of the standard J-test (J). Interestingly, the median test statistic and rejection frequency of the test statistic are larger when the null hypothesis is true. This is the case even though Hall's procedure is for estimation under the alternative hypothesis. The centered J-test may therefore over-reject a true null hypothesis.

Since the centered test has a greater size distortion than the non-centered test I compare the tests using size-power tradeoff curves as described by Davidson and MacKinnon (1998). Figure 1 to 4 present several of these curves. The dotted line represents values for the centered test and the solid line represents the standard test. The 45° line represents a test with size equal to power. A curve below the 45° line represents a biased test and a curve above the line represents a test with power greater than size. The tradeoff curves are generated by varying the critical value of the tests. For each critical value, size is measured as the percentage of rejections under the null hypothesis, and power is measured as the percentage of rejections under the alternative hypothesis. Thus power is adjusted by size. Figure 1 presents size-power tradeoff curves with $\gamma = 0.125$ for the case of $i.i.d$ observations. The centered test is only slightly more powerful than the standard test, and only at very low size. Figure 2 has the same γ but with dependent data. Again the centered test is slightly more powerful than the standard test. The overidentifying restrictions tests are more powerful with independent data than dependent data. Figure

5

<div align="center">

Table 1
Summary Statistics for Overidentifying Restrictions Tests, $\rho_1 = \rho_2 = 0$

</div>

c	γ	Med(b_T)	Med(J)	P(J)	Med(b_{cT})	Med(J_c)	P(J_c)
4	0.000	5	0.475	0.045	5	0.482	0.060
	0.125	5	1.330	0.183	5	1.386	0.219
	0.250	6	4.698	0.598	5	5.394	0.639
	0.375	7	9.048	0.926	5	12.170	0.940
	0.500	8	12.703	0.996	5	21.086	0.997
	10.00	13	18.489	1.000	5	111.21	1.0000
12	0.000	14	0.509	0.040	15	0.527	0.085
	0.125	14	1.344	0.166	15	1.508	0.250
	0.250	16	4.193	0.553	15	5.879	0.666
	0.375	20	6.784	0.897	15	13.152	0.944
	0.500	23	8.238	0.993	15	22.628	0.997
	10.00	28	9.806	1.000	15	117.92	1.000

Note: Med($\cdot$) denotes the median of the statistic, and $P(\cdot)$ denotes the probability of rejecting the null with nominal size 5 per cent. c is a truncation parameter in the estimation of the long-run covariance estimator.

<div align="center">

Table 2
Summary Statistics for Overidentifying Restrictions Tests, $\rho_1 = \rho_2 = 0.9$

</div>

c	γ	Med(b_T)	Med(J)	P(J)	Med(b_{cT})	Med(J_c)	P(J_c)
4	0.000	6	0.493	0.040	6	0.503	0.058
	0.125	6	2.254	0.294	6	2.431	0.340
	0.250	7	6.432	0.805	6	8.104	0.842
	0.375	10	9.333	0.983	7	14.163	0.990
	0.500	11	10.738	0.999	8	19.085	0.999
	10.00	13	12.298	1.000	12	30.114	1.0000
12	0.000	13	0.539	0.034	15	0.568	0.083
	0.125	15	2.195	0.248	15	2.619	0.368
	0.250	18	5.238	0.742	14	8.296	0.845
	0.375	21	6.786	0.959	14	13.991	0.988
	0.500	23	7.422	0.992	14	18.606	0.999
	10.00	26	8.147	0.999	16	29.245	1.000

Note: Med($\cdot$) denotes the median of the statistic, and $P(\cdot)$ denotes the probability of rejecting the null with nominal size 5 per cent. c is a truncation parameter in the estimation of the long-run covariance estimator.

3 presents size-power tradeoff curves with $\gamma = 0.250$ for the case of *i.i.d* observations. Figure 4 has $\gamma = 0.250$ and dependent data. The tests are more powerful for independent data than dependent data.

This may be because estimation is more difficult with serially correlated data. The Monte Carlo results indicate that once adjusted for size the centered J-test is not as powerful as initially believed.

4. Conclusion

Hall (2000) has suggested using a centered J-test to increase the power of the existing J-test for overidentification. Hall presents results that suggest that his test is 10 per cent more powerful that Hansen's (1982) original test. In this paper, I show that in finite samples the gain in power is mostly due to a greater size distortion present in the centered test.

The asymptotic gain in the divergence rate of the centered J-test over the uncentered version can nevertheless be important, as pointed out by Hall (2005), for constructing moment selection procedures as in Andrews (1999). Similar arguments can be found in Hall et al. (2003), suggesting that one should use a demeaned covariance estimator to implement Andrews' information criterion for choosing moments.

Given the performance of Hall's J-test for model specification once it is adjusted for size, however, one must search elsewhere for size-corrected power gains. One avenue is to use more robust estimators than GMM and/or more robust statistics. With respect to estimation, advances with generalized empirical likelihood (GEL) methods provide one alternative to using GMM. For example, see the statistics derived in Imbens et al. (1998). GEL has the advantage over GMM in that (i) the asymptotic bias of the parameter estimates does not increase in the number of overidentifying restrictions, and (ii) the moment conditions hold exactly in finite samples. This second point implies there is no need to re-center the moment conditions when estimating the long-run covariance matrix.

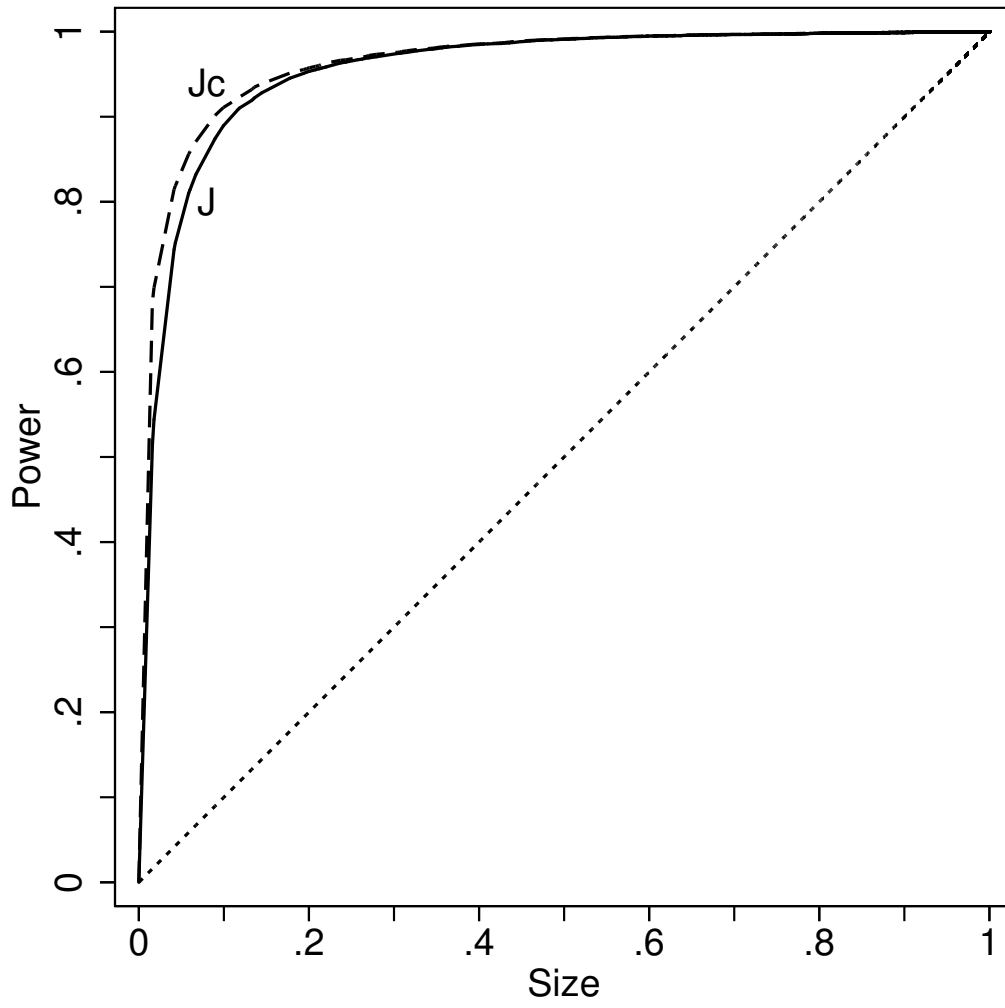**Figure 1.** Size-Power Curves: Independent data
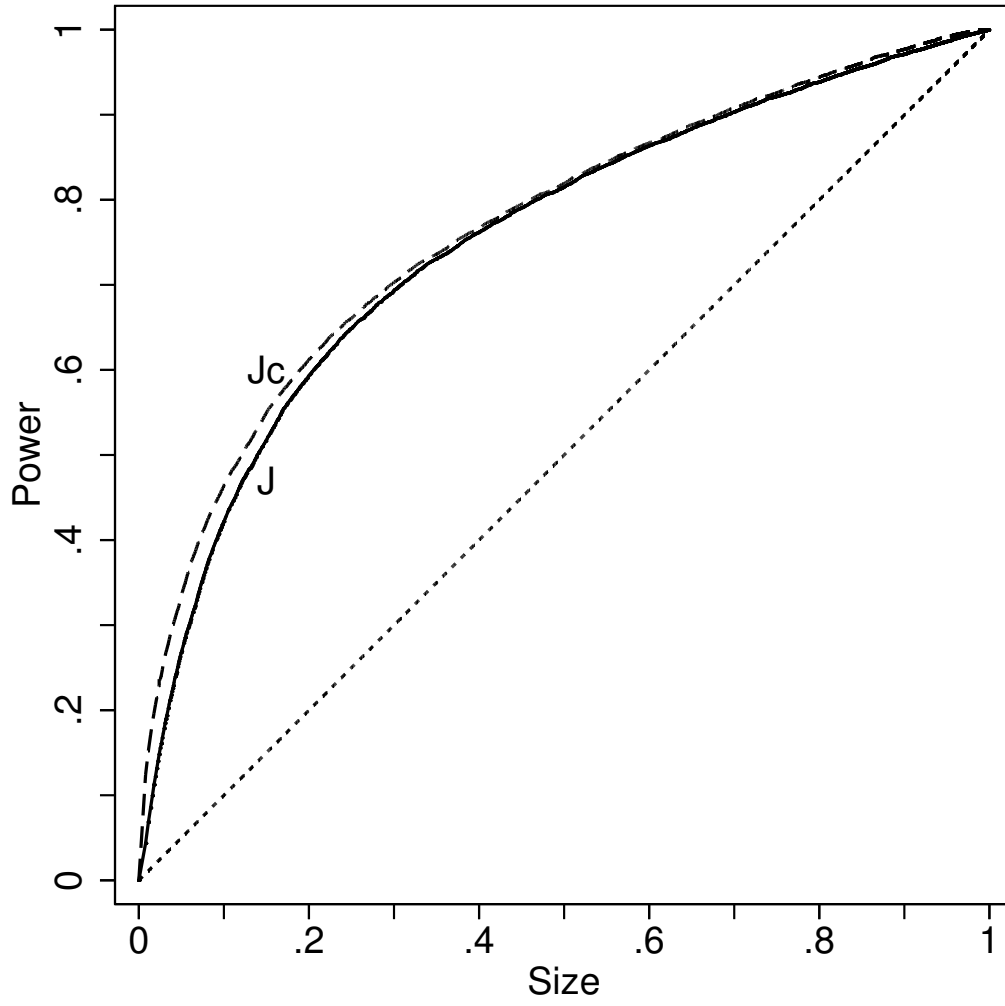
Note: $\gamma = 0.125$, $T = 300$, Replications= 10000. Jc is the centered test for overidenti-
fying restrictions and is given by the long-dashed line. J is the standard test and given
by the straight line. A test above the 45 degree line is one with power greater than size.

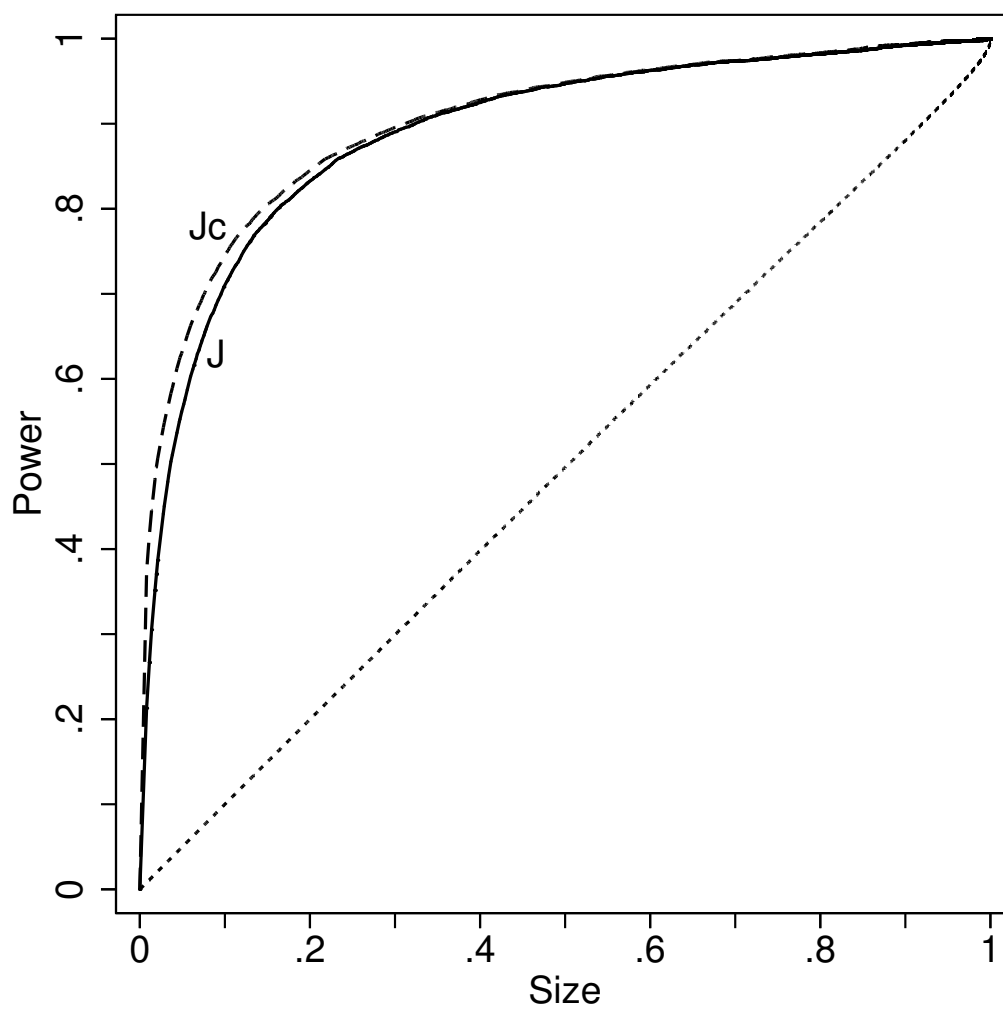**Figure 2.** Size-Power Curves: Dependent Data
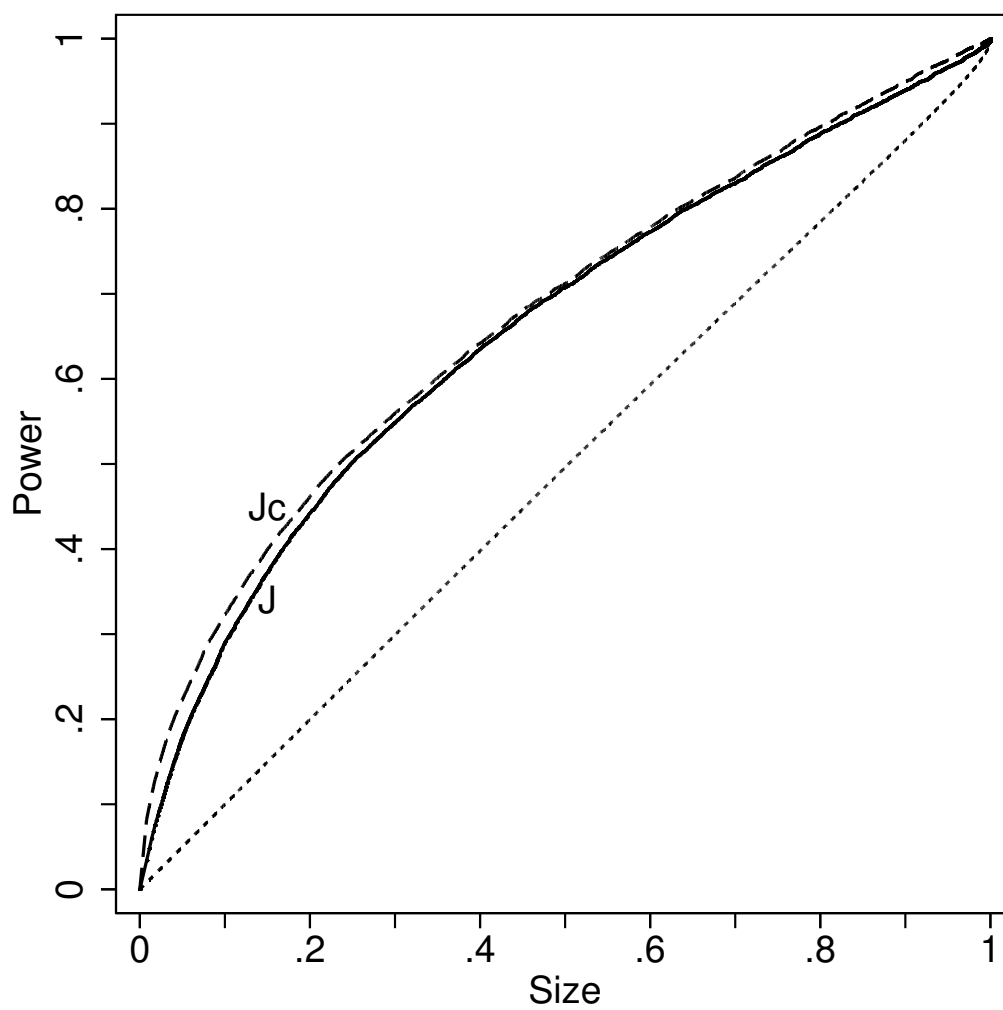
Note: $\gamma = 0.125$, $T = 300$, Replications= 10000. Jc is the centered test for overidenti-
fying restrictions and is given by the long-dashed line. J is the standard test and given
by the straight line. A test above the 45 degree line is one with power greater than size.

Figure 3. Size-Power Curves: Independent data

Note: $\gamma = 0.250$, $T = 300$, Replications= 10000. Jc is the centered test for overidenti-
fying restrictions and is given by the long-dashed line. J is the standard test and given
by the straight line. A test above the 45 degree line is one with power greater than size.

Figure 4. Size-Power Curves: Dependent Data

Note: $\gamma = 0.250$, $T = 300$, Replications= 10000. *Jc* is the centered test for overidenti-
fying restrictions and is given by the long-dashed line. *J* is the standard test and given
by the straight line. A test above the 45 degree line is one with power greater than size.

References

Altonji, J.G., and L.M. Segal (1996) 'Small-sample bias in GMM estimation of covariance structures.' *Journal of Business & Economic Statistics* 14, 353–366

Andrews, D.W.K. (1999) 'Consistent moment selection procedures for generalized method of moments estimation.' *Econometrica* 67, 543–564

Davidson, Russell, and James MacKinnon (1998) 'Graphical methods for investigating the size and power of test statistics.' *The Manchester School* 66, 1–26

Hall, A.R. (2000) 'Covariance matrix estimation and the power of the overidentifying restrictions test.' *Econometrica* 68, 1517–1527

— (2005) *Generalized Method of Moments* (New York: Oxford University Press)

Hall, A.R., A. Inoue, and F.P.M. Peixe (2003) 'Covariance matrix estimation and the limiting behavior of the overidentifying restrictions tests in the presence of neglected structural instability.' *Econometric Theory* 19, 962–983

Hall, P., and J.L. Horowitz (1996) 'Bootstrap critical values for tests based on generalized-method-of-moments estimators.' *Econometrica* 64, 891–916

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

Imbens, G.W., R.H. Spady, and P. Johnson (1998) 'Information theoretic approaches to inference in moment condition models.' *Econometrica* 66, 333–357

Newey, W.K., and K.D. West (1994) 'Automatic lag selection in covariance matrix estimation.' *Review of Economic Studies* 61, 631–654

Smith, David (1999) 'Finite-sample properties of tests of the Epstein-Zin asset pricing model.' *Journal of Econometrics* 93, 113–148