

Finite Sample Inference for GMM Estimators in Linear Panel Data Models*

Stephen Bond[†]

Nuffield College, Oxford and Institute for Fiscal Studies

Frank Windmeijer[‡]

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE

January, 2002

Abstract

We compare the finite sample performance of a range of tests of linear restrictions for linear panel data models estimated using Generalised Method of Moments (GMM). These include standard asymptotic Wald tests based on one-step and two-step GMM estimators; two bootstrapped versions of these Wald tests; a version of the two-step Wald test that uses a more accurate asymptotic approximation to the distribution of the estimator; the LM test; and three criterion-based tests that have recently been proposed. We consider both the $AR(1)$ panel model, and a design with predetermined regressors. The corrected two-step Wald test performs similarly to the standard one-step Wald test, whilst the bootstrapped one-step Wald test, the LM test, and a simple criterion-difference test can provide more reliable finite sample inference in some cases.

JEL Classification: C12, C23

Keywords: Generalised Method of Moments, Hypothesis Testing, Finite Sample Inference

***Acknowledgments:** We would like to thank Seung Ahn, Manuel Arellano, Richard Blundell, Clive Bowsher and Richard Spady for helpful comments. Financial support from ESRC research grant no. R000223068 and the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at IFS is gratefully acknowledged. The usual disclaimer applies.

[†]Email: steve.bond@economics.oxford.ac.uk

[‡]Email: f.windmeijer@ifs.org.uk

1. Introduction

Standard Wald tests based on two-step efficient Generalised Method of Moments (GMM) estimators are known to have poor finite sample properties in several contexts. A leading example occurs in panel data models estimated using sequential moment conditions, where the number of moment conditions available grows rapidly with the time dimension of the panel. In this case it is well known that the usual asymptotic Wald test based on the two-step estimator can be severely oversized, even when the estimator itself has very little bias (see, for example, Blundell and Bond, 1998). The standard expression for the asymptotic variance of this estimator ignores the presence of estimated parameters in the weight matrix, and as a result the usual asymptotic standard errors can be much too small.¹ This problem is so severe that researchers commonly report hypothesis tests based on inefficient one-step GMM estimators, for which standard asymptotic Wald tests have been found to be more reliable.

A distinct problem with inference based on GMM estimators occurs when the moment conditions used provide only weak identification of the parameters, or when the instruments available are weak. In this case the estimators can be subject to large finite sample biases, which again result in Wald tests being seriously oversized, in this case for both one-step and two-step estimators. Again this problem occurs commonly in panel data applications, where the time series are often highly persistent so that differences are difficult to forecast, but differencing-type transformations are used to remove permanent individual effects.²

¹See Arellano and Bond (1991) for simulation evidence, and Koenker and Machado (1998) and Windmeijer (2000) for analyses.

²See inter alia Nelson and Startz (1990) or Staiger and Stock (1997) for analyses of weak instruments biases in general, and Blundell and Bond (1998) for a discussion in the context of

Alternative hypothesis tests with better finite sample properties are therefore of considerable interest in the panel data context. We consider two potential remedies for the standard two-step Wald test, and a range of alternative test statistics that have recently been proposed. One approach is to bootstrap the conventional Wald test. Bootstrapping for GMM is not standard, since the population moment conditions do not hold exactly in the bootstrap samples. Hall and Horowitz (1996) propose a recentering procedure to account for this, whilst Brown, Newey and May (1999) propose drawing bootstrap samples with a reweighting of the original observations. We consider these bootstrap methods for Wald tests based on both one-step and two-step GMM estimators. The second approach constructs the Wald test using a more accurate asymptotic approximation to the distribution of the two-step GMM estimator, which accounts for the presence of estimated parameters in the weight matrix. Such a correction for linear GMM estimators has recently been developed by Windmeijer (2000).

We compare these variations on the Wald test to the LM test, and to three criterion-based tests. The LM test exploits the restrictions under the null hypothesis to reduce the number of estimated parameters used to construct the weight matrix. Criterion-based tests compare minimised values of the estimator's criterion function, or equivalently tests of overidentifying restrictions, for the restricted and the unrestricted models. Hanson, Heaton and Yaron (1996) propose such a test based on the continuously updated GMM criterion; Imbens, Spady and Johnson (1998) propose such a test based on exponential tilting parameters; whilst Bond, Bowsher and Windmeijer (2001) consider such a test based on the standard two-step GMM criterion.

dynamic panel data models.

We consider the finite sample properties of these tests for simple linear restrictions in dynamic panel data models estimated using linear moment conditions. We consider both a simple $AR(1)$ model with individual effects, and a design with predetermined regressors. In all cases our simulations confirm the poor performance of the standard two-step Wald test. The standard one-step Wald test typically provides a more reliable basis for finite sample inference. The corrected version of the two-step Wald test is found to have similar finite sample properties to this one-step test. Both these Wald tests have good size properties except in cases where the GMM estimator is subject to substantial finite sample bias, or where its small sample distribution becomes distinctly non-normal. The bootstrapped one-step Wald test, the LM test and the criterion-difference test based on the two-step GMM criterion can provide more reliable finite sample inference in cases where these Wald tests have modest size distortions due to small sample biases in the GMM estimators. The LM test performs very well except when identification is weak, whilst this simple criterion-based test appears to have the best size properties in the case of weak instruments. Bootstrapped versions of the two-step Wald test can be severely undersized in these overidentified panel data models, where the optimal weight matrix can be poorly estimated.

The paper is organised as follows. Section 2 reviews GMM estimators and distribution results; Section 3 reviews the standard Wald test, and discusses a finite sample correction for the asymptotic variance and bootstrapping procedures; Section 4 considers the LM test and Section 5 considers the criterion-based tests; Section 6 discusses related literature; Section 7 outlines our simulation designs; Section 8 presents our main Monte Carlo results and Section 9 concludes.

2. GMM

Consider the moment conditions

$$E[g(X_i, \beta_0)] = E[g_i(\beta_0)] = 0,$$

where $g(\cdot)$ is a vector of order q , X_i is a vector of observed variables for $i = 1, \dots, N$, and β_0 is a parameter vector of order k . Throughout, we assume that $k < q$. The GMM estimator $\hat{\beta}$ for β_0 then minimises³

$$J(\beta) = \left[\frac{1}{N} \sum_{i=1}^N g_i(\beta) \right]' W_N \left[\frac{1}{N} \sum_{i=1}^N g_i(\beta) \right],$$

with respect to β ; where W_N is a positive semidefinite weight matrix which satisfies $\text{plim}_{N \rightarrow \infty} W_N = W$, with W a positive definite matrix. Regularity conditions are assumed such that $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N g_i(\beta) = E[g_i(\beta)]$ and $\frac{1}{\sqrt{N}} \sum_{i=1}^N g_i(\beta_0) \rightarrow N(0, \Psi)$. Let $\Gamma(\beta) = E[\partial g_i(\beta) / \partial \beta]$ and $\Gamma_0 \equiv \Gamma(\beta_0)$, then $\sqrt{N}(\hat{\beta} - \beta_0)$ has a limiting normal distribution, $\sqrt{N}(\hat{\beta} - \beta_0) \rightarrow N(0, V_W)$, where

$$V_W = (\Gamma_0' W \Gamma_0)^{-1} \Gamma_0' W \Psi W \Gamma_0 (\Gamma_0' W \Gamma_0)^{-1}. \quad (2.1)$$

An efficient two-step GMM estimator, denoted $\hat{\beta}_2$, for a given set of moment conditions, is based on a weight matrix that satisfies $\text{plim}_{N \rightarrow \infty} W_N = \Psi^{-1}$, with $V_W = (\Gamma_0' \Psi^{-1} \Gamma_0)^{-1}$. A weight matrix that satisfies this property is given by

$$W_N(\hat{\beta}_1) = \left(\frac{1}{N} \sum_{i=1}^N g_i(\hat{\beta}_1) g_i(\hat{\beta}_1)' \right)^{-1}, \quad (2.2)$$

where $\hat{\beta}_1$ is a consistent one-step GMM estimator for β_0 based on some known weight matrix W_N . Notice that $W_N(\hat{\beta}_1)$ thus depends on estimated parameters.

³See Hansen (1982).

Denote $\bar{g}(\beta) = \frac{1}{N} \sum_{i=1}^N g_i(\beta)$. The standard test for overidentifying restrictions is based on the minimised GMM criterion, given by

$$J(\hat{\beta}_2) = \bar{g}(\hat{\beta}_2)' W_N(\hat{\beta}_1) \bar{g}(\hat{\beta}_2),$$

where $\hat{\beta}_2$ is the efficient two-step GMM estimator. In particular the test statistic $NJ(\hat{\beta}_2)$ has an asymptotic chi-squared distribution with $q - k$ degrees of freedom when the moment conditions are valid. We refer to this as the Sargan test in the sequel (see Sargan (1958, 1988)).

We will consider GMM estimators in panel data models where all the moment conditions used are linear in the parameters. Defining

$$C = \frac{\partial \bar{g}(\beta)}{\partial \beta'},$$

one-step and two-step linear GMM estimators then satisfy

$$\begin{aligned} \hat{\beta}_1 - \beta_0 &= - (C' W_N C)^{-1} C' W_N \bar{g}(\beta_0) \\ \hat{\beta}_2 - \beta_0 &= - (C' W_N(\hat{\beta}_1) C)^{-1} C' W_N(\hat{\beta}_1) \bar{g}(\beta_0), \end{aligned}$$

and the asymptotic variances are consistently estimated by

$$\begin{aligned} \widehat{\text{var}}(\hat{\beta}_1) &= \frac{1}{N} (C' W_N C)^{-1} C' W_N W_N^{-1}(\hat{\beta}_1) W_N C (C' W_N C)^{-1} \\ \widehat{\text{var}}(\hat{\beta}_2) &= \frac{1}{N} (C' W_N(\hat{\beta}_1) C)^{-1}. \end{aligned} \tag{2.3}$$

3. Wald Test

The standard Wald test for testing r linear restrictions of the form $r(\beta_0) = 0$ is calculated as

$$\text{Wald} = r(\hat{\beta})' (R' \widehat{\text{var}}(\hat{\beta}) R)^{-1} r(\hat{\beta}),$$

where $R = \partial r(\beta) / \partial \beta'$, and has an asymptotic χ_r^2 distribution under the null. Based on the two-step GMM estimator, $\hat{\beta} = \hat{\beta}_2$, and using its conventional asymptotic variance estimate, $\widehat{\text{var}}(\hat{\beta}_2)$, the Wald test has often been found to overreject correct null hypotheses severely compared to its nominal size. This can occur even when the estimator has negligible finite sample bias, due to the fact that the estimated asymptotic standard errors can be severely downward biased in small samples (see for example Arellano and Bond (1991) in the case of linear panel data models). In this section we discuss two approaches to remedy this problem. The first is to use a finite sample correction to the estimated asymptotic variance of the two-step GMM estimator, as proposed by Windmeijer (2000). The second is to use bootstrap methods as developed for GMM estimation by Hall and Horowitz (1996) and Brown, Newey and May (1999).

3.1. A Finite Sample Correction for the Asymptotic Variance

Windmeijer (2000) proposed a finite sample correction to the estimated asymptotic variance of the two-step GMM estimator that accounts for the extra variation of the estimator due to the presence of estimated parameters in the efficient weight matrix, $W_N(\hat{\beta}_1)$. This variation is neglected in the conventional asymptotic variance expression (2.3), but has a non-negligible effect on the distribution of the estimator in sample sizes that are typically encountered in practice. Accounting for this variation yields an improvement in the order of the asymptotic approximation in the case of linear GMM estimators.

In particular, the linear two-step GMM estimator satisfies

$$\hat{\beta}_2 - \beta_0 = - \left(C' W_N(\hat{\beta}_1) C \right)^{-1} C' W_N(\hat{\beta}_1) \bar{g}(\beta_0),$$

which depends on the one-step estimator $\widehat{\beta}_1$ via the weight matrix $W_N(\widehat{\beta}_1)$. A Taylor series expansion of $\widehat{\beta}_2 - \beta_0$ around $\widehat{\beta}_1 = \beta_0$ results in

$$\begin{aligned}\widehat{\beta}_2 - \beta_0 &= -(C'W_N(\beta_0)C)^{-1}C'W_N(\beta_0)\bar{g}(\beta_0) \\ &\quad - G_{\beta_0, W_N(\beta_0)}(\widehat{\beta}_1 - \beta_0) + O_p(N^{-\frac{3}{2}}),\end{aligned}\tag{3.1}$$

where

$$W_N(\beta_0) = \left(\frac{1}{N} \sum_{i=1}^N g_i(\beta_0) g_i(\beta_0)' \right)^{-1}$$

and $G_{\beta_0, W_N(\beta_0)}$ is a $k \times k$ matrix with its s -th column given by

$$\begin{aligned}G_{\beta_0, W_N(\beta_0)}[\cdot, s] &= \frac{\partial}{\partial \beta_s} \left[(C'W_N(\beta)C)^{-1}C'W_N(\beta)\bar{g}(\beta_0) \right] \Big|_{\beta_0} \\ &= A_{1s}(\beta_0)p(\beta_0) - A_{2s}(\beta_0),\end{aligned}$$

where

$$\begin{aligned}A_{1s}(\beta_0) &= (C'W_N(\beta_0)C)^{-1}C'W_N(\beta_0) \frac{\partial W_N^{-1}(\beta)}{\partial \beta_s} \Big|_{\beta_0} W_N(\beta_0)C \\ p(\beta_0) &= (C'W_N(\beta_0)C)^{-1}C'W_N(\beta_0)\bar{g}(\beta_0) \\ A_{2s}(\beta_0) &= (C'W_N(\beta_0)C)^{-1}C'W_N(\beta_0) \frac{\partial W_N^{-1}(\beta)}{\partial \beta_s} \Big|_{\beta_0} W_N(\beta_0)\bar{g}(\beta_0),\end{aligned}$$

and

$$\frac{\partial W_N^{-1}(\beta)}{\partial \beta_s} = \frac{1}{N} \sum_{i=1}^N \left(\frac{\partial g_i(\beta)}{\partial \beta_s} g_i(\beta)' + g_i(\beta) \frac{\partial g_i(\beta)'}{\partial \beta_s} \right).$$

The term $G_{\beta_0, W_N(\beta_0)}(\widehat{\beta}_1 - \beta_0)$ is of order $O_p(N^{-1})$, and incorporating it in the estimate of the variance of $\widehat{\beta}_2$ will improve the accuracy of the approximation in finite samples. The first term of $G_{\beta_0, W_N(\beta_0)}[\cdot, s]$, $A_{1s}(\beta_0)p(\beta_0)$, is a function of $p(\beta_0)$ which is the bias of an infeasible GMM estimator that uses an efficient weight matrix that is based on the true parameters β_0 . This bias tends to be small and will generally not grow with the number of moment conditions, see Newey and

Smith (2000). The second term, $A_{2s}(\beta_0)$, will in general increase with the number of moment conditions and therefore dominate the bias correction. $G_{\beta_0, W_N(\beta_0)}[\cdot, s]$ can thus be estimated by $G_{\hat{\beta}_2, W_N(\hat{\beta}_1)}[\cdot, s]$ which is given by

$$G_{\hat{\beta}_2, W_N(\hat{\beta}_1)}[\cdot, s] = - \left(C' W_N(\hat{\beta}_1) C \right)^{-1} C' W_N(\hat{\beta}_1) \frac{\partial W_N^{-1}(\beta)}{\partial \beta_s} \Big|_{\hat{\beta}_1} W_N(\hat{\beta}_1) \bar{g}(\hat{\beta}_2).$$

A feasible small sample bias corrected estimate of the variance of $\hat{\beta}_2$ is then obtained as

$$\begin{aligned} \widehat{\text{var}}_c(\hat{\beta}_2) &= \frac{1}{N} \left(C' W_N(\hat{\beta}_1) C \right)^{-1} + G_{\hat{\beta}_2, W_N(\hat{\beta}_1)} \widehat{\text{var}}(\hat{\beta}_1) G'_{\hat{\beta}_2, W_N(\hat{\beta}_1)} \\ &\quad - \frac{1}{N} \left(G_{\hat{\beta}_2, W_N(\hat{\beta}_1)} \left(C' W_N(\hat{\beta}_1) C \right)^{-1} + \left(C' W_N(\hat{\beta}_1) C \right)^{-1} G'_{\hat{\beta}_2, W_N(\hat{\beta}_1)} \right). \end{aligned} \quad (3.2)$$

This is used in place of the conventional $\widehat{\text{var}}(\hat{\beta}_2)$ to construct the corrected two-step Wald statistic.

3.2. Bootstrap

As bootstrap methods for GMM are not standard when the model is overidentified, $q > k$, Hall and Horowitz (1996), and Brown, Newey and May (1999) have developed procedures to bootstrap the distribution of test statistics, like the Wald test, in the GMM context. These bootstrap methods improve finite sample inference by reducing the error in the rejection probability of the test statistic.⁴ In principle bootstrapping can reduce size distortions in the asymptotic Wald tests that are due to finite sample biases in the GMM estimators, as well as those that are due to finite sample bias in the conventional asymptotic variance matrix.

The Hall and Horowitz (1996) procedure samples the observations (X_i) with replacement from the original sample. As the model is overidentified, the population

⁴For a comprehensive review of the bootstrap method, see Horowitz (2001).

moment conditions are not exactly matched by the sample moment conditions, and the bootstrap will not provide asymptotic refinements for the rejection probabilities of the Wald test. By recentering the moment conditions in the bootstrap sample, this problem is overcome. Let E^b denote the expectation relative to the distribution of the bootstrap sample conditional on the original sample. Then

$$E^b \left(g_{i^b}(\beta_0) - \bar{g}(\hat{\beta}) \right) = 0,$$

where i^b denotes the index in the bootstrap sample and, as before, $\bar{g}(\hat{\beta}) = \frac{1}{N} \sum_{i=1}^N g_i(\hat{\beta})$. The GMM estimator in the bootstrap sample minimises

$$\frac{1}{N} \sum_{i^b=1}^N \left(g_{i^b}(\beta) - \bar{g}(\hat{\beta}) \right) W_N \frac{1}{N} \sum_{i^b=1}^N \left(g_{i^b}(\beta) - \bar{g}(\hat{\beta}) \right),$$

with $\hat{\beta} = \hat{\beta}_1$ for the one-step GMM estimator in the bootstrap sample, and $\hat{\beta} = \hat{\beta}_2$ for the two-step estimator, and the weight matrix specified accordingly.

Brown, Newey and May (1999) propose instead to resample the data with weighting to ensure that the moment conditions hold in the weighted sample. These weights are obtained from an empirical likelihood procedure. Consider the criterion function that corresponds to the empirical likelihood estimator developed in Owen (1988). The probability weights π_i solve

$$\min_{\pi} \sum_{i=1}^N \frac{1}{N} \ln \pi_i \quad \text{subject to} \quad \sum_{i=1}^N g_i(\hat{\beta}_2) \pi_i = 0 \quad \text{and} \quad \sum_{i=1}^N \pi_i = 1,$$

and are estimated as

$$\hat{\pi}_i = \frac{1}{N} \left(1 + \hat{\gamma}' g_i(\hat{\beta}_2) \right)^{-1}$$

where $\hat{\gamma}$ are the estimated tilting parameters. Intuitively these measure how much the sample has to be reweighted in order for the moment conditions to hold

exactly, and are estimated as

$$\hat{\gamma} = \underset{\gamma}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^N \ln \left(1 + \gamma' g_i \left(\hat{\beta}_2 \right) \right).$$

The bootstrap sample is then drawn with replacement from the original sample, giving the individual observations (X_i) the probability $\hat{\pi}_i$ of being drawn.

Using the recentering method of Hall and Horowitz (1996) or the reweighting method of Brown, Newey and May (1999), we obtain bootstrap p-values for the Wald test. For example for the simple test $H_0 : \beta_j = \delta$, the standard two-step t-statistic is given by $t_j = \left(\hat{\beta}_{2j} - \delta \right) / \sqrt{\widehat{\operatorname{var}} \left(\hat{\beta}_{2j} \right)}$. For every bootstrap sample we calculate, $t_j^b = \left(\hat{\beta}_{2j}^b - \hat{\beta}_{2j} \right) / \sqrt{\widehat{\operatorname{var}} \left(\hat{\beta}_{2j}^b \right)}$, where $\hat{\beta}_{2j}^b$ is the two-step GMM estimator in the b -th bootstrap sample. The percentile in the bootstrap distribution of $|t_j|$ is then given by $p_{t_j} = \frac{1}{N_b} \sum_{b=1}^{N_b} 1 \left(|t_j^b| > |t_j| \right)$, where the indicator function $1 \left(|t_j^b| > |t_j| \right)$ equals 1 if $|t_j^b| > |t_j|$, 0 otherwise, and N_b is the number of bootstrap replications. The test rejects the null at size α if $p_{t_j} \leq \alpha$.

4. LM Test

The LM test is based on the two-step GMM estimator in the model which imposes the restrictions $r(\beta_0) = 0$, and can be calculated as

$$\text{LM} = N \bar{g} \left(\tilde{\beta}_2 \right)' W_N \left(\tilde{\beta}_1 \right) C \left(C' W_N \left(\tilde{\beta}_1 \right) C \right)^{-1} C' W_N \left(\tilde{\beta}_1 \right) \bar{g} \left(\tilde{\beta}_2 \right),$$

where $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are the one-step and two-step GMM estimators in the restricted model. The LM test has an asymptotic χ_r^2 distribution under the null.

It is further equivalent to a Wald test in the unrestricted model, constructed using a two-step GMM estimator where the restrictions imposed by the null hypothesis are exploited to reduce the dimension of the parameter vector that is

estimated to construct the weight matrix. Specifically

$$\text{LM} = r(\hat{\beta}_2)' \left(R' \widehat{\text{var}}(\hat{\beta}_2) R \right)^{-1} r(\hat{\beta}_2),$$

where $\hat{\beta}_2$ is a two-step GMM estimator in the unrestricted model, obtained using the estimate of Ψ^{-1} in the restricted model, $W_N(\tilde{\beta}_1)$, as the weight matrix. Given this relationship, one might expect the LM test to behave like the standard Wald test, especially with an increasing number of unrestricted parameters. However, in our Monte Carlo simulations, the LM test does not suffer from the poor finite sample properties found for the standard two-step Wald test. When considering simple tests of the form $H_0 : \beta_j = \delta$, we find that the small sample bias of the asymptotic standard error of $\hat{\beta}_{2j}$ is very small in a wide variety of settings, although the standard errors of the other parameters that were unrestricted under the null and thus estimated in the weight matrix $W_N(\tilde{\beta}_1)$ did have the usual small sample downward bias. This happens when $G_{\beta_0, W_N(\beta_0)}$, as defined in section 3.1, is (approximately) a diagonal matrix, as then the small sample bias of the standard error of the parameter to be tested will automatically be corrected by imposing the fixed value of this parameter when estimating the optimal weight matrix based on the restricted model.

5. Criterion-Based Tests

The standard two-step Wald test can equivalently be computed as a criterion difference (see Newey and West (1987))

$$\text{Wald} = N \left(\bar{g}(\tilde{\beta}_2)' W_N(\hat{\beta}_1) \bar{g}(\tilde{\beta}_2) - \bar{g}(\hat{\beta}_2)' W_N(\hat{\beta}_1) \bar{g}(\hat{\beta}_2) \right),$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the one-step and two-step GMM estimators in the unrestricted model, whereas $\tilde{\beta}_2$ is a two-step GMM estimator in the restricted model, but using

as a weight matrix the consistent estimate of Ψ^{-1} based on the unrestricted one-step GMM estimator, $W_N(\hat{\beta}_1)$.

The LM test can also be computed as a criterion difference

$$\text{LM} = N \left(\bar{g}(\tilde{\beta}_2)' W_N(\tilde{\beta}_1) \bar{g}(\tilde{\beta}_2) - \bar{g}(\hat{\beta}_2)' W_N(\tilde{\beta}_1) \bar{g}(\hat{\beta}_2) \right),$$

where $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are the one-step and two-step GMM estimators in the restricted model, whereas $\hat{\beta}_2$ is, as before, a two-step GMM estimator in the unrestricted model, but using as a weight matrix the consistent estimate of Ψ^{-1} , under the null, based on the restricted one-step GMM estimator, $W_N(\tilde{\beta}_1)$.

The criterion-based test statistic considered by Bond, Bowsher and Windmeijer (2001) is given by

$$\begin{aligned} D_{RU} &= N \left(\bar{g}(\tilde{\beta}_2)' W_N(\tilde{\beta}_1) \bar{g}(\tilde{\beta}_2) - \bar{g}(\hat{\beta}_2)' W_N(\tilde{\beta}_1) \bar{g}(\hat{\beta}_2) \right) \\ &= N \left(J(\tilde{\beta}_2) - J(\hat{\beta}_2) \right). \end{aligned}$$

Notice that $W_N(\tilde{\beta}_1)$ and $W_N(\hat{\beta}_1)$ are both consistent estimates of Ψ^{-1} under the null. D_{RU} is the “likelihood ratio” test equivalent for GMM, see e.g. Davidson and MacKinnon (1993, pp. 614-620), and is simply computed as the difference between the Sargan statistics in the restricted and unrestricted models.⁵ Intuitively, if the restrictions are valid, then imposing them should not disturb the validity of the overidentifying restrictions. Under the null, D_{RU} has an asymptotic χ_r^2 distribution.⁶

Hansen, Heaton and Yaron (1996) proposed the use of a criterion-based test similar to D_{RU} , but based on their continuously updated GMM estimator. This

⁵Everywhere, the instrument matrix Z is kept the same in restricted and unrestricted models.

⁶The D_{RU} statistic has recently been considered by Hansen (2000) in the context of testing non-linear restrictions in the classical linear regression model. He shows via Edgeworth expansions that the asymptotic chi-squared distribution provides a better approximation for the D_{RU} statistic than for the Wald statistic in that case.

estimator is equivalent to robust LIML and is defined as the value of β , denoted $\hat{\beta}^{CU}$, that minimizes

$$J^{CU}(\beta) = \bar{g}(\beta)' \left(\frac{1}{N} \sum_{i=1}^N g_i(\beta) g_i(\beta)' \right)^{-1} \bar{g}(\beta).$$

The test statistic D_{RU}^{CU} is then defined as

$$D_{RU}^{CU} = N \left(J^{CU}(\tilde{\beta}^{CU}) - J^{CU}(\hat{\beta}^{CU}) \right),$$

where $\hat{\beta}^{CU}$ and $\tilde{\beta}^{CU}$ are the continuously updated GMM estimators for the unrestricted model and the restricted model respectively. Under the null, D_{RU}^{CU} has an asymptotic χ_r^2 distribution.

The test proposed by Imbens, Spady and Johnson (1998) is based on the empirical likelihood method. The exponential tilting estimator for β_0 minimises the Kullback-Leibler information criterion

$$\min_{\pi, \beta} \sum_{i=1}^N \pi_i \ln \pi_i \quad \text{subject to} \quad \sum_{i=1}^N g_i(\beta) \pi_i = 0 \quad \text{and} \quad \sum_{i=1}^N \pi_i = 1.$$

The estimated probabilities are of the form

$$\pi_i(\zeta, \beta) = \frac{\exp(\zeta' g_i(\beta))}{\sum_{j=1}^N \exp(\zeta' g_j(\beta))}$$

where ζ are the tilting parameters. Again, these measure how much the sample has to be reweighted in order for the moment conditions to hold exactly.⁷

Tilting parameters can be estimated conditional on the standard GMM estimator of the parameters β . The Lagrange Multiplier test statistic

$$NJ^{ET}(\hat{\zeta}, \hat{\beta}_2) = N \hat{\zeta}' R_N(\hat{\zeta}, \hat{\beta}_2) \hat{\zeta}$$

⁷Note that these estimated probabilities can also be used to reweight the data in the bootstrap procedure of Brown, Newey and May (1999).

is a test for whether the tilting parameters, or Lagrange multipliers, are equal to zero, where

$$\hat{\zeta} = \max_{\zeta} \ln \frac{1}{N} \sum_{i=1}^N \exp(\zeta' g_i(\hat{\beta}_2)),$$

and

$$\begin{aligned} R_N(\hat{\zeta}, \hat{\beta}_2) &= \left[\sum_{i=1}^N g_i(\hat{\beta}_2) g_i(\hat{\beta}_2)' \pi_i(\hat{\zeta}, \hat{\beta}_2) \right] \left[N \sum_{i=1}^N g_i(\hat{\beta}_2) g_i(\hat{\beta}_2)' \pi_i^2(\hat{\zeta}, \hat{\beta}_2) \right]^{-1} \\ &\quad \times \left[\sum_{i=1}^N g_i(\hat{\beta}_2) g_i(\hat{\beta}_2)' \pi_i(\hat{\zeta}, \hat{\beta}_2) \right] \end{aligned}$$

is a robust estimate of the variance of the moments. $NJ^{ET}(\hat{\zeta}, \hat{\beta}_2)$ has an asymptotic χ_{q-k}^2 distribution when the moment conditions are valid. The corresponding criterion-based test for testing the hypothesis $H_0 : r(\beta_0) = 0$ is defined as

$$\begin{aligned} D_{RU}^{ET} &= N \left(\tilde{\zeta}' R_N(\tilde{\zeta}, \tilde{\beta}_2) \tilde{\zeta} - \hat{\zeta}' R_N(\hat{\zeta}, \hat{\beta}_2) \hat{\zeta} \right) \\ &= N \left(J^{ET}(\tilde{\zeta}, \tilde{\beta}_2) - J^{ET}(\hat{\zeta}, \hat{\beta}_2) \right), \end{aligned}$$

where $\tilde{\zeta}$ is the estimator of the tilting parameters based on the efficient two-step GMM estimator in the restricted model. Under the null, D_{RU}^{ET} has an asymptotic χ_r^2 distribution.

6. Related Literature

The finite sample behaviour of some of the test procedures described in the preceding sections has been analysed in different settings by various authors. Windmeijer (2000) presents Monte Carlo results for a simple linear panel data model indicating that the finite sample correction for the variance of the linear two-step GMM estimator works well in that context, leading to much more accurate inference than the standard two-step Wald test. Hall and Horowitz (1996) do not

consider panel data models, but show the improvement of their bootstrap procedure as compared to the usual asymptotic inference for various non-linear GMM estimators. Brown, Newey and May (1999) do consider panel data models and also show that their bootstrap method improves the finite sample performance of the Wald test based on the two-step GMM estimator. Bergström (1997), and Bergström, Dahlberg and Johansson (1997) analyse the performance of the two bootstrap procedures for first-differenced GMM estimators in linear autoregressive panel data models, and show that both methods improve on the standard asymptotic inference for two-step GMM. These papers do not consider the case of models with weak instruments, or additional moment conditions that are available for equations in levels under initial conditions restrictions.

Hansen, Heaton and Yaron (1996) consider the performance of the criterion-based test based on the continuously updated GMM estimator, D_{RU}^{CU} , in the context of non-linear Euler equations and show that this test has superior finite sample properties as compared to the standard Wald tests based on two-step or continuously updated GMM estimators. Imbens, Spady and Johnson (1998) focus primarily on their test of overidentifying restrictions, NJ^{ET} , and only briefly consider the performance of D_{RU}^{ET} for some specific non-linear models. Again, D_{RU}^{ET} outperforms the standard two-step Wald test. Bond, Bowsher and Windmeijer (2001) study the performance of the three criterion-based tests for linear autoregressive panel data models and conclude that these tests have quite good finite sample properties in this context and all three behave very similarly in the models they consider. The D_{RU} test then has the advantage of being extremely simple to compute. As far as we are aware, the performance of the LM test has not been

studied in the GMM context.

In the following sections, we present a systematic comparison of the finite sample performance of these various test procedures for a range of dynamic panel data models.

7. Dynamic Processes with Individual Effects

We first consider the linear first order autoregressive panel data model with individual effects (η_i)

$$y_{it} = \alpha_0 y_{it-1} + u_{it}$$

$$u_{it} = \eta_i + v_{it}$$

where $i = 1, \dots, N$ and $t = 2, \dots, T$, with N large and T fixed. Under the standard assumptions

$$E(\eta_i) = 0, E(v_{it}) = 0, E(v_{it}\eta_i) = 0 \text{ for } i = 1, \dots, N \text{ and } t = 2, \dots, T \quad (7.1)$$

$$E(v_{it}v_{is}) = 0 \text{ for } i = 1, \dots, N \text{ and } \forall t \neq s \quad (7.2)$$

and

$$E(y_{i1}v_{it}) = 0 \text{ for } i = 1, \dots, N \text{ and } t = 2, \dots, T, \quad (7.3)$$

the following $(T-1)(T-2)/2$ linear moment conditions are valid

$$E\left[y_i^{t-2} (\Delta y_{it} - \alpha_0 \Delta y_{it-1})\right] = 0 \text{ for } t = 3, \dots, T, \quad (7.4)$$

where $y_i^{t-2} = [y_{i1}, y_{i2}, \dots, y_{it-2}]$; see for example Ahn and Schmidt (1995). We call these moment conditions the DIF moment conditions, cf. Arellano and Bond (1991). Under the additional mean stationarity assumption on initial conditions

$$E(\eta_i \Delta y_{i2}) = 0, \quad (7.5)$$

the additional $(T - 2)$ linear moment conditions

$$E[\Delta y_{it-1} (y_{it} - \alpha_0 y_{it-1})] = 0 \text{ for } t = 3, \dots, T \quad (7.6)$$

are valid. The joint moment conditions (7.4) and (7.6) are the so-called SYS moment conditions, cf. Arellano and Bover (1995) and Blundell and Bond (1998).

Let Z_i be the matrix of instruments for observation i , then the moment conditions can generically be written as $E[Z_i' \omega_i(\alpha_0)] = 0$. The efficient one-step GMM weight matrix for the DIF moment conditions when the v_{it} are homoscedastic and not serially correlated is given by $W_N = \left(\frac{1}{N} \sum_{i=1}^N Z_i' H Z_i\right)^{-1}$, where H is a $(T - 2)$ square matrix which has 2's on the main diagonal, -1's on the first subdiagonals and zeros elsewhere. For the SYS moment conditions there is no simple one-step efficient weight matrix, and in our simulations the one-step weight matrix is set equal to $W_N = \left(\frac{1}{N} \sum_{i=1}^N Z_i' Z_i\right)^{-1}$. The efficient two-step weight matrix for both estimators under general conditions is given by $W_N(\hat{\alpha}_1) = \left(\frac{1}{N} \sum_{i=1}^N Z_i' \omega_i(\hat{\alpha}_1) \omega_i(\hat{\alpha}_1)' Z_i\right)^{-1}$, with $\hat{\alpha}_1$ the consistent one-step GMM DIF or SYS estimator of α .

We also present simulation results for a multivariate dynamic panel data model specified as

$$\begin{aligned} y_{it} &= \alpha_0 y_{it-1} + \beta_0 x_{it} + \eta_i + v_{it} \\ x_{it} &= \rho_0 x_{it-1} + \tau_0 \eta_i + w_{it} \\ w_{it} &= \theta_0 v_{it-1} + \varepsilon_{it}. \end{aligned}$$

The x_{it} series are correlated with η_i and predetermined but not strictly exogenous with respect to v_{it} . The DIF moment conditions in this model are the

$((T - 1)(T - 2)/2) + (T(T - 1)/2)$ moment conditions

$$\begin{aligned} E \left[y_i^{t-2} (\Delta y_{it} - \alpha_0 \Delta y_{it-1} - \beta_0 \Delta x_{it}) \right] &= 0 \text{ for } t = 3, \dots, T \\ E \left[x_i^{t-1} (\Delta y_{it} - \alpha_0 \Delta y_{it-1} - \beta_0 \Delta x_{it}) \right] &= 0 \text{ for } t = 3, \dots, T, \end{aligned}$$

where $x_i^{t-1} = [x_{i1}, x_{i2}, \dots, x_{it-1}]$. The additional $(T - 2) + (T - 1)$ moment conditions used by the SYS estimator are given by

$$\begin{aligned} E [\Delta y_{it-1} (y_{it} - \alpha_0 y_{it-1} - \beta_0 x_{it})] &= 0 \text{ for } t = 3, \dots, T \\ E [\Delta x_{it} (y_{it} - \alpha_0 y_{it-1} - \beta_0 x_{it})] &= 0 \text{ for } t = 2, \dots, T. \end{aligned}$$

8. Monte Carlo Results

8.1. Basic results

In Table 1 and Figure 1 we present some basic Monte Carlo results for the $AR(1)$ panel data process. The covariance stationary data generating process is

$$\begin{aligned} y_{it} &= \alpha_0 y_{it-1} + \eta_i + v_{it} \\ \eta_i &\sim N(0, 1) \quad ; \quad v_{it} \sim N(0, 1) \\ y_{i1} &= \frac{\eta_i}{1 - \alpha_0} + \varepsilon_i \quad ; \quad \varepsilon_i \sim N\left(0, \frac{1}{1 - \alpha_0^2}\right). \end{aligned}$$

The additional moment conditions (7.6) are satisfied in this design, and we consider inference based on both the DIF moment conditions alone, and on the combined SYS moment conditions.⁸ The initial conditions and error components are all normally distributed. The sample size is $N = 100$, $T = 6$, and we report results based on 10,000 replications. The value considered for α_0 is 0.4, which

⁸Additional linear moment conditions that depend on both homoscedasticity and covariance stationarity are also available in the design used here; see, for example, Ahn and Schmidt (1995) and Kruiniger (2000). We do not consider inference based on these estimators, as they are less commonly used in applied work.

is well below the range where problems associated with weak identification are expected to be serious. We consider tests of the null hypothesis $H_0 : \alpha_0 = 0.4$.

Table 1 presents the estimation results for both one-step and two-step versions of the SYS and DIF estimators, and Figure 1 presents p-value plots to compare the actual and nominal size properties of the various test statistics; see Davidson and MacKinnon (1996). Figure 1a shows the results for the various Wald statistics based on the SYS estimator: the standard asymptotic Wald test based on the one-step GMM results, W_1 ; the standard asymptotic Wald test based on the two-step GMM results, W_2 ; the Wald test based on the two-step estimator using the finite sample corrected estimate of the variance matrix, W_{2C} ; the bootstrapped one-step Wald test using the Hall-Horowitz method, W_{1HH} ; and the bootstrapped two-step Wald tests, W_{2HH} and W_{2BNM} , using the Hall-Horowitz and Brown-Newey-May methods respectively. The number of bootstrap replications, N_b , is equal to 500. Figure 1b shows the corresponding results for the LM test, LM , and for the three criterion-difference tests, D_{RU} , D_{RU}^{CU} and D_{RU}^{ET} , based on the standard two-step GMM criterion, the continuously updated GMM criterion and the exponential tilting criterion respectively. Results for the standard one-step Wald test, W_1 , are reproduced here to facilitate comparisons. Figures 1c and 1d then show the p-value plots for the same sets of test statistics based on the results for the DIF estimator.

Table 1. Estimation Results for $AR(1)$ Model, $N = 100$, $T = 6$

	SYS				DIF			
	one-step		two-step		one-step		two-step	
	coeff	st dev	coeff	st dev	coeff	st dev	coeff	st dev
	<i>rmse</i>	st err	<i>rmse</i>	st err	<i>rmse</i>	st err	<i>rmse</i>	st err
α_0				st errc				st errc
0.4	0.389	0.096	0.403	0.081	0.364	0.108	0.366	0.118
	<i>0.097</i>	0.094	<i>0.081</i>	0.058	<i>0.114</i>	0.106	<i>0.122</i>	0.096
				0.079				0.116

Results are based on 10,000 replications

st err is the mean of the estimated asymptotic standard errors

st errc is the mean of the estimated standard errors corrected for small sample bias

Table 1 shows that with $N = 100$, $T = 6$ and $\alpha_0 = 0.4$ there is almost no bias found for the two-step GMM estimator using the SYS moment conditions. There is a small downward bias, around 3% of the true value of α_0 , found for the one-step version, which also has a standard deviation around 19% higher than the asymptotically efficient estimator based on the same moment conditions. However whilst the finite sample standard deviation of the one-step estimator is well estimated by the mean of its usual asymptotic standard errors, this is not the case for the two-step estimator, where the mean of the uncorrected asymptotic standard errors has a downward bias of around 28%. The finite sample correction discussed in Section 3.1 virtually eliminates this bias. The pattern of results for the GMM estimators using only the DIF moment conditions is broadly similar. These estimators have a downward bias around 9% of the true value of α_0 , and a standard deviation around 40% higher than the two-step SYS estimator. Again the corrected standard error provides a much more accurate guide to the finite sample standard deviation of the two-step version than does the usual asymptotic standard error, although the downward bias in the latter, around 19%, is smaller

here than for the SYS estimator, which is based on more moment conditions. One further feature of these results is that the one-step GMM DIF estimator, which uses an optimal weight matrix in this simulation design, has a finite sample standard deviation which is around 8% smaller than the asymptotically equivalent two-step version.

Figure 1a confirms that inference based on the standard two-step Wald test could be very misleading in this context. At a nominal size of 5%, for example, this test falsely rejects the correct null hypothesis that $\alpha_0 = 0.4$ in around 17% of the replications. In contrast, despite the greater bias of the one-step GMM SYS estimator, inference based on the standard one-step Wald test is much more reliable, with a rejection frequency of around 6% at the 5% nominal size. This pattern is typical of a wide range of experiments that we have considered, a selection of which are reported below. The poor performance of the conventional two-step Wald test here is principally due to the downward bias in the usual asymptotic standard errors, which do not account for the estimation of parameters used to construct the optimal two-step weight matrix. We have seen in Table 1 that there is almost no bias in the mean of the two-step estimates in this experiment, and Figure 4a below confirms that the finite sample distribution of this estimator here is symmetric and approximately normal. Consequently we would expect that using the corrected estimates of the standard errors to construct a Wald statistic based on the two-step GMM estimator should produce a considerable improvement in the size properties of the test, and this is indeed what we find in the results of this experiment. The corrected two-step Wald statistic (W_{2C}) has rejection frequencies that are very close to those of the standard

one-step Wald statistic (W_1). Again this pattern is typical of a wide range of experiments in which we found the corrected two-step Wald test to have size properties similar to, or marginally better than, the standard one-step Wald test.

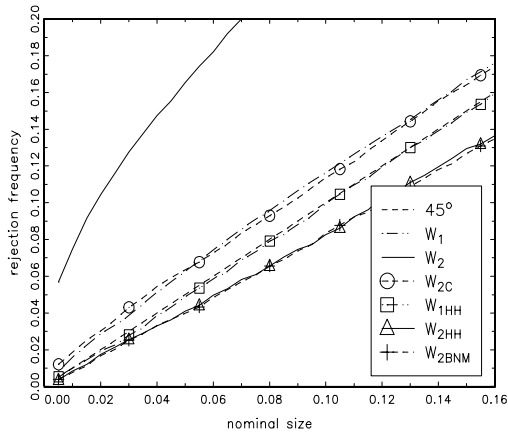


Fig 1a. P-value plot, $\alpha_0 = 0.4$, SYS, $T = 6$

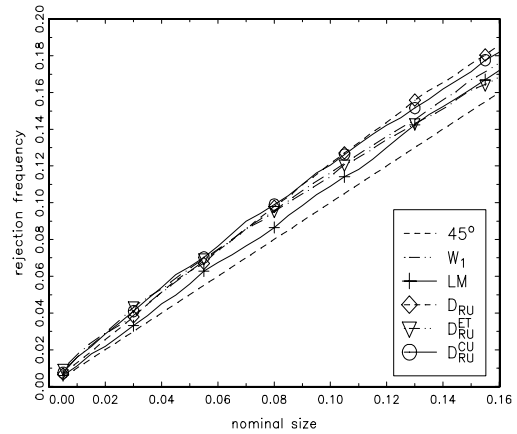


Fig 1b. P-value plot, $\alpha_0 = 0.4$, SYS, $T = 6$

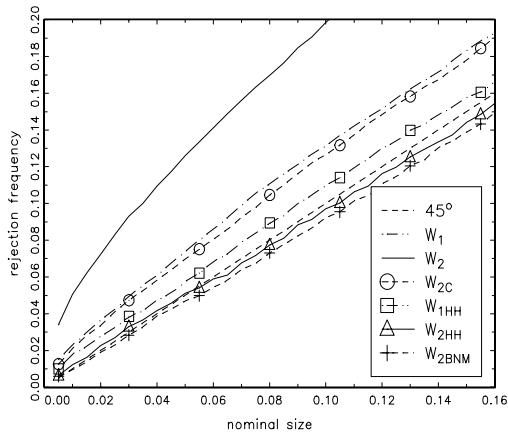


Fig 1c. P-value plot, $\alpha_0 = 0.4$, DIF, $T = 6$

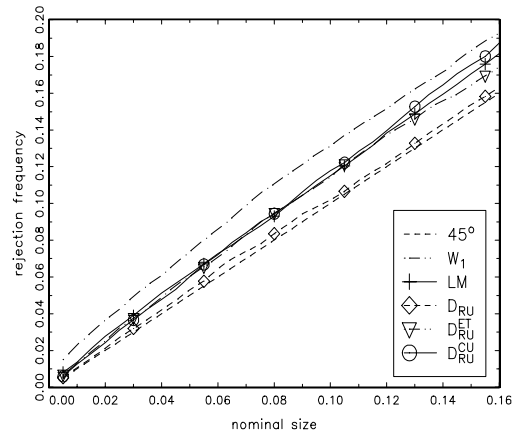


Fig 1d. P-value plot, $\alpha_0 = 0.4$, DIF, $T = 6$

Turning to the bootstrapped versions of the Wald tests, Figure 1a shows that the Hall-Horowitz procedure (W_{1HH}) does extremely well in correcting the small size distortion found for the standard one-step Wald test in this experiment. Neither the Hall-Horowitz nor the Brown-Newey-May procedures are so successful for the two-step Wald test, with both the bootstrapped two-step Wald tests tending to reject the null hypothesis too infrequently relative to their nominal size. This pattern was again common in the experiments we considered, and some cases in which the poor performance of the bootstrapped two-step Wald tests becomes more extreme will be illustrated in later sections.

Figure 1b shows that the LM test also performs well in terms of size in this experiment. We found this to be the case in all experiments where the parameter of interest is well identified; the exception being the ‘weak instruments’ case considered in Section 8.1.2 below. The three criterion-based tests are also found to have quite good size properties in this experiment, with rejection frequencies of around 6% at the 5% nominal size.

The results discussed above, based on the system GMM estimator which has little bias in this design, have tended to be relatively favourable to the standard one-step Wald test. We might expect the bootstrapped, LM or criterion-based tests to offer more improvement in settings where there is more bias present in the underlying GMM estimator. Figures 1c and 1d illustrate this by considering the same test statistics based on the first-differenced GMM estimator in the same experiment. As we saw in Table 1, this estimator does have a larger finite sample bias in this context. Again the standard two-step Wald test has a much greater size distortion than the standard one-step Wald test. In this case, though, we find

that there is a small improvement offered by the corrected two-step Wald test, and a considerable improvement offered by either the bootstrapped one-step Wald test, the LM test, or the three criterion-based tests, compared to the size performance of the standard one-step Wald test. In this experiment the bootstrapped two-step Wald tests also have rejection frequencies that are very close to the nominal size, although we would caution that this result was not typical of our findings for the bootstrapped two-step Wald tests.

8.1.1. Variations: $T = 8$ and non-normal errors

Table 2 and Figure 2 present the findings for two further experiments where interesting results were obtained. The first of these considers the effect of increasing the time dimension of the panel from $T = 6$ to $T = 8$, keeping $N = 100$ and $\alpha_0 = 0.4$. Noting that the number of moment conditions increases at the rate T^2 , this approximately doubles the number of moment conditions used by the SYS estimator from 14 to 27, whilst holding constant the cross-section dimension of the panel on which the large N , fixed T asymptotic approximations depend. The second variation maintains the $N = 100, T = 6, \alpha_0 = 0.4$ combination used in the previous section, but introduces non-normal disturbances into the $AR(1)$ process. In particular we here draw the time-varying errors (v_{it}) from a skewed χ_1^2 distribution, centred to have mean zero and scaled to have variance one. That is, we replace the standard normal random variable by

$$v_{it} \sim (\chi_1^2 - 1) / \sqrt{2}.$$

Otherwise in both cases the design is identical to that described previously.

To conserve space we focus here on the results for the GMM estimators based

on the combined set of SYS moment conditions. Comparing Tables 1 and 2, we see that adding two additional years of data to the panel reduces the finite sample standard deviation of both one-step and two-step SYS estimators by around 22%, although there is now more finite sample bias, particularly for the one-step estimator. The downward bias in the conventional asymptotic standard errors for the two-step estimator increases to around 40%, although both the conventional one-step asymptotic standard errors and the corrected two-step standard errors continue to provide reliable estimates of the finite sample standard deviations for the respective estimators.

Table 2. Estimation Results for AR(1) Model, $N = 100$

	SYS							
	$T = 8$				$T = 6, \chi_1^2$ errors			
	one-step		two-step		one-step		two-step	
	coeff	st dev	coeff	st dev	coeff	st dev	coeff	st dev
	<i>rmse</i>	st err	<i>rmse</i>	st err	<i>rmse</i>	st err	<i>rmse</i>	st err
α_0				st errc				st errc
0.4	0.374	0.076	0.391	0.063	0.387	0.093	0.400	0.075
	<i>0.078</i>	0.074	<i>0.062</i>	0.037	<i>0.094</i>	0.087	<i>0.075</i>	0.043
				0.061				0.071

Results are based on 10,000 replications

st errc is the estimated standard error corrected for small sample bias

st errc is the mean of the estimated standard errors corrected for small sample bias

Figure 2a shows that the corrected two-step Wald test improves on the size performance of the standard one-step Wald test in this experiment, and the Hall-Horowitz version of the bootstrapped one-step Wald test does extremely well again here. However the striking feature of these results is the size distortion found for the bootstrapped versions of the two-step Wald test, where we obtain rejection frequencies of around 1% at the 5% nominal size. Figure 2b shows that the LM test

again improves on the size performance of the standard one-step Wald test. The criterion-difference tests based on the standard GMM and continuously updated GMM criteria again perform quite well,⁹ although that based on the exponential tilting criterion is found to be more sensitive to the increase in the number of moment conditions used.

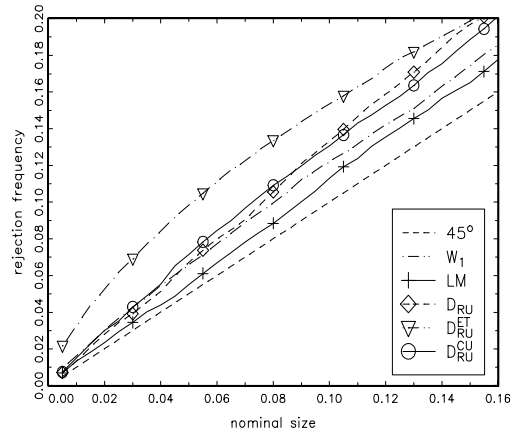
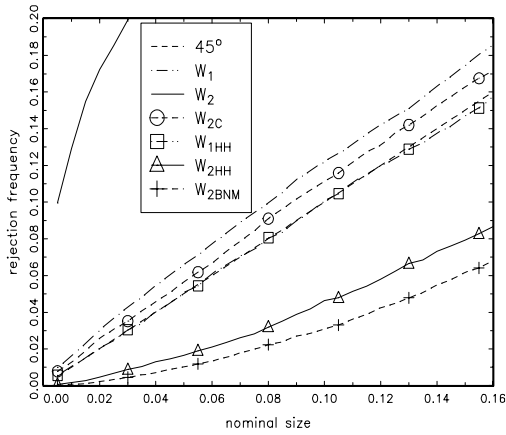


Fig 2a. P-value plot, $\alpha_0 = 0.4$, SYS, $T = 8$

Fig 2b. P-value plot, $\alpha_0 = 0.4$, SYS, $T = 8$

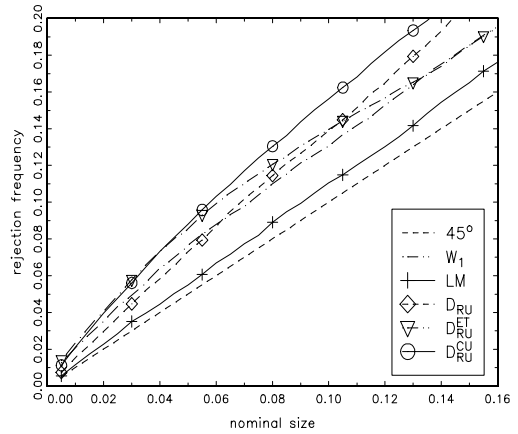
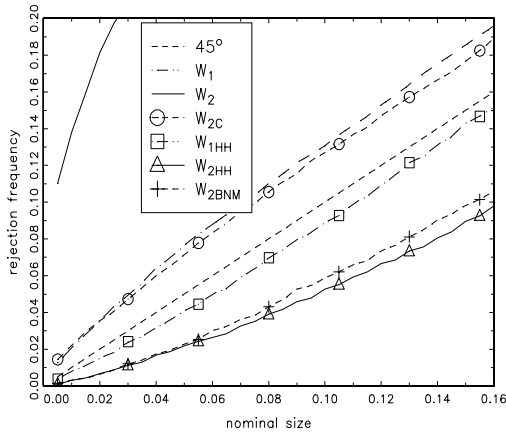


Fig 2c. P-value plot, $\alpha_0 = 0.4$, SYS, χ^2 , $T = 6$

Fig 2d. P-value plot, $\alpha_0 = 0.4$, SYS, χ^2 , $T = 6$

⁹As was the case with $T = 6$ in the previous section, these two criterion-difference tests were found to have better size properties than the standard one-step Wald test for results based on the DIF GMM estimator at $T = 8$.

The introduction of skewed, non-normal errors also results in a more pronounced downward bias in the asymptotic standard errors for the two-step estimator. Again the one-step asymptotic standard errors and the corrected two-step standard errors perform well. Figure 2c shows a similar pattern to that found in Figure 2a, with the bootstrapped one-step Wald test performing well, but with the bootstrapped two-step Wald tests under-rejecting relative to their nominal size. Figure 2d shows that the LM test offers a considerable improvement on the standard one-step Wald test in this case. The criterion-based tests are also found to be quite robust to this form of non-normality, with the simple test based on the standard GMM criterion having rejection frequencies similar to the standard one-step Wald test at conventional levels of significance.

8.1.2. Weak identification: $\alpha_0 = 0.8$

Table 3 and Figure 3 consider the reliability of these test procedures in a case where the identification of the parameter of interest becomes weak. Specifically we consider the covariance stationary $AR(1)$ process as described previously, with $N = 100$ and $T = 6$ and normally distributed error components, but we now increase the true value of the parameter α_0 to 0.8. This is in the region where Blundell and Bond (1998) reported very poor finite sample properties for the GMM estimators based on the DIF moment conditions, essentially because lagged levels of the series provide very weak instruments for subsequent first-differences as the $AR(1)$ process approaches the random walk ($\alpha = 1$) case.

Table 3 confirms the serious downward bias and imprecision previously reported for the DIF GMM estimators in this context. As stressed by Blundell and Bond (1998), the SYS GMM estimators have comparatively small finite sample

biases, around 2% of the true value of the parameter, and much greater precision. Of more interest here, we find that both the conventional asymptotic standard errors for the one-step estimators, and the corrected standard errors for the two-step estimators, continue to provide virtually unbiased estimates of the respective finite sample standard deviations, for both the SYS and the DIF moment conditions.

Table 3. Estimation Results for $AR(1)$ Model, $N = 100$, $T = 6$

	SYS				DIF			
	one-step		two-step		one-step		two-step	
	coeff	st dev	coeff	st dev	coeff	st dev	coeff	st dev
	<i>rmse</i>	st err	<i>rmse</i>	st err	<i>rmse</i>	st err	<i>rmse</i>	st err
α_0				st errc				st errc
0.8	0.819	0.100	0.818	0.096	0.579	0.237	0.553	0.279
	<i>0.102</i>	0.097	<i>0.098</i>	0.053	<i>0.324</i>	0.228	<i>0.373</i>	0.205
				0.094				0.272

Results are based on 10,000 replications

st err is the mean of the estimated asymptotic standard errors

st errc is the mean of the estimated standard errors corrected for small sample bias

Not surprisingly, both one-step and two-step Wald tests based on the seriously biased DIF estimators provide unreliable inference, as seen in Figure 3c. The bootstrapped one-step Wald test using the Hall-Horowitz procedure provides little improvement here; and the bootstrapped two-step Wald tests also remain considerably over-sized. Figure 3d shows that the LM test also offers only a small improvement over the standard one-step Wald test in this experiment; like the bootstrapped one-step Wald test, the LM test seems to be subject to a serious size distortion when the parameter of interest is only weakly identified. Two of the criterion-based tests are found to have comparatively small size distortions here. However we should note that this is likely to be of somewhat limited use

for applied research; as the parameter of interest is estimated so imprecisely using these moment conditions, well-behaved test procedures will simply tell us that the range of values for the parameter that cannot be rejected is extremely wide.

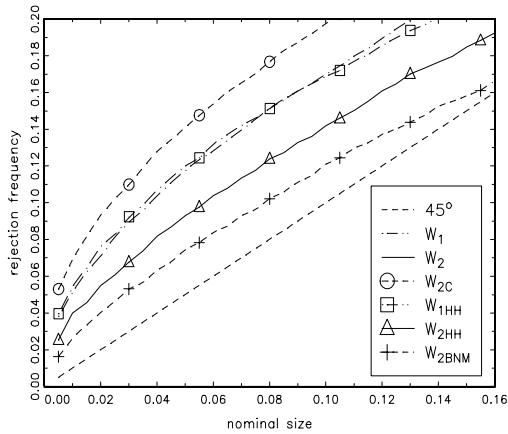


Fig 3a. P-value plot, $\alpha_0 = 0.8$, SYS, $T = 6$

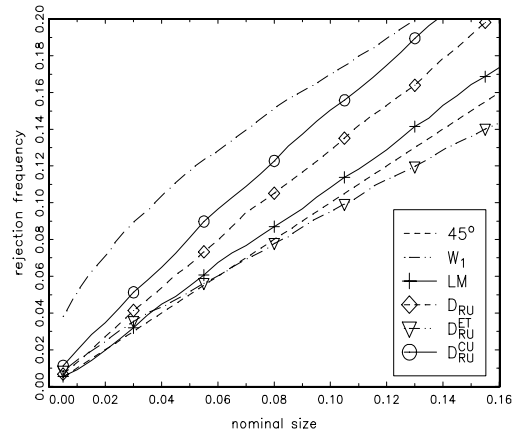


Fig 3b. P-value plot, $\alpha_0 = 0.8$, SYS, $T = 6$

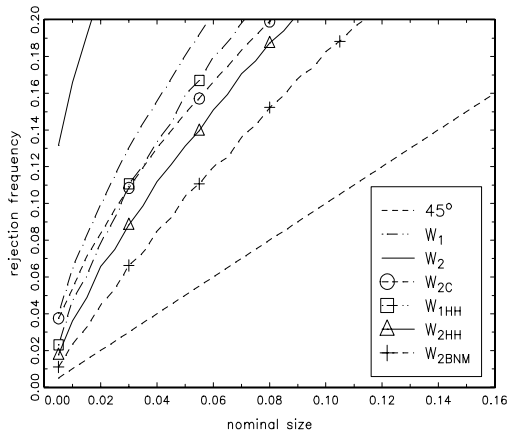


Fig 3c. P-value plot, $\alpha_0 = 0.8$, DIF, $T = 6$

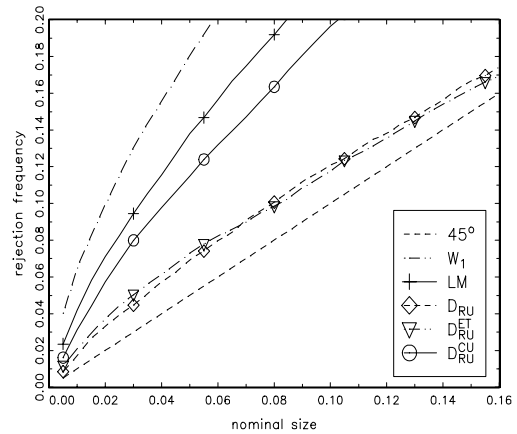


Fig 3d. P-value plot, $\alpha_0 = 0.8$, DIF, $T = 6$

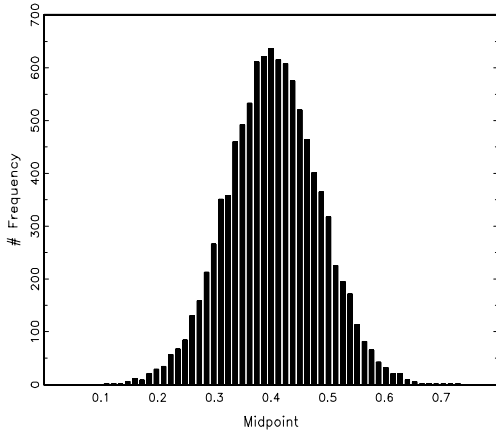


Fig 4a. SYS estimator, $\alpha_0 = 0.4$, $T = 6$

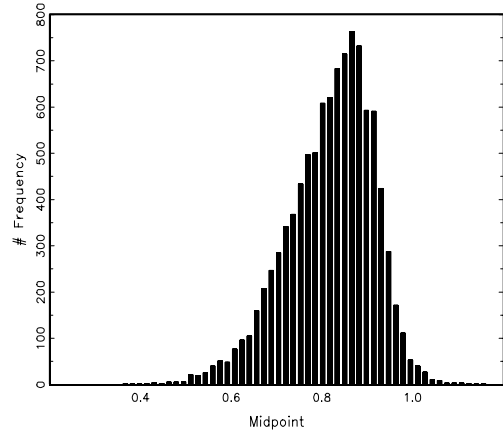


Fig 4b. SYS estimator, $\alpha_0 = 0.8$, $T = 6$

Perhaps more surprisingly, we find that both one-step and two-step Wald tests based on the much less biased SYS estimators have similarly poor size properties to those based on the DIF estimators, as seen in Figure 3a. The problem here can be traced to the shape of the finite sample distribution of the SYS estimator in the $\alpha_0 = 0.8$ case. Figure 4 shows the distribution of the two-step GMM estimators based on the SYS moment conditions in our experiments, at both $\alpha_0 = 0.4$ and $\alpha_0 = 0.8$.¹⁰ Whilst in the former case the distribution of the parameter estimates is symmetric around the mean of 0.4 and approximately normal (Figure 4a), in the latter case the distribution is highly skewed, with the mode being rather higher than the mean (Figure 4b). Thus whilst the corrected asymptotic approximation appears to provide a good guide to the first two moments of this distribution, as we saw in Table 3, it provides a poor guide to the higher moments. Inference based on the Wald tests relies on the asymptotic normality of the estimator, and hence is inaccurate in the $\alpha_0 = 0.8$ case. Perhaps related to this, we find that the

¹⁰A similar pattern was found for the one-step GMM estimator.

bootstrapped one-step Wald test offers no improvement over the standard one-step Wald test in this context. Here we find that the Brown-Newey-May version of the bootstrapped two-step Wald test performs quite well, but would caution against relying on this procedure in general. Perhaps more usefully, Figure 3b shows that the LM test based on the SYS estimator has rejection frequencies very close to nominal size in this experiment. Except for the case of weak instruments shown in Figure 3d, the LM test is found to have good size properties in all the cases we have considered. The three criterion-based tests also provide a considerable improvement on the standard one-step Wald test in this experiment, although as was the case for the DIF estimator, the version based on the continuously updated GMM criterion performs relatively poorly in this group.

8.2. A Multivariate Dynamic Model

All the designs considered in the previous section contained only one parameter of interest (α_0), and the null hypothesis being tested provided a complete specification of the parameter vector. As a result, in formulating the LM statistic as a Wald test using a two-step GMM estimator where the restrictions imposed under the null are used to construct the weight matrix (see Section 4), we have a seemingly very special case in which no parameters need to be estimated to obtain this weight matrix. In this context it may be unsurprising that the LM test avoids the size distortion found for the standard two-step Wald test, which is clearly related to neglecting the estimation of such parameters in the usual derivation of the asymptotic variance matrix. Thus it is important to consider the performance of the LM test in designs with a higher dimensional parameter vector, and particularly in relation to hypothesis tests that leave some of these

parameters unrestricted.¹¹ More generally it is useful to check that the findings presented previously are not special to the simple $AR(1)$ model.

In this section we therefore present some Monte Carlo results for a multivariate dynamic model specified as:

$$\begin{aligned} y_{it} &= \alpha_0 y_{it-1} + \beta_0 x_{it} + \eta_i + v_{it} \\ x_{it} &= \rho_0 x_{it-1} + \tau_0 \eta_i + w_{it} \\ w_{it} &= \theta_0 v_{it-1} + \varepsilon_{it} \end{aligned}$$

The parameters were chosen to be $\alpha_0 = 0.5$, $\beta_0 = 1$, $\rho_0 = 0.5$, $\tau_0 = 0.1$, $\theta_0 = 0.1$, $\eta_i \sim N(0, 1)$, $v_{it} \sim N(0, 1)$, $\varepsilon_{it} \sim N(0, 1)$, $N = 100$ and $T = 6$. We focus on the estimation of the parameters (α_0, β_0) in the first equation, and simple hypothesis tests of the form $H_0 : \alpha_0 = 0.5$ or $H_0 : \beta_0 = 1$ in this setting. The x_{it} series are correlated with η_i and predetermined but not strictly exogenous with respect to v_{it} in this equation. The choice of parameters is again intended to avoid serious problems of weak identification. Note that the inclusion of the predetermined x_{it} variable in the model to be estimated again has the effect of approximately doubling the number of moment conditions used by the SYS GMM estimator, compared to the simple $AR(1)$ models considered previously with $N = 100$ and $T = 6$.

Table 4 presents the estimation results for the SYS GMM estimators, and Figure 5 presents p-value plots based on 10,000 replications. These estimators have little bias, and both the standard one-step asymptotic standard errors and the corrected two-step standard errors perform very well in this design. Figures

¹¹The null hypothesis tested by the familiar ‘t-ratio’ in a multivariate model provides a leading example of this situation.

5a and 5c show the rejection frequencies for the various Wald tests of correct null hypotheses concerning α_0 and β_0 respectively. Both Figures have the familiar pattern, with the standard two-step Wald test seriously over-rejecting and the bootstrapped two-step Wald tests under-rejecting relative to their nominal size. The corrected two-step Wald test performs better than the bootstrapped versions, particularly in Figure 5c, with similar size properties to the standard one-step Wald test; whilst the bootstrapped one-step Wald test performs very well in both cases. Notice that by construction this is a design in which the GMM estimators have little bias, so that the size distortion of the standard one-step Wald test is already small.

Table 4. Estimation Results for multivariate model, $N = 100, T = 6$

SYS									
α					β				
	one-step		two-step		one-step		two-step		
	coeff	st dev	coeff	st dev	coeff	st dev	coeff	st dev	
	<i>rmse</i>	st err	<i>rmse</i>	st err	<i>rmse</i>	st err	<i>rmse</i>	st err	
α_0				st errc	β_0				st errc
0.5	0.503	0.057	0.512	0.052	1.0	0.999	0.086	0.998	0.082
	<i>0.058</i>	0.056	<i>0.054</i>	0.036		<i>0.086</i>	0.084	<i>0.082</i>	0.060
				0.052				0.080	

means and standard deviations of 10,000 replications

st err is the mean of the estimated asymptotic standard errors

st errc is the mean of the estimated standard errors corrected for small sample bias

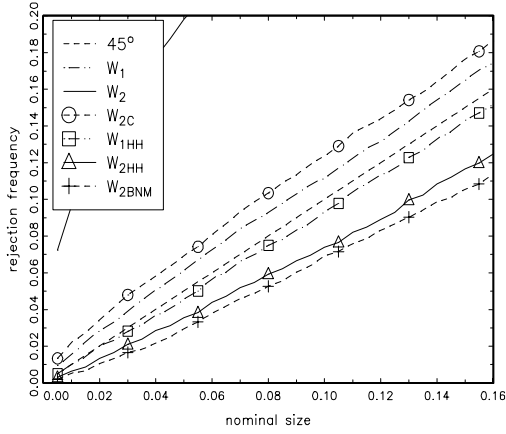


Fig 5a. P-value plot, $\alpha_0 = 0.5$, SYS, $T = 6$

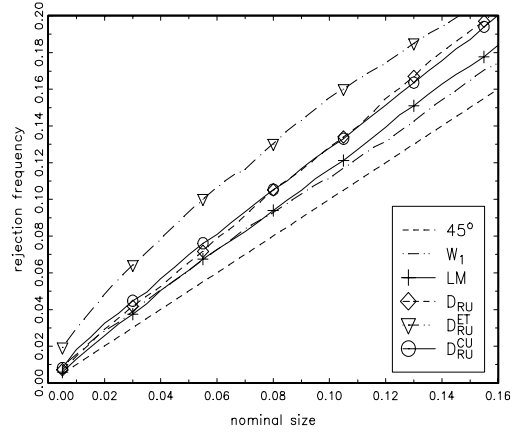


Fig 5b. P-value plot, $\alpha_0 = 0.5$, SYS, $T = 6$

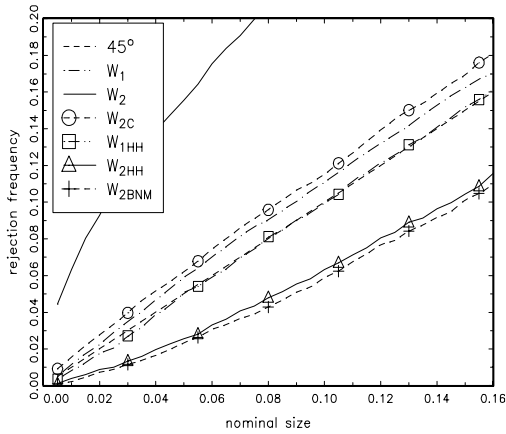


Fig 5c. P-value plot, $\beta_0 = 1$, SYS, $T = 6$

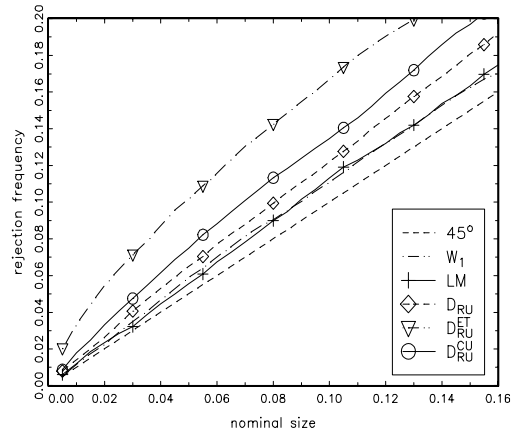


Fig 5d. P-value plot, $\beta_0 = 1$, SYS, $T = 6$

Figures 5b and 5d show the rejection frequencies for the LM test and the criterion-based tests in the same experiment. The simple criterion-difference based on the standard two-step GMM criterion performs well in both cases, whilst the sensitivity of the results for the exponential tilting criterion to an increase in the number of moment conditions, noted in Section 8.1.1, is again apparent here. Perhaps of most interest is the robustness of the good size performance of the

LM test to settings where parameters used to construct the weight matrix are left unrestricted by the null hypothesis, and therefore need to be replaced by consistent estimates in order to construct the LM test. Our suspicion was that the LM test would perform much less well in these cases, but happily that does not seem to be the case. The explanation for this convenient robustness lies in the approximate block diagonality property of the finite sample correction for the asymptotic variance matrix of the standard two-step GMM estimator, discussed in Section 4. As a result the need to estimate parameters that are not specified as part of the null hypothesis has little effect on the performance of the LM test.¹² Indeed in many cases the LM test seems to have better finite sample properties than the feasible version of the correction to the standard Wald test, implemented as W_{2C} in our simulations and discussed in Section 3.1.

8.3. Summary of Size Properties

Our results on the size properties of these alternative test statistics can now be summarised. In cases where the moment conditions used provide reasonable identification of the parameters of interest, or where the instruments are not weak, then the nominal size of the LM test provides an accurate guide to its rejection frequency in samples of the size considered here. This was found to be the case in all the experiments considered, except for the weak instruments example in Figure 3d. In almost all these cases, the bootstrapped version of the one-step Wald test, using the Hall-Horowitz recentering method, improved on the accuracy of the asymptotic one-step Wald test, and provided rejection frequencies very close to

¹²We confirmed that the LM test continued to perform well in designs with many additional explanatory variables, and was not sensitive to the correlation structure between these covariates.

the nominal size. This was particularly the case where small size distortions in the standard one-step Wald test were attributable to small finite sample biases in the GMM estimator, as in Figures 1a, 1c, 2a, 2c, 5a and 5c. The exception was in the results based on the SYS estimator at $\alpha_0 = 0.8$ (Figure 3a), where the size distortion in the standard one-step Wald test was attributable to the distinctly non-normal finite sample distribution of the estimator.

The performance of the bootstrapped versions of the two-step Wald test was comparatively disappointing, with their nominal size seriously overstating their finite sample rejection frequencies in cases such as Figure 2a, 2c and 5c. The fact that the Hall-Horowitz procedure works well for the one-step Wald test in these cases suggests that the problem is not with the method of bootstrapping but rather is related to the need to estimate a weight matrix on each replication in order to implement bootstrapping for the two-step Wald test.¹³ Unfortunately this would suggest that the bootstrapped two-step Wald tests will be least reliable in the same settings as the asymptotic two-step Wald test, where the size distortion is also related to the estimation of the weight matrix. Our simulations suggest that the feasible version of the finite sample correction suggested by Windmeijer (2000) will often result in more reliable inference based on the two-step GMM estimator than these approaches to bootstrapping, at least in the context of these estimators for dynamic panel data models.¹⁴ This corrected two-step Wald test is generally found to have similar size properties to the standard one-step Wald test. Both the bootstrapped one-step Wald test and the LM test can provide more substantial

¹³This has been noted previously in a different setting by Altonji and Segal (1996).

¹⁴Bootstrapped versions of the corrected two-step Wald test were found to have similar problems to those reported for the bootstrapped versions of the standard two-step Wald test. Again this seems likely to be related to the need to estimate the optimal weight matrix in each replication of the bootstrap.

improvements, as illustrated in Figures 2c and 2d. However none of these test procedures are found to have good size properties in the weak instruments case illustrated in Figures 3c and 3d.

Of the three criterion-based tests considered, the simplest version based on the standard two-step GMM criterion was found to be the most robust in our experiments. The version based on the exponential tilting criterion was relatively sensitive to increasing the number of moment conditions (Figures 2b, 5b and 5d), whilst the version based on the continuously updated GMM criterion was relatively poor in the case of weak identification (Figure 3d). The D_{RU} statistic is extremely simple to compute and has reasonable size properties in all the experiments we have considered. This test performs surprisingly well in the weak instruments example of Figure 3d, and can at least be expected to reveal that the parameters of interest are not estimated with any useful precision in this setting.

8.4. Power Comparisons

Having focused on the finite sample size properties of these test statistics, in this section we briefly consider their power to reject incorrect null hypotheses in this setting. Our intention is not to provide a comprehensive evaluation for all the tests previously considered, but rather to check that those tests for which we have found relatively small size distortions do not perform particularly poorly in terms of power.

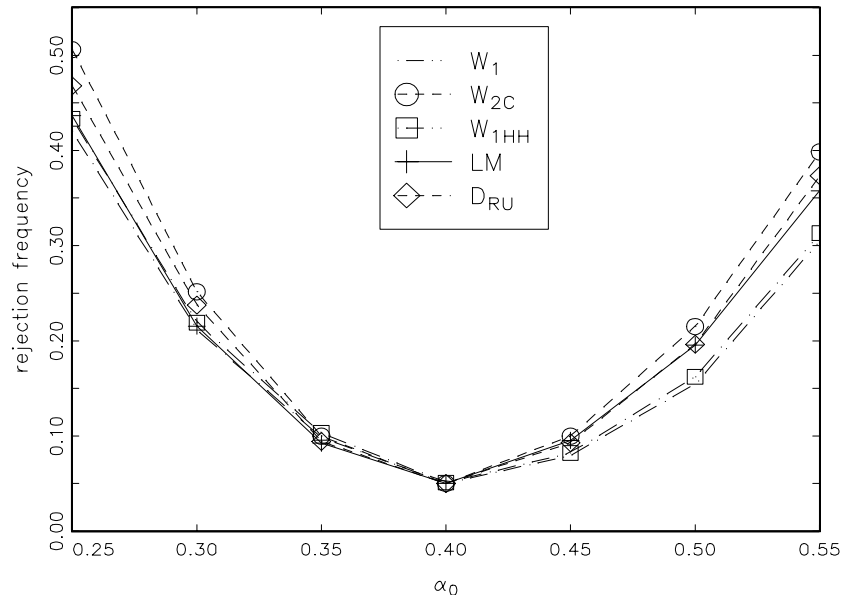


Fig 6. Power plot, $H_0 : \alpha_0 = 0.4$, SYS

To do this we return to the basic $AR(1)$ specification used in Section 8.1, with $N = 100$ and $T = 6$, and consider tests of the null hypothesis $H_0 : \alpha_0 = 0.4$ for a range of true values of the parameter α_0 between 0.25 and 0.55.¹⁵ We focus on the tests previously found to have reasonable size properties in a wide range of experiments: the standard one-step Wald test, the corrected two-step Wald test, the bootstrapped one-step Wald test, the LM test, and the simple criterion-difference test (D_{RU}) based on the standard two-step GMM criterion. Figure 6 presents the power plots for these tests based on the GMM SYS estimator, using the critical values that are found to give a 5% rejection frequency for each test at the value of $\alpha_0 = 0.4$ considered in the null hypothesis.¹⁶ In this example

¹⁵Specifically we use the parameter values (0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55).

¹⁶The results are based on 5000 replications.

we find a small improvement in power for the corrected two-step Wald test, the LM test and the D_{RU} test, compared to the standard one-step Wald test, whilst the bootstrapped one-step Wald test has very similar power to its asymptotic counterpart. In a wider range of experiments we have found no cases in which these alternative test statistics have lower power than the standard one-step Wald test unless the estimator itself is subject to quite serious finite sample bias.

9. Conclusions

We study the finite sample properties of tests of linear restrictions in the context of dynamic panel data models estimated using linear moment conditions. Our results confirm the very poor properties of the standard asymptotic Wald test based on the optimal two-step GMM estimator, even when this estimator has little bias and is approximately normally distributed. This is explained by the neglect of sampling variance in the estimated parameters used to construct the weight matrix, which can result in serious size distortions in panels with dimensions that are often encountered in practice.

Wald tests based on one-step GMM estimators that use a fixed weight matrix typically provide a more reliable basis for finite sample inference. We have considered a range of alternative test procedures: bootstrapped one-step and two-step Wald tests, an alternative two-step Wald test that uses a finite sample correction for the asymptotic variance matrix, the LM test, and three criterion-based tests. We find that the bootstrapped one-step Wald test using the recentering method proposed by Hall and Horowitz (1996) works extremely well, except in cases where the parameters of interest are only weakly identified or where the size distortion in

the standard one-step Wald test results from the small sample distribution of the GMM estimator becoming noticeably non-normal. In contrast the bootstrapped versions of the two-step Wald test perform comparatively poorly, which seems to be related to the problem of estimating the optimal weight matrix. In general these bootstrapping procedures do not provide a reliable correction for Wald tests based on the two-step GMM estimator in these overidentified panel data models. The corrected two-step Wald test developed by Windmeijer (2000) often provides more reliable inference based on the two-step GMM estimator, with size properties similar to those of the standard one-step Wald test. The LM test is found to have smaller size distortions than the one-step Wald test in a number of our experiments, and performs well except in the case of weak identification. The simple criterion-difference test based on the standard two-step GMM criterion is found to perform quite well in all the cases we have considered, and seems to provide the most reliable inference in the case of weak instruments. The tests based on the continuously updated or exponential tilting criteria are found to be less robust.

In conclusion we find that the finite sample corrected two-step Wald test provides a reasonable alternative to the standard one-step Wald test. The bootstrapped one-step Wald test, the LM test and the simple criterion-based test can provide more reliable finite sample inference in cases where these Wald tests have modest size distortions due to small sample biases in the GMM estimators.

References

- [1] Ahn, S.C. and P. Schmidt (1995), Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, 68, 5-28.
- [2] Alonso-Borrego, C. and M. Arellano (1999), Symmetrically Normalised Instrumental-Variable Estimation using Panel Data, *Journal of Business & Economic Statistics* 17, 36-49.
- [3] Altonji, J.G. and L.M. Segal (1996), Small Sample Bias in GMM Estimation of Covariance Structures, *Journal of Business and Economic Statistics* 14, 353-366.
- [4] Arellano, M. and S.R. Bond (1991), Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies* 58, 277-98.
- [5] Arellano, M. and S.R. Bond (1998), Dynamic Panel Data Estimation using DPD98 for GAUSS, http://www.ifs.org.uk/staff/steve_b.shtml.
- [6] Arellano, M. and O. Bover (1995), Another Look at the Instrumental-Variable Estimation of Error-Components Models, *Journal of Econometrics* 68, 29-51.
- [7] Bergström P. (1997), On Bootstrap Standard Errors in Dynamic Panel Data Models, Working Paper No. 1997/23, University of Uppsala.
- [8] Bergström P., M. Dahlberg and E. Johansson (1997), GMM Bootstrapping and Testing in Dynamic Panels, Working Paper 1997/10, University of Uppsala.

- [9] Blundell, R.W. and S.R. Bond (1998), Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, 87, 115-143.
- [10] Bond, S.R., C. Bowsher and F. Windmeijer (2001), Criterion-Based Inference for GMM in Autoregressive Panel Data Models, *Economics Letters* 73, 379-388.
- [11] Brown, B.W., W.K. Newey and S. May (1999), Efficient Bootstrapping for GMM, mimeo, MIT.
- [12] Davidson, R. and J.G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York.
- [13] Davidson, R. and J.G. MacKinnon (1996), Graphical Methods for Investigating the Size and Power of Hypothesis Tests, *Manchester School* 66, 1-26.
- [14] Hall, P. and J.L. Horowitz (1996), Bootstrap Critical Values for Tests based on Generalized Method of Moments Estimators, *Econometrica*, 64, 891-916.
- [15] Hansen, B.E., (2000), Edgeworth Expansions for the Wald and GMM Statistics for Nonlinear Restrictions, mimeo, University of Wisconsin.
- [16] Hansen, L.P., (1982), Large Sample Properties of Generalised Method of Moments Estimators, *Econometrica* 50, 1029-1054.
- [17] Hansen, L.P., J. Heaton and A. Yaron (1996), Finite-Sample Properties of some Alternative GMM Estimators, *Journal of Business & Economic Statistics* 14, 262-280.

- [18] Horowitz, J.L. (2001), The Bootstrap, in J.J. Heckman and E.E. Leamer (eds.), *Handbook of Econometrics Volume 5*, Elsevier Science.
- [19] Imbens, G.W., R.H. Spady and P. Johnson (1998), Information Theoretic Approaches to Inference in Moment Condition Models, *Econometrica* 66, 333-357.
- [20] Koenker, R. and J.A.F. Machado (1999), GMM Inference when the Number of Moment Conditions is Large, *Journal of Econometrics* 93, 327-344.
- [21] Kruiniger, H. (2000), GMM Estimation of Dynamic Panel Data Models with Persistent Data, QMW, University of London, Dept. of Economics Working Paper Series No. 428.
- [22] Nelson, C.R. and R. Startz (1990), The Distribution of the Instrumental Variable Estimator and Its t-ratio when the Instrument is a Poor One, *Journal of Business & Economic Statistics* 63, 5125-5140.
- [23] Newey, W.K. and R.J. Smith (2000), Asymptotic Bias and Equivalence of GMM and GEL Estimators, Working Paper, MIT Department of Economics.
- [24] Newey, W.K. and K.D. West (1987), Hypothesis Testing with Efficient Method of Moments Estimation, *International Economic Review* 28, 777-787.
- [25] Owen, A. (1988), Empirical Likelihood Ratio Confidence Intervals for a Single Functional, *Biometrika* 75, 237-249.
- [26] Sargan, J.D. (1958), The Estimation of Economic Relationships using Instrumental Variables, *Econometrica*, 26, 329-338.

- [27] Sargan, J.D. (1988), Testing for Misspecification after Estimation Using Instrumental Variables, in: E. Maasoumi (ed.), *Contributions to Econometrics: John Denis Sargan*, Volume I, Cambridge University Press.
- [28] Staiger, D. and J.H. Stock (1997), Instrumental Variables Regression with Weak Instruments, *Econometrica* 65, 557-586.
- [29] Windmeijer, F. (2000), A Finite Sample Correction for the Variance of Linear Two-Step GMM Estimators, Institute for Fiscal Studies Working Paper Series No. W00/19, London, <http://www.ifs.org.uk/workingpapers/wp0019.pdf>.