GMM Based Tests for Locally Misspecified Models

Anil K. Bera
Department of Economics
University of Illinois, USA

Walter Sosa Escudero*
University of San Andrés and
National University of La Plata, Argentina

August 29, 2002

Abstract

This paper shows that the standard Newey-West GMM based test is sensitive to the presence of locally misspecified alternatives. In particular, such test is shown to have incorrect size when the null model is locally contaminated, making the test spuriously reject the null hypothesis even when it is correct. After obtaining the distribution of the Newey-West test statistic under the misspecified alternative, the paper derives a Bera-Yoon modified test based on a restricted GMM estimator, which has the correct size under local misspecification. The paper can be seen as extending the likelihood based results of Saikkonen (1989), Davidson and MacKinnon (1987) and Bera and Yoon (1993) to the GMM framework, which does not require full specification of the underlying probabilistic model.

JEL Classification: C12, C23, C52
Keywords: Generalized method of moments, testing, misspecification.

*Corresponding author: Walter Sosa Escudero, Department of Economics, University of San Andres, Vito Dumas 284, B1644BID Victoria, Argentina, Ph/Fax: (54-11) 4725-7000, e-mail: wsosa@udesa.edu.ar. This paper was written while the second author visited the Department of Economics of the University of Illinois at Urbana-Champaign, whose hospitality is greatly acknowledged. The usual disclaimer applies.

1 Introduction

Like any other, econometric models are useful simplifications of a complex reality. The simplification process that leads to a particular specification usually involves a trade-off between accuracy with reality, and parsimony. Hence one particular type of modeling error consists in proposing a model that it is too parsimonious, in the sense that the model ignores some relevant aspect of reality, which might negatively affect its use. In the context where the small model is a particular case of a larger model, in many practical cases restricted models are easier to handle and hopefully more efficient than larger, unrestricted models, so in such cases the benefits associated to being able to impose restrictions have both methodological and practical advantages.

A common practice when there are strong preferences for a restricted model is to start with a small model and then check whether particular departures from this null model are supported or rejected by the data. A well known methodological problem with this approach is that the design of an appropriate test requires some specific knowledge of both the null model and the departure being examined. In the context of maximum-likelihood estimation and nested hypotheses, the restricted model usually implies a certain restriction on a parameter or group of parameters of a general, unrestricted model of which the null model is a particular case. When the null model is easier to estimate than the unrestricted model, this type of parametric restriction can be easily (and in some cases, optimally) tested based on the Rao-Score (or Lagrange Multiplier) test which is based on the estimation of a restricted model.

An obvious problematic situation arises when the null is false and the alternative 'true' model does not correspond to the one believed by the researcher. In such cases, it is natural to expect that tests designed to detect particular departures from the null do not behave correctly under other possible departures. In a more concrete setup, and without loss of generality, suppose that the econometric model consists in a probability distribution that can be fully characterized by three parameters γ, ψ and ϕ. Suppose that the model can be easily estimated under appropriate restrictions on ψ and ϕ, like $\psi = \psi_*$ and $\phi = \phi_*$. Let γ_0, ψ_0 and ϕ_0 be the true, unknown values of the parameters, and suppose that the researcher discards the possibility that $\phi_0 \neq \phi_*$ and proceeds by estimating γ under the restriction that $\psi = \psi_*$, and then tests the null $\psi = \psi_*$ using a Rao-Score (RS) test. When the alternative model is correctly specified, that is, when $\phi_0 = \phi_*$, a well known result is that the RS test is locally most powerful to detect departures from the null in the direction implicit in $\psi \neq \psi_*$ (Cox and Hinkley (1974)).

A natural question is what happens to the RS test when the alternative model is incorrectly specified, that is, when $\phi_0 \neq \phi_*$? Davidson and MacKinnon (1987) and Saikkonen (1989) have studied the asymptotic distribution of the RS test under *local misspecification*. They have found that even under the null hypothesis $\psi = \psi_*$ the RS test no longer has an asymptotic central chi-square distribution and consequently, even under the null hypothesis the test tends to spuriously reject it too often. In this case, even though rejections can still be informative about the falseness of the null model, RS tests are of a rather limited use when trying to explore the appropriate nature of the misspecification, since the test rejects when the null hypothesis of interest is not true but also when the alternative hypothesis is misspecified.

Based on the results of Saikkonen (1987) and Davidson and MacKinnon (1989), Bera and Yoon (1993) propose a modified RS test that though still based on the restricted maximum likelihood estimator, it is insensitive to local misspecification, that is, the proposed test statistic has the central chi-squared distribution under $H_0 : \psi = \psi_0$ independently of whether $\phi = \phi_*$ or not, in a local sense. This principle has been succesfully implemented in recent research. For example, Bera, Sosa-Escudero and Yoon (2001) have found that the presence of first order serial correlation makes the standard Breusch and Pagan (1980) test for random effects reject the null hypothesis too often, implying that rejections of the null may be due to the presence of random effects but also due to the presence of first order serial correlation. Based on the principle mentioned before, they derive a RS modified test for random effects that is not affected by the presence of local serial correlation. In a similar fashion, Anselin, Bera, Florax and Yoon (1996) derive tests for spatial autocorrelation that are not sensitive to the presence of local lag dependence. Baltagi and Li (1999) use the Bera-Yoon principle to obtain tests for functional form misspecification and spatial correlation.

A well known restrictive feature of maximum-likelihood based procedures is that they require full specification of the underlying probabilistic model, which limits the scope of the modified test by Bera and Yoon (1993) in situations where researchers cannot guarantee such a detailed knowledge. The final goal of this paper is to derive a Bera-Yoon principle that is based on a restricted generalized method of moments (GMM) estimation framework which does not require full specification of the probabilistic structure but only some moment conditions. We begin by exploring a standard testing framework based on restricted GMM estimation with correctly specified alternatives proposed by Newey and West (1987), which is a GMM equivalent of the RS principle in the likelihood context. We then derive a result similar to that of Saikkonen (1989) that shows that the Newey-West test behaves incorrectly under local misspecification. Section 3 presents the main result of the paper: a modified test that is insensitive to local misspecification of the alternative hypothesis. Section 4 presents some concluding remarks and directions for future work.

2 The effect of locally misspecified alternatives

Following Hall (2002, Chapter 3), in the context of GMM estimation a *model* is a particular set of assumptions about the data generation process for a (possibly vector valued) random variable of interest y. For the purposes of our problem, and without loss of generality, suppose that the underlying statistical model can be parametrized by a 3×1 vector of parameters $\theta = (\gamma, \psi, \phi)'$, $\theta \in \Theta \subseteq \Re^3$. We will assume that there is a vector of m functions $g(y, \theta)$ for which the following moment conditions are satisfied:

$$E\ g(y, \theta) = 0 \qquad \text{if and only if } \theta = \theta_0 \tag{1}$$

These are the moment conditions of the problem, and for identification purposes we will require $m \geq 3$. We will refer to $\theta_0 = (\gamma_0, \psi_0, \phi_0)'$ as the 'true' values of the parameters.

There is available an i.i.d. sample $y_1, y_2, \ldots, y_n$ of n observations from the same data generation process. Define the sample analog of the left-hand side of (1) as follows:

$$g_n(\theta) \equiv \frac{1}{n} \sum_{i=1}^{n} g(y_t, \theta)$$

and let Ω_n be any $m \times m$ positive definite symmetric matrix. The (unrestricted) GMM estimator of θ_0 is defined as:

$$\hat{\theta}_n \equiv \text{argmin } Q_n(\theta)$$

with $Q_n(\theta) \equiv g_n(\theta)'\Omega_n^{-1}g_n(\theta)$. Let $\Omega \equiv E[g(y,\theta_0)g(y,\theta_0)']$. Then, according to Hansen's (1982) results, efficiency requires that we use Ω_n such that $\Omega_n \rightarrow \Omega$ as n goes to infinity, where '$\rightarrow$' denotes convergence in probability. Let $\nabla_\theta g(y,\theta)$ be the $m \times 3$ Jacobian matrix of $g(y,\theta)$. Let $G \equiv E[\nabla_\theta g(y,\theta_0)]$, and define

$$G_n(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} \nabla_\theta g(y_t,\theta)$$

and

$$\Omega_n(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} g(y_i,\theta)g(y_i,\theta)'$$

It will be useful to use label the gradient of the GMM objetive function as follows:

$$\nabla_\theta Q_n(\theta) \equiv G_n(\theta)\Omega_n^{-1}(\theta)g(y,\theta) \tag{2}$$

Also, let $B \equiv G'\Omega^{-1}G$, and $B_n(\theta) \equiv G_n(\theta)'\Omega_n^{-1}(\theta)G_n(\theta)$

The analytic framework of this paper is a situation where a researcher is interested in a null model based on the estimation of the parameter γ after which, for specification check or model search purposes, the interest is in testing $H_0 : \psi = \psi_*$ against $H_A : \psi \neq \psi_*$. That is, estimation and inference proceeds by ignoring the possible presence of the parameter ϕ. Suppose that for some value $\phi = \phi_*$ the following relationship holds:

$$g_1(\gamma,\psi) = g(\gamma,\psi,\phi_*)$$

That is, g_1 is a restricted version of the moment conditions. Under some standard regularity conditions (described in detail in Appendix 1) a restricted GMM estimator for γ is readily available and it is given by:

$$\tilde{\theta} = \text{argmin } Q_1(\gamma,\psi) \qquad \text{,subject to } \psi = \psi_*$$

where, according to the previous assumptions:

$$Q_1(\psi,\gamma) = g_1(\gamma,\psi)'\Omega_n^{-1}g_1(\gamma,\psi) = g(\gamma,\psi,\phi_*)'\Omega_n^{-1}g(\gamma,\psi,\phi_*).$$

A standard procedure to test the null hypothesis $H_0 : \psi = \psi_*$ can be based on the gradient test proposed by Newey and West (1987), which is the GMM equivalent of the Rao-Score test in the context of maximum-likelihood estimation:

$$LM_\psi(\tilde{\theta}) = n\ \nabla_\psi Q_1(\tilde{\theta})'\ B_{\psi\cdot\gamma}^{-1}\ \nabla_\psi Q_1(\tilde{\theta})$$

where $B_{\psi\cdot\gamma} \equiv B_{\psi\psi} - B_{\psi\gamma}B_{\gamma\gamma}^{-1}B_{\gamma\psi}$. According to the results in Newey and West (1987) when the model is correctly specified, that is, when $\phi_* = \phi_0$, when $\psi = \psi_* + \eta/\sqrt{n}$, $LM_\psi(\tilde{\theta})$ has an asymptotic non-central chi-square distribution $\chi^2(\lambda_\eta)$ with non-centrality parameter $\lambda_\eta = \eta' B_{\psi\cdot\gamma}\eta$. In particular, when the null hypothesis is true ($\eta = 0$) the test has asymptotic central chi-square distribution and, hence, when the adopted simplification on the nuisance parameter is correct, the test has power only in the alternative it was designed to detect.

A natural question is what happens with $LM_\psi(\tilde{\theta})$ when $\phi_* \neq \phi_0$, that is, when it is based on a false restriction on a parameter that is not of direct interest to the testing procedure. A general answer to this question would require to explore the nature of the misspecification in detail, but a particular and commonly proposed situation can be analyzed. Consider the case where the model is *locally misspecified* in the sense that $\phi_0 = \phi_* + \delta/\sqrt{n}$. In this setup, we are interested in exploring what happens to $LM_\psi(\tilde{\theta})$ when we proceed under the assumption $\phi = \phi_*$ and hence the assumed alternative differs from the true one in a local sense. Let '$\sim$' denote convergence in distribution. The following Proposition shows that in general the $LM_\psi(\tilde{\theta})$ becomes contaminated by the presence of this local misspecification:

Proposition 1 *Under $H_0 : \psi = \psi_0$ and when $\phi_0 = \phi_* + \delta/\sqrt{n}$, $LM_\psi(\tilde{\theta}) \sim \chi_r^2(\lambda(\delta))$, with $\lambda(\delta) \equiv \delta' B_{\psi\phi\cdot\gamma}B_{\psi\cdot\gamma}^{-1}B_{\psi\phi\cdot\gamma}\delta$.*

Proof: See Appendix 2.

The Lemma says that under local misspecification of the nuisance parameter, even when the null hypothesis of interest is correct, the test for $H_0 : \psi = \psi_0$ will not have a central chi-square distribution as it would have when the alternative is correctly specified. Then, even when the null is correct, the testing procedure will tend to reject it due to the locally misspecified parameter ϕ. Consequently, under this particular type of misspecification, if such test is used in practice, rejections are not informative about the nature of the departure from the null model since both, the fact that the null hypothesis $H_0 : \psi = \psi_*$ is not true, and/or the presence of local misspecifitation ($\phi_0 = \phi_* + \delta/\sqrt{n}$) would induce the test to reject the null. This result can be seen as an extension of that of Davidson and MacKinnon (1986) and Saikkonnen (1989) for the GMM framework.

3 Testing with locally misspecified alternatives

The search for a valid testing procedure that does not depend on the concurrent estimation of the nuisance parameter ϕ will be based on the fact that an alternative procedure that would allow us to test $H_0 : \psi = \psi_*$, and which does not depend on imposing restrictions on ϕ, can be based on the *partially restricted* estimator $\bar{\theta} = (\bar{\gamma}, \psi_*, \bar{\phi})$, which solves:

$$\bar{\theta} = \text{argmin } Q_n(\theta) \qquad \text{subject to } \psi = \psi_*$$

Let $\beta \equiv (\gamma, \phi)'$ so θ is now partitioned as $\theta = (\psi, \beta')'$. Again based on the results of Newey and West (1987), under $H_0 : \psi = \psi_*$ the gradient test

$$LM_\psi(\bar{\theta}) = n \nabla_\psi Q_n(\bar{\theta})' \, B_{\psi \cdot \beta}^{-1} \, \nabla_\psi Q_n(\bar{\theta})$$

has a limiting central chi-square distribution with r degrees of freedom, where $B_{\psi \cdot \beta} \equiv B_{\psi\psi} - B_{\psi\beta} B_{\beta\beta}^{-1} B_{\beta\psi}$. This second procedure is based on the estimation of γ and ϕ, but the purpose of this paper is to derive a valid test for $H_0 : \psi = \psi_*$ that does not require the concurrent estimation of the nuisance parameter ϕ but it can be based on the 'fully restricted' estimator $\tilde{\theta}$. By 'valid' we mean a testing prodecure for the null hypothesis $H_0 : \psi = \psi_*$ against $H_A : \psi \neq \psi_*$ whose asymptotic distribution is not affected by whether $\phi_0 = \phi_*$ or $\phi_0 \neq \phi_*$, at least in a local sense, so that rejections indicate departures from $H_0 : \psi = \psi_*$ exclusively.

According to the results in Newey and McFadden (1994), $LM_\psi(\bar{\theta})$ is asympotically equivalent to a test where the partially restricted GMM estimator $\bar{\theta}$ is replaced by an optimal one-step estimator derived from an initial $\sqrt{n}$-consistent estimator. Call $\hat{\theta}$ the optimal one-step estimator. Then, an asymptotically equivalent version of $LM_\psi(\bar{\theta})$ is:

$$LM_\psi(\hat{\theta}) = \nabla_\psi Q_n(\hat{\theta})' \, B^{-1}_{\psi\cdot\gamma\phi} \, \nabla_\psi Q_n(\hat{\theta}) \sim \chi^2_r(0) \tag{3}$$

In order to derive the one-step estimator we will follow the arguments in the derivation of Bera and Yoon's (1993) procedure within the context of maximum-likelihood estimation. Under the null hypothesis $H_0 : \psi = \psi_*$ and when $\phi_0 = \phi_* + \delta/\sqrt{n}$, a trivial initial $\sqrt{n}$-consistent estimator that can be proposed under the assumptions of this paper is $\theta_I = (\tilde{\gamma}, \psi_0, \phi_*)'$, where $\tilde{\gamma}$ is the GMM estimator of γ from the fully restricted estimation problem. From $\phi_0 = \phi_* + \delta/\sqrt{n}$ we immediately see that ϕ_* is also a $\sqrt{n}$-consistent estimator.

When the null hypothesis $H_0 : \psi = \psi_*$ is true, consider a GMM model for the estimation of γ and ϕ under the restriction implied by the null hypothesis. Following Newey and McFadden (1994) the optimal one-step estimator is given by:

$$\begin{bmatrix} \hat{\gamma} \\ \hat{\phi} \end{bmatrix} = \begin{bmatrix} \tilde{\gamma} \\ \phi_* \end{bmatrix} + \begin{bmatrix} B_{\gamma\gamma}(\theta) & B_{\gamma\phi}(\theta) \\ B_{\phi\gamma}(\theta) & B_{\phi\phi}(\theta) \end{bmatrix}^{-1}_{\theta=\tilde{\theta}} \begin{bmatrix} \nabla_\gamma Q_n(\theta) \\ \nabla_\phi Q_n(\theta) \end{bmatrix}_{\theta=\tilde{\theta}}$$

From the first order conditions used to derive $\tilde{\gamma}$, $\nabla_\gamma Q_n(\tilde{\theta}) = 0$, we obtain:

$$(\hat{\phi} - \phi_*) = B^{-1}_{\phi\cdot\gamma}(\tilde{\theta}) \nabla_\phi Q_n(\tilde{\theta}) \tag{4}$$

where $B_{\phi\cdot\gamma}(\theta) = B_{\phi\phi} - B_{\phi\gamma} B^{-1}_{\gamma\gamma} B_{\phi\gamma}$.

Following the stragegy of the proof of Proposition 1, we take a Taylor expansion of $g_n(\tilde{\theta})$ around $\bar{\theta}$ and then replace in the gradient $\nabla_\phi Q_n(\tilde{\theta})$ to get:

$$\nabla_\psi Q_n(\hat{\theta}) = \nabla_\psi Q_n(\bar{\theta}) + B_{\psi\phi}(\hat{\phi} - \phi_*) + o_p \tag{5}$$

Now replacing the right hand side of (4) in (5) we get:

$$\nabla_\psi Q_n(\hat{\theta}) = \nabla_\psi Q_n(\bar{\theta}) - B_{\psi\phi} B^{-1}_{\phi\cdot\gamma} \nabla_\phi Q_n(\bar{\theta}) + o_p \equiv \hat{\nabla}_\psi Q_n(\bar{\theta}) + o_p$$

Then, replacing in (3), and eliminating asymptotically irrelevant terms we obtain the asymptotically equivalent version of (3) which is:

$$LM_{\psi}^{*}(\tilde{\theta}) = \hat{\nabla}_{\psi} Q_n(\bar{\theta})' \, B_{\psi\cdot\gamma\phi}^{-1} \, \hat{\nabla}_{\psi} Q_n(\bar{\theta})$$

To get the final result, note that $B_{\psi\cdot\gamma\phi}$ can be seen as the sum of squared errors of projecting $\nabla_{\psi} Q_n$ on $\nabla_{\gamma} Q_n$ and $\nabla_{\phi} Q_n$. By standard linear projection algebra, this should be exactly the same as the sum of squared residuals of regressing $\nabla_{\psi} Q_n$ on $\nabla_{\phi} q_n$ after eliminating the linear effect of $\nabla_{\gamma} Q_n$, that is, $B_{\psi\phi\cdot\gamma}$, then, replacing we get:

$$LM_{\psi}^{*}(\tilde{\theta}) = \hat{\nabla}_{\psi} Q_n(\bar{\theta})' \, B_{\psi\phi\cdot\gamma}^{-1} \, \hat{\nabla}_{\psi} Q_n(\bar{\theta})$$

which has a central chi-square distribution under $H_0 : \psi = \psi_*$ and when $\phi_0 = \phi_* + \delta/\sqrt{n}$. We summarize the main result of this section in the next Proposition.

Proposition 2 *Define a test statistic* $LM_{\psi}^{*}(\tilde{\theta}) = \hat{\nabla}_{\psi} Q_n(\bar{\theta})' \, B_{\psi\phi\cdot\gamma}^{-1} \, \hat{\nabla}_{\psi} Q_n(\bar{\theta})$*, where* $\hat{\nabla}_{\psi} Q_n(\bar{\theta}) \equiv \nabla_{\psi} Q_n(\bar{\theta}) - B_{\psi\phi} B_{\phi\cdot\gamma}^{-1} \nabla_{\phi} Q_n(\bar{\theta})$*. Then, under* $H_0 : \psi = \psi_*$ *and when* $\psi_0 = \psi_* + \delta/\sqrt{n}$*,* $LM_{\psi}^{*}(\tilde{\theta}) \sim \chi^2(0)$*.*

This result proposes a GMM based statistic that, unlike the Newey-West test, has asymptotic central χ^2 distribution under the null and when the alternative is possibly locally misspecified. The derived procedure is still based on the fully restricted null (small) model, which is makes it computationally convenient since all the elements to construct it are usually available after GMM estimation. The obtained result extends the likelihood-based Bera-Yoon (1993) principle to the GMM family, and hence it replaces the distributional requirements of the likelihood framework by moment conditions. Of course, the Newey-West test is a special case of it, and it corresponds to the case where $\delta = 0$, that is, to the case where the alternative is correctly specified.

4 Concluding remarks and future work

This paper studies the behavior of standard GMM based specification tests after estimating null models. In particular, the paper shows that when the alternative model is locally misspecified, the usual Newey-West test does not have a central χ^2 distribution under the null hypothesis, which leads to spurious rejections of the null. This limits the use of such test since rejections are not informative about the source of the misspecification. The paper derives explicitely the distribution of the Newey-West test under the misspecified alternative for the case of local misspecification, and based on this result it proposes a Bera-Yoon type of modified test that is insensitive to local misspecification. A convenient computational feature is that the proposed modified test can still be computed based on the restricted GMM estimator. The results of this paper can be seen as extending those of Saikkonen (1989), Davidson and MacKinnon (1987) and Bera and Yoon (1993) to the more general GMM framework that, unlike the original maximum-likelihood setup, does not require full specification of the probability model.

A potential source of applications of the proposed modified procedure is the modified tests recently proposed based on the Bera-Yoon principle in the likelihood framework, which includes recent work by Bera et. al. (2001), Anselin et al. (1996) and Baltagi and Li (1999). All these results are based on strict distributional assumptions, so it would be important to derive GMM based equivalent of those tests that do not require full specification of the underlying likelihood. Another source of applicability is to produce resistant versions of recent GMM based test. For example, Saavedra (2002) derives GMM versions of spatial lag and autocorrelation tests, and it would be interesting to derive a spatial lag test that is insensitive to serial correlation.

Another important route to explore is the evaluation of the small sample performance of the proposed test statistics through the design of Monte Carlo experiments. In particular, it would be important to explore the severity of local misspecification on the size of standard Newey-West tests. Also, it would be relevant to explore how restrictive is the local nature of the proposed solution, though the results in Bera et.al. (2001) look promising, in the sense that their local procedure is shown to provide a useful correction even in non-local frameworks.

References

Anselin, L., Bera, A., Florax, R. and Yoon, M., 1996, Simple diagnostic tests for spatial dependence, *Regional Science and Urban Economics*, 26, 77-104.

Baltagi, B. and Li, D., 1999, LM tests for functional form and spatial correlation, mimeo, Texas A&M University.

Bera, A. and Yoon, M., 1993, Specification testing with locally misspecified alternatives, *Econometric Theory*, 9, 649-658.

Bera, A. and Yoon, M., 2001, Adjustments of Rao's score test for distributional and local parametric misspecifications, mimeo, University of Illinois at Urbana-Champaign.

Bera, A., Sosa-Escudero, W. and Yoon, M., 2001, Tests for the error component model in the presence of local misspecification, *Journal of Econometrics*, 101, 1-23.

Breusch, T. and Pagan, A., 1980, The Lagrange multiplier test and its applications to model specification in econometrics, *Review of Economic Studies*, 47, 239-253.

Cox, D.R. and Hinkley, D.V., 1974, *Theoretical Statistics*, Champan & Hall/CRC, New York.

Godfrey, L. and Orme, C., 1996, On the behavior of conditional moment tests in the presence of unconsidered local alternatives, *International Economic Review*, 37, 2, 263-281.

Hall, A., 2002, *Generalized Method of Moments*, Oxford University Press, Oxford, forthcoming.

Hansen, L.P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica*, 50, 1029-1054.

Newey, W. and West, K., 1987, Hypothesis testing with efficient method of moments estimation, *International Economic Review*, 28,3, 777-787.

Newey, W. and McFadden, D., 1994, Large sample estimation and hypothesis testing, in Engle, R. and McFadden, D. (ed.), *Handbook of Econometrics*, Volume 4, 2111-2245, North Holland, Amsterdam.

Saavedra, L., 2002, Tests for spatial lag dependence based on method of moments estimation, *Regional Science and Urban Economics*, forthcoming.

Saikkonen, P., 1989, Asymptotic relative efficiency of the classical tests under misspecification, *Journal of Econometrics*, 42, 351-369.

Appendix 1: Regularity conditions

The following regularity conditions will be assummed. They correspond to the regularity conditions in Theorems 2.6 (consistency of GMM estimator), 3.4 (asymptotic normality of the GMM estimator) and 4.5 (variance estimation) in Newey and McFadden (1994).

1. Ω_n is a sequence of positive semi-definite matrices that converge to a positive definite symmetric matrix Ω, with $\Omega E[g(y, \theta_0) = 0$ if and only if $\theta = \theta_0$.

2. $\theta_0 \in \Theta$, which is a compact set.

3. $g(y, \theta)$ is continuous at each $\theta \in \Theta$ with probability one.

4. $E\left[\sup_{\theta \in \Theta} \parallel g(y, \theta) \parallel\right] < \infty$

5. $g_n(\theta)$ is continuously differentiable in a neighborhood N of θ_0

6. $\sqrt{n} g_n(\theta_0) \xrightarrow{D} N(0, \Omega)$, where $\xrightarrow{D}$ denotes convergence in distribution.

7. There is $G(\theta)$ that is continuous at θ_0 and $\sup_{\theta \in N} \parallel \nabla_\theta g_n(\theta) - G(\theta) \parallel \xrightarrow{p} 0$, where '$\xrightarrow{p}$' denotes convergence in probability.

8. For $G = G(\theta_0)$, $B = G' \Omega^{-1} G$ is non-singular.

Conditions (1)-(4) guarantee the consistency of the unrestricted GMM estimator ($\hat{\theta} \xrightarrow{p} \theta_0$). Conditions (1) to (8) imply asymptotic normality. See Newey and McFadden (1994) for proofs.

Appendix 2: Proof of Proposition 1

The logic of the proof follows closely Saikkonen (1989), who bases the proof in standard Taylor expansions. In the case of GMM the strategy, as in Newey and McFadden (1994), is to expand only the part of the gradient of the objetive function that corresponds to the moment conditions.

Consider the following version of the gradient of the GMM criterion function:

$$\nabla_\theta Q_n(\theta) = G'\Omega^{-1} g_n(\theta) \tag{6}$$

that is, (6) is simply (2) where G_n and Ω_n have been replaced by their population limits. The elements of the matrix B will be denoted by $B_{\psi\psi}, B_{\psi\gamma}$, etc. We will use B_ψ to denote the row of B corresponding to ψ. For example, $B_\psi = (B_{\psi\gamma}, B_{\psi\psi}, B_{\psi\phi})'$. B_γ and B_ϕ are defined alike.

Take a first order Taylor expansion of $g_n(\tilde{\theta})$ about $g_n(\theta_0)$ and replace in (6) to get:

$$\begin{aligned} \nabla_\psi Q_n(\tilde{\theta}) &= \nabla_\psi Q_n(\theta_0) + B_\psi(\tilde{\theta} - \theta_0) + o_p \\ &= \nabla_\psi Q_n(\theta_0) + B_{\psi\gamma}(\tilde{\gamma} - \gamma_0) - B_{\psi\phi}(\theta_1)\ \ \delta/\sqrt{n} + o_p \end{aligned} \tag{7}$$

Let $\theta_* \equiv (\gamma_0, \psi_*, \phi_*)'$, and now consider a Taylor expansion of $g_n(\theta_*)$ about $g_n(\theta_0)$ to get:

$$\begin{aligned} \nabla_\gamma Q_n(\theta_*) &= \nabla_\gamma Q_n(\theta_0) + B_\gamma(\theta_* - \theta_0) + o_p \\ &= \nabla_\gamma Q_n(\theta_0) - B_{\gamma\phi}\ \ \delta/\sqrt{n} + o_p \end{aligned} \tag{8}$$

Finally, a third Taylor expansion of $g_n(\theta_*)$ about $g_n(\tilde{\theta})$ leads to:

$$\begin{aligned} \nabla_\gamma Q_n(\theta_*) &= \nabla_\gamma Q_n(\tilde{\theta}) + B_\gamma(\theta_* - \tilde{\theta}) + o_p \\ &= -B_{\gamma\gamma}(\tilde{\gamma} - \gamma_0) + o_p \end{aligned} \tag{9}$$

From (9) we obtain:

$$\tilde{\gamma} - \gamma_0 = -B_{\gamma\gamma}^{-1} \nabla_\gamma Q_n(\theta_*) + o_p$$

Replacing in (7) we get:

$$\nabla_\psi Q(\tilde{\theta}) = \nabla_\psi Q(\theta_0) - B_{\psi\gamma} B_{\gamma\gamma}^{-1} \nabla_\gamma Q_n(\theta_*) - B_{\psi\phi} \;\; \delta/\sqrt{n} + o_p$$

Now replace $\nabla_\gamma Q_n(\theta_*)$ by the expression in (8):

$$\begin{aligned}
\nabla_\psi Q_n(\tilde{\theta}) &= \nabla_\psi Q_n(\theta_0) - B_{\psi\gamma} B_{\gamma\gamma}^{-1} \left[\nabla_\gamma Q_n(\theta_0) - B_{\gamma\phi} \;\; \delta/\sqrt{n} \right] - B_{\psi\phi} \;\; \delta/\sqrt{n} + o_p \\
&= \nabla_\psi Q_n(\theta_0) - B_{\psi\gamma} B_{\gamma\gamma}^{-1} \nabla_\gamma Q_n(\theta_0) + \\
&\qquad + B_{\psi\gamma} B_{\gamma\gamma}^{-1} B_{\gamma\phi} \;\; \delta/\sqrt{n} - B_{\psi\phi} \;\; \delta/\sqrt{n} + o_p \\
&= \nabla_\psi Q_n(\theta_0) - B_{\psi\gamma} B_{\gamma\gamma}^{-1} \nabla_\gamma Q_n(\theta_0) - \\
&\qquad - \left[B_{\psi\phi} - B_{\psi\gamma} B_{\gamma\gamma}^{-1} B_{\gamma\phi} \right] \delta/\sqrt{n} + o_p \\
&= \nabla_\psi Q_n(\theta_0) - B_{\psi\gamma} B_{\gamma\gamma}^{-1} \nabla_\gamma Q_n(\theta_0) - B_{\psi\phi\cdot\gamma} \delta/\sqrt{n} + o_p \qquad (10)
\end{aligned}$$

Now by the assumed regularity of GMM

$$\nabla_\theta Q_n(\theta_0) \sim N(0, B)$$

Then the asymptotic distribution of (10) is:

$$\nabla_\psi Q(\tilde{\theta}) \sim N(-B_{\psi\phi\cdot\gamma} \delta/\sqrt{n}, V)$$

with:

$$\begin{aligned}
V &= E \left[\nabla_\psi Q_n(\theta_0) - B_{\psi\gamma} B_{\gamma\gamma}^{-1} \nabla_\gamma Q_n(\theta_0) \right] \left[\nabla_\psi Q_n(\theta_0) - B_{\psi\gamma} B_{\gamma\gamma}^{-1} \nabla_\gamma Q_n(\theta_0) \right]' \\
&= E \left[\nabla_\psi Q_n(\theta_0) - B_{\psi\gamma} B_{\gamma\gamma}^{-1} \nabla_\gamma Q_n(\theta_0) \right] \left[\nabla_\psi Q_n(\theta_0)' - \nabla_\gamma Q_n(\theta_0)' B_{\gamma\gamma}^{-1} B_{\psi\gamma} \right] \\
&= B_{\psi\psi} - B_{\psi\gamma} B_{\gamma\gamma}^{-1} B_{\psi\gamma} - B_{\psi\gamma} B_{\gamma\gamma}^{-1} B_{\gamma\psi} - B_{\psi\gamma} B_{\gamma\gamma}^{-1} B_{\gamma\gamma} B_{\gamma\gamma}^{-1} B_{\gamma\psi} \\
&= B_{\psi\psi} - B_{\psi\gamma} B_{\gamma\gamma}^{-1} B_{\psi\gamma} \;\; \equiv \;\; B_{\psi\cdot\gamma}
\end{aligned}$$

Then,

$$\nabla_\psi Q_n(\tilde{\theta}) \sim N(-B_{\psi\phi\cdot\gamma} \delta/\sqrt{n}, B_{\psi\cdot\gamma})$$

and consequently,

$$LM_\psi(\tilde{\theta}) = \nabla_\psi Q_n(\tilde{\theta})' \; B_{\psi\cdot\gamma}^{-1} \; \nabla_\psi Q_n(\tilde{\theta}) \sim \chi^2(\lambda(\delta))$$

with $\lambda(\delta) \equiv \delta' B_{\psi\phi\cdot\gamma} B_{\psi\cdot\gamma}^{-1} B_{\psi\phi\cdot\gamma} \delta$, which completes the proof.