Increasing the Power of Moment-based Tests

TIEMEN WOUTERSEN*
UNIVERSITY OF ARIZONA

December 29th, 2022

ABSTRACT. This paper shows how to increase the power of the Hansen (1982) test for the case where only a subset of the exclusion restrictions is used. The 'ignored' exclusion restrictions are used to derive a new estimator for the covariance matrix, which has a different probability limit than the usual one when the model is false. If the null hypothesis is true, then the proposed test has the same distribution as the existing ones in large samples. If the hypothesis is false, then the proposed test statistic is larger with probability approaching one as the sample size increases in several important examples. Simulations show that the improvement can be dramatic in some cases. As the Hansen (1982) test is very popular in empirical work, including testing the validity of Euler equations, we expect the current results to be useful as well.

Keywords: Overidentification test, Hansen test, Power of tests.
JEL Codes: C01, C14, C18, C41

1. INTRODUCTION

Testing restrictions is important in empirical economics and other empirical research. Such tests help the researcher to evaluate whether an economic model is credible. Many restrictions can be stated using the generalized method of moments framework, i.e. Hansen's (1982) overidentification test or his test for whether the moments holds for a particular parameter value.

*I thank Jim Powell for insightful discussions about generalized method of moments and other semiparametric methods. I thank Donald Andrews, Xiaohong Chen, Hide Ichimura, Jerry Hausman, Michael Jansson, and Whitney Newey for helpful discussions, and Miriam Arden and Paige Pearcy for excellent research assistance. Further, I thank seminar participants at MIT, University of Wisconsin, and Yale University. All remaining errors are my own. Comments are welcome: woutersen@email.arizona.edu.

The purpose of this paper is to derive moment-based tests that are more powerful than the existing ones in important applications. In particular, we consider the case where only a subset of the exclusion restrictions is used as moments. We impose the exclusion restrictions and the restrictions of the null hypothesis when estimating the asymptotic variation of the test statistic. In particular, in large samples, our estimator for the asymptotic variance has the same probability limit as existing ones when the model is true, but the asymptotic variance estimate is smaller when the model is false in several important applications.[1] This causes the test statistic to be larger so that we reject a false hypothesis more often. We then use the new estimator of the asymptotic covariance matrix when calculating the Sargan (1958) and Hansen (1982) test statistics.

These moment-based tests are very popular in empirical economics and other empirical research because they are linked to economic theory. In particular, many economic models imply an error term that is zero in expectation given an information set. Empirical researchers then use this error term to build moments in, for example, Euler equations. Hansen (2014) provides an overview.

A feature of the proposed estimator of the covariance matrix is that it can be inconsistent under the alternative. In particular, Hansen (1982) uses the sample analogue of the covariance matrix. This estimates an unconditional covariance matrix. This paper estimates a conditional covariance matrix. When the null hypothesis is true then these two matrices are the same. However, in general these matrices may be different so the estimand of Hansen (1982) covariance estimator and the one proposed here may differ. Newey (1985) considers the power of the Hansen (1982) test and its special case, the Sargan (1958) test, when the sample analogue of the covariance matrix is used. Further, Lehmann and Romano (2005) and Romano, Shaikh, and Wolf (2010) provide overviews of testing in statistics and econometrics.

This paper is organized as follows: Section 2 presents an example and simulations, Section 3 gives the theorem, and Section 4 concludes.

[1] Smaller in this context means that the asymptotic variance is smaller for the scalar case and that the difference between the asymptotic covariance matrices is negative semi-definite with at least one strictly negative element on the diagonal for the vector case.

2. EXAMPLE: CONSUMPTION-BASED ASSET PRICING MODEL

This section presents an example that shows how imposing the restrictions of the null hypothesis can yield a more powerful test. Consider the Consumption-based Asset Pricing Model. This model is used by financial economists to explain how assets are priced, and it is used by macro economists to explain the evolution of consumption spending. Hansen and Singleton (1982) use this model and assume a Constant Relative Risk Aversion (CRRA) utility function,

$$U(c_t) = \{ \begin{array}{l} \frac{c_t^{1-\kappa}-1}{1-\kappa}, \text{ for } \kappa \neq 1, \kappa > 0 \\ \ln(c_t), \text{ for } \kappa = 1. \end{array}$$

Suppose that the consumer maximizes expected discounted utility given an information set Υ_t and using discount factor δ. Further assume that the consumer buys a unit of a portfolio in period t at price p_t, and that the payoff of this unit is r_{t+1} in period $t+1$. Then, the first order condition, or Euler equation, is

$$E[(\delta r_{t+1}/p_t)(c_{t+1}^{-\kappa}/c_t^{-\kappa})|\Upsilon_t] - 1 = 0.$$

The information set Υ_t contains all the variables that are known to the consumer at time t. In other words, the consumer takes these variables into account when making her consumption decision. Thus, the following residual is orthogonal to the variables in Υ_t. In particular, we can define the residual $\varepsilon_t(\kappa, \delta)$ as

$$\varepsilon_t(\kappa, \delta) = (\delta r_{t+1}/p_t)(c_{t+1}^{-\kappa}/c_t^{-\kappa}) - 1.$$

The expectation of this residual, given the information set Υ_t, is zero. This suggests the following moment vector function,

$$g(\kappa, \delta) = \frac{\sum_t}{T} g_t(\kappa, \delta) \text{ where } g_t(\kappa, \delta) = Q_t' \varepsilon_t(\kappa, \delta), \tag{1}$$

where Q_t is a column vector with regressors that are contained in Υ_t. Havránek (2015) reviews 169 published papers that estimate the elasticity of intertemporal substitution in consumption, i.e. $1/\kappa$ in our notation. After correcting for publication bias, Havránek (2015) establishes a range for these elasticities that implies that κ is between 2.5 and 3.3. Thus, an applied researcher may want to test the Consumption-based Asset Pricing Model with $\kappa = 3$ while choosing the discount factor $\delta = 0.99$ per year. More generally, she may want to test the parameter values κ_{H_0} and δ_{H_0}. We can now state the null hypothesis that the moments in equation (1) has expectation zero at $\{\kappa_{H_0}, \delta_{H_0}\}$,

$$\mathrm{H}_0 : E\{g_t(\kappa_{H_0}, \delta_{H_0})|\Upsilon_t\} = 0 \text{ for all } t,$$
$$\mathrm{H}_1 : E\{g_t(\kappa_{H_0}, \delta_{H_0})|\Upsilon_t\} \neq 0 \text{ for some } t.$$

The set Υ_t can contain many variables. Researchers may use a subset of the regressors that are in Υ_t since they have some idea which variables are relevant. This selection can increase the power of their test, as illustrated by our simulations. For example, one chooses 2 regressors out of 30 regressors in Υ_t and ignores the other 28. The main contribution of this paper is that it uses the 28 variables that are ignored to increase the power of the specification test. To illustrate this, consider the following data generating process,

$$\varepsilon_t(\kappa_{H_0}, \delta_{H_0})|\Upsilon_t \sim N(0, \sigma^2), \tag{2}$$

where $\varepsilon_t(\kappa_{H_0}, \delta_{H_0})$ is independently distributed across time periods. Consider the moment vector function that only uses X_t,

$$\tilde{g}(\kappa, \delta) = \frac{\sum_t}{T} X_t' \varepsilon_t(\kappa, \delta).$$

We can test the null hypothesis using the Hansen (1982) test. In particular, using only two variables (assuming that one of the variables is a constant) and assuming homoscedasticity yields

$$T_{\text{Hansen, 2 variables}} = \frac{\tilde{g}(\kappa_{H_0}, \delta_{H_0})' \{\frac{\sum_t}{T} X_t X_t'\}^{-1} \tilde{g}(\kappa_{H_0}, \delta_{H_0})/2}{\tilde{e}'\tilde{e}/(T-2)}, \tag{3}$$

where $\tilde{e}$ is the vector of residuals from regressing $\varepsilon_t(\kappa_{H_0}, \delta_{H_0})$ on X_t. In this case, $T_{\text{Hansen, 2 variables}}$ has an F-distribution with $\{2, T-2\}$ degrees of freedom under H_0 (see the Appendix for details).

An alternative to this is to use more variables from Υ_t. For example, one can choose $K + 2$ variables, including the ones contained in X_t. Let S_t denote the values of these $K + 2$ variables. This gives

$$T_{\text{Hansen, K+2 variables}} = \frac{\dot{g}(\kappa_{H_0}, \delta_{H_0})' \{\frac{\sum_t}{T} S_t S_t'\}^{-1} \dot{g}(\kappa_{H_0}, \delta_{H_0})/(K+2)}{e'e/(T-K-2)}, \tag{4}$$

where $\dot{g}(\kappa_{H_0},\delta_{H_0}) = \frac{\sum_t}{T} S_t' \varepsilon_t(\kappa_{H_0},\delta_{H_0})$, and e is the vector of residuals from regressing $\varepsilon_t(\kappa_{H_0},\delta_{H_0})$ on S_t. In this case, $T_{\text{Hansen, K+2 variables}}$ has an F-distribution with $\{K+2, T-K-2\}$ degrees of freedom under H_0. A problem with using many regressors from Υ_t is that the Hansen test loses power, even if $\varepsilon_t(\kappa_{H_0},\delta_{H_0})$ is normally distributed and homoscedastic. If a researcher adjusts the Hansen test in equation (4) to allow for heteroscedasticity, then a further complication is that the size may be incorrect, even if $\varepsilon_t(\kappa_{H_0},\delta_{H_0})$ is normally distributed and homoscedastic. Our simulations illustrate these complications.

Note that under the null hypothesis, the denominators in equation (3) and (4), i.e. $\tilde{e}'\tilde{e}/(T-2)$ and $e'e/(T-K-2)$, have the same expectation for any $T > K+2$ and converge to the same probability limit, σ^2, as T increases. However, if the variables in S_t that are not in X_t are correlated with $\varepsilon_t(\kappa_{H_0},\delta_{H_0})$, then the probability limit of $e'e/(T-K-2)$ is smaller than the probability limit of $\tilde{e}'\tilde{e}/(T-2)$. This suggests that we should replace the denominator in equation (3) by $e'e/(T-K-2)$, and that is what our new test does in this case, giving

$$T_{\text{New}} = \frac{\tilde{g}(\kappa_{H_0},\delta_{H_0})'\{\frac{\sum_t}{T} X_t X_t'\}^{-1}\tilde{g}(\kappa_{H_0},\delta_{H_0})/2}{e'e/(T-K-2)}. \tag{5}$$

Here, T_{New} has an F-distribution with $\{2, T-K-2\}$ degrees of freedom under H_0. Note that the critical value of T_{New} is very close to the critical value of $T_{\text{Hansen, 2 variables}}$ when the degrees of freedom, $T-K-2$, is greater than or equal to 100. These critical values converge to each other (and to half of the the critical value of the χ^2-distribution with two degrees of freedom).

The following simulations illustrate that T_{New} is, in general, more powerful than $T_{\text{Hansen, 2 variables}}$ and $T_{\text{Hansen, K+2 variables}}$. The error term $\varepsilon_t(\kappa_{H_0},\delta_{H_0})$ is a function of observables and the parameter values that are under consideration. So we consider $\varepsilon_t(\kappa_{H_0},\delta_{H_0})$ to be observed. In the example, $\varepsilon_t(\kappa_{H_0},\delta_{H_0})$ has a martingale difference sequence property with respect to the information set Υ_t, and for the simulations we use this property and also make additional assumptions. Consider the following data generating process,

$$\varepsilon_t(\kappa_{H_0},\delta_{H_0}) = \alpha + \beta B_t + W_t\gamma + \eta_t, \tag{6}$$

where the data is independently distributed across time periods, $B_t \sim N(0,1)$, $W_t|B_t \sim N(0, I_K)$, and $\eta_t|B_t, W_t \sim N(0, \sigma^2)$. We first test whether the coefficients in the lasts equation are zero. In particular, we use the test statistics $T_{\text{Hansen, 2 variables}}$ and $T_{\text{Hansen, K+2 variables}}$ to test whether $\varepsilon_t(\kappa_{H_0}, \delta_{H_0})$ has mean zero and is uncorrelated with the regressors (two respectively $K+2$ regressors, including the constant). Thus, $T_{\text{Hansen, 2 variables}}$ tests $H_0 : \alpha = \beta = 0$ versus $H_1 : \alpha \neq 0$, or $\beta \neq 0$, and $T_{\text{Hansen, 2 variables}}$ tests $H_0 : \alpha = \beta = \gamma = 0$ versus $H_1 : \alpha \neq 0$, $\beta \neq 0$, or $\gamma \neq 0$. Both test whether the Euler equation holds and so does the proposed test (using the same null and alternative hypothesis as $T_{\text{Hansen, 2 variables}}$). For the overidentification test, we estimate α and then test whether the moments are valid. When we generate the data, we use $\alpha = 0$, and $\sigma^2 = 1 - \gamma'\gamma$. Using $\sigma^2 = 1 - \gamma'\gamma$ yields that the variation of the residual term when only a constant and B_t are used, $W_t\gamma + \eta_t$, is constant in γ. The first three elements of γ can be nonzero, as denoted in the table, and the other elements are zero. The simulations below show that the proposed test improves on the Hansen test.

Table 1: Rejection Frequencies, size is 0.05

N	K	β	γ	Hansen K+2	Hansen 2 moments	New Test
50	15	0.4	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	0.54695	0.72873	0.77007
100	30	0.2828	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	0.64540	0.74878	0.79627
200	60	0.2	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	0.77940	0.75884	0.80695
50	30	0.4	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	0.25943	0.72842	0.73945
100	60	0.2828	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	0.30943	0.74715	0.78080
200	120	0.2	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	0.40082	0.75982	0.80197

Results based on 100,000 simulations.

In Table 1, the size of the tests is 5%, i.e. the critical value is such that the probability of falsely rejecting the null hypothesis is 5%. The simulations show that the new test is more powerful than the existing ones. Further, the performance of the Hansen (1982) test with $K + 2$ moments worsens when the number of moments is increased, as shown in the last three rows. The Hansen (1982) test with 2 moments does not depend[2] on K, while the new test is robust against doubling the number of moments.

[2] The slight differences between the first three and last three rows for the Hansen (1982) test with two moments reflects a slight randomness from simulating. In the Appendix, we report the simulation results with size 0.01, and these are very simular to the results in the main text with size 0.05.

Table 2: Rejection Frequencies, Allowing for Heteroscedasticity

N	K	β	γ	Hansen K+2	Hansen 2 moments	New Test
50	15	0.4	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	size: 0.46971	0.74843	0.78774
100	30	0.2828	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	size: 0.61505	0.75987	0.80467
200	60	0.2	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	size: 0.80602	0.76334	0.81092

Results based on 100,000 simulations.

Table 2 shows that the Hansen test with two moments and the new test have good properties when we allow for heteroscedasticity (including good size properties; see the Appendix). The size of the Hansen test with $K + 2$ moments is too large; it varies from about 47% to 80%, i.e. well above 5%.

Table 3: Rejection Frequencies; Overidentification Test

N	K	β	γ	Hansen K+2	Hansen 2 moments	New Test
50	15	0.4	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	0.55981	0.81260	0.84085
100	30	0.2828	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	0.64736	0.83203	0.86122
200	60	0.2	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	0.78318	0.83922	0.86911
50	30	0.4	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	size: 0.46128	0.82768	0.85362
100	60	0.2828	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	size: 0.61116	0.83932	0.86648
200	120	0.2	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	size: 0.80535	0.84367	0.87231

Results based on 100,000 simulations.

Table 3 shows that the new overidentification test has more power than the existing ones. The last three rows allow for heteroscedasticity. The size of the Hansen overidentification test with $K + 2$ moments is too large; it varies from about 46% to 80%, i.e. well above 5%. Using a size of 1% yields very similar results as table 1-3 and we report these in the Appendix.

In summary, a test in our simulations is more powerful when we impose the conditions that the contributions to the moment functions are uncorrelated with the regressors. More generally, the idea is to impose the restrictions of the null and the exclusion restrictions when estimating the covariance matrix. In the following section, we generalize the above example and simulations with a theorem.

3. Theorem

The last section gave an example where the power of the Hansen (1982) test could be improved. We now generalize this example and state our theorem. We first assume that the normalized moments converges in distribution to a normally distributed random variable as in Hansen (1982). The vector constant c allows for local misspecification.

Assumption 1

Let

$$\sqrt{T}\{g(\theta_{H_0}) - \frac{c}{\sqrt{T}}\} \underset{d}{\rightarrow} N(0, \Omega) \tag{7}$$

for some vector constant c and positive definite Ω, where T is the sample size. Also, let there exist a consistent estimator for Ω, given by $\widehat{\Omega} = \Omega + o_p(1)$.

We use the information set Υ_t, which consists of all data available at time t. In the example, Υ_t consists of regressors that, under the null hypothesis, cannot predict $\varepsilon_t(\kappa_{H_0}, \delta_{H_0})$ in the sense that $E\{\varepsilon_t(\kappa_{H_0}, \delta_{H_0})|\Upsilon_t\} = 0$. We can generalize this aspect of the example by assuming that the elements of the moment vector function $g_t(\theta_{H_0})$ have expectation zero given Υ_t, that is, $E\{g_t(\theta_{H_0})|\Upsilon_t\} = 0$ for all Υ_t under the null hypothesis. More generally, let ψ_t denote this conditional expectation, i.e. $\psi_t = E\{g_t(\theta_{H_0})|\Upsilon_t\}$. The condition $E\{g_t(\theta_{H_0})|\Upsilon_t\} = 0$ for all Υ_t implies that $E(\frac{1}{T}\sum_t \psi_t'\psi_t) = 0$, and this is the version that we use. The alternative hypothesis is that some elements of the moment vector function have nonzero expectation, i.e. $E\{g_t(\theta_{H_0})|\Upsilon_t\} = \psi_t \neq 0$ for some t and Υ_t, so $E(\frac{1}{T}\sum_t \psi_t'\psi_t) > 0$, and

$$H_0 : E(\frac{1}{T}\sum_t \psi_t'\psi_t) = 0,$$
$$H_1 : E(\frac{1}{T}\sum_t \psi_t'\psi_t) > 0.$$

Note that the null hypothesis implies that $c = 0$ in equation (7). The hypothesis H_1 states that a conditional expectation is nonzero, and we can use this to reduce the variation of the Hansen test under H_1, as in the example. Further, note that

$$E[\frac{\sum_{s,t}}{T}\{g_s(\theta_{H_0}) - \psi_s\}\{g_t(\theta_{H_0}) - \psi_t\}'] = E\{\frac{\sum_t}{T} g_t(\theta_{H_0})g_t(\theta_{H_0})'\} - E(\frac{\sum_t}{T}\psi_t\psi_t')$$

under the assumption that $\{g_t(\theta_{H_0}) - \psi_t\}$ is a martingale difference sequence, i.e. $\{g_s(\theta_{H_0}) - \psi_s\} \in \Upsilon_t$ for $s < t$. Of course, this martingale difference sequence assumption also holds if the data from period s and t are independently distributed for $s \neq t$.[3] More generally, this simplification of the covariance matrix holds if $\{g_s(\theta_{H_0}) - \psi_s\}$ and $\{g_t(\theta_{H_0}) - \psi_t\}$

[3] In that case, Υ_s is not informative about data in period t given knowledge of the information set Υ_t for $s \neq t$.

are uncorrelated for every $s \neq t$. Note that $E(\frac{\sum_t}{T}\psi_t\psi_t{}')$ is positive semidefinite. Further, an estimator for the variation of the moments is $\frac{\sum_t}{T}g_t(\theta_{H_0})g_t(\theta_{H_0})'$; see for example Hansen (1982) and Newey and McFadden (1994). Using $\frac{\sum_t}{T}g_t(\theta_{H_0})g_t(\theta_{H_0})' - \frac{\sum_t}{T}\psi_t\psi_t{}'$ as an estimator for the variation of the moments yields a smaller estimate of the variation if $\psi_t'\psi_t > 0$ for some t, as in the example above.

Neither the martingale difference sequence assumption nor the no correlation assumption is necessary for the main argument of this paper, but it holds in many models. For example, the contributions to the score function in likelihood models generally have expectation zero. De Jong and Woutersen (2011) give examples of estimating functions in a dynamic model, and these functions also have this property. Newey and McFadden (1994) provide generalized method of moments examples where this property holds.

Alternatively, rather than making the martingale difference sequence assumption or the no correlation assumption, we may assume that an error term in an Euler equation is independent of regressors but can be correlated over time. This also accommodates time series models with dependent errors. In this case, $E\{\frac{\sum_{s,t}}{T}(g_s - \psi_s)(g_t - \psi_t)'\} = E(\frac{\sum_{s,t}}{T}g_s g_t') - E(\frac{\sum_{s,t}}{T}\psi_s\psi_t{}')$ under the assumption that $\Upsilon_s \subset \Upsilon_t$ for $s < t$. The assumption $\Upsilon_s \subset \Upsilon_t$ means that the information available at period s is still available in period t. This assumption holds in the example about the Euler equation. The term $E(\frac{\sum_{s,t}}{T}\psi_s\psi_t{}')$ reduces the variation of the moments if $E(\frac{\sum_{s,t}}{T}\psi_s\psi_t{}')$ is positive semidefinite and has at least one nonzero element on its diagonal. This is considerably weaker than requiring that $E(\frac{\sum_{s,t}}{T}\psi_s\psi_t{}')$ is positive definite.

A simple way to approximate the conditional expectation $\psi_t = E\{g_t(\theta_{H_0})|\Upsilon_t\}$ is by using a projection. In particular, we can calculate the moment vector functions for all periods, $\{g_1(\theta_{H_0}), g_2(\theta_{H_0}), ..., g_T(\theta_{H_0})\}$. We can then use regressors that are in the information set for all periods and regress the vectors $\{g_1(\theta_{H_0}), g_2(\theta_{H_0}), ..., g_T(\theta_{H_0})\}$ on these regressors to generate the predictions $\{\hat{g}_1(\theta_{H_0}), \hat{g}_2(\theta_{H_0}), ..., \hat{g}_T(\theta_{H_0})\}$ and the residuals $\{g_1(\theta_{H_0}) - \hat{g}_1(\theta_{H_0}), g_2(\theta_{H_0}) - \hat{g}_2(\theta_{H_0}), ..., g_T(\theta_{H_0}) - \hat{g}_T(\theta_{H_0})\}$. Using these residuals we can then estimate the matrix $var[\frac{\sum_t}{T}\{g_t(\theta_{H_0}) - \hat{g}_t(\theta_{H_0})\}]$ by the Newey-West estimator.[4] More

[4] For the Hansen (1982) overidentification test, we estimate the generalized method of moments estimator for θ, denoted by $\hat{\theta}$, and use this to calculate the moment vector functions for all periods, $\{g_1(\hat{\theta}), g_2(\hat{\theta}), ..., g_T(\hat{\theta})\}$. We then regress these on regressors that are in the information set for all periods

generally, let ϕ_t be the probability limit of an estimator for the conditional expectation, ψ_t. If we use a projection, then $\phi_t = S_t\pi$, where $\pi = \underset{T\to\infty}{plim}\ [\frac{\sum_t}{T} S_t' S_t]^{-1}[\frac{\sum_t}{T} S_t' g_t(\theta_{H_0})]$. Rather then assuming that ψ_t can be consistently estimated, it seems useful to allow for the conditional expectation ψ_t to be approximated by ϕ_t, and Assumption 2 (ii) and Assumption 2 (iv) below allow for this. In the example in the last section the regressors S_t can be used to estimate the conditional expectation of $g_t(\kappa_{H_0}, \delta_{H_0}) = \{\varepsilon_t(\kappa_{H_0}, \delta_{H_0}), X_t{\cdot}\varepsilon_t(\kappa_{H_0}, \delta_{H_0})\}'$. As X_t is part of S_t in the example, the goal is to estimate the conditional expectation of $\varepsilon_t(\kappa_{H_0}, \delta_{H_0})$. This estimate may not be consistent as is the case in a linear projection and, therefore, ϕ_t is allowed to be different from the conditional expectation ψ_t. Our final alternative assumption states that the variation of $g_t(\theta_{H_0}) - \phi_t$ is smaller than the variation of $g_t(\theta_{H_0})$. This means that ϕ_t can be a projection or other approximation of $g_t(\theta_{H_0})$ and is not required to be a conditional expectation. Rather, it assumes that under H_1, $(\Omega - var[\frac{\sum_t}{T}\{g_t(\theta_{H_0}) - \phi_t\}])$ is positive semidefinite and has at least one nonzero element on its diagonal. This condition can hold under mixing conditions; see Pötscher and Prucha (1997) for an overview of such mixing conditions.

Assumption 2

Let one of the following hold:

(i) $\{g_s(\theta_{H_0}) - \psi_s\}$ and $\{g_t(\theta_{H_0}) - \psi_t\}$ are uncorrelated for every $s \neq t$; or

(ii) $\{g_s(\theta_{H_0}) - \phi_s\}$ and $\{g_t(\theta_{H_0}) - \phi_t\}$ are uncorrelated for every $s \neq t$; if H_0 holds, then $E(\frac{1}{T}\sum_t \phi_t\phi_t') = 0$, and if H_1 holds, then $M = E[\frac{\sum_t}{T}\{2g_t(\theta_{H_0}) - \phi_t\}\phi_t']$ is positive semidefinite and $M_{kk} > 0$ for some k; or

(iii) $\Upsilon_s \subset \Upsilon_t$ for all $s < t$, $s,t = 1,...,T$, $\dot{M} = \Omega - var[\frac{\sum_t}{T}\{g_t(\theta_{H_0}) - \psi_t\}] = E(\frac{\sum_{s,t}}{T}\psi_s\psi_t')$ is positive semidefinite; if H_1 holds, then $\dot{M}_{kk} > 0$ for some k; or

(iv) $\ddot{M} = \Omega - var[\frac{\sum_t}{T}\{g_t(\theta_{H_0}) - \phi_t\}]$ is positive semidefinite; if H_0 holds, then $E(\frac{1}{T}\sum_t \phi_t\phi_t') = 0$, and if H_1 holds, then $\ddot{M}_{kk} > 0$ for some k.

The theorem that follows states that T_{New} and T_{Hansen} have the same asymptotic distribution under H_0, but that under the conditions of the theorem, T_{New} is more powerful against violations of H_0.

and construct $\{\hat{g}_1(\hat{\theta}), \hat{g}_2(\hat{\theta}), ..., \hat{g}_T(\hat{\theta})\}$ and residuals $\{g_1(\hat{\theta}) - \hat{g}_1(\hat{\theta}), g_2(\hat{\theta}) - \hat{g}_2(\hat{\theta}), ..., g_T(\hat{\theta}) - \hat{g}_T(\hat{\theta})\}$.

Theorem 1 (Specific Parameter Values)

Let Assumptions 1 and 2 hold. Let $\Lambda = var\{\frac{\sum_t}{T}(g_t(\theta_{H_0}) - \psi_t)\}$ if Assumption 2(i) or 2(iii) hold, and let $\Lambda = var\{\frac{\sum_t}{T}(g_t(\theta_{H_0}) - \phi_t)\}$ if Assumption 2(ii) or 2(iv) hold. Let Λ be positive definite, and let $\hat{\Lambda}$ be a consistent estimator of Λ, i.e. $\hat{\Lambda} = \Lambda + o_p(1)$.

(i) If H_0 is true, then (a) $T_{Hansen} = T \cdot g(\theta_{H_0})'\hat{\Omega}^{-1}g(\theta_{H_0}) \underset{d}{\rightarrow}$ χ^2-distribution with $\dim(\theta_{H_0})$ degrees of freedom, and (b) $T_{New} = T \cdot g(\theta_{H_0})'\hat{\Lambda}^{-1}g(\theta_{H_0}) \underset{d}{\rightarrow}$ χ^2-distribution with $\dim(\theta_{H_0})$ degrees of freedom.

(ii) If H_1 is true then T_{New} is more powerful then T_{Hansen} in the sense that $T_{New} > T_{Hansen}$ with probability approaching one.

(iii) If H_1 is true, and $\Lambda = \varphi\Omega$, then (a) $T_{Hansen} \underset{d}{\rightarrow}$ noncentral χ^2-distribution with $\dim(\theta_{H_0})$ degrees of freedom and noncentrality parameter $c'\Omega^{-1}c$, and (b) $0 < \varphi < 1$ and $T_{New} = \frac{T_{Hansen}}{\varphi} + o_p(1)$.

Proof: See Appendix.

In the example of the last section, Assumption 1 holds and so does Assumption 2 (i)-(iv). That is, the projection estimates the conditional expectation consistently in this example. Leaving out one of the regressors on which the conditional expectation depends would change the example but Assumption 1 and Assumption 2 (ii) and (iv) still hold so that the theorem still applies. For Hansen's (1982) overidentification test, one evaluates the moments at the generalized method of moments estimator[5] $\hat{\theta}$ rather than at θ_{H_0}. Thus, for the overidentification test, we assume that Assumption 2 holds, but with $g_t(\theta_{H_0})$ replaced by $g_t(\hat{\theta})$, ψ_t replaced by $\dot{\psi}_t = E\{g_t(\hat{\theta})|\Upsilon_t\}$, and the approximation ϕ_t (for ψ_t) replaced by $\dot{\phi}_t$ (for $\dot{\psi}_t$). We call this Assumption 2* and state it in the Appendix.

Theorem 2 (Overidentification)

Let Assumptions 1 and 2* hold. Let $\Lambda = var\{\frac{\sum_t}{T}(g_t(\hat{\theta}) - \dot{\psi}_t)\}$ if Assumption 2*(i) or 2*(iii) hold, and let $\Lambda = var\{\frac{\sum_t}{T}(g_t(\hat{\theta}) - \dot{\phi}_t)\}$ if Assumption 2*(ii) or 2*(iv) hold. Let Λ be positive definite, and let $\hat{\Lambda}$ be a consistent estimator of Λ, i.e. $\hat{\Lambda} = \Lambda + o_p(1)$.

(i) If H_0 is true, and $T_{Hansen} = T \cdot g(\hat{\theta})'\hat{\Omega}^{-1}g(\hat{\theta}) \underset{d}{\rightarrow}$ χ^2-distribution with $\dim(s)$

[5] $\hat{\theta} = \underset{\theta}{\arg\min}\; g(\theta)'\hat{\Omega}^{-1}g(\theta)$ where $\hat{\Omega} = \Omega + o_p(1)$.

degrees of freedom, then $T_{New} = T \cdot g(\hat{\theta})'\hat{\Lambda}^{-1}g(\hat{\theta}) \underset{d}{\rightarrow} \chi^2$-distribution with $\dim(s)$ degrees of freedom where s is the degree of overidentification.

(ii) If H_1 is true, and all elements of the vector $\sqrt{T}g(\hat{\theta})$ are nonzero with probability approaching one, then T_{New} is more powerful then T_{Hansen} in the sense that
$T_{New} > T_{Hansen}$ with probability approaching one.

(iii) If H_1 is true, and $\Lambda = \varphi\Omega$, then $0 < \varphi < 1$ and $T_{New} = \frac{T_{Hansen}}{\varphi} + o_p(1)$.

Proof: See Appendix.

Adjusting for the degrees of freedom does not effect the results in the theorem. However, we suggest making such a correction if the number of regressors that is used in the projection is large. Our example makes such corrections. The theorem is stated in terms of covariance matrices and conditional expectations and allows for unobservables to be dependent under Assumptions 1 and 2(iii) (or 2*(iii)), and Assumptions 1 and 2(iv) (or 2*(iv)). That is, the theorem does not require a martingale difference sequence assumption.

Further, the goal of the current paper is to create a more powerful test, not a consistent estimator for the asymptotic covariance matrix under the alternative. In particular, our estimator for the asymptotic covariance matrix can be inconsistent under the alternative. For example, in the simulation design of table 1 the asymptotic covariance matrix of the moments $\tilde{g}(\kappa_0, \delta_0) = \frac{\sum_t}{T} X_t'\varepsilon_t(\kappa_0, \delta_0)$ is $\text{plim}\{\frac{\sum_t}{T} X_t X_t'\} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. This is the true asymptotic covariance of the moments and it holds for any value of α, β, and γ. In contrast, the proposed test uses the asymptotic covariance matrix $\text{plim}\{(1-\gamma'\gamma)\frac{\sum_t}{T} X_t X_t'\} = (1 - \gamma'\gamma)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Clearly, for $0 < \gamma'\gamma < 1$, $\Lambda = (1-\gamma'\gamma)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is smaller than Ω in the sense that $\Omega - \Lambda$ is positive definite. This causes the proposed test to be more powerful.

Finally, we do not use local asymptotics in the example, since we have the exact distribution. For the simulations, we use fixed values or values that somewhat resemble local asymptotics. The motivation to use local asymptotics in Assumption 1 is to ensure that the Hansen test and the proposed test have the same distribution under H_0, and therefore, have the same critical values. The simulation design in table 1 is sparse in the sense that only three values of the vector γ are nonzero for any value of N. An alternative simulation design would be to use the same values of the parameters as in table 1 line 1, but to change the values in lines 2 and 3. In particular replacing the values of γ by $\gamma_j = 0.14$ for $j = 1, ..., 6$, (and $\gamma_j = 0$ for $j = 7, ..., 30$) in line 2 and $\gamma_j = 0.1$ for $j = 1, ..., 12$ (and $\gamma_j = 0$ for $j = 13, ..., 60$) in line 3 yields the same value of $\gamma'\gamma$ as in table 1 and approximately the same results for the new test. This local to zero simulation design is in the spirit of the many instruments asymptotics proposed by Kunitomo (1980), expanded by Bekker (1994), and more recently used by Hausman, Newey, Woutersen, Chao, and Swanson (2012).

4. Conclusion

This paper shows how to increase the power of the Hansen (1982) test by using a new estimator of the asymptotic covariance matrix. We impose the restrictions of the null hypothesis and the model when estimating this asymptotic covariance matrix. In large samples, our new estimator has the same probability limit as existing ones when the model is true but has a different probability limit when the model is false. We then use this new estimate of the asymptotic covariance matrix when calculating the Sargan (1958) and Hansen (1982) test statistics. If the null hypothesis is true, then the proposed test has the same distribution as the existing ones in large samples. If the null hypothesis is false, then the proposed test statistic is larger with probability approaching one as the sample size increases in several important examples. We consider a version of the Consumption-based Asset Pricing Model. The simulations show that the improvement can be dramatic in some cases. A test that is related to the Hansen (1982) test is the Hausman (1978) test. Woutersen and Hausman (2019) show that applying the tools of this paper to the Hausman test improves the power of the Hausman test as well. As the Hansen (1982) test is very popular in empirical work, including testing the validity of Euler equations, we expect the current results to be useful as well.

References

[1] Abadir, K. M. and J. R. Magnus (2005): *Matrix Algebra, Econometric Exercises 1*, Cambridge University Press, Cambridge.

[2] Bekker, P. A. (1994), "Alternative approximations to the distributions of instrumental variables estimators." *Econometrica*, 62: 657–681.

[3] De Jong, R. M., and T. Woutersen (2011): "Dynamic Time Series Binary Choice," 2011, *Econometric Theory,* 27, 673-702

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

[5] Hansen, L.P. (2014): "Uncertainty Outside and Inside Economic Models," Nobel Prize Lecture, *Journal of Political Economy*, 122: 945-987.

[6] Hansen, L. P. and K. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 1982, 50: 1269-86.

[7] Hausman, J.A. (1978): "Specification Tests in Econometrics", *Econometrica*, 46: 1251-1272.

[8] Hausman, J.A., W.K. Newey, T. Woutersen, J. Chao, and N.R. Swanson, (2012): "IV estimation with heteroskedasticity and many instruments," *Quantitative Economics* 3: 211-255.

[9] Havránek, T. (2015): "Measuring Intertemporal Substitution: The Importance Of Method Choices and Selective Reporting," *Journal of the European Economic Association*, 13: 1180-1204.

[10] Kunitomo, N. (1980): "Asymptotic expansions of distributions of estimators in a linear functional relationship and simultaneous equations." *Journal of the American Statistical*

Association, 75:693–700.

[11] Lehmann, E. L. and J. P. Romano (2005): *Testing Statistical Hypotheses*, Third Edition, Springer-Verlag, New York.

[12] Mohammadi, M. (2016): "On the Bounds for Diagonal and Off-Diagonal Elements of the Hat Matrix in the Linear Regression Model," *REVSTAT Statistical Journal*, 14: 75-87.

[13] Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in the *Handbook of Econometrics,* Vol. 4, ed. by R. F. Engle and D. MacFadden. North-Holland, Amsterdam.

[14] Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica,* 55: 703–708.

[15] Pötscher, B.M. and I.R. Prucha (1997): *Dynamic nonlinear econometric models,* Springer-Verlag; Berlin.

[16] Romano, J. P., A. M. Shaikh, and M. Wolf (2010): "Hypothesis Testing in Econometrics," *Annual Review of Economics,* 75-104.

[17] Ruud, Paul A. (2000): *An Introduction to Classical Econometric Theory.* Oxford University Press, New York.

[18] Sargan, J.D. (1958): "The Estimation of Economic Relationships using Instrumental Variables," *Econometrica,* 26: 393-415.

[19] Woutersen, T. and J. A. Hausman (2019): "Increasing the Power of Specification Tests," *Journal of Econometrics*, 211: 166-175.

5. Appendix

Appendix 1: F-distributions in the Example

Let ε be a column vector with ε_t, $t = 1, ..., T$, as its elements. Let X be a T by 2 matrix with X_t as its rows, $t = 1, ..., T$. Let

$$\varepsilon | X \sim N(0, \sigma^2 I_T). \tag{8}$$

Consider the moment vector function that only uses X_t,

$$\tilde{g}(\kappa_{H_0}, \delta_{H_0}) = \frac{\sum_t}{T} X_t' \varepsilon_t,$$

and note that

$$\tilde{g}(\kappa_{H_0}, \delta_{H_0}) | X \sim N(0, \sigma^2 \frac{\sum_t}{T} X_t X_t'),$$

and

$$\frac{1}{\sigma^2} \tilde{g}(\kappa_{H_0}, \delta_{H_0})' \{\frac{\sum_t}{T} X_t X_t'\}^{-1} \tilde{g}(\kappa_{H_0}, \delta_{H_0}) \sim \chi^2(2).$$

Define $M_X = I - X(X'X)^{-1}X'$. Note that $\tilde{e}$ is the vector of residuals from regressing $\varepsilon_t(\kappa_{H_0}, \delta_{H_0})$ on X_t, i.e.

$$\tilde{e} = M_X \varepsilon, \text{ and } \tilde{e} | X \sim N(0, \sigma^2 M_X).$$

This gives

$$\frac{1}{\sigma^2} \tilde{e}' \tilde{e} = \frac{1}{\sigma^2} \varepsilon' M_X \varepsilon, \text{ and } \frac{1}{\sigma^2} \tilde{e}' \tilde{e} \sim \chi^2(N-2).$$

The next step is to show that the vectors $\tilde{e} = M_X \varepsilon$ and $\tilde{g}(\kappa_{H_0}, \delta_{H_0}) = \frac{X'\varepsilon}{T}$ are independently distributed. Note that $\tilde{e}$ and $\tilde{g}(\kappa_{H_0}, \delta_{H_0})$ are jointly normally distributed, so we only have to show that every element of $\tilde{e}$ is uncorrelated with $\tilde{g}(\kappa_{H_0}, \delta_{H_0})$. Consider

$$E\{M_X \varepsilon \cdot \frac{\varepsilon' X}{T}\} = \frac{\sigma^2}{T} E\{M_X X\} = 0,$$

since $M_X X = 0$. This gives that $T_{\text{Hansen, 2 variables}}$ has an F-distribution with $\{2, T-2\}$ degrees of freedom. The same reasoning gives that $T_{\text{Hansen, } K+2 \text{ variables}}$ has an F-distribution with $\{K+2, T-K-2\}$ degrees of freedom.

For the new test, we have that

$$\frac{1}{\sigma^2}\tilde{g}(\kappa_{H_0},\delta_{H_0})'\{\frac{\sum_t}{T}X_tX_t'\}^{-1}\tilde{g}(\kappa_{H_0},\delta_{H_0}) \sim \chi^2(2),$$

as shown above. Let S be a T by $(K+2)$ matrix with S_t as its rows, $t=1,...,T$. Define $M_S = I - S(S'S)^{-1}S'$. Note that e is the vector of residuals from regressing $\varepsilon_t(\kappa_{H_0},\delta_{H_0})$ on S_t, i.e.

$$e = M_S\varepsilon, \text{ and that } e|S \sim N(0,\sigma^2 M_S).$$

Further,

$$\frac{1}{\sigma^2}e'e = \varepsilon' M_S\varepsilon, \text{ and } \frac{1}{\sigma^2}e'e \sim \chi^2(N-K-2).$$

The last step is to show that the vectors $e = M_S\varepsilon$ and $\tilde{g}(\kappa_{H_0},\delta_{H_0}) = \frac{X'\varepsilon}{T}$ are independently distributed. Note that e and $\tilde{g}(\kappa_{H_0},\delta_{H_0})$ are jointly normally distributed, so we only have to show that every element of e is uncorrelated with every element of $\tilde{g}(\kappa_{H_0},\delta_{H_0})$. Consider

$$E\{M_S\varepsilon \cdot \frac{\varepsilon' X}{T}\} = \frac{\sigma^2}{T}E\{M_S X\} = 0,$$

since S contains the regressors X so that $M_S X = 0$. This gives that T_{New} has an F-distribution with $\{2, T-K-2\}$ degrees of freedom.

Appendix 2: Simulations with 0.01 Rejection Frequency

The tables A1 through A3 use the same test statistics and data generating process as tables 1 through 3 in the main text, but now the rejection frequency is 0.01.

Table A1: 0.01 Rejection Frequencies

N	K	β	γ	Hansen K+2	Hansen 2 moments	New Test
50	15	0.4	$\gamma_1=\gamma_2=\gamma_3=0.2$	0.28235	0.49198	0.55065
100	30	0.2828	$\gamma_1=\gamma_2=\gamma_3=0.2$	0.37906	0.52120	0.59308
200	60	0.2	$\gamma_1=\gamma_2=\gamma_3=0.2$	0.53841	0.53503	0.61401
50	30	0.4	$\gamma_1=\gamma_2=\gamma_3=0.2$	0.08231	0.49051	0.48817
100	60	0.2828	$\gamma_1=\gamma_2=\gamma_3=0.2$	0.11120	0.51923	0.56180
200	120	0.2	$\gamma_1=\gamma_2=\gamma_3=0.2$	0.16820	0.53430	0.59993

Results based on 100,000 simulations.

In Table A1, the size of the tests is 1%, i.e. the critical value is such that the probability of falsely rejecting the null hypothesis is 1%. The critical values are derived from the F-distribution.

Table A2: 0.01 Rejection Frequencies: Allowing for Heteroscedasticity

N	K	β	γ	Hansen K+2	Hansen 2 moments	New Test
50	15	0.4	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	size: 0.27438	0.53417	0.58859
100	30	0.2828	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	size: 0.40839	0.54640	0.61286
200	60	0.2	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	size: 0.62965	0.54736	0.62381

Results based on 100,000 simulations.

Table A2 shows that the new overidentification test has more power than the Hansen test with 2 moments. Here, the test statistics allow for heteroscedasticity. The size of the Hansen overidentification test with $K + 2$ moments is too large; it varies from about 27% to 62%, i.e. well above 1%.

Table A3: 0.01 Rejection Frequencies: Overidentification Test

N	K	β	γ	Hansen K+2	Hansen 2 moments	New Test
50	15	0.4	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	0.29292	0.60242	0.65189
100	30	0.2828	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	0.38049	0.63203	0.68873
200	60	0.2	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	0.54491	0.64452	0.70505
50	30	0.4	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	size: 0.26726	0.63863	0.68293
100	60	0.2828	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	size: 0.40159	0.64970	0.70338
200	120	0.2	$\gamma_1 = \gamma_2 = \gamma_3 = 0.2$	size: 0.62877	0.65664	0.71507

Results based on 100,000 simulations.

Table A3 rows 1-3 show that the new overidentification test has more power than the existing ones for these data generating processes. The last three rows allow for heteroscedasticity. The size of the Hansen overidentification test with $K + 2$ moments is too large; it varies from about 26% to 62%, i.e. well above 1%.

Appendix 3: The size of the tests

Table A4: Frequency of Rejecting a true H_0, Exact Coverage

N	K	β	γ	Hansen K+2	Hansen 2 moments	New Test	Size
50	15	0	$\gamma_j = 0$ for all j	0.05115	0.05006	0.04991	0.05
100	30	0	$\gamma_j = 0$ for all j	0.05016	0.05066	0.05115	0.05
200	60	0	$\gamma_j = 0$ for all j	0.05009	0.05160	0.05140	0.05
50	30	0	$\gamma_j = 0$ for all j	0.01054	0.01056	0.01036	0.01
100	60	0	$\gamma_j = 0$ for all j	0.00991	0.01047	0.01029	0.01
200	120	0	$\gamma_j = 0$ for all j	0.00981	0.01027	0.01022	0.01

Results based on 100,000 simulations.

The size of all tests in Table 1 is exact. Rows 1-3 of this table illustrate the exact coverage of rows 1-3 for the tests with size 0.05. Rows 4-6 illustrate that the exact coverage when the size of the test is decreased to 0.01.

Table A5: Frequency of Rejecting a true H_0, Allowing for Heteroscedasticity

N	K	β	γ	Hansen K+2	Hansen 2 moments	New Test	Size
50	15	0	$\gamma_j = 0$ for all j	0.46971	0.06626	0.06604	0.05
100	30	0	$\gamma_j = 0$ for all j	0.61505	0.05828	0.05842	0.05
200	60	0	$\gamma_j = 0$ for all j	0.80602	0.05493	0.05484	0.05
50	30	0	$\gamma_j = 0$ for all j	0.27438	0.01744	0.01660	0.01
100	60	0	$\gamma_j = 0$ for all j	0.40839	0.01399	0.01391	0.01
200	120	0	$\gamma_j = 0$ for all j	0.62965	0.01211	0.01226	0.01

Results based on 100,000 simulations.

Table A6 Overidentification Test: Frequency of Rejecting a true H_0, Exact Coverage

N	K	β	γ	Hansen K+2	Hansen 2 moments	New Test	Size
50	15	0	$\gamma_j = 0$ for all j	0.04923	0.04955	0.04940	0.05
100	30	0	$\gamma_j = 0$ for all j	0.04988	0.05134	0.05106	0.05
200	60	0	$\gamma_j = 0$ for all j	0.04885	0.04968	0.05001	0.05
50	15	0	$\gamma_j = 0$ for all j	0.00960	0.01027	0.01021	0.01
100	30	0	$\gamma_j = 0$ for all j	0.01038	0.01022	0.01027	0.01
200	60	0	$\gamma_j = 0$ for all j	0.00949	0.01035	0.01021	0.01

Results based on 100,000 simulations.

Table A7 Overidentification Test: Frequency of Rejecting a true H_0, Heteroscedasticity

N	K	β	γ	Hansen K+2	Hansen 2 moments	New Test	Size
50	15	0	$\gamma_j = 0$ for all j	0.46128	0.06924	0.06917	0.05
100	30	0	$\gamma_j = 0$ for all j	0.61116	0.06055	0.06032	0.05
200	60	0	$\gamma_j = 0$ for all j	0.80535	0.05399	0.05404	0.05
50	15	0	$\gamma_j = 0$ for all j	0.26726	0.01803	0.01800	0.01
100	30	0	$\gamma_j = 0$ for all j	0.40159	0.01494	0.01453	0.01
200	60	0	$\gamma_j = 0$ for all j	0.62877	0.01226	0.01221	0.01

Results based on 100,000 simulations.

Appendix 4: Lemmas

Lemma A1 (Abadir and Magnus, 2005, exercise and solution 12.16): Let the matrices Ω and Λ (i) be positive definite, (ii) be symmetric, and (iii) have the same dimensions. Then there exists a nonsingular matrix R such that

$$\Omega = RR', \text{ and } \Lambda = R\Sigma R',$$

where the diagonal matrix Σ contains the eigenvalues of $\Omega^{-1}\Lambda$.

Proof: Abadir and Magnus, 2005, exercise and solution 12.16.

Lemma A2: Let the matrices Ω and Λ (i) be positive definite, (ii) be symmetric, and (iii) have the same dimensions. Then there exists a nonsingular matrix R such that

$$\Lambda^{-1} - \Omega^{-1} = R^{'-1}\{\Sigma^{-1} - I\}R^{-1},$$

where the diagonal matrix Σ contains the eigenvalues of $\Omega^{-1}\Lambda$.

Proof: Lemma A1 implies that

$$\Omega^{-1} = R'^{-1}R^{-1} \text{ and } \Lambda^{-1} = R'^{-1}\Sigma^{-1}R^{-1},$$

and the result follows.

Lemma A3: Let the matrices Ω and Λ (i) be positive definite, (ii) be symmetric, and (iii) have the same dimensions. Further, let $\Omega - \Lambda$ be positive semidefinite. Then $\Omega^{-1} - \Lambda^{-1}$ is positive semidefinite.

Proof: Lemma A1 shows that

$$\Lambda = R\Sigma R',$$

where Σ is a diagonal matrix, and Λ is positive definite by assumption. R is nonsingular and has full rank so that all the diagonal elements of Σ are strictly positive, i.e. $\Sigma_j > 0$ for $j = 1, ..., J$, where J is the number of rows of the diagonal matrix Σ. Further, note that $\Omega - \Lambda = R\{I - \Sigma\}R'$ is positive semidefinite by assumption. This implies that $1 - \Sigma_j \geq 0$ for $j = 1, ..., J$. Thus, $0 < \Sigma_j \leq 1$ for $j = 1, ..., J$. Lemma A2 states

$$\Lambda^{-1} - \Omega^{-1} = R'^{-1}\{\Sigma^{-1} - I\}R^{-1},$$

and by using $0 < \Sigma_j \leq 1$ for $j = 1, ..., J$, this yields that $\Lambda^{-1} - \Omega^{-1}$ is positive semidefinite since R is nonsingular and has full rank.

Under H_1, we have that the matrix $E(\frac{\sum_t}{T}\phi_t\phi_t')$ contains nonzero elements, and we use this in the following lemma.

Lemma A4: Let the matrices Ω and Λ (i) be positive definite, (ii) be symmetric, and (iii) be J by J matrices. Further, let $\Omega - \Lambda = E(\frac{\sum_t}{T}\phi_t\phi_t')$, where the matrix $E(\frac{\sum_t}{T}\phi_t\phi_t')$ contains nonzero elements. Let s be a column vector with J elements that are all nonzero (i.e. $s_k \neq 0$ for all k). Let the diagonal matrix Σ contain the eigenvalues of $\Omega^{-1}\Lambda$. Then

(a) $0 < \Sigma_j \leq 1$ for all $j = 1, ..., J$, and $\Sigma_k < 1$ for some k, $k = 1, ..., J$, and

(b) $s'\{\Lambda^{-1} - \Omega^{-1}\}s > 0$.

Proof: Let b be a column vector with J elements. Note that

$$b'(\Omega - \Lambda)b = E(\frac{\sum_t}{T} b'\phi_t{\phi_t}'b) \geq 0$$

for any b. Thus, $\Omega - \Lambda$ is positive semidefinite, and Lemma A3 applies. This gives that $b'\{\Lambda^{-1} - \Omega^{-1}\}b \geq 0$ for any b, and that $0 < \Sigma_j \leq 1$ for $j = 1, ..., J$. Next, Lemma A1 implies that $\Omega - \Lambda = R\{I - \Sigma\}R'$ for a nonsingular R. By assumption, $\Omega - \Lambda = E(\frac{\sum_t}{T}\phi_t{\phi_t}')$ so that

$$R\{I - \Sigma\}R' = E(\frac{\sum_t}{T}\phi_t{\phi_t}').$$

Next, consider $\dot{s} = R^{-1}s$ where $\dot{s}$ and s are column vectors with J elements. Using this in the last equation gives

$$\dot{s}'R\{I - \Sigma\}R'\dot{s} = s'\{I - \Sigma\}s = E(\frac{\sum_t}{T}s'\phi_t{\phi_t}'s) = \frac{\sum_t}{T}s'E(\phi_t{\phi_t}')s.$$

Note that $s'E(\phi_t{\phi_t}')s \geq 0$ for every t. By assumption, $M = \frac{\sum_t}{T}E(\phi_t{\phi_t}')$ has nonzero elements. Let M_{lk} denote the element on the l^{th} row and k^{th} column. Suppose that $M_{lk} \neq 0$. Then, by the Cauchy–Schwarz inequality, M_{ll} and M_{kk} are also nonzero. Let $s[k]$ denote a column vector with J elements. Further, let the k^{th} element of this vector be one, and let all the other elements be zero. Then $s[k]'M \cdot s[k] = M_{kk}$, which is nonzero. Since $s'E(\phi_t{\phi_t}')s \geq 0$ for every s, we have that $s[k]'M \cdot s[k] = M_{kk} > 0$. Define $\dot{s}[k] = R^{-1\prime}s[k]$. Thus, if $M_{kk} > 0$, then

$$\begin{aligned}\dot{s}[k]'\{\Omega - \Lambda\} \cdot \dot{s}[k] &= \dot{s}[k]'\{R(I - \Sigma)R'\} \cdot \dot{s}[k] \\ &= s[k]'(I - \Sigma) \cdot s[k] \\ &= 1 - \Sigma_k > 0.\end{aligned}$$

This yields part (a) and (b) of this Lemma.

Lemma A5: Let the matrices Ω and Λ (i) be positive definite, (ii) be symmetric, and (iii) be J by J matrices. Further, let $M = \Omega - \Lambda$ be positive semidefinite, and $M_{kk} \neq 0$ for some k. Let s be a column vector with J elements that are all nonzero (i.e. $s_k \neq 0$ for all k). Then

(a) $0 < \Sigma_j \leq 1$ for all $j = 1, ..., J$, and $\Sigma_k < 1$ for some k, $k = 1, ..., J$, and

(b) $s'\{\Lambda^{-1} - \Omega^{-1}\}s > 0$.

Proof: M is positive semidefinite. The same reasoning as in Lemma A4 yields that $0 < \Sigma_j \leq 1$ for all $j = 1, ..., J$. Further, since M is positive semidefinite, and $M_{kk} \neq 0$ for some k, we have that $M_{kk} > 0$. As in the proof of Lemma A4, let $s[k]$ denote a column vector with J elements. Further, let the k^{th} element of this vector be one, and let all the other elements be zero. Then $s[k]'M \cdot s[k] = M_{kk}$, which is strictly positive. Thus, we have that $s[k]'M \cdot s[k] = M_{kk} > 0$. Define $\dot{s}[k] = R^{-1\prime}s[k]$. This gives

$$\begin{aligned} \dot{s}[k]'\{\Omega - \Lambda\} \cdot \dot{s}[k] &= \dot{s}[k]'\{R(I - \Sigma)R'\} \cdot \dot{s}[k] \\ &= s[k]'(I - \Sigma) \cdot s[k] \\ &= 1 - \Sigma_k > 0. \end{aligned}$$

This yields that $\Sigma_k < 1$ for some k, $k = 1, ..., J$, and that $s'\{\Lambda^{-1} - \Omega^{-1}\}s > 0$.

Appendix 5 Proof of Theorem 1

(i) We first consider the case where Assumption 1 holds, and Assumption 2(i) or 2(iii) hold. In that case, if H_0 is true, then $E(\frac{\sum_t}{T}\psi_t\psi_t')$ equals zero so that $\Omega = \Lambda$. The result then follows from the properties of the χ^2-distribution (see, e.g., Lehmann and Romano (2005)).

Now consider the case where Assumption 1 holds, and Assumption 2(ii) or 2(iv) hold. In that case, if H_0 is true, then $E(\frac{\sum_t}{T}\phi_t\phi_t')$ equals zero so that, again, $\Omega = \Lambda$. The result then follows from the properties of the χ^2-distribution (see, e.g., Lehmann and Romano (2005)).

(ii) We first consider the case where Assumption 1 and Assumption 2(i) hold. Thus, $\Omega - \Lambda = E(\frac{\sum_t}{T}\psi_t\psi_t')$ where $E(\frac{\sum_t}{T}\psi_t\psi_t')$ is positive semidefinite by inspection. The matrices Ω and Λ are symmetric and positive definite. Thus, the conditions of Lemma A2 and Lemma A4 are satisfied so that

$$\Lambda^{-1} - \Omega^{-1} = R'^{-1}\{\Sigma^{-1} - I\}R^{-1},$$

where $0 < \Sigma_j \leq 1$ for all $j = 1, ..., J$, and $\Sigma_k < 1$ for some k, $k = 1, ..., J$. Now consider

$$T \cdot \{g(\theta_{H_0})'\Lambda^{-1}g(\theta_{H_0}) - g(\theta_{H_0})'\Omega^{-1}g(\theta_{H_0})\} = T \cdot [g(\theta_{H_0})'R'^{-1}\{\Sigma^{-1} - I\}R^{-1}g(\theta_{H_0})].$$

Define $h(\theta_{H_0}) = R^{-1}g(\theta_{H_0})$. This gives

$$T \cdot \{g(\theta_{H_0})'\Lambda^{-1}g(\theta_{H_0}) - g(\theta_{H_0})'\Omega^{-1}g(\theta_{H_0})\} = T \cdot [h(\theta_{H_0})\{\Sigma^{-1} - I\}h(\theta_{H_0})].$$

Using the fact that Σ^{-1} is a diagonal matrix yields

$$T \cdot \{g(\theta_{H_0})'\Lambda^{-1}g(\theta_{H_0}) - g(\theta_{H_0})'\Omega^{-1}g(\theta_{H_0})\} = T \cdot \sum_j \{h_j(\theta_{H_0})\}^2(\frac{1}{\Sigma_j} - 1).$$

$\sqrt{T}h_j(\theta_{H_0})$ is nonzero with probability approaching one, as $\sqrt{T}h_j(\theta_{H_0})$ converges to a normal distribution with a zero or nonzero mean. Since $0 < \Sigma_j \leq 1$ for all $j = 1, ..., J$, and $\Sigma_k < 1$ for some k, $k = 1, ..., J$, we have that $T \cdot \{g(\theta_{H_0})'\Lambda^{-1}g(\theta_{H_0}) - g(\theta_{H_0})'\Omega^{-1}g(\theta_{H_0})\} > 0$ with probability approaching one. The same inequality holds with probability approaching one if we replace Λ^{-1} by $\hat{\Lambda}^{-1}$ and Ω^{-1} by $\hat{\Omega}^{-1}$.

Next, we consider the cases where Assumption 1 and Assumption 2(ii), Assumption 1 and Assumption 2(iii), or Assumption 1 and Assumption 2(iv) hold. In these cases, $\Omega - \Lambda$ is positive semidefinite by assumption. Further, the matrix $[\Omega - \Lambda]$ has at least one diagonal element that is nonzero if H_1 holds. The matrices Ω and Λ are symmetric positive definite. Thus, conditions of Lemma A2 and Lemma A4 are satisfied. The remainder of the proof is the same as above.

(iii) Assumption 1 states that

$$\sqrt{T}\{g(\theta_{H_0}) - \frac{c}{\sqrt{T}}\} \underset{d}{\rightarrow} N(0, \Omega) \tag{9}$$

for some vector constant c and positive definite Ω, and $\widehat{\Omega} = \Omega + o_p(1)$. If H_1 is true, then $c \neq 0$ so that $T \cdot g(\theta_{H_0})'\hat{\Omega}^{-1}g(\theta_{H_0})$ converges to a noncentral χ^2-distribution with $\dim(\theta_{H_0})$ degrees of freedom and noncentrality parameter $c'\Omega^{-1}c$ (see, e.g., Lehmann and Romano (2005)). Let $\tilde{M} = \Omega - \Lambda$ so that $\tilde{M}$ is symmetric. By Assumption 2, we have that $\tilde{M}$ is positive semidefinite, and that $\tilde{M}_{kk} > 0$ for some k. Thus, by Lemma A4 or Lemma A5, we have that $s'\{\Lambda^{-1} - \Omega^{-1}\}s > 0$ where all the elements of s are nonzero. By assumption, $\Lambda = \varphi\Omega$ so that $s'\{\Lambda^{-1} - \Omega^{-1}\}s = s'\Omega^{-1}s(\frac{1}{\varphi} - 1) > 0$. The matrices Ω and Ω^{-1} are positive definite so that $0 < \varphi < 1$. Q.E.D.

Appendix 6 Assumption 2* and Proof of Theorem 2

As in the main text, let $\dot{\psi}_t = E\{g_t(\hat{\theta})|\Upsilon_t\}$ and $\dot{\phi}_t$ be an approximation for $\dot{\psi}_t$.

Assumption 2*

Let one of the following hold:

(i) $\{g_s(\hat{\theta}) - \dot{\psi}_s\}$ and $\{g_t(\hat{\theta}) - \dot{\psi}_t\}$ are uncorrelated for every $s \neq t$; or

(ii) $\{g_s(\hat{\theta}) - \dot{\phi}_s\}$ and $\{g_t(\hat{\theta}) - \dot{\phi}_t\}$ are uncorrelated for every $s \neq t$; if H_0 holds, then $E(\frac{1}{T}\sum_t \dot{\phi}_t' \dot{\phi}_t) = 0$, and if H_1 holds, then $M = E[\frac{\sum_t}{T}\{2g_t(\hat{\theta}) - \dot{\phi}_t\}\dot{\phi}_t']$ is positive semidefinite, and $M_{kk} > 0$ for some k; or

(iii) $\Upsilon_s \subset \Upsilon_t$ for all $s < t$, $s,t = 1,...,T$, and $\dot{M} = \Omega - var[\frac{\sum_t}{T}\{g_t(\hat{\theta}) - \dot{\psi}_t\}] = E(\frac{\sum_{s,t}}{T}\dot{\psi}_s\dot{\psi}_t')$ is positive semidefinite; if H_1 holds, then $\dot{M}_{kk} > 0$ for some k; or

(iv) $\ddot{M} = \Omega - var[\frac{\sum_t}{T}\{g_t(\hat{\theta}) - \dot{\phi}_t\}]$ is positive semidefinite; if H_0 holds, then $E(\frac{1}{T}\sum_t \dot{\phi}_t'\dot{\phi}_t) = 0$, and if H_1 holds, then $\ddot{M}_{kk} > 0$ for some k.

Remark: Assumption 1 assumes a consistent estimator for Ω. In the simulations, we based the consistent estimator on $g_t(\hat{\theta})$, i.e. on the relevant residuals rather then on $g_t(\theta_{H_0})$, since using $g_t(\hat{\theta})$ yields a larger value of T_{Hansen}.

Proof of Theorem 2:

(i) We first consider the case where Assumption 1 and Assumption 2*(i) hold, or Assumption 1 and Assumption 2*(iii) hold. In that case, if H_0 is true, then $E(\frac{\sum_t}{T}\dot{\psi}_t\dot{\psi}_t{}')$ equals zero so that $\Omega = \Lambda$. The result then follows from the properties of the χ^2-distribution (see, e.g., Lehmann and Romano (2005)).

Now consider the case where Assumption 1 holds and Assumption 2*(ii) or 2*(iv) hold. In that case, if H_0 is true, then $E(\frac{\sum_t}{T}\dot{\phi}_t\dot{\phi}_t{}')$ equals zero so that, again, $\Omega = \Lambda$. The result then follows from the properties of the χ^2-distribution (see, e.g., Lehmann and Romano (2005)).

(ii) Under the assumptions of Theorem 2 (ii) we have that the matrices Ω and Λ (i) are positive definite, (ii) are symmetric, and (iii) are J by J matrices. Further, let $M = \Omega - \Lambda$ be positive semidefinite, and $M_{kk} \neq 0$ for some k. Thus, the assumptions of Lemma A5 are satisfied. Thus, for a column vector s with J elements that are all nonzero (i.e. $s_k \neq 0$ for all k) we have that $s'\{\Lambda^{-1} - \Omega^{-1}\}s > 0$. Further, all elements of the vector $\sqrt{T}g(\hat{\theta})$

are nonzero with probability approaching one so that $Tg(\hat{\theta})'\{\Lambda^{-1} - \Omega^{-1}\}g(\hat{\theta}) > 0$ with probability approaching one.

(iii) Note that $\hat{\theta} = \arg\min_{\theta} g(\theta)'\hat{\Omega}^{-1}g(\theta)$ so that $g(\hat{\theta})'\hat{\Omega}^{-1}g(\hat{\theta}) \leq g(\theta_{H_0})'\hat{\Omega}^{-1}g(\theta_{H_0})$, where $g(\theta_{H_0})'\hat{\Omega}^{-1}g(\theta_{H_0})$ is bounded in probability by Assumption 1. Thus, T_{Hansen} is bounded in probability. Let $\tilde{M} = \Omega - \Lambda$ so that $\tilde{M}$ is symmetric. By Assumption 2*, we have that $\tilde{M}$ is positive semidefinite, and that $\tilde{M}_{kk} > 0$ for some k. Thus, by Lemma A4 or Lemma A5, we have that $s'\{\Lambda^{-1} - \Omega^{-1}\}s > 0$ where all the elements of s are nonzero. By assumption, $\Lambda = \varphi\Omega$ so that $s'\{\Lambda^{-1} - \Omega^{-1}\}s = s'\Omega^{-1}s(\frac{1}{\varphi} - 1) > 0$. The matrices Ω and Ω^{-1} are positive definite so that $0 < \varphi < 1$. Thus, $T_{New} = \frac{T_{Hansen}}{\hat{\varphi}} + o_p(1)$. Q.E.D.

Discussion: Theorem 2 (ii) assumes that "all elements of the vector $\sqrt{T}g(\hat{\theta})$ are nonzero with probability approaching one". This assumption is implied by assumptions that are used to derive the asymptotic distribution of $T \cdot g(\hat{\theta})'\hat{\Omega}^{-1}g(\hat{\theta})$, see for example Hansen (1982), Newey and McFadden (1994), or Ruud (2000). In particular, let $\hat{\theta} = \underset{\theta\in\Theta}{argmin}\; T \cdot g(\theta)'\hat{\Omega}^{-1}g(\theta)$ and let the true value θ_0 be an element of the interior of the parameter space Θ, which is compact. Let $G(\theta)$ denote the derivative of $g(\theta)$ with respect to θ. Assuming continuous differentiability of the moment vector function $g(\theta)$ yields the following first order condition,

$$G(\hat{\theta})'\hat{\Omega}^{-1}g(\hat{\theta}) = 0.$$

Using a Taylor expansion around θ_0 yields

$$(\hat{\theta} - \theta_0) = -\{G(\hat{\theta})'\hat{\Omega}^{-1}G(\dot{\theta})\}^{-1}G(\hat{\theta})\hat{\Omega}^{-1}g(\theta_0), \tag{10}$$

where $\dot{\theta}$ is an intermediate value. Using a Taylor expansion of $g(\hat{\theta})$ around θ_0 yields

$$g(\hat{\theta}) = g(\theta_0) + G(\ddot{\theta})(\hat{\theta} - \theta_0), \tag{11}$$

where $\ddot{\theta}$ is an intermediate value that may differ from $\dot{\theta}$. Combining the last two equations yields

$$g(\hat{\theta}) = g(\theta_0) - G(\ddot{\theta})\{G(\hat{\theta})\hat{\Omega}^{-1}G(\dot{\theta})\}^{-1}G(\hat{\theta})\hat{\Omega}^{-1}g(\theta_0) \tag{12}$$

$$= g(\theta_0) - G\{G'\hat{\Omega}^{-1}G\}^{-1}G'\hat{\Omega}^{-1}g(\theta_0) + o_p(1), \tag{13}$$

where $G = G(\theta_0)$. This gives

$$\sqrt{T}g(\hat{\theta}) = \sqrt{T}(I - H)g(\theta_0)o_p(1), \tag{14}$$

where $H = G\{G'\hat{\Omega}^{-1}G\}^{-1}G'\hat{\Omega}^{-1}$. Notice that H is symmetric and idempotent while $(I - H)$ has rank equal to the degree of overidentification. Therefore, Mohammadi (2016, lemma 2.1) applies so that $0 \leq H_{jj} < 1$ for all j. Thus, the diagonal elements of $(I - H)$ are all strictly larger than zero. In other words, the first element of $g(\hat{\theta})$ is a linear function of the first element of $g(\theta_0)$ (and could be a linear function of other elements as well). We now show that the first element of $g(\hat{\theta})$ being a linear function of the first element of $g(\theta_0)$ implies that the first element of $\sqrt{T}g(\hat{\theta})$ is nonzero with probability approaching one. Note that Ω is positive definite so that none of elements of the moment vector function, a random variable, is a linear function of the other elements. Further, by assumption 1, $\sqrt{T}g(\theta_0)$ is asymptotically normally distributed so that the first element of $\sqrt{T}g(\hat{\theta})$ are nonzero with probability approaching one. The same reasoning applies to all other elements of $\sqrt{T}g(\hat{\theta})$.