Criterion-Based Inference for GMM in Autoregressive Panel Data Models

Stephen Bond
Clive Bowsher
Frank Windmeijer

THE INSTITUTE FOR FISCAL STUDIES
WP 01/02

Criterion-Based Inference for GMM in Autoregressive Panel Data Models

Stephen Bond[a,b], Clive Bowsher[b], Frank Windmeijer[a,*]

[a] *Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE, UK*
[b] *Nuffield College, Oxford, OX1 1NF, UK*

Abstract

In this paper we examine the properties of a simple criterion-based, likelihood ratio type test of parameter restrictions for standard GMM estimators in autoregressive panel data models. A comparison is made with recent test proposals based on the continuously-updated GMM criterion (Hansen, Heaton and Yaron, 1996) or exponential tilting parameters (Imbens, Spady and Johnson, 1998). The likelihood ratio type statistic is computed simply as the difference between the standard GMM tests of overidentifying restrictions in the restricted and unrestricted models. In Monte Carlo simulations we find this test has similar properties to the two criterion-based alternatives, whilst being much simpler to compute. All three criterion-based tests outperform conventional Wald tests in this context.

Key Words: Generalised Method of Moments; Hypothesis testing; Panel data
JEL Classification: C12, C23

*Corresponding author. Tel.: +44-20-7291-4800; fax: +44-20-7323-4780. *E-mail address:* f.windmeijer@ifs.org.uk (F. Windmeijer)

1. Introduction

The problems of doing inference based on the efficient two-step GMM estimator for panel data are well known. The asymptotic standard errors underestimate the variability of this estimator in small samples, and standard Wald tests of parameter restrictions are seriously oversized (see, for example, Arellano and Bond (1991) and Koenker and Machado (1999)). Because of this, it has become common practice to report the one-step GMM results for more reliable inference. However, as this estimator is generally not asymptotically efficient, one would expect the power properties of tests based on it to be sub-optimal.

In this paper we compare the size and power properties of some alternative tests of parameter restrictions, based on the minimised values of different criterion functions for estimation in the restricted and unrestricted models. Recent papers by Hansen, Heaton and Yaron (1996), (HHY), and Imbens, Spady and Johnson (1998), (ISJ), have proposed such criterion-based tests and shown that their finite sample properties are superior to those of the standard GMM Wald test, albeit not in the context of panel data models. HHY propose a test of parameter restrictions using the minimised values of the criterion function for their continuously-updated GMM estimator, which is equivalent to robust LIML. This estimator requires numerical methods for optimisation even in linear models, and is documented to have convergence problems and multi-modality (see HHY, ISJ and Alonso-Borrego and Arellano (1999)). ISJ use the empirical likelihood framework and advocate use of a weighted optimisation criterion, exponential tilting. This also requires numerical optimisation methods even in linear models, although the particular test proposed by ISJ avoids many of the computational problems that have been found for the HHY test. ISJ show that their exponential tilting test of parameter restrictions has superior size properties compared to the two-step GMM Wald statistic, but do not compare it to the HHY test.

We compare the properties of these test statistics with a particularly simple alternative based on the standard GMM criterion. This test is also of the "likelihood ratio" form, comparing the minimised value of the GMM criterion function under the null to the criterion under the alternative. Our Monte Carlo results suggest that this simple criterion-based test behaves well in the context of autoregressive panel data models. It is found to have similar size and power properties to the computationally more burdensome alternatives based on the continuously-updated GMM criterion or the exponential tilting parameters. All three criterion-based tests perform well compared to the standard GMM Wald tests in this context.

2. GMM and Test Statistics

Consider the moment conditions

$$E\left[g\left(X_i,\theta_0\right)\right]=E\left[g_i\left(\theta_0\right)\right]=0,$$

where $g\left(.\right)$ is a vector of order q and θ_0 is a parameter vector of order $k<q$. The GMM estimator $\widehat{\theta}$ for θ_0 then minimises the weighted quadratic distance[1]

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

with respect to θ; 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 $\lim_{N\rightarrow\infty}\frac{1}{N}\sum_{i=1}^{N}g_i\left(\theta\right)=E\left[g_i\left(\theta\right)\right]$ and $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}g_i\left(\theta_0\right)\rightarrow N\left(0,\Psi\right)$. Let $\Gamma\left(\theta\right)=E\left[\partial g_i\left(\theta\right)/\partial\theta\right]$ and $\Gamma_0\equiv\Gamma\left(\theta_0\right)$, then $\sqrt{N}\left(\widehat{\theta}-\theta_0\right)$ has a limiting normal distribution, $\sqrt{N}\left(\widehat{\theta}-\theta_0\right)\rightarrow N\left(0,V_W\right)$, where

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

[1] See Hansen (1982).

The efficient GMM estimator is based on a weight matrix that satisfies $\text{plim}_{N\to\infty} W_N = \Psi$, with $V_W = \left(\Gamma_0'\Psi^{-1}\Gamma_0\right)^{-1}$. A weight matrix that satisfies this property is given by

$$W_N\left(\widehat{\theta}_1\right) = \frac{1}{N}\sum_{i=1}^{N} g_i\left(\widehat{\theta}_1\right) g_i\left(\widehat{\theta}_1\right)', \tag{2.2}$$

where $\widehat{\theta}_1$ is a consistent initial (one-step) estimator for θ_0.

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

$$J\left(\widehat{\theta}_2\right) = \overline{g}\left(\widehat{\theta}_2\right)' W_N^{-1}\left(\widehat{\theta}_1\right)\overline{g}\left(\widehat{\theta}_2\right),$$

where $\widehat{\theta}_2$ is the efficient two-step GMM estimator. In particular the test statistic $NJ\left(\widehat{\theta}_2\right)$ has an asymptotic chi-squared distribution with $q-k$ degrees of freedom when the moment conditions are valid.

For testing r restrictions of the form

$$r(\theta_0) = 0,$$

the criterion-based test statistic we consider is given by

$$D_{RU} = N\left(J\left(\widetilde{\theta}_2\right) - J\left(\widehat{\theta}_2\right)\right),$$

where $\widehat{\theta}_2$ is the two-step GMM estimator in the unrestricted model and $\widetilde{\theta}_2$ is the two-step GMM estimator in the restricted model, based on the same set of moment conditions. Notice that $\widetilde{\theta}_2$ uses a weight matrix $W_N\left(\widetilde{\theta}_1\right)$, where $\widetilde{\theta}_1$ is a consistent initial estimator for the restricted model. D_{RU} is the "likelihood ratio" test equivalent for GMM (see, for example, Newey and West (1987) and Davidson and MacKinnon (1993, pp. 614-620)). Under the null hypothesis that the restrictions are valid, D_{RU} has an asymptotic chi-squared distribution with r degrees of freedom.[2]

[2]The D_{RU} statistic has recently been considered by Hansen (2000) in the context of testing

HHY (1996) proposed the use of a statistic similar to D_{RU} for the continuously-updated GMM estimator. This estimator is equivalent to robust LIML and is defined as the value of θ, denoted $\widehat{\theta}^{CU}$, that minimizes

$$J^{CU}(\theta) = \overline{g}(\theta)' \left(\frac{1}{N} \sum_{i=1}^{N} g_i(\theta) g_i(\theta)' \right)^{-1} \overline{g}(\theta). \qquad (2.3)$$

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

$$D_{RU}^{CU} = N \left(J^{CU}\left(\widetilde{\theta}^{CU}\right) - J^{CU}\left(\widehat{\theta}^{CU}\right)\right),$$

where $\widehat{\theta}_N^{CU}$ and $\widetilde{\theta}_N^{CU}$ are the continuously-updated GMM estimators for the unrestricted and restricted models respectively.

The ISJ (1998) approach to testing is based on the empirical likelihood method. The "exponential tilting" estimator for θ_0, proposed by Imbens (1997), minimises the Kullback-Leibler information criterion and is given by the solution to the following optimisation problem

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

The estimated probabilities have the form

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

where γ is the vector of tilting parameters, of order q. Intuitively these measure how much the sample has to be re-weighted in order for the moment conditions to hold exactly. ISJ (1998) consider tests of overidentifying restrictions based on the magnitude of γ. Tilting parameters can also be estimated conditional on the standard GMM estimator $\left(\widehat{\theta}_2\right)$, rather than the exponential tilting estimator of θ_0. ISJ (1998, p.349) report that the tilting parameter test of overidentifying

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.

restrictions based on the GMM estimator performs similarly to that based on the exponential tilting estimator, and therefore recommend the former which is simpler to compute.

The corresponding test of parameter restrictions based on the tilting parameters evaluated using the two-step GMM estimator in the restricted and unrestricted models is given by the difference

$$D_{RU}^{ET} = N\left(\widetilde{\gamma}' R_N\left(\widetilde{\theta}_2, \widetilde{\gamma}\right)\widetilde{\gamma} - \widehat{\gamma}' R_N\left(\widehat{\theta}_2, \widehat{\gamma}\right)\widehat{\gamma}\right)$$

where

$$\widehat{\gamma} = \max_{\gamma} \ln \frac{1}{N}\sum_{i=1}^{N} \exp\left(\gamma' g_i\left(\widehat{\theta}_2\right)\right); \tag{2.4}$$

$\widetilde{\gamma}$ is the equivalent vector of tilting parameters based on the efficient two-step GMM estimator in the restricted model, $\widetilde{\theta}_2$; and

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

is a robust estimate of the variance of the moments.

Both D_{RU}^{CU} and D_{RU}^{ET} have an asymptotic chi-squared distribution with r degrees of freedom, and have been shown by respectively HHY (1996) and ISJ (1998) to have better finite sample properties than the conventional two-step Wald tests in particular contexts. However both require the use of numerical optimisation procedures even to test linear restrictions in linear models estimated using linear moment conditions.[3] This is not the case for the criterion-based test based on the standard GMM criterion (D_{RU}). So far as we are aware, the HHY $\left(D_{RU}^{CU}\right)$ and ISJ $\left(D_{RU}^{ET}\right)$ tests of linear restrictions have not been compared to the simpler D_{RU} test. We consider this here in the context of linear dynamic panel data models.

[3] We note that computing the ISJ test is more straightforward than computing the HHY test, given that the objective function in (2.4), unlike (2.3), is strictly concave with derivatives that can be calculated easily.

3. AR(1) Process with Individual Effects

To evaluate the finite sample behaviour of the test statistics described in the previous section, we consider the linear first order autoregressive panel data model with individual effects (η_i)

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

where $i = 1, ..., N$ and $t = 2, ..., T$; N is large, T is fixed and $|\alpha_0| < 1$. If the observations are independent across individuals and the error term satisfies

$$E(\eta_i) = 0,\ E(u_{it}) = 0,\ E(y_{i1}u_{it}) = 0 \quad \text{for } i = 1, ..., N \text{ and } t = 2, ..., T \tag{3.1}$$

and

$$E(u_{it}u_{is}) = 0 \quad \text{for } i = 1, ..., N \text{ and } t \neq s \tag{3.2}$$

then the complete set of second order moment conditions available are the $(T-1)(T-2)/2$ linear moment conditions

$$E\left[y_i^{t-2}(\Delta y_{it} - \alpha_0 \Delta y_{it-1})\right] = 0; \quad t = 3, ..., T, \tag{3.3}$$

where $y_i^{t-2} = [y_{i1}, y_{i2}, ..., y_{it-2}]$. We call these the DIF moment conditions, see Arellano and Bond (1991).[4]

Under the additional error components assumption

$$E(\eta_i u_{it}) = 0, \quad \text{for } i = 1, ..., N \text{ and } t = 2, ..., T \tag{3.4}$$

and the initial conditions assumption

$$E(\eta_i \Delta y_{i2}) = 0 \quad \text{for } i = 1, ..., N \tag{3.5}$$

[4]If the error components assumption, $E(\eta_i u_{it}) = 0$, for $i = 1, ..., N$ and $t = 2, ..., T$, is added, then further non-linear moment conditions become available, see Ahn and Schmidt (1995).

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

$$E\left[\Delta y_{it-1}\left(y_{it}-\alpha_0 y_{it-1}\right)\right]=0; \quad t=3,...,T \tag{3.6}$$

are valid. The joint moment conditions (3.3) and (3.6) comprise the complete set of second order moment conditions available under assumptions (3.1), (3.2), (3.4) and (3.5). We call these the SYS moment conditions, see 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 $E\left[Z_i'v\left(\alpha_0\right)\right]=0$. For the estimator which uses the DIF moment conditions only, there is an efficient one-step GMM weight matrix in the special case when the u_{it} are homoscedastic and not serially correlated. This is given by $W_N=\frac{1}{N}\sum_{i=1}^N Z_i'HZ_i$, 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 estimator which uses the SYS moment conditions there is no simple one-step efficient weight matrix, and often the one-step weight matrix is set to $W_N=\frac{1}{N}\sum_{i=1}^N Z_i'Z_i$. We use these initial weight matrices for the DIF and SYS moment conditions respectively in the Monte Carlo study below. The efficient weight matrix for both estimators under general conditions is given by $W_N\left(\widehat{\alpha}_1\right)=\frac{1}{N}\sum_{i=1}^N Z_i'v_i\left(\widehat{\alpha}_1\right)v_i\left(\widehat{\alpha}_1\right)'Z_i$, with $\widehat{\alpha}_1$ the consistent one-step GMM estimator of α.

In Figures 1-8 we present some Monte Carlo results for the AR(1) panel data process. The data generating process is

$$\begin{aligned}
y_{it} &= \alpha_0 y_{it-1}+\eta_i+u_{it}\\
\eta_i &\sim iidN\left(0,1\right) \quad ; \quad u_{it}\sim iidN\left(0,\sigma_{it}^2\right)\\
y_{i1} &= \frac{\eta_i}{1-\alpha_0}+e_i \;\; ; \;\; e_i\sim iidN\left(0,\sigma_i^2\right),
\end{aligned}$$

with η_i, e_i and u_{it} mutually independent. The sample size is $N=100$, $T=6$, and we report the size properties of the test statistics for $10,000$ samples. We consider tests of the null hypothesis $H_0 : \alpha_0 = 0$. We report results based on the DIF moment conditions and on the SYS moment conditions, and in each case two different designs are considered. A covariance stationary process sets $\sigma_{it}^2 = 1$ for all i and t, and $\sigma_i^2 = \frac{1}{1-\alpha_0^2}$ for all i. For the second design the u_{it} disturbances are conditionally heteroskedastic with $u_{it} \sim N\left(0, 0.4 + 0.3y_{it-1}^2\right)$. For this case fifty initial time periods are generated before the estimation sample is drawn, with $y_{i,-49} \sim N\left(\frac{\eta_i}{1-\alpha_0}, \frac{1}{1-\alpha_0^2}\right)$.

Figures 1-4 compare size properties of the various test statistics for $\alpha_0 = 0$. The statistic W_2 is the Wald test based on the efficient two-step GMM estimator, W_1 is the Wald test based on the one-step GMM estimator, and D_{RU}, D_{RU}^{CU} and D_{RU}^{ET} are the three criterion-based tests described above.[5,6] Note that in this one-parameter model the criterion function under the null, as required for the computation of D_{RU} and D_{RU}^{CU}, is the Anderson-Rubin statistic.[7]

These results confirm that the two-step Wald test is severely oversized in this context, particularly when conditional heteroskedasticity is present. This test will reject a correct null hypothesis much more frequently than indicated by its nominal size. In contrast, the one-step Wald test and the three criterion-based tests have rejection frequencies approximately equal to their nominal size in the designs with i.i.d. disturbances (Figures 1 and 2), and are only slightly oversized in the designs with conditional heteroskedasticity (c.h.) (Figures 3 and 4). The test based on the exponential tilting parameters (D_{RU}^{ET}) appears to have the best size properties

[5] The test statistics D_{RU} and D_{RU}^{ET} can be negative in finite samples. When a statistic is negative, we interpret this as a non-rejection of the null hypothesis.

[6] For the calculation of the continuously-updated estimator and the exponential tilting parameters we used Maxlik 4.0 in Gauss with analytical derivatives.

[7] We have obtained similar results for higher-order autoregressive models in which the null hypothesis tested does not completely specify the parameter vector. These can be obtained from the authors on request.

in the i.i.d. designs, though not necessarily when conditional heteroskedasticity is present.

Figures 5-8 display the power of the three criterion-based tests and the one-step Wald test at the 5% level of significance testing $H_0 : \alpha_0 = 0$, for various true values of α_0, again based on $10,000$ Monte Carlo replications. The power functions are corrected for size distortions. The properties of the three criterion-based tests are strikingly similar. They have more power than the one-step Wald test except in the design with i.i.d. disturbances when only the DIF moment conditions are used - this is the special case in which the one-step GMM estimator is asymptotically efficient.

Overall the results of these Monte Carlo simulations indicate that these criterion-based tests offer advantages in terms of size and power compared to standard Wald tests for GMM estimators in the context of autoregressive panel data models. Moreover the simple to compute D_{RU} test, based on the standard GMM criterion, is found to have very similar properties to the computationally more burdensome D_{RU}^{CU} and D_{RU}^{ET} test statistics in this context.

4. Conclusions

In this paper we have considered the properties of a simple test of parameter restrictions based on the standard two-step efficient GMM estimator. The test is computed simply as the difference between the minimised values of the GMM criterion function in the restricted and unrestricted models. We compared this to criterion-based tests of parameter restrictions based on the continuously-updated GMM estimator of Hansen, Heaton and Yaron (1996) and the exponential tilting parameters of Imbens, Spady and Johnson (1998), as well as to standard asymptotic Wald tests.

We investigated the properties of these tests using Monte Carlo experiments in the context of simple parameter restrictions in linear dynamic panel data models. Our main findings are that the criterion-based tests appear to be useful alternatives to standard Wald tests, and the simple test based on the standard GMM criterion function has very similar properties to the computationally more burdensome alternatives. In future research we will investigate whether this finding holds in more general settings.

Acknowledgements

We thank Richard Spady and an anonymous referee for suggestions and comments. Financial support from ESRC grant R000223068 is gratefully acknowledged. This work has benefited from a longer term programme of research supported by the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at IFS. The usual disclaimer applies.

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] Arellano, M. and S. 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.

[4] Arellano, M. and O. Bover (1995), Another Look at the Instrumental-Variable Estimation of Error-Components Models, *Journal of Econometrics* 68, 29-51.

[5] Blundell, R.W. and S.R. Bond (1998), Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics* 87, 115-143.

[6] Davidson, R. and J.G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York..

[7] Hansen, B.E., (2000), Edgeworth Expansions for the Wald and GMM Statistics for Nonlinear Restrictions, mimeo, University of Wisconsin.

[8] Hansen, L.P., (1982), Large Sample Properties of Generalised Method of Moments Estimators, *Econometrica* 50, 1029-1054.

[9] 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.

[10] Imbens, G.W., (1997), One Step Estimators for Over-Identified Generalized Method of Moments Models, *Review of Economic Studies* 64, 359-383.

[11] Imbens, G.W., R.H. Spady and P. Johnson (1998), Information Theoretic Approaches to Inference in Moment Condition Models, *Econometrica* 66, 333-357.

[12] Koenker, R. and J.A.F. Machado (1999), GMM Inference when the Number of Moment Conditions is Large, *Journal of Econometrics* 93, 327-344.

[13] Newey, W.K. and K.D. West (1987), Hypothesis Testing with efficient Method of Moments Estimation, *International Economic Review* 28, 777-787.

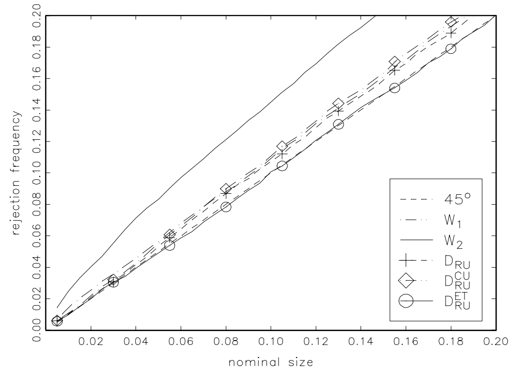

Fig 1. P-value plot, $\alpha_0 = 0$, i.i.d., DIF

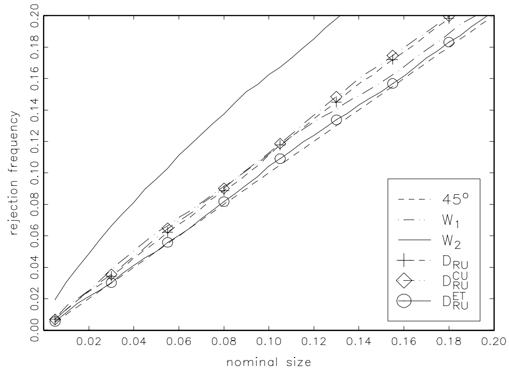

Fig 2. P-value plot, $\alpha_0 = 0$, i.i.d., SYS

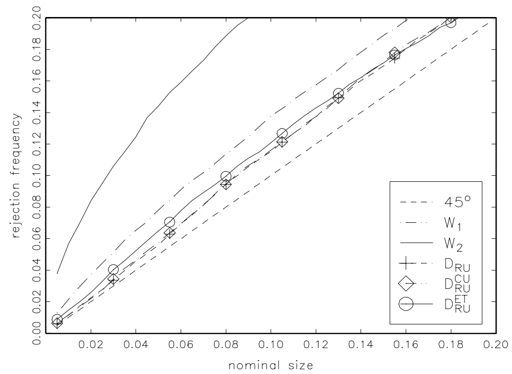

Fig 3. P-value plot, $\alpha_0 = 0$, c.h., DIF

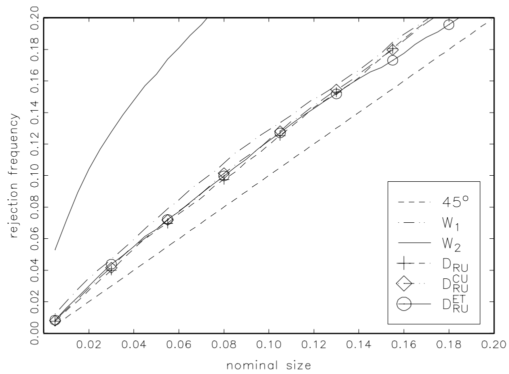

Fig 4. P-value plot, $\alpha_0 = 0$, c.h., SYS

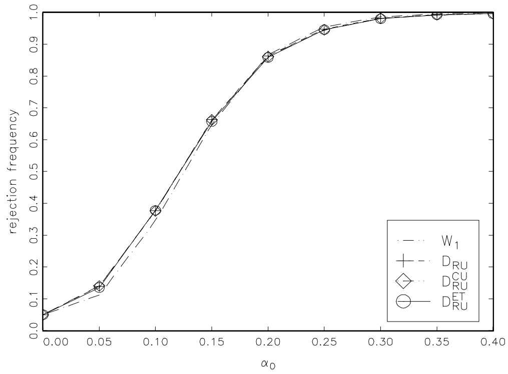

Fig 5. Power-plot, $H_0 : \alpha_0 = 0$, i.i.d., DIF

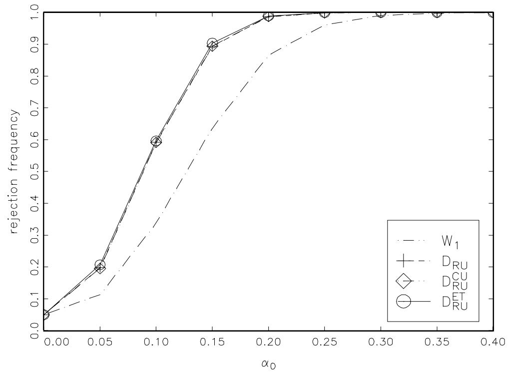

Fig 6. Power-plot, $H_0 : \alpha_0 = 0$, i.i.d., SYS

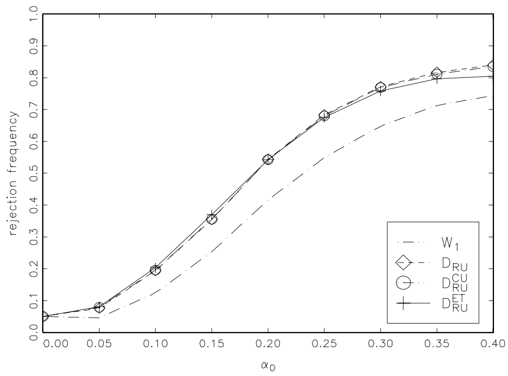

Fig 7. Power plot, $H_0 : \alpha_0 = 0$, c.h., DIF

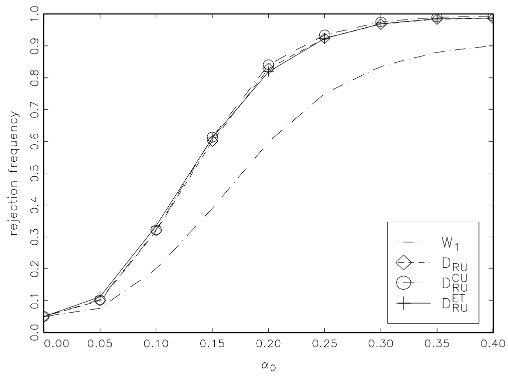

Fig 8. Power plot, $H_0 : \alpha_0 = 0$, c.h., SYS