Tests for Violations of Moment Conditions

Fallaw Sowell *
Graduate School of Industrial Administration
Carnegie Mellon University

August 1995
Revised January 1996

Abstract

This paper presents optimal tests for violations of the moment conditions in the GMM framework. New tests are presented for the validity of the overidentifying restrictions.

The usefulness of these tests is highlighted by a simple AR(1) example with instability in the moment conditions. For this example, the J test in Hansen (1982) does not reject the overidentifying restrictions but the rejection by a parameter stability test incorrectly suggests a time varying parameter model. The new tests correctly show the instability in the moment conditions and would direct the researcher to consider alternative moment conditions.

KEYWORDS: optimal tests, overidentifying restrictions

*Helpful comments were provided by Brent Boning, Harold Zhang and seminar participants at the University of Pennsylvania.

1 Introduction

GMM estimation and inference are based on the assumption that certain moment conditions should equal zero at the population parameter values:

$$\underset{m \times 1}{E f(x_t, \theta_0)} = 0 \quad \text{for all} \quad t, \tag{1}$$

where x_t is an $r \times 1$ vector of observed data and θ_0 are $k \leq m$ parameters to be estimated. This type of moment condition can result from structural economic models via stochastic Euler equations or from the first order conditions of estimation problems such as least squares or maximum likelihood. Because violations of this assumption can produce inconsistent parameter estimates and incorrect inference, researchers must be concerned with its validity. To assist researchers in this task, this paper presents new tests for violations of the moment conditions.

The moment condition in equation (1) has two important features: (i) the expectation is stable (constant) across every time period and (ii) that stable expectation is zero. The new optimal tests presented in this paper will give researchers power against every possible violation of this assumption for which a test is feasible. Previously proposed optimal tests only have power for some alternatives. For example, the J test in Hansen (1982) does not have power for instability of any of the moment conditions and the optimal tests for parameter instability in Sowell (1994) do not have power for violation of the overidentifying restrictions.

The new tests are LM tests and hence do not require additional nonlinear optimization beyond what is performed in a standard GMM estimation problem. The new tests are optimal for local alternatives that imply a one time structural jump in the moment conditions when the location of the break is unknown. However, the tests are consistent for any testable alternative.

The m moment conditions can be split into two orthogonal sets: the identifying restrictions and the overidentifying restrictions. The k identifying restrictions are the moment conditions that are used to estimate the parameters. The remaining $m - k$ moment conditions can be used to test the validity of the overidentifying restrictions implied by the model. The new tests are directed at testing the stability of the overidentifying restrictions. This can be contrasted with both the J test in Hansen (1982) which only tests whether the average value of the overidentifying restrictions is zero and the optimal tests for parameter instability in Sowell (1994) which test the instability of the identifying restrictions.

Recently, Ghysels, Guay and Hall (1995) presented two predictive failure tests for structural change. These tests will have power against alternatives with instability in the overidentifying restrictions. However, these are not optimal tests and require the repeated solutions of nonlinear estimation problems to evaluate a single test. The tests also simultaneously test instability of both the overidentifying and identifying restrictions. A problem with these tests is that a rejection does not give the researcher guidance about the source of the model's rejection.

The optimal tests for instability of the overidentifying restrictions presented in this paper are independent of the optimal tests for parameter instability presented in Sowell (1994, 1995). When considered together these tests give researchers guidance when a model is rejected. A Monte Carlo simulation in section 3 demonstrates this guidance. For a simple AR(1) model the J test does not reject the model but a parameter instability test does. Hence a researcher might incorrectly model the series with a time varying parameter model. Fortunately, the test for the instability of the overidentifying restriction also rejects this model, signaling that inappropriate moment conditions were used. For the simulated model, changing the moment conditions can result in consistent estimates.

The optimal tests are presented in the paper. Proofs, derivations and sufficient conditions are presented in the Appendix. The following notation will be used in this paper: $b(s)$ ($b^{\circ}(s)$) defined on $s \in [0,1]$ denotes the univariate standard Brownian motion (bridge) process, $B_j(s)$ ($B_j^{\circ}(s)$) denotes the j-dimensional vectors of independent standard Brownian motions (bridges), $\Rightarrow$ denotes weak convergence, $\rightarrow^p$ denotes convergence in probability, $\|\bullet\|$ denotes the Euclidian norm, and $[\bullet]$ is the greatest integer function.

2 The Model and the Tests

In GMM estimation functions of the observed data summarize the information in the model. This model may be suggested by economic theory or an estimation problem. The functions are of the form $f(x_t, \theta)$, where $f : R^r \times R^k \rightarrow R^m$. The relevant theory will imply the functions satisfy equation (1). The sample estimate of $Ef(x_t, \theta)$ is the function $F_{sT}(\theta) = \frac{1}{T}\sum_{t=1}^{[sT]} f(x_t, \theta)$ evaluated at $s = 1$, where $s \in [0,1]$. For notational simplicity, $F_T(\theta)$ will be used to denote $F_{1T}(\theta)$. The GMM estimator of θ is selected to make an estimate of $Ef(x_t, \theta)$ close to the zero vector in some metric. A sequence of weighting matrices, $\{W_T\}$, determines this metric. The GMM estimator, $\hat{\theta}_T$, is defined as a sequence of random vectors that solves

$$\hat{\theta}_T = \underset{\theta}{\operatorname{argmin}} \; F_T(\theta)' W_T F_T(\theta).$$

The asymptotic variance of $F_T(\theta_0)$ is defined by $\Sigma \equiv \lim_{T\rightarrow\infty} E\left[T F_T(\theta_0) F_T(\theta_0)'\right]$. The probability limit of the gradient of the sample moments will be denoted $M \equiv \text{plim}_{T\rightarrow\infty} \frac{\partial F_T(\theta_0)}{\partial \theta}$. This gradient is often normalized by the symmetric matrix $\Sigma^{-1/2}$. This matrix will be denoted $\overline{M} \equiv \Sigma^{-1/2}M$. Below, $\widehat{M}$ will be used to denote a consistent estimate of M.

The parameters are estimated by setting k linear combinations of the normalized sample moments equal to zero. Because Σ is of full rank and because M has full column rank, asymptotically the identifying subspace spanned by the k restrictions is also spanned by the columns of $\overline{M}$. The $m-k$ dimension orthogonal complement of the space spanned by the columns of $\overline{M}$ is the space spanned by the overidentifying restrictions.

Alternative hypotheses that imply violations of the moment conditions can be decomposed into violations of the identifying restrictions and the overidentifying restrictions. This decomposition can be achieved by projecting onto the two orthogonal subspaces of identifying and overidentifying restrictions. A projection matrix for the space of identifying restrictions is $P_{\overline{M}} = \overline{M}(\overline{M}'\overline{M})^{-1}\overline{M}'$ and a projection matrix for the space of overidentifying restrictions is $P^c_{\overline{M}} = I_m - \overline{M}(\overline{M}'\overline{M})^{-1}\overline{M}'$. Consistent estimates of these matrices will be denoted by $\widehat{P}_{\overline{M}}$ and $\widehat{P}^c_{\overline{M}}$, eg.

$$\widehat{P}_{\overline{M}} = W_T^{1/2}\widehat{M}(\widehat{M}'W_T\widehat{M})^{-1}\widehat{M}'W_T^{1/2} \quad \text{and} \quad \widehat{P}^c_{\overline{M}} = I_m - W_T^{1/2}\widehat{M}(\widehat{M}'W_T\widehat{M})^{-1}\widehat{M}'W_T^{1/2}.$$

A violation of the moment condition in equation (1) can be embodied in the local alternatives of[1]

Assumption 3

$$Ef(x_t,\theta_0) = \frac{h\left(\eta,\pi,\frac{t}{T}\right)}{\sqrt{T}} \tag{2}$$

where $h(\eta,\pi,s)$, for $s \in [0,1]$, is an m-dimensional function that can be expressed as the uniform limit of step functions, $\eta \in R^i$, $\pi \in [0,1]$, and θ_0 is in the interior of Θ.

[1] A full set of sufficient conditions for the optimality of the tests presented in this paper are presented in Appendix 1.

The function $h(\eta, \pi, s)$ allows a wide range of alternative hypotheses including multiple jumps and gradual shifts. The parameter π denotes the times of the structural changes as fractions of the sample size. The vector η parameterizes the function that defines the local alternatives. Note that the dimension of η is different from the dimension of θ, the structural parameters. For example, if the alternative of interest has n different jumps in the value of a single moment condition then $i = n$. There would be one parameter in η for each change in the moment condition. Similarly, if the alternative of interest is a one-time jump in the values of each of the moment condition then $i = k$. To reduce notation $h(\eta, \pi, s)$ will sometimes be denoted $h(s)$, its definite integral will be denoted $H(s) = \int_0^s h(r)dr$ and $H^{\circ}(s)$ will denote $H(s) - sH(1)$.

Assumption 3 implies that the orthogonality conditions are incorrect. This can occur for a variety of reasons, such as: (1) residuals correlated with the regressors, (2) misspecification of the Euler equations so the expectational error is correlated with an instrument, (3) the calculation of the reported data series may have changed over time, (4) the structural model may not capture learning by economic agents or (5) the moment conditions may simply be misspecified.

As suggested above, the local alternative places restrictions on the two independent sets of moment conditions: the identifying and the overidentifying conditions. Because $I_m = P_{\overline{M}} + P_{\overline{M}}^c$, the sequence of local alternatives can be decomposed into the two independent sets of restrictions

$$Ef(x_t, \theta_0) = \frac{P_{\overline{M}} h\left(\frac{t}{T}\right)}{\sqrt{T}} \tag{3}$$

and

$$Ef(x_t, \theta_0) = \frac{P_{\overline{M}}^c h\left(\frac{t}{T}\right)}{\sqrt{T}} \tag{4}$$

where $\frac{P_{\overline{M}} h\left(\frac{t}{T}\right)}{\sqrt{T}}$ is the local alternative on the identifying moments and $\frac{P^c_{\overline{M}} h\left(\frac{t}{T}\right)}{\sqrt{T}}$ is the local alternative on the overidentifying moments.

This general class of alternatives includes some special cases that have already been presented in the literature. Newey (1985) shows the J statistics presented in Hansen (1982) to be the optimal test for the alternative where $h(s)$ is a constant function. The optimal tests for parameter instability presented in Sowell (1994) can be interpreted as optimal tests for instability of the identifying restrictions. Sowell (1994) considered the GMM model where $E_t f(x_t, \theta_{t,T}) = 0$, for t = 1, ..., T, and local alternatives of the form

$$\theta_{t,T} = \theta_0 + \frac{g\left(\frac{t}{T}\right)}{\sqrt{T}}. \tag{5}$$

A Taylor series expansion of $f(x_t, \theta)$ shows that the alternative in (5) implies the moment restrictions with local alternatives

$$Ef(x_t, \theta_0) = -M\frac{g\left(\frac{t}{T}\right)}{\sqrt{T}} + O(T^{-1}).$$

Substituting these local alternatives into (3) and (4) shows that the parameter instability alternative only puts restrictions on the identifying restrictions.

Unfortunately, the J test and the tests for parameter instability are not optimal for alternatives with instability of the overidentifying restrictions. More importantly, for the subset of these alternatives where $P^c_{\overline{M}} H(1) = 0$ the J test and the tests for parameter instability have local power equal to the size of the tests. Optimal tests for instability of the overidentifying restrictions have not previously been considered in the literature. These tests are a special case of the optimal tests for arbitrary $h(s)$ functions presented in Appendix 2 of this paper.[2]

The remainder of this paper concerns tests that are optimal for the particular alternative of a one time jump in all the moments where the location of the jump is unknown. This alternative is represented by

$$h(\eta, \pi, s) = U(\pi - s)\eta \text{ where } U(x) = \begin{cases} 0, & x < 0 \\ 1, & 0 \leq x \end{cases} .$$

Though the tests presented below are only <u>optimal</u> for the one time break alternatives the tests are <u>consistent</u> for arbitrary alternatives. Appendix 2 contains the derivation of two optimal tests for a one time jump in all the moment condition where the location of the jump is unknown. The test statistics are

$$E = \frac{1}{T}\sum_{t=1}^{T} \exp\left\{\frac{1}{2} T F_t(\hat{\theta}_T)' W_T F_t(\hat{\theta}_T)\right\},$$

which is optimal for arbitrarily large jumps and

$$L = \sum_{t=1}^{T} F_t(\hat{\theta}_T)' W_T F_t(\hat{\theta}_T)$$

which is optimal for arbitrarily small jumps[3].

These tests simultaneously test for instability of the identifying and overidentifying restrictions. This causes a problem for a researcher because when a test rejects it does not

[2]One set of alternatives does not posses a most powerful test. Optimal tests are not feasible for alternatives that are constant over the identifying restrictions. For these alternatives the asymptotic distributions are the same under both the null and the sequence of alternatives. This is because the estimated parameters will be the sum of the population parameter value and the parameters of the constant alternative. These are the only alternatives for which a most powerful test cannot be determined. Every possible test will have power equal to its size for these alternatives.

[3]The L statistic and its exact distribution under the null were first presented in Sowell (1993). However, no optimality properties were presented.

give guidance for the source of the rejection. The two alternative sources of instability can be separated by projecting the normalized sample moments onto the identifying subspace and the overidentifying subspace. The resulting optimal test statistics are[4]:

$$
\begin{aligned}
E_A &= \frac{1}{T}\sum_{t=1}^{T}\exp\left\{\frac{1}{2}T F_t(\hat{\theta}_T)' W_T^{1/2}\widehat{P}_{\overline{M}}W_T^{1/2}F_t(\hat{\theta}_T)\right\}, && (6)\\
&= \frac{1}{T}\sum_{t=1}^{T}\exp\left\{\frac{1}{2}T F_t(\hat{\theta}_T)' W_T\widehat{M}(\widehat{M}'W_T\widehat{M})^{-1}\widehat{M}'W_T F_t(\hat{\theta}_T)\right\}, && (7)
\end{aligned}
$$

which is the optimal test for alternatives with one time jumps of the identifying restrictions where the level of the jump is extremely large,

$$
\begin{aligned}
E_B &= \frac{1}{T}\sum_{t=1}^{T}\exp\left\{\frac{1}{2}T F_t(\hat{\theta}_T)' W_T^{1/2}\widehat{P}^{c}_{\overline{M}}W_T^{1/2}F_t(\hat{\theta}_T)\right\}, && (8)\\
&= \frac{1}{T}\sum_{t=1}^{T}\exp\left\{\frac{1}{2}T F_t(\hat{\theta}_T)'\left(W_T - W_T\widehat{M}(\widehat{M}'W_T\widehat{M})^{-1}\widehat{M}'W_T\right)F_t(\hat{\theta}_T)\right\}, && (9)
\end{aligned}
$$

which is the optimal test for alternatives with one time jumps of the overidentifying restrictions where the level of the jump is extremely large,

$$
\begin{aligned}
L_A &= \sum_{t=1}^{T} F_t(\hat{\theta}_T)' W_T^{1/2}\widehat{P}_{\overline{M}}W_T^{1/2}F_t(\hat{\theta}_T) && (10)\\
&= \sum_{t=1}^{T} F_t(\hat{\theta}_T)' W_T\widehat{M}(\widehat{M}'W_T\widehat{M})^{-1}\widehat{M}'W_T F_t(\hat{\theta}_T) && (11)
\end{aligned}
$$

which is the optimal test for alternatives with one time jumps of the identifying restrictions

[4] An alternative derivation of these optimal tests would start with local alternatives restricted to one of the subspaces. This was the approach used in Sowell (1994, 1995) where local alternatives of parameter instability were considered. As noted above such alternatives only imply instability of the identifying restrictions.

where the level of the jump is extremely small and

$$L_B = \sum_{t=1}^{T} F_t(\hat{\theta}_T)' W_T^{1/2} \widehat{P}_{\overline{M}}^{c} W_T^{1/2} F_t(\hat{\theta}_T) \tag{12}$$

$$= \sum_{t=1}^{T} F_t(\hat{\theta}_T)' \left(W_T - W_T \widehat{M} (\widehat{M}' W_T \widehat{M})^{-1} \widehat{M}' W_T \right) F_t(\hat{\theta}_T) \tag{13}$$

which is the optimal test for alternatives with one time jumps of the overidentifying restrictions where the level of the jump is extremely small[5]. Because $P_{\overline{M}}$ and $P_{\overline{M}}^{c}$ are orthogonal L_A and L_B are independent and E_A and E_B are independent. This implies that in practice it is appropriate to report significance values for either pair of statistics.

The asymptotic distributions under the local alternatives can be characterized as[6]

$$E_A \sim^a \int_0^1 \exp \left\{ \frac{B_k^\circ(s)' B_k^\circ(s) + H^\circ(s)' P_{\overline{M}} H^\circ(s)}{2} \right\} ds$$

$$E_B \sim^a \int_0^1 \exp \left\{ \frac{B_{m-k}(s)' B_{m-k}(s) + H(s)' P_{\overline{M}}^{c} H(s)}{2} \right\} ds,$$

$$L_A \sim^a \int_0^1 \left\{ B_k^\circ(s)' B_k^\circ(s) + H^\circ(s)' P_{\overline{M}} H^\circ(s) \right\} ds,$$

and

$$L_B \sim^a \int_0^1 \left\{ B_{m-k}(s)' B_{m-k}(s) + H(s)' P_{\overline{M}}^{c} H(s) \right\} ds.$$

This shows that L_A and E_A only have local power against violations of the identifying restrictions and L_B and E_B only have local power against violations of the overidentifying restrictions. The distributions under the null can be characterized by setting $h(s)$ (and

[5] Using the notation of Sowell (1995), E_A and L_A are the test statistics associated with $TS_{\infty,0}$ and $TS_{0,0}$, respectively.

[6] This characterization follows by applying the functional that defines the test statistics to the limiting stochastic process presented in Theorem 1. The functionals are presented in Appendix 2 and Theorem 1 is presented in Appendix 1.

Table 1: The critical values for distributions.

	$\int_0^1 \exp\left\{\frac{B_p^\circ(s)'B_p^\circ(s)}{2}\right\} ds$			$\log\left(\int_0^1 \exp\left\{\frac{B_p(s)'B_p(s)}{2}\right\} ds\right)$		
	α			α		
p	.10	.05	.01	.10	.05	.01
1	1.2129	1.2998	1.5416	0.7374	1.0783	2.0330
2	1.4025	1.5336	1.9050	1.3734	1.8441	2.9926
3	1.6091	1.7805	2.2749	1.9859	2.5851	3.9697
4	1.8380	2.0744	2.7142	2.5763	3.2400	4.7358
5	2.0977	2.3816	3.1751	3.1670	3.9392	5.6175
6	2.3873	2.7323	3.7456	3.7860	4.5768	6.3655
7	2.7238	3.1594	4.4667	4.3477	5.1907	7.1274
8	3.0969	3.6312	5.1355	4.9514	5.8657	7.8991
9	3.5343	4.2026	6.1884	5.5467	6.5204	8.6780
10	4.0225	4.8368	7.0989	6.1171	7.1100	9.2913

Notes: The critical values are based on 40,000 realizations. Each realization was constructed by approximating the integrals with the average over the discrete grid of 4,000 equal intervals on $[0, 1]$. For each realization the integrands were calculated by simulating the p-dimensional Brownian bridge with the partial sums of the deviations from mean for 4,000 normal random variables with variance $(4,000)^{-1}$.

hence $H(s)$) equal to zero in the above characterizations. Critical values for these null distributions of E_A and $\log(E_B)$ are presented in Table 1 and critical values for these null distributions of L_A and L_B are presented in Table 2.

The intuition for the tests is that under the null, (stable moments with expectation equal to zero) the normalized partial sum of the sample moments projected onto the space of overidentifying restrictions converges to linear combinations of $m-k$ Brownian motions with no drift. Both tests build on the inner product of the normalized partial sum of the sample moments projected onto the space of overidentifying restrictions which in the limit is the square of a Bessel process with dimension $m-k$ with drift. The L_B statistic is the average value of the inner product and E_B is the average value of the exponential of the inner product.

These tests have many attractive features. They have optimality properties. They only

Table 2: The critical values for distributions.

	$\int_0^1 B_p^\circ(s)'B_p^\circ(s)ds$			$\int_0^1 B_p(s)'B_p(s)ds$		
	α			α		
p	.10	.05	.01	.10	.05	.01
1	0.3473	0.4641	0.7435	1.1958	1.6557	2.7875
2	0.6070	0.7475	1.0737	2.0622	2.6241	3.9286
3	0.8412	1.0002	1.3586	2.8256	3.4596	4.8907
4	1.0631	1.2373	1.6226	3.5410	4.2339	5.7704
5	1.2777	1.4651	1.8740	4.2273	4.9716	6.6004
6	1.4872	1.6864	2.1167	4.8939	5.6841	7.3962
7	1.6930	1.9030	2.3529	5.5458	6.3781	8.1667
8	1.8958	2.1159	2.5840	6.1864	7.0577	8.9174
9	2.0964	2.3258	2.8111	6.8179	7.7256	9.6522
10	2.2950	2.5333	3.0348	7.4417	8.3840	10.3738

Notes: The critical values are derived by evaluating the distributions functions presented in Nyblom (1989). Note a typographical error in equation (3.4) of Nyblom (1989). The subscript of the parabolic cylinder function should be $\frac{(p-2)}{2}$ not $\frac{(p-1)}{2}$.

require the single standard GMM optimization. They are easily computed from the terms typically calculated in GMM estimation.

3 An Example

A simple demonstration will highlight the usefulness of these new tests. Consider an AR(1) model where the error is correlated with the two most recent lagged values and this correlation changes over time:

$$y_t = .1y_{t-1} + \epsilon_t \qquad \text{where } \epsilon_t = \frac{h\left(\frac{t}{T}\right)}{\sqrt{T}} y_{t-1} + \frac{3h\left(\frac{t}{T}\right)}{\sqrt{T}} y_{t-2} + u_t$$

where

$$h(s) = 1 - 2U(s - 1/2) \tag{14}$$

and $u_t \sim N(0,1)$. Because $\int_0^1 h(s)ds = 0$ this model is constructed so that the J statistic will have local power equal to size.

For this model, 10,000 samples of sample size $T = 200$ were simulated. The initial values (y_{-1} and y_0) were drawn from the stationary distribution under the null: $N\left(0, \frac{1}{1-.1^2}\right)$. For each simulated sample, the AR parameter was estimated by efficient GMM using the two moments

$$\begin{bmatrix} y_{t-1}(y_t - \rho y_{t-1}) \\ y_{t-2}(y_t - \rho y_{t-1}) \end{bmatrix}.$$

The results of the simulations were that the mean of the GMM estimates was .1195 with a standard deviation of .0060.

The empirical local power for the J statistics were[7]

$$Pr[J > k_{.1}] = .1183 \text{ and } Pr[J > k_{.05}] = .0606.$$

The empirical local power for the L_A statistics were

$$Pr[L_A > k_{.1}] = .3073 \text{ and } Pr[L_A > k_{.05}] = .2075.$$

The empirical local power for the L_B statistics were

$$Pr[L_B > k_{.1}] = .3533 \text{ and } Pr[L_B > k_{.05}] = .2323.$$

The estimated parameter is biased because the moment conditions are invalid. However, the local power of the J statistic is only marginally higher than the size of the test. So for this model typically a researcher would conclude that the moments are correctly specified. The local power of the L_A statistic is three to four times greater than the size of the test. This implies the identifying moment conditions are unstable. So for this model it is not uncommon for a researcher to observe the J statistic not rejecting the overidentifying moment conditions but an optimal test for parameter instability, i.e. L_A, rejecting. This

[7]The symbol k_α denotes the critical value so that the associated test will have size $\alpha \in (0,1)$.

might lead the researcher to incorrectly fit a time varying parameter model to the data.

Fortunately, the local power of the L_B statistic is three and a half to five times greater than the size of the test. The L_B statistic will regularly show that the overidentifying moments are unstable. This will direct the researcher to look at different moment conditions that are stable instead of fitting a model with time varying parameters. If the researcher used the different moments

$$\begin{bmatrix} y_{t-3}(y_t - \rho y_{t-1}) \\ y_{t-4}(y_t - \rho y_{t-1}) \end{bmatrix}$$

the parameter of interest can be consistently estimated.

This simple example highlights the usefulness of the new statistics in selecting moment conditions and avoiding incorrectly selecting a time varying parameter model. Though the L_B statistic was not optimal for the alternative (14), it still has local power and is a consistent test.

4 Conclusion

This paper presents new specification tests for the GMM framework. These tests are optimal for alternatives that allows for instability in the overidentifying restrictions. These tests are independent of recently presented optimal tests for parameter instability. Together the new tests and the parameter instability tests can give researchers insights into the reasons for model rejection.

A simple Monte Carlo simulation highlighted how the new tests can help researchers avoid incorrectly concluding parameter instability when the models' moments are unstable.

Appendix 1

The Asymptotic Results: The Sufficient Conditions and Theorem.

The following are not the weakest assumptions possible. Rather, these assumptions are relatively straightforward to verify and general enough to be of interest.

Assumption 1 *For each T, the sequence $\{x_{t,T}\}$ consists of the first T elements of an r-dimensional stationary and ergodic stochastic process $\{x_{t,T} : t = 1, 2, \ldots\}$.*

For notational simplicity, x_t will be used to denote $x_{t,T}$.

Assumption 2 *The parameter space Θ is a compact subset of R^k.*

As discussed in the paper, to allow the calculation of power against local alternatives, a sequence of alternatives will be considered. This class of alternatives allows for structural changes.

Assumption 3

$$Ef(x_t, \theta_0) = \frac{h\left(\eta, \pi, \frac{t}{T}\right)}{\sqrt{T}}$$

where $h(\eta, \pi, s)$, for $s \in [0, 1]$, is an m-dimensional function that can be expressed as the uniform limit of step functions, $\pi \in [0, 1]$, $\eta \in R^i$, and θ_0 is in the interior of Θ.

Assumption 4 *The matrix Σ is positive definite and the matrix M has full column rank.*

An identification assumption is required to assure that the sequence of GMM estimator has a unique limit.

Assumption 5 $\lim_{T\to\infty} EF_T(\theta) = 0$, *only when $\theta = \theta_0$.*

The functions of the data must satisfy smoothness and boundedness regularity conditions.

Assumption 6 *$f(x, \theta)$ is continuously partially differentiable in θ in a neighborhood of θ for every $\theta \in \Theta^*$ where Θ^* is some convex or open subset that contains Θ. The functions $f(x, \theta)$ and $\frac{\partial f(x,\theta)}{\partial \theta}$ are measurable functions of x for each $\theta \in \Theta^*$, and $E \sup_{\theta\in\Theta^*} \|\frac{\partial f(x_t,\theta)}{\partial \theta}\| < \infty$. $Ef(x_t, \theta_0) = 0$, $Ef(x_t, \theta_0)' f(x_t, \theta_0) < \infty$ and $\sup_{\theta\in\Theta} \|f(x_t, \theta)\| < \infty$ for all $t = 1, \ldots, T$ and $T = 1, 2, \ldots$. Each element of $f(x_t, \theta_{t,T})$ is uniformly square integrable, for all $t = 1, \ldots, T$ and $T = 1, 2, \ldots$.*

Only optimal GMM is considered, i.e., attention is restricted to efficient GMM estimators. This is achieved by restricting the choice of the weighting matrix in the next assumption.

Assumption 7 *The sequence of positive definite weighting matrices $\{W_T\}_{T=k}^{\infty}$ converge in probability to Σ^{-1}.*

The next assumption imposes restrictions on the amount of heteroskedasticity and autocorrelation allowed in the observed series. See Phillips and Durlauf (1986) for the definitions of strong and uniform mixing for multivariate processes, which generalizes the univariate work of McLeish (1975).

Assumption 8 *Either*

1. $\{x_t\}$ *is strong mixing with strong mixing coefficients* $\{\alpha(n)\}$, $\sum_{n=1}^{\infty} \alpha(n)^{1-2/\beta} < \infty$ *with* $\beta > 2$ *or*

2. $\{x_t\}$ *is uniform mixing with uniform mixing coefficients* $\{\phi(n)\}$, $\sum_{n=1}^{\infty} \phi(n)^{1-1/\beta} < \infty$ *with* $\beta \geq 2$

and the individual elements of $f(x_t, \theta_{t,T})$ *have the finite absolute moment* $E\left|f^{(i)}(x_t, \theta_{t,T})\right|^{\beta} < \infty$ *for* $i = 1, \ldots, m$.

The asymptotic results needed for optimal testing are the weak convergence of the partial sums of the sample moments under the null and alternative hypotheses. Theorem 1 gives this convergence.

Theorem 1 *If assumptions 1-8 are satisfied then*

$$\sqrt{T} W_T^{1/2} F_{sT}(\hat{\theta}_T) \Rightarrow P_{\overline{M}}^{c}(B_m(s) + H(s)) + P_{\overline{M}}((B_m(s) - sB_m(1)) + (H(s) - sH(1))), \quad (15)$$

where $H(s) = \int_0^s h(\eta, \pi, r) dr$.

Under the local alternative hypothesis, Theorem 1 shows the normalized partial sums of the sample moment conditions converge to linear combinations of k Brownian bridges with drift and $m-k$ Brownina motions with drift. The drift of the Brownian motions are the partial sum of the alternative function, i.e. $H(s) = \int_0^s h(r)dr$, projected onto the overidentifying subspace. The drift of the Brownian bridges are the partial sum of the devitaions from the mean of the alternative function, i.e. $H^{\circ}(s) = H(s) - sH(1) = \int_0^s [h(r) - \int_0^1 h(v)dv]dr$, projected onto the identifying subspace.

Proof of Theorem 1

The convergence, $\hat{\theta}_T \to^p \theta_0$, results from the identification assumption (Assumption 5) and the uniform convergence of $F_T(\theta)$ to $\lim_{T\to\infty} EF_T(\theta)$. The uniform convergence is established by verifying assumptions A1, B1 and A5 in Andrews (1987). A1 follows from Assumptions 2. B1 follows from Assumption 8. A5 follows from Assumption 2 and Assumption 6.

Phillips and Durlauf (1986) presents a multivariate generalization of the univariate results in McLeish (1975). Assumptions 1, 4 and 8 of this paper imply the assumptions of Corollary 2.2 in Phillips and Durlauf (1986); hence, the sample moments evaluated at $\theta_{t,T}$ satisfy the multivariate invariance principle

$$\frac{1}{\sqrt{T}} W^{1/2} \sum_{t=1}^{[sT]} \left(f(x_t, \theta_0) - \frac{h\left(\frac{t}{T}\right)}{\sqrt{T}} \right) \Rightarrow B_m(s). \quad (16)$$

To reduce notation define $f_t(\theta) \equiv f(x_t, \theta)$ and $h(s) \equiv h(\eta, \pi, s)$.

The left hand side of (16) can be written

$$\frac{1}{\sqrt{T}} W^{1/2} \sum_{t=1}^{[sT]} f(x_t, \theta_0) - \frac{1}{\sqrt{T}} W^{1/2} \sum_{t=1}^{[sT]} \frac{h\left(\frac{t}{T}\right)}{\sqrt{T}}$$

The second term converges in probability to $\Sigma^{-1/2} \int_0^s h(r)dr$. The first term is by definition

$\sqrt{T}W^{1/2}F_{sT}(\theta_0)$. Using (16) now gives the result

$$\sqrt{T}W^{1/2}F_{sT}(\theta_0) \Rightarrow B_m(s) + \Sigma^{-1/2}H(s) \tag{17}$$

where $H(s) = \int_0^s h(r)dr$. (Note that this is a Brownian motion with drift.)

Now derive the weak convergence when the sample moments are evaluated at the GMM estimates of the parameter values. Expand $F_{sT}(\theta)$ about θ_0 and evaluate the expansion at $\hat{\theta}_T$

$$F_{sT}(\hat{\theta}_T) = F_{sT}(\theta_0) + \frac{\partial F_{sT}(\overline{\overline{\theta}}_{s,T})}{\partial\theta}(\hat{\theta}_T - \theta_0) \tag{18}$$

where $\overline{\overline{\theta}}_{s,T} = \left[\begin{array}{ccc} \overline{\overline{\theta}}_{s,T}^{(1)} & \ldots & \overline{\overline{\theta}}_{s,T}^{(k)} \end{array} \right]$ and $\overline{\overline{\theta}}_{s,T}^{(i)} = k_s^{(i)}\hat{\theta}_T^{(i)} + \left(1 - k_s^{(i)}\right)\theta_0^{(i)}$ for some $k_s^{(i)} \in [0,1]$ and every $s \in [0,1]$ and $i = 1,\ldots,k$. Because $\hat{\theta}_T$ is consistent for θ_0, $\overline{\overline{\theta}}_{s,T} \rightarrow^p \theta_0$.

Calculate an alternative form for $(\hat{\theta}_T - \theta_0)$. Expand $f_t(\theta)$ about θ_0 and evaluate the expansion at $\hat{\theta}_T$

$$\begin{aligned} f_t(\hat{\theta}_T) &= f_t(\theta_0) + \frac{\partial f_t(\overline{\theta})}{\partial\theta}(\hat{\theta}_T - \theta_0) \\ &= \left(f_t(\theta_0) - \frac{h\left(\frac{t}{T}\right)}{\sqrt{T}} \right) + \frac{\partial f_t(\overline{\theta})}{\partial\theta}(\hat{\theta}_T - \theta_0) + \frac{h\left(\frac{t}{T}\right)}{\sqrt{T}} \end{aligned} \tag{19}$$

where $\overline{\theta} = \left[\begin{array}{ccc} \overline{\theta}^{(1)} & \ldots & \overline{\theta}^{(k)} \end{array} \right]$ and $\overline{\theta}^{(i)} = \overline{k}^{(i)}\theta_0^{(i)} + \left(1 - \overline{k}^{(i)}\right)\hat{\theta}_T^{(i)}$ for some $\overline{k}^{(i)} \in [0,1]$ and each $t = 1,\ldots,T$ and $i = 1,\ldots,k$. Because $\hat{\theta}_T$ is consistent for θ_0, $\overline{\theta} \rightarrow^p \theta_0$. Sum these terms from 1 to T and divide by T to get

$$F_T(\hat{\theta}_T) = \frac{1}{T}\sum_{t=1}^{T}\left(f_t(\theta_0) - \frac{h\left(\frac{t}{T}\right)}{\sqrt{T}} \right) + \frac{1}{T}\sum_{t=1}^{T}\frac{\partial f_t(\overline{\theta})}{\partial\theta}(\hat{\theta}_T - \theta_0) + \frac{1}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h\left(\frac{t}{T}\right)$$

Multiply both sides by

$$\frac{\partial F_T(\hat{\theta}_T)'}{\partial\theta}W_T.$$

Assumptions 4 and 7 imply that there exists a T^* such that for all $T^* < T$

$$\begin{aligned} (\hat{\theta}_T - \theta_0) &= \left[\frac{\partial F_T(\hat{\theta}_T)'}{\partial\theta} W_T \frac{1}{T}\sum_{t=1}^{T}\frac{\partial f_t(\overline{\theta})}{\partial\theta} \right]^{-1} \frac{\partial F_T(\hat{\theta}_T)'}{\partial\theta}W_T \\ &\quad \times \left[-\frac{1}{T}\sum_{t=1}^{T}\left(f_t(\theta_0) - \frac{h\left(\frac{t}{T}\right)}{\sqrt{T}} \right) - \frac{1}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h\left(\frac{t}{T}\right) \right]. \end{aligned} \tag{20}$$

This equality holds because with probability one the first order condition

$$F_T(\hat{\theta}_T)'W_T\frac{\partial F_T(\hat{\theta}_T)}{\partial\theta} = 0$$

will be satisfied and

$$\frac{\partial F_T(\hat{\theta}_T)'}{\partial\theta}W_T\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f_t(\overline{\theta})}{\partial\theta} \rightarrow^p M'\Sigma^{-1}M$$

where Σ is nonsingular and M is of full rank. Substitute (20) into (18) to get

$$F_{sT}(\hat{\theta}_T) = F_{sT}(\theta_0) + \frac{\partial F_{sT}(\overline{\overline{\theta}})}{\partial\theta}\left[\frac{\partial F_T(\hat{\theta}_T)'}{\partial\theta}W_T\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f_t(\overline{\theta})}{\partial\theta}\right]^{-1}\frac{\partial F_T(\hat{\theta}_T)'}{\partial\theta}W_T \times \left[-\frac{1}{T}\sum_{t=1}^{T}\left(f_t(\theta_0) - \frac{h\left(\frac{t}{T}\right)}{\sqrt{T}}\right) - \frac{1}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}h\left(\frac{t}{T}\right)\right].$$

Multiply both sides by $\sqrt{T}W_T^{1/2}$. The convergence of the first term on the right hand side is given by (17). The remaining terms converge as:

$$W_T^{1/2}\frac{\partial F_{sT}(\overline{\overline{\theta}})}{\partial\theta}\left[\frac{\partial F_T(\hat{\theta}_T)'}{\partial\theta}W_T\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f_t(\overline{\theta})}{\partial\theta}\right]^{-1}\frac{\partial F_T(\hat{\theta}_T)'}{\partial\theta}W_T^{1/2} \to^p s\overline{M}(\overline{M}'\overline{M})^{-1}\overline{M}'$$

$$\sqrt{T}W_T^{1/2}\frac{1}{T}\sum_{t=1}^{T}\left(f_t(\theta_0) - \frac{h\left(\frac{t}{T}\right)}{\sqrt{T}}\right) \Rightarrow B_m(1)$$

$$\sqrt{T}W_T^{1/2}\frac{1}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}h\left(\frac{t}{T}\right) \to^p \Sigma^{-1/2}\int_0^1 h(\nu)d\nu = \Sigma^{-1/2}H(1)$$

To simplify the presentation define $P_{\overline{M}} = \overline{M}(\overline{M}'\overline{M})^{-1}\overline{M}'$ and $P^c_{\overline{M}} = (I_m - P_{\overline{M}})$. The above results imply

$$\begin{aligned}\sqrt{T}W_T^{1/2}F_{sT}(\hat{\theta}_T) &\Rightarrow B_m(s) - sP_{\overline{M}}B_m(1) + \Sigma^{-1/2}H(s) - sP_{\overline{M}}\Sigma^{-1/2}H(1)\\ &= P^c_{\overline{M}}B_m(s) + P_{\overline{M}}(B_m(s) - sB_m(1)) + P^c_{\overline{M}}H(s) + P_{\overline{M}}(H(s) - sH(1))\\ &= P^c_{\overline{M}}(B_m(s) + H(s)) + P_{\overline{M}}((B_m(s) - sB_m(1)) + (H(s) - sH(1))).\end{aligned}$$

Appendix 2
Derivation of optimal tests

Sowell (1993) presented a generic procedure to obtain specification tests in the GMM framework. The foundation of the procedure is the weak convergence of the normalized partial sums of sample moment conditions. A statistical test is characterized by a functional defined on the space of possible sample paths:

1. The functional applied to the normalized partial sums of observed sample moment conditions gives a test statistic.
2. The functional applied to the weak convergence limit under the null characterizes the test statistic's distribution under the null hypothesis.
3. The functional applied to the weak convergence limit under the sequence of local alternatives characterizes the test statistic's distribution under the sequence of local alternatives.

For a given local alternative the optimal test is defined in terms of the optimal functional. The asymptotically optimal test is determined by the Radon-Nikodym derivative of the measures[8] implied by the null hypothesis and the local alternatives. The Radon-Nikodym derivative is the functional that implies an asymptotically optimal test for the sequence of local alternatives. For composite alternatives the functional that implies the optimal test, with the greatest weighted average power, is the weighted average of the Radon-Nikodym derivative. The weighting distribution is defined on the alternatives consistent with the composite alternative hypothesis. This general procedure is used to derived the optimal tests. The derivation will first considering general local alternatives $h(\eta, \pi, s)$, then attention will be focused on the one time structural break alternatives with unknown breakpoint: $h(\eta, \pi, s) = U(\pi - s)\eta$. To reduce notation $h(s)$ will denote $h(\eta, \pi, s)$ and $H(s)$ will denote $H(\eta, \pi, s)$.

Let C denote a normalized set of eigenvectors such that

$$P_{\overline{M}} = C'\Lambda C \quad \text{and } P_{\overline{M}} = C'(I_m - \Lambda)C,$$

where $\Lambda = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}$. **Theorem 1** shows that under the null the normalized partial sum of sample moments converge to

$$CdZ(s) = \begin{bmatrix} dB_k^{\circ}(s) \\ dB_{m-k}(s) \end{bmatrix}$$

and under the alternative hypotheses it converges to

$$CdZ(s) = \left[CP_{\overline{M}}(h(s) - H(1)) + CP_{\overline{M}}^{c}h(s)\right] ds + \begin{bmatrix} dB_k^{\circ}(s) \\ dB_{m-k}(s) \end{bmatrix} \tag{21}$$

These limiting stochastic processes only differ in terms of the drift. Asymptotically, testing between the hypotheses is equivalent to testing for the drift of a stochastic process. Given

[8] The measures are on the space of sample paths for the limiting stochastic process.

weighting distribution function $J(\pi)$ for the location of the instability and weighting distribution $R(\eta, \pi)$ for the magnitude of the instability. Theorem 3 in Sowell (1994) implies that the greatest weighted average power test will reject the null hypothesis of moment stability if

$$\int\int \exp\left\{\int_0^1 \mu(\eta,\pi,s)'dZ(s) - \frac{1}{2}\int_0^1 \mu(\eta,\pi,s)'\mu(\eta,\pi,s)ds\right\} dR(\eta,\pi)dJ(\pi) \geq k_\alpha$$

where

$$\mu(\eta,\pi,s) = CP_{\overline{M}}^{c} h(\eta,\pi,s) + CP_{\overline{M}}(h(\eta,\pi,s) - H(\eta,\pi,1))$$

and k_α is defined so that the test has size alpha. The one time break alternative is characterized by the function

$$h(s) = U(\pi - s)\eta$$

where η is the parameter vector that indicates the magnitude of the onetime break and

$$U(x) \equiv \begin{cases} 0, & x < 0 \\ 1, & 0 < x \end{cases}.$$

For this alternative the drift of the asymptotic stochastic process becomes

$$\mu(\eta,\pi,s) = P_{\overline{M}}(U(\pi - s) - \pi)\eta + P_{\overline{M}}^{c} U(\pi - s)\eta.$$

The tests statistic can be written

$$\begin{aligned} & \int\int \exp\left\{\eta' P_{\overline{M}} Z(\pi) - \eta'\pi P_{\overline{M}} Z(1) + \eta' P_{\overline{M}}^{c} Z(\pi) - \frac{1}{2}\eta' P_{\overline{M}}\pi(1-\pi)\eta - \frac{1}{2}\eta' P_{\overline{M}}^{c}\pi\eta\right\} dR(\eta,\pi)dJ(\pi) \\ = & \int\int \exp\left\{\eta' Z(\pi) - \frac{1}{2}\eta'\left(P_{\overline{M}}\pi(1-\pi) + P_{\overline{M}}^{c}\pi\right)\eta\right\} dR(\eta,\pi)dJ(\pi) \end{aligned}$$

For the weighting distributions

1. $R(\eta,\pi)$, a normal weighting distribution with zero mean and covariance matrix $U(\pi)$ where $U(\pi)^{-1} = \frac{1+c}{c} I_m - \left(P_{\overline{M}}\pi(1-\pi) + P_{\overline{M}}^{c}\pi\right)$,
2. and $J(\pi)$, with density proportional to $\left(\frac{1+c}{c}\right)^{m/2} |U(\pi)|^{1/2}$,

the test statistic reduces to

$$\int_0^1 \exp\left\{\frac{1}{2}\frac{c}{1+c} Z(\pi)' Z(\pi)\right\} d\pi.$$

Note the implied hypothesis test is unaffected by multiplicative constants. The parameter c controls the variance of the normal weighting density. Small values of c imply that the covariance matrix is approaching a matrix of zeros, this implies more weight to alternatives closer to the null. Conversely, large values of c imply more weight to alternatives further from the null.

For alternatives arbitrarily "close to" and "far away from" the null, the test statistics converge to

$$\tilde{E} = \lim_{c\to\infty} \int_0^1 \exp\left\{\frac{1}{2}\frac{c}{1+c} Z(\pi)' Z(\pi)\right\} d\pi = \int_0^1 \exp\left\{\frac{1}{2} Z(\pi)' Z(\pi)\right\} d\pi$$

and

$$\tilde{L} = \lim_{c \downarrow 0} \frac{2\left(\int_0^1 \exp\left\{\frac{1}{2}\frac{c}{1+c} Z(\pi)' Z(\pi)\right\} d\pi - 1\right)}{c} = \int_0^1 Z(\pi)' Z(\pi) d\pi.$$

These are the functionals that imply asymptotically optimal tests. The functionals applied to the empirical stochastic process created from the sample moments,

$$\sqrt{T} W_T^{1/2} F_{[sT]}(\hat{\theta}_T),$$

give the test statistics

$$E = \frac{1}{T} \sum_{t=1}^{T} \exp\left\{\frac{1}{2} T F_t(\hat{\theta}_T)' W_T F_t(\hat{\theta}_T)\right\},$$

and

$$L = \sum_{t=1}^{T} F_t(\hat{\theta}_T)' W_T F_t(\hat{\theta}_T).$$

In the paper these two statistics are each decomposed into two independent tests: one for the overidentifying restrictions and one for the overidentifying restrictions. The functional associated with the specialized statistics are

$$\tilde{E}_A = \int_0^1 \exp\left\{\frac{1}{2} Z(\pi)' P_{\overline{M}} Z(\pi)\right\} d\pi$$

and

$$\tilde{L}_A = \int_0^1 Z(\pi)' P_{\overline{M}} Z(\pi) d\pi$$

are optimal for instability in the identifying restrictions. The functionals

$$\tilde{E}_B = \int_0^1 \exp\left\{\frac{1}{2} Z(\pi)' P_{\overline{M}}^{c} Z(\pi)\right\} d\pi$$

and

$$\tilde{L}_B = \int_0^1 Z(\pi)' P_{\overline{M}}^{c} Z(\pi) d\pi$$

are optimal for instability in the overidentifying restrictions. These functionals could be derived individually by start with local alternatives restricted to one of the subspace, see footnote 4 for more details.

The distributions under the null and under the sequence of local alternatives are characterized by applying these functionals to the limiting stochastic process given in Theorem 1.

References

Andrews, Donald W.K. (1987), "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica* **55**, 1465-1471.

Ghysels, Eric, Alain Guay and Alastair Hall (1995), "Predictive Tests for Structural Change with Unknown Breakpoint," C.R.D.E., University of Montreal.

Hansen, Lars Peter (1982), " Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* **50**, 1029-1054.

McLeish, D. L. (1975), "Invariance Principles for Dependent Variables," *Z. Wahrscheinlichkeitstheorie und Verw. Gibiete* **32**, 165-178.

Newey, Whitney K. (1985), "Generalized Method of Moments Specification Testing," *Journal of Econometrics* **29**, 229-256.

Phillips, P.C.B. and S.N. Durlauf (1986), "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies* **53**, 473-495.

Sowell, Fallaw (1993), "A New Approach to Testing Parameter Instability in the Generalized Method of Moments Framework," Carnegie Mellon University, Graduate School of Industrial Administration.

Sowell, Fallaw (1994), "An Approach to Specification Tests in the Generalized Method of Moments Framework," Carnegie Mellon University, Graduate School of Industrial Administration.

Sowell, Fallaw (1995), "Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework," Carnegie Mellon University, Graduate School of Industrial Administration.

forthcoming in *Econometrica*