Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Econometrics 126 (2005) 25–51

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

A finite sample correction for the variance of linear efficient two-step GMM estimators

Frank Windmeijer*

Centre for Microdata Methods and Practice, Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE, UK

Accepted 9 February 2004

Abstract

Monte Carlo studies have shown that estimated asymptotic standard errors of the efficient two-step generalized method of moments (GMM) estimator can be severely downward biased in small samples. The weight matrix used in the calculation of the efficient two-step GMM estimator is based on initial consistent parameter estimates. In this paper it is shown that the extra variation due to the presence of these estimated parameters in the weight matrix accounts for much of the difference between the finite sample and the usual asymptotic variance of the two-step GMM estimator, when the moment conditions used are linear in the parameters. This difference can be estimated, resulting in a finite sample corrected estimate of the variance. In a Monte Carlo study of a panel data model it is shown that the corrected variance estimate approximates the finite sample variance well, leading to more accurate inference.
© 2004 Elsevier B.V. All rights reserved.

JEL classification: C12; C23

Keywords: Generalized method of moments; Variance correction; Panel data

1. Introduction

In Monte Carlo studies it has often been found that the estimated asymptotic standard errors of the efficient, two-step, generalized method of moments (GMM) estimator are severely downward biased in small samples, see e.g. Arellano and Bond (1991), whereas the asymptotic standard errors of one-step GMM estimators are virtually

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

0304-4076/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2004.02.005

unbiased.[1] One-step GMM estimators use weight matrices that are independent of estimated parameters, whereas the efficient two-step GMM estimator weighs the moment conditions by a consistent estimate of their covariance matrix. This weight matrix is constructed using an initial consistent estimate of the parameters in the model. In this paper it is shown that the extra variation due to the presence of these estimated parameters in the efficient weight matrix accounts for much of the difference between the finite sample and the estimated asymptotic variance for two-step GMM estimators based on moment conditions that are linear in the parameters. This difference can be estimated, resulting in finite sample corrected estimates of the variance. The proposed feasible correction to the estimate of the asymptotic variance is very simple to implement and is shown to approximate the finite sample variance of the two-step GMM estimator well in a Monte Carlo study of a panel data model, leading to more accurate inference. As this inference is based on a Wald test using the corrected variance estimate, it will be reliable when the finite sample distribution of the estimator is symmetric and the estimator itself not seriously biased. The variance correction is illustrated using panel data models and data from Arellano and Bond (1991) and Blundell and Bond (2000).

Other proposed solutions to the inference problem for efficient GMM have been based on nonlinear procedures, like the continuously updated GMM approach (Hansen et al., 1996), or empirical likelihood (e.g Imbens et al., 1998). Alternatively, bootstrap methods for GMM have been developed by Hall and Horowitz (1996) and Brown and Newey (2002). Bond and Windmeijer (2002) evaluate these various test procedures in the context of dynamic panel data models. They report problems with the bootstrap procedures when the weight matrix is a poor estimate of the covariance matrix of the moment conditions, which occurs for example when there are a large number of overidentifying restrictions. In such cases inference using the normal asymptotic approximation with the finite sample corrected variance estimate proposed in this paper can be more reliable, which is further illustrated in the Monte Carlo results below.

Section 2 analyses the influence of estimated parameters in the weight matrix of efficient two-step GMM estimators on their asymptotic variance, and derives a finite sample correction that is feasible to implement. Section 3 illustrates this for a bivariate panel data model, and Section 4 presents Monte Carlo results for this model. The effect of estimated weight matrix parameters on the finite sample behaviour of the Sargan/Hansen test of overidentifying restrictions is also considered in this section. Section 5 presents the empirical applications. Section 6 presents some Monte Carlo results that show that the Wald test using two-step estimation results and the corrected variance estimate can have considerable more power than the Wald test based on the one-step results. Section 7 presents some Monte Carlo results that show that inference based on the finite sample corrected variance can still be unreliable when the finite sample distribution of the two-step GMM estimator is asymmetric. Finally, Section 8 concludes.

[1] The same observation has been made for alternative GMM estimators, like the continuously updated and iterated GMM estimators (see Hansen et al., 1996). A finite sample variance correction for the iterated GMM is given in Windmeijer (2000).

2. GMM and finite sample variance correction

Consider the moment conditions

$$\mathrm{E}[g(X_i,\theta_0)] = \mathrm{E}[g_i(\theta_0)] = 0,$$

where $g(.)$ is vector of order q, X_i is a vector of variables for $i = 1,\ldots,N$, and θ_0 is a parameter vector of order k, with $k < q$. The GMM estimator $\hat{\theta}$ for θ_0 minimizes[2]

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

with respect to θ; where W_N satisfies $\mathrm{plim}_{N\to\infty} W_N = W$, with W a positive definite matrix. Regularity conditions are assumed such that $\lim_{N\to\infty}(1/N)\sum_{i=1}^{N} g_i(\theta) = \mathrm{E}[g_i(\theta)]$ and $(1/\sqrt{N})\sum_{i=1}^{N} g_i(\theta_0) \to \mathrm{N}(0,\Psi)$. Let $\Gamma(\theta) = \mathrm{E}[\partial g_i(\theta)/\partial\theta']$ and $\Gamma_{\theta_0} \equiv \Gamma(\theta_0)$, then $\sqrt{N}(\hat{\theta}-\theta_0)$ has a limiting normal distribution, $\sqrt{N}(\hat{\theta}-\theta_0) \to \mathrm{N}(0, V_W)$, where

$$V_W = (\Gamma_{\theta_0}' W^{-1}\Gamma_{\theta_0})^{-1}\Gamma_{\theta_0}' W^{-1}\Psi W^{-1}\Gamma_{\theta_0}(\Gamma_{\theta_0}' W^{-1}\Gamma_{\theta_0})^{-1}. \tag{2.1}$$

The efficient two-step GMM estimator, denoted $\hat{\theta}_2$, is based on a weight matrix that satisfies $\mathrm{plim}_{N\to\infty} W_N = \Psi$, with $V_W = (\Gamma_{\theta_0}'\Psi^{-1}\Gamma_{\theta_0})^{-1}$. A weight matrix that satisfies this property is given by

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

where $\hat{\theta}_1$ is an initial consistent estimator for θ_0.

Let

$$\bar{g}(\theta) = \frac{1}{N}\sum_{i=1}^{N} g_i(\theta),$$

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

$$G(\theta) = \frac{\partial C(\theta)}{\partial \theta} = \begin{bmatrix} \dfrac{\partial C(\theta)}{\partial \theta_1} \\ \dfrac{\partial C(\theta)}{\partial \theta_2} \\ \vdots \\ \dfrac{\partial C(\theta)}{\partial \theta_k} \end{bmatrix}$$

[2] See Hansen (1982).

and

$$b_{\theta_0,W_N} = \frac{1}{2}\frac{\partial Q_{W_N}}{\partial\theta}|_{\theta_0} = C(\theta_0)'W_N^{-1}\bar{g}(\theta_0),$$

$$A_{\theta_0,W_N} = \frac{1}{2}\frac{\partial^2 Q_{W_N}}{\partial\theta\partial\theta'}|_{\theta_0} = C(\theta_0)'W_N^{-1}C(\theta_0) + G(\theta_0)'(I_k \otimes W_N^{-1}\bar{g}(\theta_0)).$$

A standard first-order Taylor series approximation of $\hat{\theta}_2$ around θ_0, conditional on $W_N(\hat{\theta}_1)$, then results in

$$\hat{\theta}_2 - \theta_0 = -A^{-1}_{\theta_0,W_N(\hat{\theta}_1)} b_{\theta_0,W_N(\hat{\theta}_1)} + \mathrm{O_p}(N^{-1})$$

and an estimate for the asymptotic variance of $\hat{\theta}_2$ is given by

$$\widehat{\mathrm{var}}(\hat{\theta}_2) = \frac{1}{N} A^{-1}_{\hat{\theta}_2,W_N(\hat{\theta}_1)} C(\hat{\theta}_2)'W_N^{-1}(\hat{\theta}_1)C(\hat{\theta}_2)A^{-1}_{\hat{\theta}_2,W_N(\hat{\theta}_1)}.$$

However, a further expansion of $\hat{\theta}_1$ around θ_0 results in

$$\hat{\theta}_2 - \theta_0 = -A^{-1}_{\theta_0,W_N(\theta_0)} b_{\theta_0,W_N(\theta_0)} + D_{\theta_0,W_N(\theta_0)}(\hat{\theta}_1 - \theta_0) + \mathrm{O_p}(N^{-1}), \tag{2.3}$$

where

$$W_N(\theta_0) = \frac{1}{N}\sum_{i=1}^{N} g_i(\theta_0)g_i(\theta_0)'$$

and

$$D_{\theta_0,W_N(\theta_0)} = \frac{\partial}{\partial\theta'}(-A^{-1}_{\theta_0,W_N(\theta)} b_{\theta_0,W_N(\theta)})|_{\theta_0}$$

is a $k \times k$ matrix. The jth column of $D_{\theta_0,W_N(\theta_0)}$ is given by[3]

$$D_{\theta_0,W_N(\theta_0)}[.,j] = -A^{-1}_{\theta_0,W_N(\theta_0)} F_{1j,\theta_0,W_N(\theta_0)} A^{-1}_{\theta_0,W_N(\theta_0)} b_{\theta_0,W_N(\theta_0)} + F_{2j,\theta_0,W_N(\theta_0)}, \tag{2.4}$$

where

$$\begin{aligned} F_{1j,\theta_0,W_N(\theta_0)} = {} & C(\theta_0)'W_N^{-1}(\theta_0)\frac{\partial W_N(\theta)}{\partial\theta_j}|_{\theta_0} W_N^{-1}(\theta_0)C(\theta_0) \\ & + G(\theta_0)'\left(I_k \otimes W_N^{-1}(\theta_0)\frac{\partial W_N(\theta)}{\partial\theta_j}|_{\theta_0} W_N^{-1}(\theta_0)\bar{g}(\theta_0)\right), \\ F_{2j,\theta_0,W_N(\theta_0)} = {} & A^{-1}_{\theta_0,W_N(\theta_0)} C(\theta_0)'W_N^{-1}(\theta_0)\frac{\partial W_N(\theta)}{\partial\theta_j}|_{\theta_0} W_N^{-1}(\theta_0)\bar{g}(\theta_0) \end{aligned}$$

and

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

The first term of $D_{\theta_0,W_N(\theta_0)}$ is a function of $A^{-1}_{\theta_0,W_N(\theta_0)} b_{\theta_0,W_N(\theta_0)}$ which is the bias of an infeasible GMM estimator that uses an efficient weight matrix that is based on

[3] Using results given in Magnus and Neudecker (1988, p. 151).

the true parameters θ_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, $F_{2j,\theta_0,W_N(\theta_0)}$, which in general does increase with the number of moment conditions, will therefore dominate.

Let $\hat{\theta}_1$ be a one-step GMM estimator that uses a weight matrix W_N that does not depend on estimated parameters. An estimate of the variance of $\hat{\theta}_2$ that incorporates the term involving the one-step parameter estimates used in the weight matrix can then be obtained as

$$\begin{aligned}\widehat{\mathrm{var}}_c(\hat{\theta}_2) = &\frac{1}{N} A^{-1}_{\hat{\theta}_2,W_N(\hat{\theta}_1)} C(\hat{\theta}_2)' W_N^{-1}(\hat{\theta}_1) C(\hat{\theta}_2) A^{-1}_{\hat{\theta}_2,W_N(\hat{\theta}_1)} \\ &+ \frac{1}{N} D_{\hat{\theta}_2,W_N(\hat{\theta}_1)} A^{-1}_{\hat{\theta}_1,W_N} C(\hat{\theta}_1)' W_N^{-1} C(\hat{\theta}_2) A^{-1}_{\hat{\theta}_2,W_N(\hat{\theta}_1)} \\ &+ \frac{1}{N} A^{-1}_{\hat{\theta}_2,W_N(\hat{\theta}_1)} C(\hat{\theta}_2)' W_N^{-1} C(\hat{\theta}_1) A^{-1}_{\hat{\theta}_1,W_N} D'_{\hat{\theta}_2,W_N(\hat{\theta}_1)} \\ &+ D_{\hat{\theta}_2,W_N(\hat{\theta}_1)} \widehat{\mathrm{var}}(\hat{\theta}_1) D'_{\hat{\theta}_2,W_N(\hat{\theta}_1)},\end{aligned} \tag{2.5}$$

where $D_{\hat{\theta}_2,W_N(\hat{\theta}_1)}$ is as defined in (2.4) with θ_0 and $W_N(\theta_0)$ substituted by $\hat{\theta}_2$ and $W_N(\hat{\theta}_1)$, respectively,[4] and the estimated variance of the one-step estimator is given by

$$\widehat{\mathrm{var}}(\hat{\theta}_1) = \frac{1}{N} A^{-1}_{\hat{\theta}_1,W_N} C(\hat{\theta}_1)' W_N^{-1} W_N(\hat{\theta}_1) W_N^{-1} C(\hat{\theta}_1) A^{-1}_{\hat{\theta}_1,W_N}.$$

The term $D_{\theta_0,W(\theta_0)}(\hat{\theta}_1 - \theta_0)$ in (2.3) is itself $O_p(N^{-1})$ and in this general setting, incorporating non-linear models and/or non-linear moment conditions, whether taking account of it will improve the estimation of the small sample variance substantially depends on the other remainder terms which are of the same order.

2.1. Linear moment conditions

An improvement of the variance estimate will be obtained in models where all the moment conditions used are linear in the parameters, as in this case

$$\begin{aligned}\hat{\theta}_2 - \theta_0 &= -(C' W_N^{-1}(\hat{\theta}_1) C)^{-1} C' W_N^{-1}(\hat{\theta}_1) \bar{g}(\theta_0) \\ &= -(C' W_N^{-1}(\theta_0) C)^{-1} C' W_N^{-1}(\theta_0) \bar{g}(\theta_0) \\ &\quad + D_{\theta_0,W_N(\theta_0)}(\hat{\theta}_1 - \theta_0) + o_p(N^{-1}),\end{aligned}$$

[4] As $A^{-1}_{\hat{\theta}_2,W_N(\hat{\theta}_1)} b_{\hat{\theta}_2,W_N(\hat{\theta}_1)} = 0$, the jth column of $D_{\hat{\theta}_2,W_N(\hat{\theta}_1)}$ is equal to $F_{2j,\hat{\theta}_2,W_N(\hat{\theta}_1)}$.

where the jth column of $D_{\theta_0,W_N(\theta_0)}$ is given by

$$D_{\theta_0,W_N(\theta_0)}[.,j] = -(C'W_N^{-1}(\theta_0)C)^{-1}C'W_N^{-1}(\theta_0)\frac{\partial W_N(\theta)}{\partial\theta_j}|_{\theta_0}W_N^{-1}(\theta_0)C$$
$$\times(C'W_N^{-1}(\theta_0)C)^{-1}C'W_N^{-1}(\theta_0)\bar{g}(\theta_0)$$
$$+(C'W_N^{-1}(\theta_0)C)^{-1}C'W_N^{-1}(\theta_0)\frac{\partial W_N(\theta)}{\partial\theta_j}|_{\theta_0}W_N^{-1}(\theta_0)\bar{g}(\theta_0).$$

In this case taking account of the $O_p(N^{-1})$ term $D_{\theta_0,W_N(\theta_0)}(\hat{\theta}_1-\theta_0)$ will result in a more accurate approximation of the variance of $\hat{\theta}_2$ in finite samples.

A one-step linear estimator satisfies

$$\hat{\theta}_1-\theta_0 = -(C'W_N^{-1}C)^{-1}C'W_N^{-1}\bar{g}(\theta_0)$$

and the finite sample corrected estimate of the variance of $\hat{\theta}_2$ can be obtained as

$$\widehat{\text{var}}_c(\hat{\theta}_2) = \frac{1}{N}(C'W_N^{-1}(\hat{\theta}_1)C)^{-1}$$
$$+\frac{1}{N}(D_{\hat{\theta}_2,W_N(\hat{\theta}_1)}(C'W_N^{-1}(\hat{\theta}_1)C)^{-1} + (C'W_N^{-1}(\hat{\theta}_1)C)^{-1}D'_{\hat{\theta}_2,W_N(\hat{\theta}_1)})$$
$$+D_{\hat{\theta}_2,W_N(\hat{\theta}_1)}\widehat{\text{var}}(\hat{\theta}_1)D'_{\hat{\theta}_2,W_N(\hat{\theta}_1)}, \quad (2.6)$$

where the first term is the conventional estimate of the asymptotic variance;[5]

$$D_{\hat{\theta}_2,W_N(\hat{\theta}_1)}[.,j] = (C'W_N^{-1}(\hat{\theta}_1)C)^{-1}C'W_N^{-1}(\hat{\theta}_1)\frac{\partial W_N(\theta)}{\partial\theta_j}|_{\hat{\theta}_1}W_N^{-1}(\hat{\theta}_1)\bar{g}(\hat{\theta}_2)$$

and

$$\widehat{\text{var}}(\hat{\theta}_1) = \frac{1}{N}(C'W_N^{-1}C)^{-1}C'W_N^{-1}W_N(\hat{\theta}_1)W_N^{-1}C(C'W_N^{-1}C)^{-1}.$$

2.2. Discussion

The above derivation of the finite sample variance correction for linear efficient two-step GMM is similar in spirit to the variance adjustment for models with generated regressors where one of the explanatory variables is a function of estimated parameters, see e.g. Pagan (1984) and Murphy and Topel (1985). A crucial difference is that the presence of generated regressors affects the variance of the limiting distribution of an estimator, whereas the estimated parameters in the weight matrix for efficient two-step GMM does not affect the limiting distribution of the estimator, as

$$\sqrt{N}(\hat{\theta}_2-\theta_0) = -\sqrt{N}(C'W_N^{-1}(\theta_0)C)^{-1}C'W_N^{-1}(\theta_0)\bar{g}(\theta_0)$$
$$+D_{\theta_0,W_N(\theta_0)}\sqrt{N}(\hat{\theta}_1-\theta_0)+o_p(N^{-1/2}) \quad (2.7)$$

[5] Note that $D_{\hat{\theta}_2,W_N(\hat{\theta}_1)}\widehat{\text{var}}(\hat{\theta}_1)D'_{\hat{\theta}_2,W_N(\hat{\theta}_1)}$ is of lower order than the covariance term $(1/N)$ $D_{\hat{\theta}_2,W_N(\hat{\theta}_1)}(C'W_N^{-1}(\hat{\theta}_1)C)^{-1}$, and is of the same order as the remainder term.

and $\text{plim}_{N\to\infty} D_{\theta_0,W_N(\theta_0)}=0$. The variance correction then simply incorporates the finite sample value of $D_{\hat{\theta}_2,W_N(\hat{\theta}_1)}$ which is not equal to zero, unless the model is just identified. This type of expansion would therefore not work in settings like just identified feasible GLS estimation that uses parameter estimates in the calculation of the variance matrix of the residuals. Related to this is the derivation by Koenker et al. (1994) of exact moment expansions for a minimum distance estimator on the basis of OLS estimated parameters. Their first term in the expansion for the variance of $\sqrt{N}(\hat{\theta}-\theta_0)$ is of order $O(N^{-1})$. They do not address the issue of the exact finite sample distribution of the estimator, which is also beyond the scope of this paper. Note that expansion (2.7) is not a higher order Edgeworth approximation. Based on the standard first-order asymptotic approximation, the limiting distribution of $\hat{\theta}_2$ is normal with a particular variance. Correction (2.6) will provide a better finite sample estimate of this variance by taking into account the finite sample variation of $\hat{\theta}_1$. If θ_0 were known for the estimation of the efficient weight matrix, this adjustment would clearly not occur.

A related issue is the finite sample bias of the GMM estimator itself. Altonji and Segal (1996) find large finite sample biases for their optimal minimum distance estimator of the variance of certain distributions, using grouped observations. This bias is due to the fact that the second order moments are correlated with the fourth order moments in the efficient weight matrix. Newey and Smith (2000) use higher order asymptotic expansions to show that in some cases the estimation of the efficient weight matrix may be a significant source of bias. This bias is not present when the third moments of $g_i(\theta_0)$ are zero. Further sources of bias are the degree of overidentification in the model, with the bias increasing with increasing numbers of instruments, given the number of estimated parameters, see for example Hahn et al. (2001),[6] and weak instruments, see for example Bound et al. (1995) and Staiger and Stock (1997). It is clear that the finite sample variance correction will only be useful for improving inference if the GMM estimator itself does not suffer from a large finite sample bias in a given application, and when the finite sample distribution of the GMM estimator is symmetric. In the next sections, the finite sample variance correction is evaluated for a simple linear panel data model by means of a Monte Carlo study with a design for which the GMM estimator does not have a large finite sample bias and is symmetrically distributed. Section 6 provides an example of an asymmetrically distributed GMM estimator.

3. A panel data model

Consider the panel data model specification

$$y_{it}=\beta_0 x_{it}+u_{it},$$

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

[6] Koenker and Machado (1999) investigate the asymptotic behaviour of linear GMM estimators when the number of moment conditions increases with the sample size. They establish that a sufficient condition for the usual limiting distribution of the GMM estimator is that the number of moment conditions is of the order o $(N^{-1/3})$.

for $i=1,\ldots,N$, $t=1,\ldots,T$. The single regressor x_{it} is correlated with η_i and predetermined with respect to v_{it}, meaning that $\mathrm{E}(x_{it}v_{it+s})=0$, $s=0,\ldots,T-t$, but $\mathrm{E}(x_{it}v_{it-r}) \neq 0$, $r=1,\ldots,t-1$. A commonly used estimator is the GMM estimator in the model in first differences, see Arellano and Bond (1991),

$$\Delta y_{it} = \beta_0 \Delta x_{it} + \Delta u_{it}, \quad t=2,\ldots,T$$

with $T(T-1)/2$ sequential instruments

$$Z_i = \begin{bmatrix} x_{i1} & 0 & 0 & 0 & 0 & 0 \\ 0 & x_{i1} & x_{i2} & 0 & 0 & 0 \\ & & \ddots & & & \\ 0 & 0 & 0 & x_{i1} & \ldots & x_{iT-1} \end{bmatrix}.$$

The moment conditions are then given by $\mathrm{E}(Z_i'\Delta u_i)=0$, where Δu_i is the $(T-1)$ vector $(\Delta u_{i2},\ldots,\Delta u_{iT})'$. The covariance matrix of the moment conditions is denoted Ψ.

A one-step GMM estimator is given by

$$\hat{\beta}_1 = (\Delta x' Z W_N^{-1} Z' \Delta x)^{-1} \Delta x' Z W_N^{-1} Z' \Delta y,$$

where Z' is the $T(T-1)/2 \times N(T-1)$ matrix $(Z_1', Z_2', \ldots, Z_N')$, Δx_i is the $(T-1)$ vector $(\Delta x_{i2},\ldots,\Delta x_{iT})'$, Δy_i is the $(T-1)$ vector $(\Delta y_{i2},\ldots,\Delta y_{iT})'$, Δx and Δy are $N(T-1)$ vectors $(\Delta x_1', \Delta x_2', \ldots, \Delta x_N')'$ and $(\Delta y_1', \Delta y_2', \ldots, \Delta y_N')'$, respectively, and W_N^{-1} is an initial positive definite weight matrix. For example, 2SLS sets $W_N = (1/N) Z'Z$. An initial weight matrix that is efficient when the v_{it} are i.i.d. is $W_N = (1/N)\sum_{i=1}^{N} Z_i' H Z_i$, where H is a matrix with 2's on the main diagonal, -1's on the first off-diagonals and zeros elsewhere.

The asymptotic variance of $\hat{\beta}_1$ is estimated by

$$\widehat{\mathrm{var}}(\hat{\beta}_1) = N(\Delta x' Z W_N^{-1} Z' \Delta x)^{-1} \Delta x' Z W_N^{-1} W_N(\hat{\beta}_1) W_N^{-1} Z' \Delta x (\Delta x' Z W_N^{-1} Z' \Delta x)^{-1},$$

where

$$W_N(\hat{\beta}_1) = \frac{1}{N} \sum_{i=1}^{N} Z_i' \Delta \hat{u}_{i1} \Delta \hat{u}_{i1}' Z_i,$$

$$\Delta \hat{u}_{i1} = \Delta y_i - \hat{\beta}_1 \Delta x_i$$

with $W_N(\hat{\beta}_1)$ a consistent estimate of Ψ. Given the estimate $\hat{\beta}_1$, the efficient two-step GMM estimator is given by

$$\hat{\beta}_2 = \left(\Delta x' Z W_N^{-1}\left(\widehat{\beta}_1\right) Z' \Delta x \right)^{-1} \Delta x' Z W_N^{-1}\left(\hat{\beta}_1\right) Z' \Delta y.$$

Standard theory implies that the asymptotic variance of $\hat{\beta}_2$ is estimated by

$$\widehat{\text{var}}(\hat{\beta}_2) = N(\Delta x' Z W_N^{-1}(\hat{\beta}_1) Z' \Delta x)^{-1}, \tag{3.1}$$

which is an estimate of $(1/N)(\Gamma'_{Z\Delta x}\Psi^{-1}\Gamma_{Z\Delta x})^{-1}$, with $\Gamma_{Z\Delta x} = \text{plim}_{N\to\infty}(1/N)Z'\Delta x$.

Applying the Taylor series expansion developed in the previous section to account for the presence of $\hat{\beta}_1$ in the estimated weight matrix results in

$$\begin{aligned}\hat{\beta}_2 - \beta_0 = {} & (\Delta x' Z W_N^{-1}(\beta_0) Z' \Delta x)^{-1} \Delta x' Z W_N^{-1}(\beta_0) Z' \Delta u \\ & + D_{\beta_0, W_N(\beta_0)}(\hat{\beta}_1 - \beta_0) + \text{o}_\text{p}(N^{-1}),\end{aligned} \tag{3.2}$$

where Δu is the $N(T-1)$ vector $(\Delta u'_1, \Delta u'_2, \ldots, \Delta u'_N)'$, and $D_{\beta_0, W_N(\beta_0)}$ is given by

$$\begin{aligned}D_{\beta_0, W_N(\beta_0)} = {} & (\Delta x' Z W_N^{-1}(\beta_0) Z' \Delta x)^{-1} \Delta x' Z W_N^{-1}(\beta_0) \frac{\partial W_N(\beta)}{\partial \beta}|_{\beta_0} W_N^{-1}(\beta_0) Z' \Delta x \\ & \times (\Delta x' Z W_N^{-1}(\beta_0) Z' \Delta x)^{-1} \Delta x' Z W_N^{-1}(\beta_0) Z' \Delta u \\ & - (\Delta x' Z W_N^{-1}(\beta_0) Z' \Delta x)^{-1} \Delta x' Z W_N^{-1}(\beta_0) \frac{\partial W_N(\beta)}{\partial \beta}|_{\beta_0} W_N^{-1}(\beta_0) Z' \Delta u\end{aligned}$$

with

$$W_N(\beta_0) = \frac{1}{N} \sum_{i=1}^{N} Z'_i \Delta u_i \Delta u'_i Z_i$$

and

$$\frac{\partial W_N(\beta)}{\partial \beta}|_{\beta_0} = -\frac{1}{N} \sum_{i=1}^{N} Z'_i (\Delta x_i \Delta u'_i + \Delta u_i \Delta x'_i) Z_i.$$

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

$$\begin{aligned}\widehat{\text{var}}_c(\hat{\beta}_2) = {} & N(\Delta x' Z W_N^{-1}(\hat{\beta}_1) Z' \Delta x)^{-1} \\ & + N D_{\hat{\beta}_2, W_N(\hat{\beta}_1)} (\Delta x' Z W_N^{-1}(\hat{\beta}_1) Z' \Delta x)^{-1} \\ & + N(\Delta x' Z W_N^{-1}(\hat{\beta}_1) Z' \Delta x)^{-1} D'_{\hat{\beta}_2, W_N(\hat{\beta}_1)} \\ & + D_{\hat{\beta}_2, W_N(\hat{\beta}_1)} \widehat{\text{var}}(\hat{\beta}_1) D'_{\hat{\beta}_2, W_N(\hat{\beta}_1)}.\end{aligned} \tag{3.3}$$

Again, as $(\Delta x' Z W_N^{-1}(\hat{\beta}_1) Z' \Delta x)^{-1} \Delta x' Z W_N^{-1}(\hat{\beta}_1) Z' \Delta \hat{u}_2 = 0$, where $\Delta \hat{u}_2 = \Delta y_i - \hat{\beta}_2 \Delta x$, the expression of $D_{\hat{\beta}_2, W_N(\hat{\beta}_1)}$ simplifies to

$$D_{\hat{\beta}_2, W_N(\hat{\beta}_1)} = -(\Delta x' Z W_N^{-1}(\hat{\beta}_1) Z' \Delta x)^{-1} \Delta x' Z W_N^{-1}(\hat{\beta}_1) \frac{\partial W_N(\beta)}{\partial \beta}|_{\hat{\beta}_1} W_N^{-1}(\hat{\beta}_1) Z' \Delta \hat{u}_2.$$

4. Monte Carlo results

A panel data process is generated as

$$y_{it} = \beta_0 x_{it} + \eta_i + v_{it},$$

$$x_{it} = 0.5x_{it-1} + \eta_i + 0.5v_{it-1} + \varepsilon_{it},$$

$$\eta_i \sim \mathrm{N}(0,1) \quad \varepsilon_{it} \sim \mathrm{N}(0,1),$$

$$v_{it} = \delta_i \tau_t \omega_{it} \quad \omega_{it} \sim (\chi_1^2 - 1),$$

$$\delta_i \sim \mathrm{U}[0.5, 1.5] \quad \tau_t = 0.5 + 0.1(t-1).$$

Fifty time periods are generated, with $\tau_t = 0.5$ for $t = -49, \ldots, 0$ and $x_{i,-49} \sim N(\eta_i/0.5 + 1/0.75)$, before the estimation sample is drawn. This model design corresponds to the features of the panel data model described in the previous section, the x_{it} are correlated with the unobserved heterogeneity η_i and are predetermined with respect to v_{it}. The design is further such that the v_{it} are skewed and heteroscedastic over both time t and individuals i. The parameters are estimated by first differenced GMM as described in the previous section.

Table 1 reports estimation results for $\beta_0 = 1$, $N = 100$, $T = 4$ and 8. Reported are means and standard deviations of a one-step GMM estimator with $W_N = (1/N)\sum_{i=1}^{N} Z_i' H Z_i$, which is not efficient in this case, the two-step GMM estimator using $W_N(\hat{\beta}_1)$, and an infeasible estimator that uses the true parameter β_0 to evaluate the weight matrix, $W_N(\beta_0)$. This latter estimator is denoted $\hat{\beta}_{W_N(\beta_0)}$. For all three GMM estimators, means of the conventional asymptotic standard errors are reported, denoted $\mathrm{se}\hat{\beta}_1$, $\mathrm{se}\hat{\beta}_2$, and

Table 1
Monte Carlo results

	$T = 4$	$T = 8$
$\hat{\beta}_1$	0.9800	0.9784
$\mathrm{sd}\hat{\beta}_1$	0.1534	0.0832
$\mathrm{se}\hat{\beta}_1$	0.1471	0.0809
$\hat{\beta}_2$	0.9868	0.9810
$\mathrm{sd}\hat{\beta}_2$	0.1423	0.0721
$\mathrm{se}\hat{\beta}_2$	0.1244	0.0477
$\mathrm{sec}\hat{\beta}_2$	0.1391	0.0715
$\hat{\beta}_{W_N(\beta_0)}$	0.9895	0.9915
$\mathrm{sd}\hat{\beta}_{W_N(\beta_0)}$	0.1278	0.0481
$\mathrm{se}\hat{\beta}_{W_N(\beta_0)}$	0.1229	0.0474

Notes: $N = 100$, $\beta_0 = 1$, means and standard deviations of 10,000 replications. $\mathrm{sec}\hat{\beta}_2$ is the finite sample corrected estimated standard error of $\hat{\beta}_2$. $\hat{\beta}_{W_N(\beta_0)}$ is the GMM estimator for β_0 using $W_N(\beta_0)$.

se$\hat{\beta}_{W_N(\beta_0)}$. The means of the feasible estimated corrected standard errors, calculated from (3.3), is denoted sec$\hat{\beta}_2$.

The means of the estimates show that the GMM estimators are only slightly downward biased, less so for the two-step GMM than for the one-step GMM, whereas the infeasible GMM estimator has the smallest bias. This is due in part to the fact that the parameters in the data generating process for x_{it} are chosen such that there is no weak instrument problem and so the lagged values of x_{it} are informative for Δx_{it}. Also, the correlation between the moments and the estimated weight matrix is small in this case, inducing only a small finite sample bias in the mean. Increasing T from 4 to 8, and the number of instruments from 6 to 28, has only a very small impact on the bias of the estimators.

The estimated asymptotic standard errors of the GMM estimators that do not have estimated parameters present in the weight matrix, $\hat{\beta}_1$ and $\hat{\beta}_{W_N(\beta_0)}$, are on average only slightly smaller than their standard deviations, less so at $T=8$ than at $T=4$. For the two-step GMM estimator, however, the means of the estimated asymptotic standard errors are considerably smaller than the standard deviations of $\hat{\beta}_2$, especially at $T=8$. At $T=4$, se$\hat{\beta}_2$ accounts for 87% of sd$\hat{\beta}_2$, whereas when $T=8$, se$\hat{\beta}_2$ accounts for only 66% of sd$\hat{\beta}_2$. When $T=8$, there are 28 instruments, whereas there are only 6 instruments when $T=4$.

The standard deviations of the Taylor series expansion (3.2) evaluated at the true parameter values are given by 0.1414 and 0.0717 for $T=4$ and 8, respectively, and therefore almost equal to the standard deviations of $\hat{\beta}_2$. The standard deviations of the leading term in (3.2), $(\Delta x' Z W_N^{-1}(\beta_0) Z' \Delta x)^{-1} \Delta x' Z W_N^{-1}(\beta_0) Z' \Delta u$, are given by sd$\hat{\beta}_{W_N(\beta_0)}$, and so the term involving $(\hat{\beta}_1 - \beta_0)$ accounts for 10% and 33% of the standard deviation of $\hat{\beta}_2$ for $T=4$ and 8, respectively. As can be seen, the conventional estimated asymptotic standard error of $\hat{\beta}_2$, se$\hat{\beta}_2$, is in fact a good estimate of the standard deviation of the estimator $\hat{\beta}_{W_N(\beta_0)}$ rather than the standard deviation of $\hat{\beta}_2$, and the difference between sd$\hat{\beta}_2$ and se$\hat{\beta}_2$ is due to the presence of the estimated $\hat{\beta}_1$ in the weight matrix.

The means of the feasible estimated standard errors that correct for this extra variation due to the estimation of weight matrix parameters, as estimated from (3.3), are also very close to the standard deviations of $\hat{\beta}_2$. The means of the corrected standard errors now account for 98% and 99% of the standard deviation of $\hat{\beta}_2$, for $T=4$ and 8, respectively.

In order to evaluate the behavior of the Wald test statistics for the test H_0: $\beta_0 = 1$, based on the one-step and two-step estimators and associated standard errors, Figs. 1 and 2 show p-value plots (see Davidson and MacKinnon, 1996) for these Wald statistics, for $T=4$ and 8, respectively. W_1 is based on the one-step estimator and its asymptotic standard error. W_2 is based on the conventional two-step estimation results, whereas W_{2C} uses the corrected variance estimate. Also shown are the p-value plots for the symmetric t-tests using critical values from the bootstrap procedure for two-step GMM proposed by Hall and Horowitz (1996), on the basis of 500 bootstrap samples per Monte Carlo replication. This test procedure is denoted W_{2HH} in the graphs.

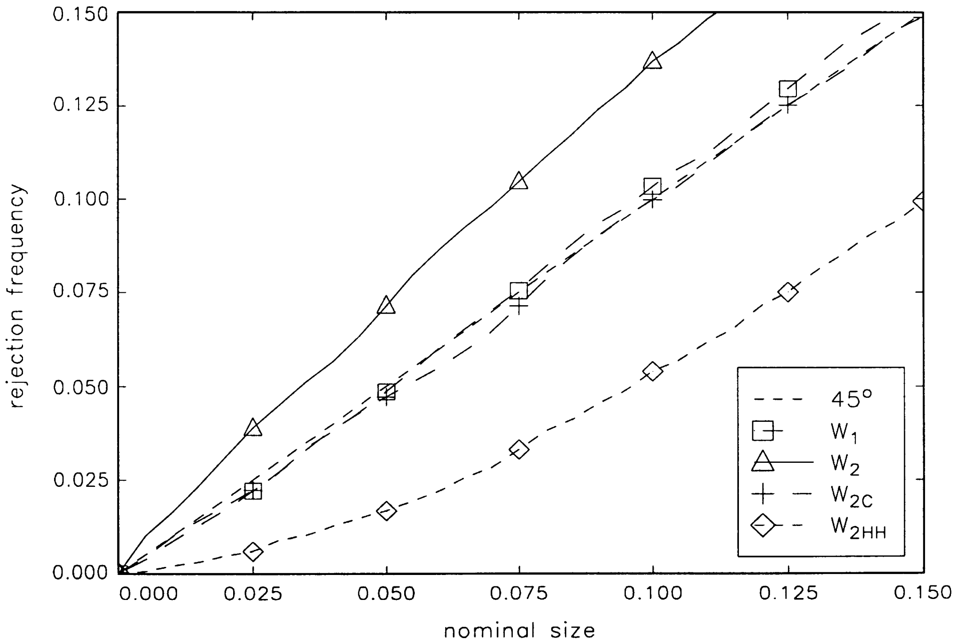

Fig. 1. p-Value plot, H_0: $\beta_0 = 1$, $T = 4$.

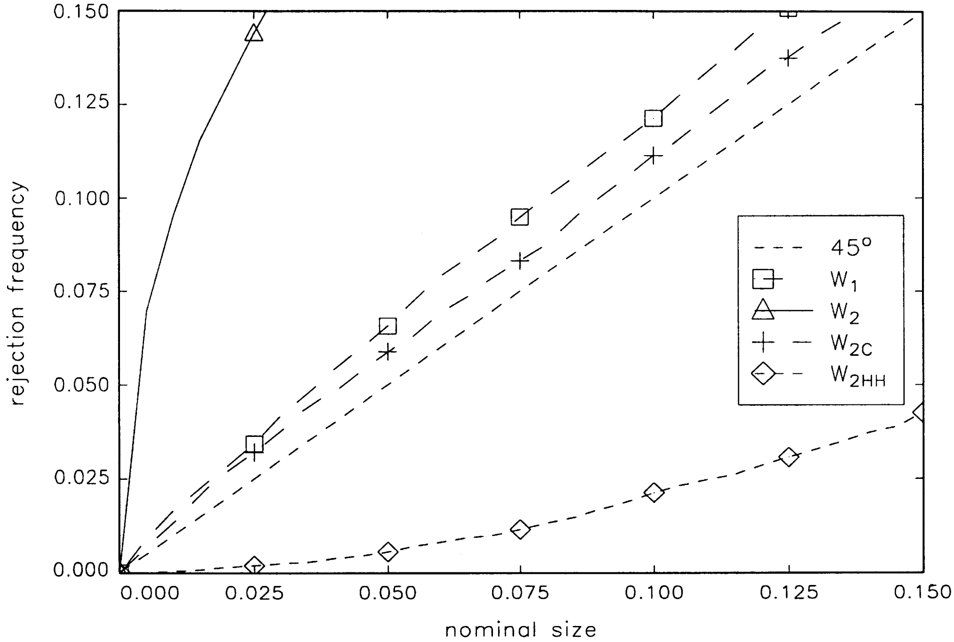

Fig. 2. p-Value plot, H_0: $\beta_0 = 1$, $T = 8$.

For $T=4$, W_2 is moderately oversized, whereas W_1 and W_{2C} have good size properties. Perhaps surprisingly, using the bootstrap method results in a test that is quite severely undersized. For $T=8$, W_2 is severely oversized. Using the corrected standard errors improves the size of the test dramatically and W_{2C} is only slightly oversized. W_1 is more oversized here than W_{2C} as the one-step estimator has a larger small sample bias. It is clear that using the corrected variance estimate for the two-step estimator improves the finite sample inference considerably. The Hall–Horowitz bootstrap procedure again results in a severely undersized test.[7] Similar results were found by Bond and Windmeijer (2002) in their analysis of various test procedures for dynamic panel data models estimated by GMM. They suggest that the reliability of this bootstrap procedure is related to how well the weight matrix estimates the covariance of the moment conditions, so that the performance of the bootstrap deteriorates with an increasing number of moment conditions. That there is a problem with estimating the covariance of the moment conditions in this case will also become apparent in Section 4.1 where the behaviour of the test of overidentifying restrictions is considered. The fragility of this bootstrap procedure is clearly an issue that merits further future investigation.

Due to the fact that the bias of the GMM estimator is related to the number of overidentifying restrictions, it could in practice be better not to use the full set of 28 instruments in the case of $T=8$, see e.g. Tauchen (1986) and Koenker and Machado (1999). For example, the instruments used in each of the first-differenced equations could be limited to (at most) x_{it-1} and x_{it-2}. In the Monte Carlo experiment, reducing the total number of instruments to 13 in this way, results in a mean of the estimated two-step parameters of 0.9886, thus decreasing the average bias of the two-step GMM estimator by 0.0076, or 40%. The empirical standard deviation for this estimator is 0.0774, which is an increase of about 7.5%. The mean of the usual estimated asymptotic standard errors is now 0.0644, which is much closer to the empirical standard deviation than for the GMM estimator that uses all the moment conditions. The mean of the finite sample corrected standard errors is 0.0775, almost identical to the empirical standard deviation.

Fig. 3 shows the p-value plot for the Wald statistics based on the GMM estimators that use this reduced instrument set. W_2 performs better than in the case of the full instrument set, but is still severely oversized. W_1 and W_{2C} have good size properties and they are less oversized than in the case of the full instrument set. Again, W_{2HH} is undersized.

4.1. The Sargan/Hansen test of overidentifying restrictions

The test statistic of overidentifying restrictions in the simple linear panel data model based on the two-step GMM estimator is given by

$$S_{W_N(\hat{\beta}_1)} = \frac{1}{N}\,\Delta\hat{u}_2' Z W_N^{-1}(\hat{\beta}_1) Z' \Delta\hat{u}_2,$$

[7] An alternative bootstrap procedure proposed by Brown and Newey (2002) results in almost identical size properties of the test in these experiments.

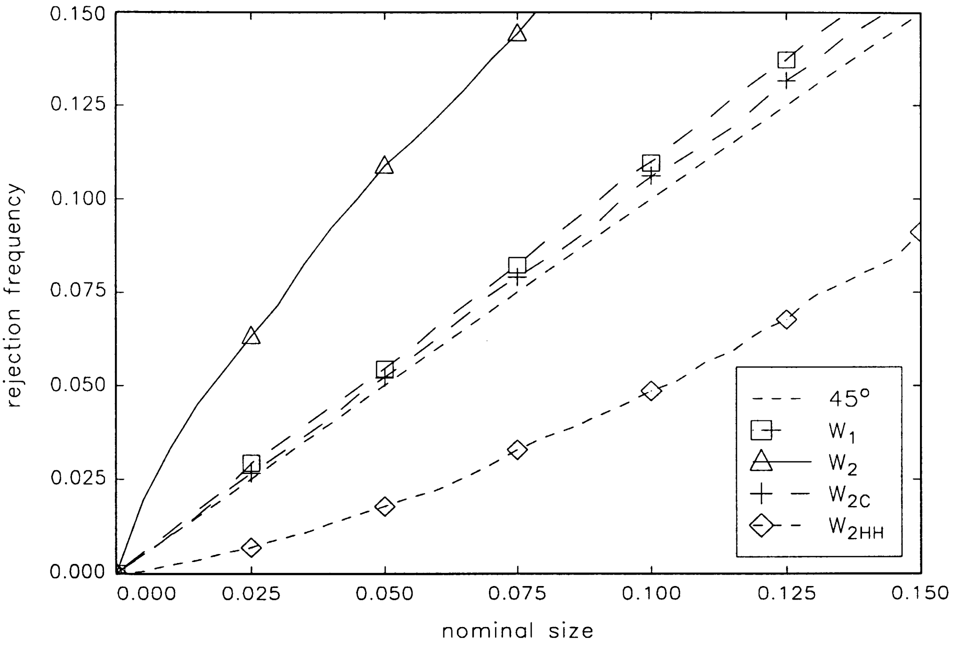

Fig. 3. p-Value plot, H_0: $\beta_0 = 1$, $T = 8$, reduced instrument set.

whereas the test statistic of overidentifying restrictions based on the infeasible GMM estimator $\hat{\beta}_{W_N(\beta_0)}$ is given by

$$S_{W_N(\beta_0)} = \frac{1}{N} \Delta\hat{u}_0' Z W_N^{-1}(\beta_0) Z' \Delta\hat{u}_0,$$

where $\Delta\hat{u}_0 = \Delta y - \hat{\beta}_{W_N(\beta_0)}\Delta x$. Under the null that the moment conditions are valid, $S_{W_N(\beta_0)}$ and $S_{W_N(\hat{\beta}_1)}$ both have a limiting χ^2_{q-k} distribution.

Figs. 4 and 5 depict the p-value plots for the Sargan/Hansen tests for overidentification based on the two-step and infeasible GMM estimators from the Monte Carlo experiments described in Section 4, for $T = 4$ and 8, respectively. In the figures $S_{W_N(\hat{\beta}_1)}$ is denoted SAR2 and $S_{W_N(\beta_0)}$ is denoted SAR0. It is clear that the two statistics have almost exactly the same size properties, and so the size performance of the Sargan/Hansen test based on the two-step GMM estimator is not affected by the estimation of $\hat{\beta}_1$ used to construct the weight matrix. For $T = 8$, when there are many overidentifying restrictions, both test statistics are severely undersized.[8] This is due to the fact that both $W_N(\beta_0)$ and $W_N(\hat{\beta}_1)$ are poor estimates of the covariance of the moment conditions in this case, which has a much more profound impact on

[8] This has also been documented by Bowsher (2002), who shows that the test of overidentifying restrictions becomes severely undersized with an increasing number of overidentifying moment conditions in autoregressive panel data models with normal errors. In contrast, Ziliak (1997) finds the test for overidentifying restrictions to be oversized within the context of a linear panel data setting with increasing number of instruments, using data from the PSID and bootstrap Monte Carlo methods.

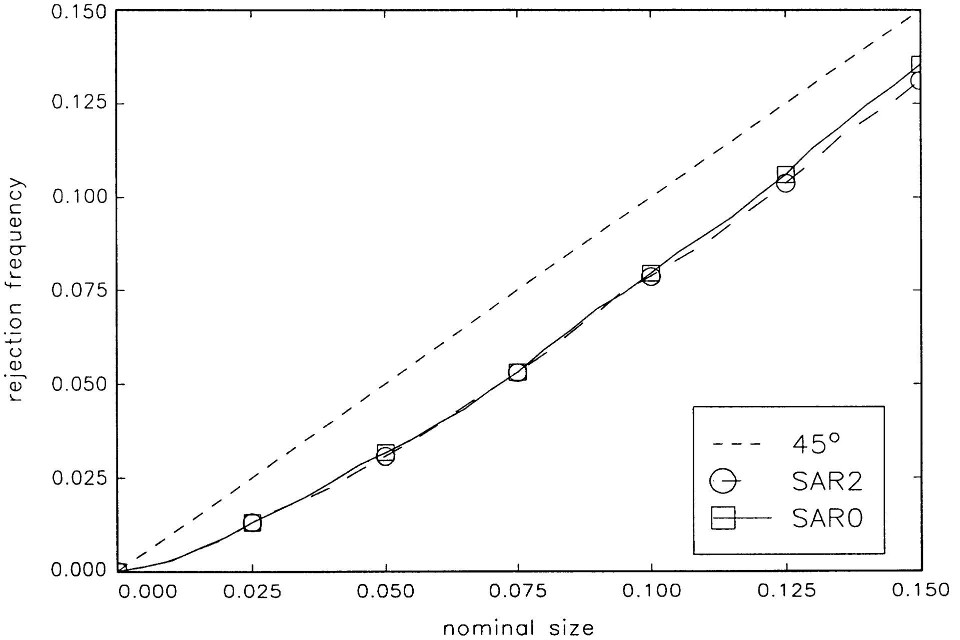

Fig. 4. p-Value plot, $T = 4$.

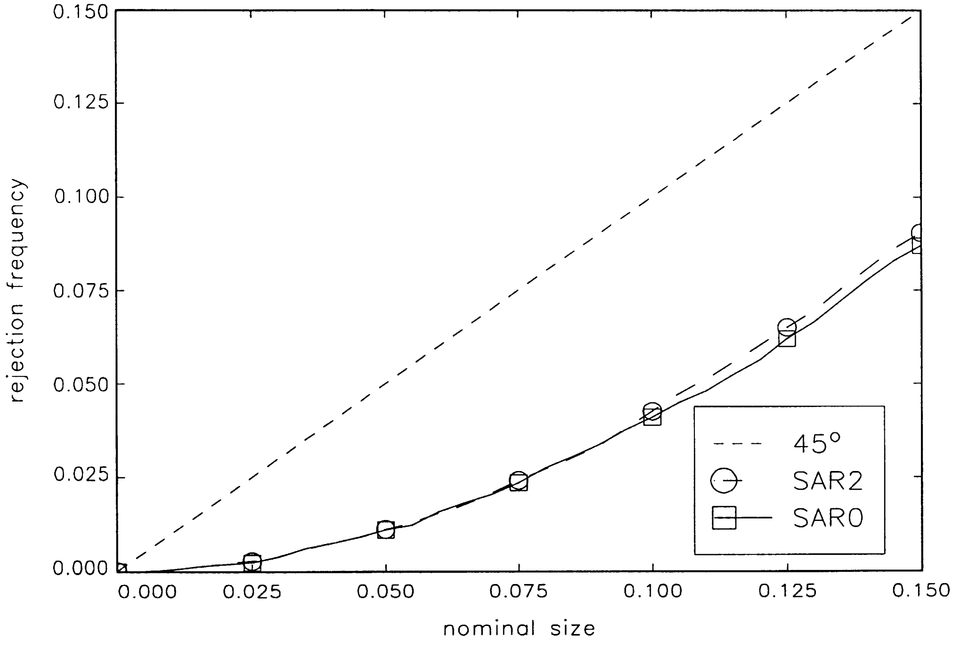

Fig. 5. p-Value plot, $T = 8$.

the distribution of the test of overidentifying restrictions than on the distribution of $\hat{\beta}_2$ itself, as is clear from Fig. 2.[9]

The relation between $S_{W_N(\hat{\beta}_1)}$ and $S_{W_N(\beta_0)}$ is given by

$$\begin{aligned}
S_{W_N(\hat{\beta}_1)} &= \Delta u' Z W_N^{-1}(\hat{\beta}_1) Z' \Delta u \\
&\quad - \Delta u' Z W_N^{-1}(\hat{\beta}_1) Z' \Delta x (\Delta x' Z W_N^{-1}(\hat{\beta}_1) Z' \Delta x)^{-1} \Delta x' Z W_N^{-1}(\hat{\beta}_1) Z' \Delta u \\
&= \Delta u' Z W_N^{-1}(\beta_0) Z' \Delta u \\
&\quad - \Delta u' Z W_N^{-1}(\beta_0) Z' \Delta x (\Delta x' Z W_N^{-1}(\beta_0) Z' \Delta x)^{-1} \Delta x' Z W_N^{-1}(\beta_0) Z' \Delta u \\
&\quad + P_{\beta_0, W_N(\beta_0)}(\hat{\beta}_1 - \beta_0) + \mathrm{o_p}(N^{-1/2}) \\
&= S_{W_N(\beta_0)} + P_{\beta_0, W_N(\beta_0)}(\hat{\beta}_1 - \beta_0) + \mathrm{o_p}(N^{-1/2}),
\end{aligned}$$

where $P_{\beta_0, W_N(\beta_0)}$ is given by

$$\begin{aligned}
& -\Delta u' Z W_N^{-1}(\beta_0) \frac{\partial W_N(\beta)}{\partial \beta} |_{\beta_0} W_N^{-1}(\beta_0) Z' \Delta u \\
& \quad + 2 \Delta u' Z W_N^{-1}(\beta_0) \frac{\partial W_N(\beta)}{\partial \beta} |_{\beta_0} W_N^{-1}(\beta_0) Z' \Delta x (\Delta x' Z W_N^{-1}(\beta_0) Z' \Delta x)^{-1} \\
& \quad \times \Delta x' Z W_N^{-1}(\beta_0) Z' \Delta u \\
& \quad - \Delta u' Z W_N^{-1}(\beta_0) Z' \Delta x (\Delta x' Z W_N^{-1}(\beta_0) Z' \Delta x)^{-1} \Delta x' Z W_N^{-1}(\beta_0) \\
& \quad \times \frac{\partial W_N(\beta)}{\partial \beta} |_{\beta_0} W_N^{-1}(\beta_0) Z' \Delta x \\
& \quad \times (\Delta x' Z W_N^{-1}(\beta_0) Z' \Delta x)^{-1} \Delta x' Z W_N^{-1}(\beta_0) Z' \Delta u.
\end{aligned}$$

$P_{\beta_0, W_N(\beta_0)}(\hat{\beta}_1 - \beta_0)$ is $\mathrm{O_p}(N^{-1/2})$. However, the terms tend to cancel each other out. For example, in the Monte Carlo simulations, the means of the three terms of $P_{\beta_0, W_N(\beta_0)}(\hat{\beta}_1 - \beta_0)$ are given by -0.346, 0.384 and -0.098 for $T = 4$, and -1.035, 0.958 and -0.067 for $T = 8$.

As the size properties of the test deteriorates with the number of overidentifying restrictions, Fig. 6 shows the p-value plots for the test statistics based on GMM estimators that use the reduced instrument set x_{it-1} and x_{it-2} per time period. As also noted by Bowsher (2002), the size properties of these tests improve upon those based on the full instrument set, but here the tests remain undersized. As before, the size properties of the two statistics are virtually identical.

[9] Asymmetries in the finite sample distribution of $Z' \Delta \hat{u}_2$ will also lead to poor properties of the Sargan/Hansen test.

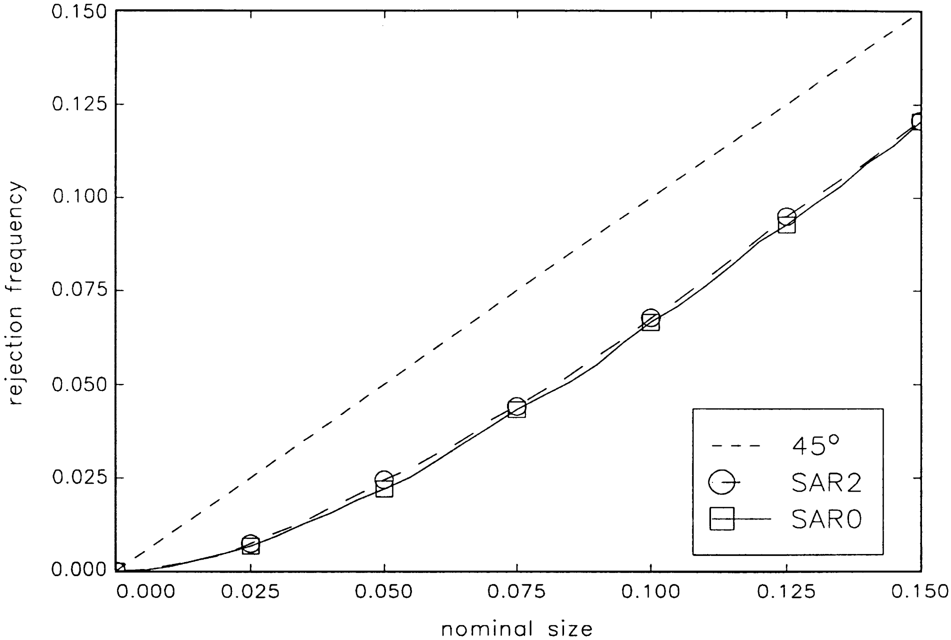

Fig. 6. p-Value plot, $T = 8$, reduced instrument set.

5. Empirical illustrations

In this section results of the two-step GMM variance correction are illustrated for two examples from the literature. The first example is taken from Arellano and Bond (1991), who used a sample of 140 UK quoted firms over the years 1976–1984. The sample is unbalanced with observations varying between 7 and 9 records per company. Arellano and Bond (1991) estimated dynamic employment equations, one of which was specified as

$$n_{it} = \alpha_1 n_{it-1} + \alpha_2 n_{it-2} + \beta w_{it} + \beta_1 w_{it-1} + \gamma k_{it} + \delta ys_{it} + \delta_1 ys_{it-1} + \lambda_t + \eta_i + u_{it},$$

where n_{it} is the logarithm of UK employment in company i at the end of period t, w_{it} is the log of the real product wage, k_{it} is the log of gross capital and ys_{it} is the log of industry output. The model is estimated in first differences, with an instrument set of the form

$$Z_i = \begin{bmatrix} 1 & n_{i1} & n_{i2} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \Delta x'_{i4} \\ 0 & 0 & 0 & 1 & n_{i1} & n_{i2} & n_{i3} & & 0 & 0 & & 0 & \Delta x'_{i5} \\ \vdots & \vdots & & & & & & \ddots & & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & n_{i1} & \cdots & n_{i7} & \Delta x'_{i9} \end{bmatrix},$$

Table 2
Estimation results for Arellano and Bond (1991) data

	One step		Two step		
	Coeff	Std err	Coeff	Std err	Std errc
n_{it-1}	0.5346	0.1664	0.4742	0.0853	0.1854
n_{it-2}	−0.0751	0.0680	−0.0523	0.0273	0.0517
w_{it}	−0.5916	0.1679	−0.5132	0.0493	0.1456
w_{it-1}	0.2915	0.1416	0.2246	0.0801	0.1420
k_{it}	0.3585	0.0538	0.2927	0.0395	0.0626
ys_{it}	0.5972	0.1719	0.6098	0.1085	0.1562
ys_{it-1}	−0.6117	0.2118	−0.4464	0.1248	0.2173
m_1		−2.493		−2.826	−1.999
m_2		−0.359		−0.327	−0.316
Wald		219.6		372.0	142.0

The dependent variable is n_{it}. No. of firms 140. No. of observations 611. Time dummies included. Std err are asymptotic standard errors, std errc are corrected for the estimation of $\hat{\beta}_1$ in the efficient weight matrix. m_1 and m_2 are N(0, 1) tests for first- and second-order serial correlation. Wald is a χ^2_7 test of joint significance of the coefficients.

where $\Delta x'_{it}=[\Delta w_{it}, \Delta w_{it-1}, \Delta k_{it}, \Delta ys_{it}, \Delta ys_{it-1}]$. There are a total of 25 overidentifying moment conditions in this model.

Table 2 presents estimation results for the one-step estimator, using the weight matrix $(1/N)\sum_{i=1}^{N} Z_i' H Z_i$, and the two-step estimator.[10] The two-step estimation results are identical to those presented in column (b) of Table 4 in Arellano and Bond (1991). Both the asymptotic standard errors and some tests based on the asymptotic variance, and the corrected versions of these, are reported.

The usual asymptotic standard errors for the two-step estimator are much smaller than the standard errors for the one-step estimator. However, the standard errors that adjust for the estimation of parameters used to construct the efficient weight matrix indicate that this perceived increase in precision is due to the downward bias of the usual estimates of the standard errors. The corrected standard errors are very similar to, and sometimes even larger than, those of the one-step estimator. Similarly, the corrected Wald test for joint significance of the reported parameters is much smaller than the test based on the usual asymptotic covariance matrix.

The next example uses data from Blundell and Bond (2000), who investigated estimation of production functions using the so-called system GMM estimator. A

[10] The estimation was performed using the DPD98 program for Gauss, see Arellano and Bond (1998). Standard error corrections were implemented in Gauss by the author and the adjusted programme is available upon request. DPD for Ox (Doornik et al., 2001), PcGive 10 (Doornik and Hendry, 2001) and XTABOND2 for STATA (Roodman, 2003) report the corrected standard errors.

Table 3
Estimation results for Blundell and Bond (2000) data

	One step		Two step		
	Coeff	Std err	Coeff	Std err	Std errc
First differences					
$(y-k)_{it-1}$	0.4600	0.0740	0.4146	0.0574	0.1000
$(n-k)_{it}$	0.5272	0.1024	0.5731	0.0698	0.0993
$(n-k)_{it-1}$	−0.2041	0.1086	−0.1607	0.0746	0.1158
m_1		−6.139		− 6.210	−4.711
m_2		−0.612		− 0.623	−0.583
Wald		129.5		236.02	120.1
System					
$(y-k)_{it-1}$	0.5618	0.0790	0.6292	0.0371	0.0759
$(n-k)_{it}$	0.5158	0.1009	0.5389	0.0598	0.0829
$(n-k)_{it-1}$	−0.2876	0.1169	−0.3155	0.0609	0.0946
m_1		−6.800		− 8.788	−7.737
m_2		−0.364		− 0.209	−0.202
Wald		416.4		1254.7	532.5

The dependent variable is $(y-k)_{it}$. No. of firms 509. No. of observations 2545. Time dummies included. Std err are asymptotic standard errors, std errc are corrected for the estimation of $\hat{\beta}_1$ in the efficient weight matrix. m_1 and m_2 are N(0, 1) tests for first- and second-order serial correlation. Wald is a χ^2_3 test of joint significance of the coefficients.

specification that was estimated is

$$(y-k)_{it} = \alpha(y-k)_{it-1} + \beta(n-k)_{it} + \gamma(n-k)_{it-1} + \delta_t + u_{it},$$
$$u_{it} = \eta_i + v_{it}, \tag{5.1}$$

where y_{it}, n_{it} and k_{it} are the logs of sales, employment and capital stock of firm i in year t, respectively. This specification accommodates first-order autocorrelation in productivity shocks and imposes constant returns to scale. The data used are a balanced panel of 509 R&D-performing US manufacturing companies observed for 8 years, 1982–1989, similar to that used in Mairesse and Hall (1996).

Table 3 presents estimation results for both the first differenced and system GMM estimators. The first differenced estimator in this case uses the $3(T-2)(T-3)/2+(T-3)$ moment conditions

$$\mathrm{E}((1, x_i^{t-3})\Delta u_{it}) = 0, \quad t = 4, \ldots, T,$$

where $x_i^{t-3} = (x_{i1}, \ldots, x_{it-3})$ and $x_{is} = (y_{is}, n_{is}, k_{is})$.

The system GMM estimator uses the $3(T-2)(T-3)/2$ moment conditions for the differenced equations

$$\mathrm{E}(x_i^{t-3}\Delta u_{it}) = 0, \quad t = 4, \ldots, T$$

plus $4(T-3)$ moment conditions for the levels equations

$$\mathrm{E}((1, \Delta x_{it-2})u_{it}) = 0, \quad t = 4, \ldots, T. \tag{5.2}$$

The additional $3(T-3)$ moment conditions $\mathrm{E}(u_{it}\Delta x_{it-2}) = 0$ are valid under the additional mean stationarity assumption on initial conditions $\mathrm{E}(\eta_i \Delta x_{it}) = 0$, see Arellano and Bover (1995) and Blundell and Bond (1998). When the data are persistent, and the instruments potentially weak in the first-differenced equations, Blundell and Bond (1998) show that the additional moment conditions (5.2) remain informative, resulting in estimates that have a much smaller finite sample bias and are also more efficient. The gain in precision from using the two-step GMM estimator rather than the one-step GMM estimator is also likely to be greater in this case, since there is no feasible one-step weight matrix that yields an asymptotically equivalent estimator to two-step GMM, even in the case of i.i.d. disturbances. The two-step system GMM estimator is therefore the preferred estimator in terms of mean squared error, see Blundell and Bond (1998, 2000) and Blundell et al. (2000) for Monte Carlo simulation evidence.

The one-step estimation results presented in Table 3 are identical to those presented in columns 3 and 4 of Table 6 in Blundell and Bond (2000). For the first-differenced GMM estimator there are 42 overidentifying moment conditions, whereas there are 57 overidentifying moment conditions for the system GMM estimator. Although the number of firms is quite large, again the corrected standard errors of the two-step differenced GMM estimator are much larger than the uncorrected ones. The one-step standard errors are actually smaller than the corrected two-step standard errors for two of the three coefficients. For the system two-step GMM estimator, again the corrected standard errors are larger than the uncorrected ones, but here they are smaller than the corresponding one-step standard errors. So, as expected, in this case there does appear to be a genuine gain in precision from using the efficient weight matrix. The corrected standard errors for the two-step system estimator are further smaller than those for the two-step differenced GMM estimator, indicating a gain in precision from using the additional moment conditions arising from mean stationarity. The estimated coefficient on the lagged dependent variable for the two-step system estimator is about 50% larger than that for two-step differenced estimator, an indication that the system GMM estimator corrects a downward bias of the differenced GMM estimator, due to the use of a more informative set of instruments.

6. Power

The correction to the estimated standard errors for the two-step GMM estimator as developed in this paper has been shown to improve inference considerably in terms of size of the Wald test. One advantage of this correction is therefore that use of it will guard against the reporting of the hugely inflated values of the two-step Wald test based on the usual asymptotic standard errors. Another advantage is that as the two-step GMM estimator is expected to be more efficient also in finite samples, the Wald test based on the two-step estimation results using the variance correction can have more power than the one-step Wald test. Figs. 7a and b show size and power properties of

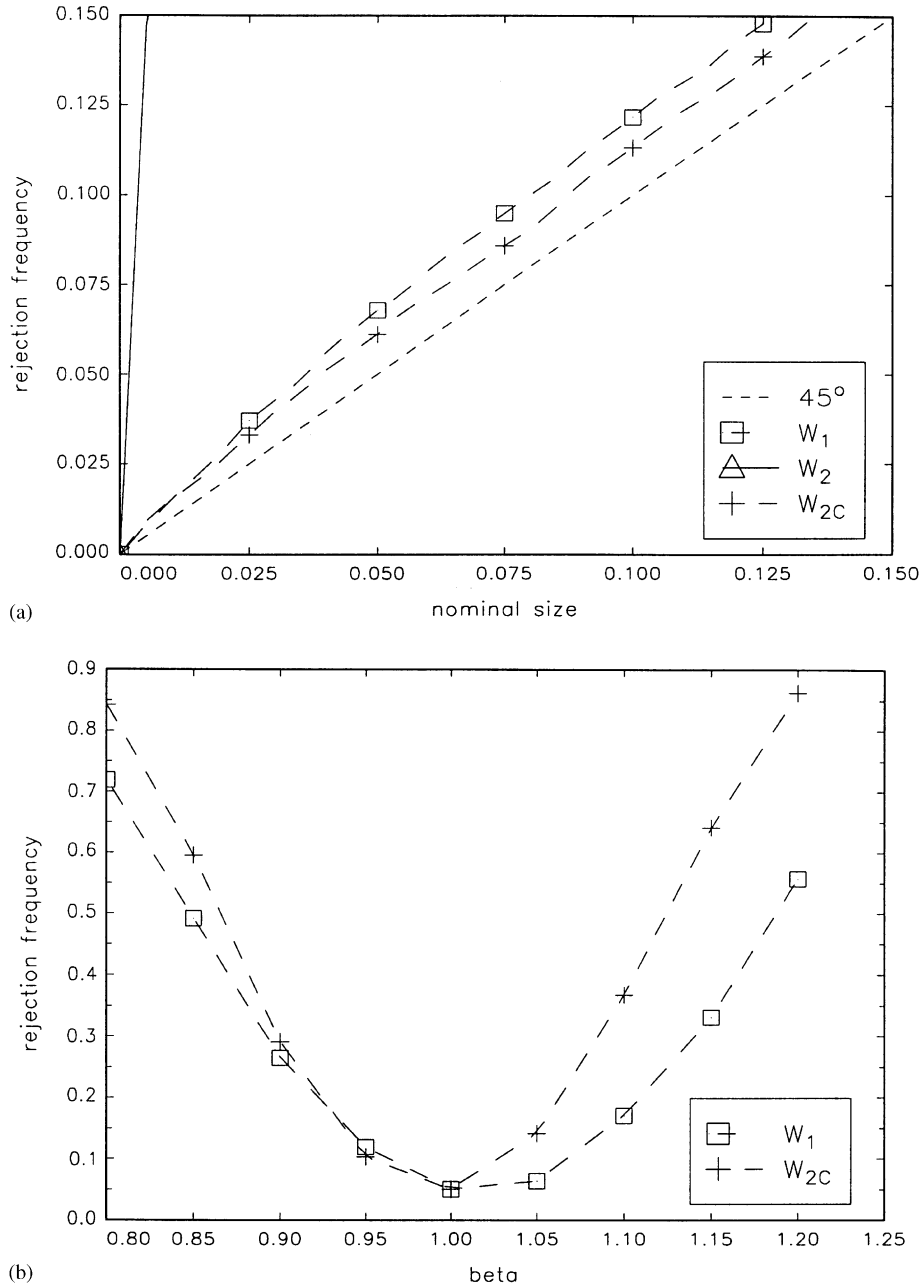

Fig. 7. (a) p-Value plot, system estimator, $T = 8$. (b) Power, system estimator, $T = 8$.

W_1 and W_{2C} for the system GMM estimator using the same design as in Section 4, for $T = 8$. The power of the tests is calculated for values of $\beta = 0.8, 0.85, \ldots, 1.20$, using as critical values the 95th percentiles of the distribution of the test statistics when $\beta = 1$. Again, all results are based on 10,000 replications. The figures show that W_{2C} has almost everywhere better power than W_1 with considerable larger rejection frequencies for values of β larger than 1. W_{2C} also has slightly better size properties in this case, as shown in Fig. 7a.

7. Bias and/or asymmetry

The Monte Carlo simulations presented so far have shown that the corrected two-step GMM variance estimate approximates the finite sample variance well. The resulting Wald test statistics also performed well, due to the fact that the finite sample distribution of the two-step estimator was symmetric, close to normal and centered around the true parameter. Obviously, use of the corrected variance estimate will not result in well-behaved Wald test statistics when the estimator is biased and/or when its finite sample distribution is non-normal, especially when it is asymmetric. As an example of the latter, Table 4 presents results for the system GMM estimator using the same design as in Section 4, with ρ increased from 0.5 to 0.6.

The means of the corrected standard errors are again very close to the standard deviation of the two-step GMM estimates, which are considerably more efficient than the one-step estimates. For $T = 4$, the two-step estimator is upward biased by about 3%, and the ratio of bias to standard deviation is not very dissimilar from the results in Table 1 for $T = 8$. The corrected Wald test is therefore expected to overreject somewhat. However, the finite sample distribution of the estimator is skewed to the left for this design, as displayed in Fig. 8a, whereas the finite sample distribution of the corrected t-statistic is skewed to the right as shown in Fig. 8b. Because of this, the performance of the corrected Wald test is quite poor in this case when $T = 4$ as shown in Fig. 9a. The test overrejects the null considerably, more so than the one-step Wald test which

Table 4
Monte Carlo results for system GMM

	$T = 4$	$T = 8$
$\hat{\beta}_1$	1.0161	0.9829
$\mathrm{sd}\hat{\beta}_1$	0.1163	0.0891
$\mathrm{se}\hat{\beta}_1$	0.1125	0.0846
$\hat{\beta}_2$	1.0308	1.0086
$\mathrm{sd}\hat{\beta}_2$	0.0977	0.0675
$\mathrm{se}\hat{\beta}_2$	0.0793	0.0349
$\mathrm{sec}\hat{\beta}_2$	0.0961	0.0657

Notes: $N = 100$, $\beta_0 = 1$, $\rho = 0.6$, means and standard deviations of 10,000 replications. $\mathrm{sec}\hat{\beta}_2$ is the finite sample corrected estimated standard error of $\hat{\beta}_2$.

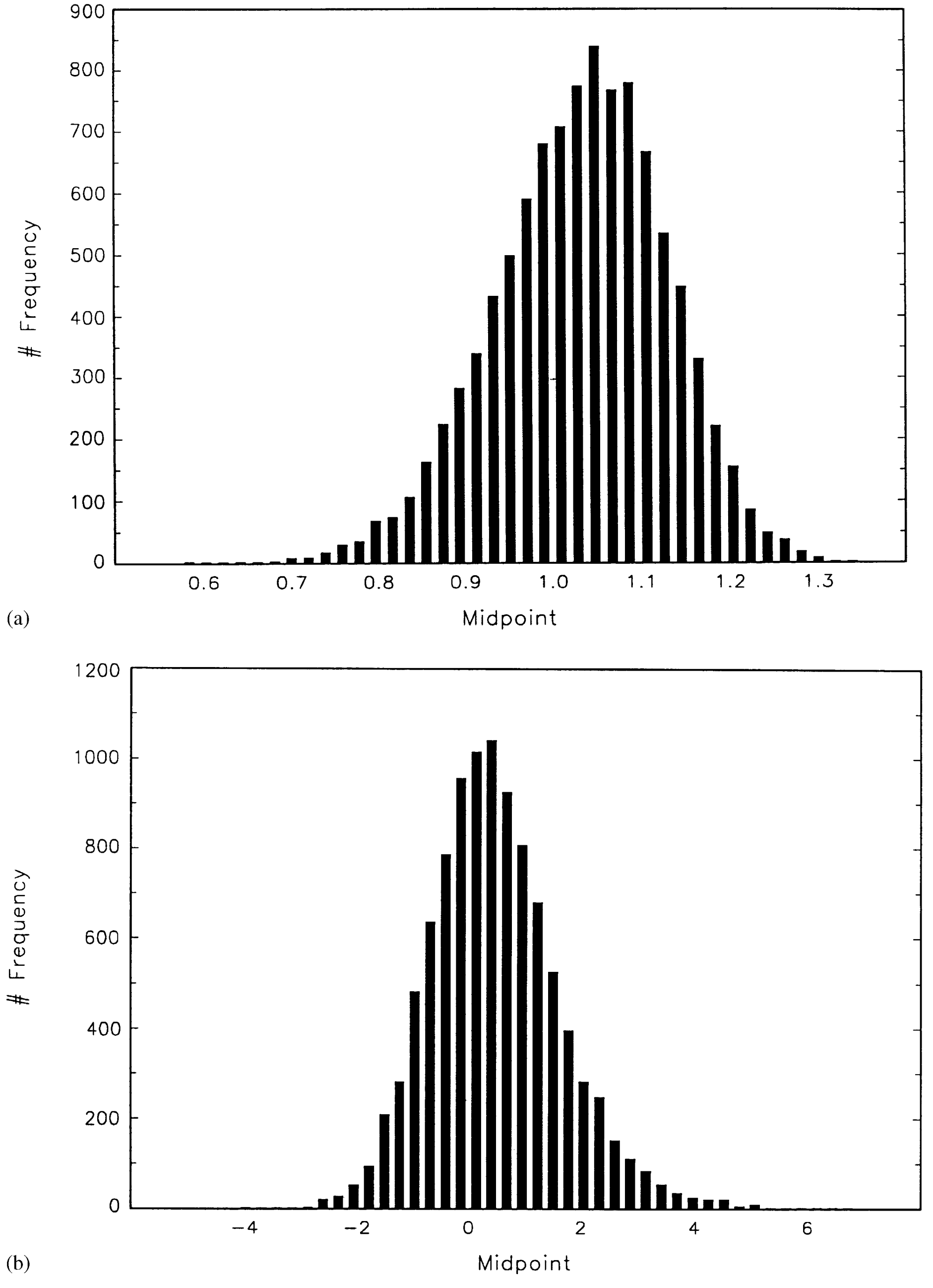

Fig. 8. (a) Two-step system GMM estimator, $T = 4$. (b) Corrected two-step t-statistic, $T = 4$.

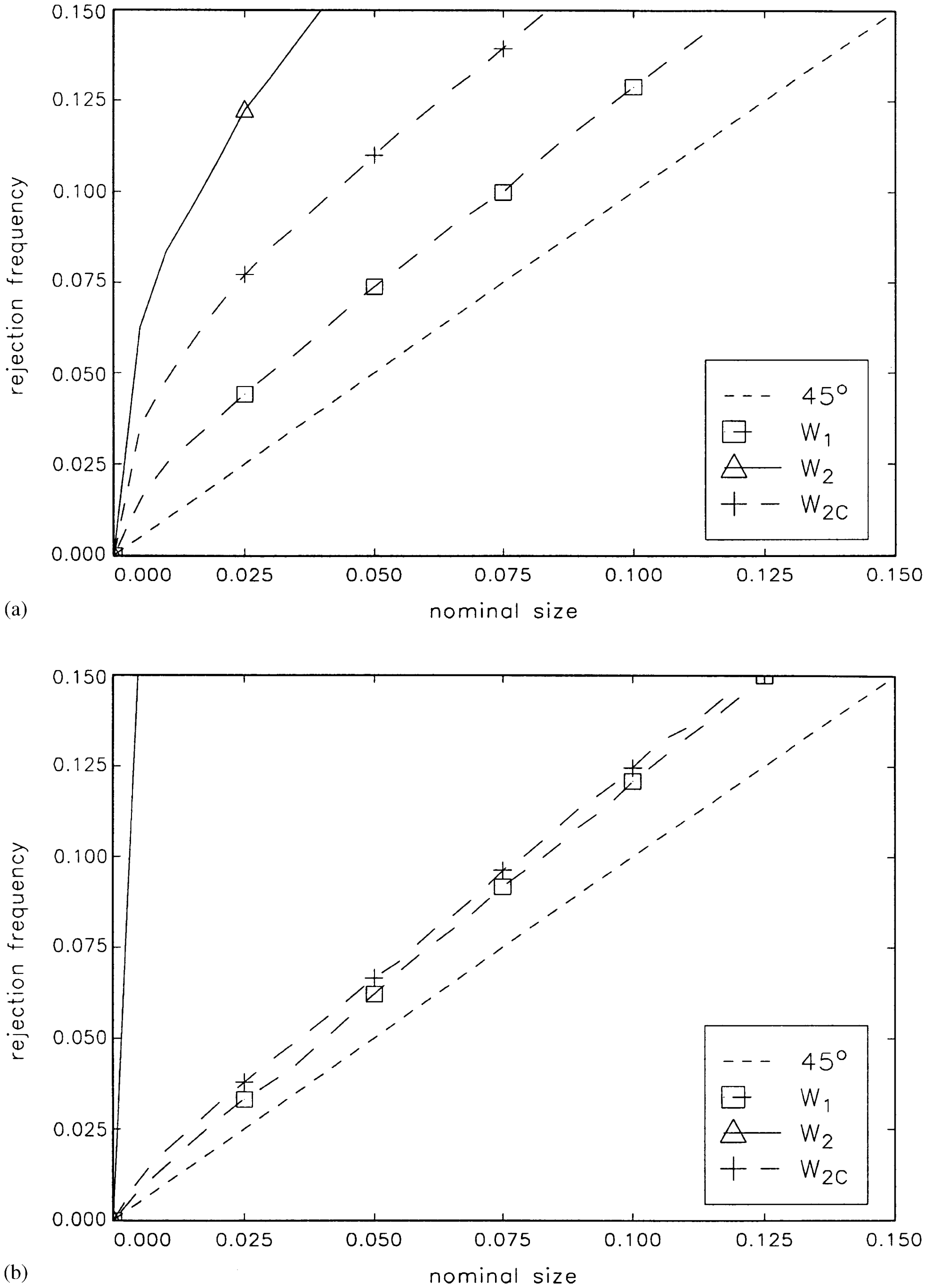

Fig. 9. (a) p-Value plot, H_0: $\beta_0 = 1$, $T = 4$. (b) p-Value plot, H_0: $\beta_0 = 1$, $T = 8$.

has a more symmetric finite sample distribution, although it still considerably improves upon the performance of the usual asymptotic Wald test. These results indicate that one should be cautious when applying the Wald test in small samples even when using the corrected variance estimate. When $T = 8$, the Wald test based on the corrected variance estimates is much better behaved, as shown in Fig. 9b, as there is almost no bias and the finite sample distribution (not displayed here) is again very close to being symmetric.

8. Conclusions

This paper has shown that the commonly found small sample downward bias of the estimated asymptotic standard errors of the efficient two-step GMM estimator in linear models can be attributed to the fact that the usual asymptotic standard errors do not take account of the extra variation in small samples which is due to the use of estimated parameters in constructing the efficient weight matrix. A simple first order Taylor series expansion generates an extra term that accounts for the estimation of these parameters. This correction term vanishes with increasing sample size, but provides a more accurate approximation in finite samples when all the moment conditions are linear. This extra term can be estimated and in a Monte Carlo study of a panel data model it is shown that this feasible corrected estimate of the variance is close to the finite sample variance of the efficient two-step GMM estimator.

The Monte Carlo results further show that the conventional asymptotic variance estimate of the two-step GMM estimator is a good estimate of the variance of an infeasible GMM estimator that uses the true values of the parameters to calculate the efficient weight matrix. The difference between the variances of the infeasible and feasible two-step GMM estimators can be quite large in finite samples. The estimated corrected variance of the two-step GMM estimator captures this difference well, and results in more accurate inference.

Acknowledgements

I am grateful to two anonymous referees, Stephen Bond, Manuel Arellano, Richard Blundell, Chris Skeels and Richard Smith for helpful comments. Financial support of the ESRC, grant no. 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

Altonji, J.G., Segal, L.M., 1996. Small sample bias in GMM estimation of covariance structures. Journal of Business and Economic Statistics 14, 353–366.

Arellano, M., Bond, S.R., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58, 277–298.

Arellano, M., Bond, S.R., 1998. Dynamic panel data estimation using DPD98 for GAUSS. http://www.ifs.org.uk/staff/steve_b.shtml.

Arellano, M., Bover, O., 1995. Another look at the instrumental-variable estimation of error-components models. Journal of Econometrics 68, 29–52.

Blundell, R.W., Bond, S.R., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115–143.

Blundell, R.W., Bond, S.R., 2000. GMM estimation with persistent panel data: an application to production functions. Econometric Reviews 19, 321–340.

Blundell, R.W., Bond, S.R., Windmeijer, F., 2000. Estimation in dynamic panel data models: improving on the performance of the standard GMM estimator. In: Baltagi, B. (Ed.), Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Advances in Econometrics, Vol. 15. JAI Press, Elsevier Science, Amsterdam, pp. 53–91.

Bond, S.R., Windmeijer, F., 2002. Finite sample inference for GMM estimators in linear panel data models. CEMMAP Working Paper No. CWP04/02, Institute for Fiscal Studies, London.

Bound, J., Jaeger, D.A., Baker, R.M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American Statistical Association 90, 443–450.

Bowsher, C.G., 2002. On testing overidentifying restrictions in dynamic panel data models. Economics Letters 77, 211–220.

Brown, B.W., Newey, W.K., 2002. Generalized method of moments, efficient bootstrapping, and improved inference. Journal of Business and Economic Statistics 20, 507–517.

Davidson, R., MacKinnon, J.G., 1996. Graphical methods for investigating the size and power of hypothesis tests. Manchester School 66, 1–26.

Doornik, J.A., Hendry, D.F., 2001. Econometric Modelling Using PcGive 10, Vol. III. Timberlake Consultants Ltd, London.

Doornik, J.A., Arellano, M., Bond, S., 2001. Panel Data Estimation using DPD for Ox. Nuffield College, Oxford.

Hahn, J., Hausman, J., Kuersteiner, G., 2001. Bias corrected instrumental variables estimation for dynamic panel data models with fixed effects. Mimeo, MIT.

Hall, P., Horowitz, J.L., 1996. Bootstrap critical values for tests based on generalized method of moments estimators. Econometrica 64, 891–916.

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

Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14, 262–280.

Imbens, G.W., Spady, R.H., Johnson, P., 1998. Information theoretic approaches to inference in moment condition models. Econometrica 66, 333–357.

Koenker, R., Machado, J.A.F., 1999. GMM inference when the number of moment conditions is large. Journal of Econometrics 93, 327–344.

Koenker, R., Machado, J.A.F., Skeels, C.L., Welsh, A.H., 1994. Momentary lapses: moment expansions and the robustness of minimum distance estimation. Econometric Theory 10, 172–197.

Magnus, J.R., Neudecker, H., 1988. Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, Chichester.

Mairesse, J., Hall, B.H., 1996. Estimating the productivity of research and development in French and US manufacturing firms: an exploration of simultaneity issues with GMM methods. In: Wagner, K., Van Ark, B. (Eds.), International Productivity Differences and Their Explanations. Elsevier, Amsterdam, pp. 285–315.

Murphy, K.M., Topel, R.H., 1985. Estimation and inference in two-step econometric models. Journal of Business and Economic Statistics 3, 370–379.

Newey, W.K., Smith, R.J., 2000. Asymptotic bias and equivalence of GMM and GEL estimators. Mimeo, MIT.

Pagan, A.R., 1984. Econometric issues in the analysis of regressions with generated regressors. International Economic Review 25, 221–247.

Roodman, D., 2003. XTABOND2, STATA Module to Estimate Dynamic Panel Data Models. Center for Global Development, Washington.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.

Tauchen, G., 1986. Statistical properties of generalized method-of moments estimators of structural parameters obtained from financial market data. Journal of Business and Economic Statistics 4, 397–416.

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.

Ziliak, J.P., 1997. Efficient estimation with panel data when instruments are predetermined: an empirical comparison of moment-condition estimators. Journal of Business and Economic Statistics 15, 419–431.