Computational Statistics and Data Analysis 100 (2016) 734–762

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Approximating and reducing bias in 2SLS estimation of dynamic simultaneous equation models

Garry D.A. Phillips [a], Gareth Liu-Evans [b,*]

[a] Cardiff Business School, Cardiff University, Cardiff, CF10 3EU, UK

[b] Management School, University of Liverpool, Liverpool, L69 7ZH, UK

ARTICLE INFO

Article history:
Received 13 May 2015
Received in revised form 13 November 2015
Accepted 20 November 2015
Available online 2 December 2015

Keywords:
2SLS
Bias correction
Simultaneous equation model
Asymptotic approximation
Dynamic
Weak instruments

ABSTRACT

An order O(1/T) approximation is made to the bias in 2SLS estimation of a dynamic simultaneous equation model, building on similar large-T moment approximations for non-dynamic models. The expression is long because it contains two distinct parts: a part due to the simultaneity which is directly related to the Nagar bias and a part due to the dynamics which has many component terms. However, the analytically corrected 2SLS estimators resulting from this approximation perform well in terms of remaining estimation bias. The biases of these estimators are compared with the Quenouille half-sample jackknife and the residual bootstrap for 2SLS in dynamic models, and are found to be competitive. The Monte Carlo and bias approximation also suggest that the bias in estimating endogenous variable coefficients in dynamic simultaneous equation models is non monotonic in the sample size, contrary to the well known theoretical result for static models. The effect of using weaker instruments on our numerical and Monte Carlo results is explored.

1. Introduction

The issues of bias approximation and reduction have been previously addressed in relation to static simultaneous equation models. Recent examples of bias approximation are Phillips (2000), Hahn and Hausman (2002), Hahn et al. (2004), Phillips (2007), Iglesias and Phillips (2010), and Bun and Windmeijer (2011). On bias reduction see MacKinnon and Davidson (2006), Dahlberg and Blomquist (2006), Davidson and MacKinnon (2007) and Ackerberg and Devereux (2009), who consider the JIVE method and its variants, see also Phillips and Hale (1977), Iglesias and Phillips (2012) who construct estimators that are unbiased up to orders $O(T^{-1})$ and $O(T^{-2})$, where T is the sample size, and Hsu et al. (1986), who assess the bootstrap method due to Freedman (1984) and the standard delete-1 jackknife for static models. The Freedman (1984) method is asymptotically valid in the dynamic setting, and performs well in Ip (1991) for dynamic models. Freedman and Peters (1984a,b) use the method to obtain bootstrap estimates of the bias in GLS and 3SLS coefficient estimators, respectively. Freedman and Peters (1984a) also conduct a Monte Carlo simulation study to assess the performance of the bootstrap in estimating standard errors, and MacKinnon (2002) presents Monte Carlo evidence for its use in hypothesis testing in static models. Also in the context of dynamic models, Kiviet and Phillips (1995) present a small-σ approximation to the 2SLS coefficient bias, where, following Kadane (1971), σ is a small scalar multiple of the variance of the structural equation disturbance, and examine its use in bias reduction, showing that certain results for the static model do not carry over to the dynamic case.

* Corresponding author.
E-mail address: gareth.liu-evans@liverpool.ac.uk (G. Liu-Evans).

http://dx.doi.org/10.1016/j.csda.2015.11.011
0167-9473/

Given a sample size T and an estimate $\hat{\alpha}$ of a coefficient vector α, the large-T approach in Nagar (1959) starts by expanding the estimation error as follows:

$$\sqrt{T}(\hat{\alpha}-\alpha)=\sum_{s=1}^{p}\frac{e_s}{T^{\frac{1}{2}(s-1)}}+\frac{r_p}{T^{\frac{1}{2}p}}, \quad (1)$$

where e_s, for $s=1,\ldots,p$, and r_p are all $O_p(1)$ as $T\to\infty$. The last term is the remainder in an expansion of $\sqrt{T}(\hat{\alpha}-\alpha)$ to order $O_p(T^{-\frac{1}{2}(p-1)})$. In the small-$\sigma$ approach the general expansion is

$$\frac{1}{\sigma}(\hat{\alpha}-\alpha)=\sum_{s-1}^{p}\sigma^{s-1}\dot{e}_p+\sigma^p\dot{r}_p, \quad (2)$$

where $\dot{e}_s$, for $s=1,\ldots,p$, and $\dot{r}_p$ are also bounded in probability, this time as σ, the standard deviation of the equation disturbance, tends to zero. The bias is then approximated to order $O(T^{-1})$ or $O(\sigma^2)$ by calculating the first moment of the approximate estimation error in each case.

Kadane (1971) shows that the large-T and small-σ approaches yield essentially equivalent results for the static SEM. In particular, it is shown that the large-T result in Nagar (1959) can be obtained by taking the limit of the small-σ result as $T\to\infty$. Kiviet and Phillips (1989) and Kiviet and Phillips (1993) show that the same is not true in dynamic settings.

A large-T moment approximation for a dynamic simultaneous equation model ("DSEM") is presented here under a Normality assumption, building on the above results for static models and on the small-σ approximations for dynamic models, see also Phillips and Liu-Evans (2009). The simulation experiments in Section 3 investigate the remaining bias and the mean squared error after using this for bias reduction. The performance of the analytically corrected estimator, C2SLS, is compared with the bootstrap method due to Freedman (1984) and the half-sample jackknife in Quenouille (1956). Though Freedman (1984) provides a consistency result for the bootstrap in DSEMs, there is no theoretical result for bootstrap bias correction in this context, though the favourable simulation results in Hsu et al. (1986) for bias-corrected estimation of 2SLS estimation of static models suggest that a correction is likely. Finally, the behaviour of the bias correction numerically is explored as the instruments grow weak, and the three bias correction methods are compared in a situation where the instruments are relatively weak.

The jackknife method considered is due to Quenouille (1956). Dhaene and Jochmans (2010) find that it performs well in terms of bias correction in large-T dynamic panel data modelling with fixed effects. It is referred to as the Quenouille jackknife (QJ) here. Rather than creating subsamples by deleting one observation at a time for each subsample, two subsamples are obtained from the first and second halves of the whole sample with the ordering intact. This has the benefit of retaining the dynamics of the data, and it means that the 2SLS bias does not need to be monotonically decreasing in the sample size for a bias correction to occur. The related delete-d jackknife in Shao (1989) can be applied with $d=\lceil T/2\rceil$, but it does not retain the dynamics and will not work here.

2. The model and bias approximation

The complete system is assumed to be as follows:

$$YB+Y_{-1}\Lambda+XC=\bar{U}, \quad (3)$$

where Y is a $T\times G$ matrix of observations on G endogenous variables, Y_{-1} is a $T\times G$ matrix of observations on the endogenous variables lagged one time period, X is a $T\times K$ matrix of observations on K stationary exogenous variables and $\bar{U}$ is a $T\times G$ matrix of structural disturbances. The matrices B, Λ and C are of dimension $G\times G$, $G\times G$ and $K\times G$, respectively, while B is assumed to be non-singular. The rows of $\bar{U}$ are assumed to be normally and independently distributed with zero mean and fixed covariance matrix Σ.

The reduced form of the model is

$$\begin{aligned}Y&=-Y_{-1}\Lambda B^{-1}-XCB^{-1}+\bar{U}B^{-1}\\&=Y_{-1}\Gamma+X\Pi+\bar{V},\end{aligned} \quad (4)$$

where $\Gamma=-\Lambda B^{-1}$, $\Pi=-CB^{-1}$ and $\bar{V}=\bar{U}B^{-1}$. Here the rows of $\bar{V}$ are normally distributed with zero mean and covariance matrix $\Omega=(B^{-1})'\Sigma B^{-1}$, and as a stationarity condition it is assumed that the eigenvalues of Γ are inside the unit circle.

It will be assumed that the rows of the Y matrix are generated from a fixed value $Y_{0,.}$ at time $t=0$ so that by successive substitution the matrix may be separated into stochastic and non-stochastic parts. This is done by noting that the $t-th$ row of Y may be written as

$$y_{t,.}=y_{0,.}\Gamma^t+\sum_{i=1}^{t}X_{i,.}\Pi\Gamma^{t-i}+\sum_{i=1}^{t}\bar{V}_{i,.}\Gamma^{t-i}, \quad (5)$$