GMM Estimation with Noncausal Instruments

Markku Lanne and Pentti Saikkonen

University of Helsinki, HECER

September 2009

Helsinki Center of Economic Research

Discussion Papers

GMM Estimation with Noncausal Instruments

Markku Lanne
University of Helsinki, RUESG and HECER

and

Pentti Saikkonen
University of Helsinki, RUESG and HECER

Discussion Paper No. 274
September 2009

ISSN 1795-0562

HECER – Helsinki Center of Economic Research, P.O. Box 17 (Arkadiankatu 7), FI-00014 University of Helsinki, FINLAND, Tel +358-9-191-28780, Fax +358-9-191-28781,
E-mail info-hecer@helsinki.fi, Internet www.hecer.fi

HECER
Discussion Paper No. 274

GMM Estimation with Noncausal Instruments*

Abstract

Lagged variables are often used as instruments when the generalized method of moments (GMM) is applied to time series data. We show that if these variables follow noncausal autoregressive processes, their lags are not valid instruments and the GMM estimator is inconsistent. Moreover, in this case, endogeneity of the instruments may not be revealed by the J-test of overidentifying restrictions that may be inconsistent and, as shown by simulations, its finite-sample power is, in general, low. Although our explicit results pertain to a simple linear regression, they can be easily generalized. Our empirical results indicate that noncausality is quite common among economic variables, making these problems highly relevant.

JEL Classification: C12, C22, C51

Keywords: Noncausal autoregression, instrumental variables, test of overidentifying restrictions

Markku Lanne

Department of Economics,
P.O. Box 17 (Arkadiankatu 7)
FI-00014 University of Helsinki
FINLAND

e-mail: markku.lanne@helsinki.fi

Pentti Saikkonen

Department of Mathematics and Statistics
P.O. Box 68 (Gustaf Hällströmn katu 2b)
FI-00014 University of Helsinki
FINLAND

e-mail: pentti.saikkonen@helsinki.fi

* We thank Joose Sauli for excellent research assistance. Financial support from the Academy of Finland and the Okobank Group Research Foundation is gratefully acknowledged.

1 Introduction

The generalized method of moments (GMM) is widely used in different fields of economics, including macroeconomics and finance. Among other things, its popularity presumably follows from the development of more and more complicated theoretical models which would in practice be impossible to take to data by alternative methods, such as the method of maximum likelihdood (ML). Even if ML estimation were possible, the GMM may be considered more robust in that it allows the researcher to concentrate on the central implications of the theory without the need to specify an empirical model in detail. In their survey, Hansen and West (2002) list the three most common uses of the GMM in economics: estimation of *a* first-order condition or *a* decision rule from dynamic optimization problem, examination of forecasting ability of survey data or of a financial variable, and setups with efficiency gains from the use of many moments. The first two of these are ubiquitous in the empirical analysis of asset pricing models, while all of them pertain to macroeconomic applications.

For instrumental variable methods to be applicable, a sufficiently large number of instruments are needed that satisfy the relevance and exogeneity requirements. The former has received more attention in the burgeoning weak instrument literature (see, e.g., Stock, Wright and Yogo, 2002), while it has been thought that the exogeneity of candidate instruments can reliably be determined by tests such as Hansen's (1982) J-test of overidentifying restrictions. Moreover, in applications using time series data, lagged values of economic variables, especially those included in the model, have been considered natural instruments that should be predetermined by construction. Provided the dynamics of such instruments can be described by causal autoregressive (AR) processes, the exogeneity requirement is indeed satisfied. However, while economic variables typically can be adequately modeled as AR processes, noncausality seems to be quite common among them (see Section 2.3) and, as we argue in this paper, in that case lags are not, in general, valid instruments. The difference between these two types of AR processess is that a causal AR process only depends on the past, whereas a noncausal AR process allows for dependence on the future.

Our theoretical (asymptotic) results pertain to the simple special case of univariate linear regression with a conditionally homoskedastic error term. In addition, we report results on simulation experiments to illustrate the finite-sample behavior of the GMM estimator and the J-test in the presence of noncausal instruments. The GMM estimator is shown to be inconsistent in our simple setup, and the simulations show that the biases of the ordinary least squares (OLS) estimator and the GMM estimator are very close to each other, especially in the case where the instruments follow purely noncausal AR processes. We also show that Hansen's J-test can be inconsistent in some cases and, therefore, futile in checking the exogeneity of the instruments when noncausality is present. Even in cases where the test is not inconsistent, it may have low finite-sample power, as suggested by our simulation results. Although our findings explicitly concern relatively simple special cases, it is easy to see that lagged values of variables following noncausal AR processes are, in general, never valid instuments.

The plan of the paper is as follows. In Section 2, the noncausal AR process is introduced and checking for its presence is discussed. In Subsection 2.3, we also present evidence that economic time series are quite often better described as noncausal than causal AR processes. Section 3 contains our main results concerning the asymptotic and finite-sample properties of the GMM estimator and the J-test. Finally, Section 4 concludes.

2 Noncausal autoregression

In this section, we briefly discuss noncausal AR processes as a prelude to the results concerning the GMM estimation in Section 3. In addition to presenting one parametrization of the noncausal autoregression to be used throughout the paper, we pick up on various aspects of model selection. Finally, we show evidence based on an extensive data set consisting of 343 macroeconomic and financial time series in favor of the prevalence of noncausality, attesting to the practical significance of the concerns put forth in this paper.

2.1 Model

The literature on noncausal AR models is not voluminous, and their economic applications are almost nonexistent. For a brief survey covering most of this literature, see Lanne and Saikkonen (2008), who introduced a new formulation of the model, developed the related likelihood-based theory of estimation and statistical inference, and devised a model selection procedure. In particular, they considered a stochastic process x_t $(t = 0, \pm 1, \pm 2, ...)$ generated by

$$\varphi\left(B^{-1}\right)\phi\left(B\right)x_t = \epsilon_t, \tag{1}$$

where $\phi\left(B\right) = 1 - \phi_1 B - \cdots - \phi_r B^r$, $\varphi\left(B^{-1}\right) = 1 - \varphi_1 B^{-1} - \cdots - \varphi_s B^{-s}$, and ϵ_t is a sequence of independent, identically distributed (continuous) random variables with mean zero and variance σ^2 or, briefly, $\epsilon_t \sim i.i.d.\left(0, \sigma^2\right)$. Moreover, B is the usual backward shift operator, that is, $B^k y_t = y_{t-k}$ $(k = 0, \pm 1, ...)$, and the polynomials $\phi\left(z\right)$ and $\varphi\left(z\right)$ have their zeros outside the unit circle so that

$$\phi\left(z\right) \neq 0 \quad \text{for} \quad |z| \leq 1 \quad \text{and} \quad \varphi\left(z\right) \neq 0 \quad \text{for} \quad |z| \leq 1. \tag{2}$$

We use the abbreviation AR(r, s) for the model defined by (1) and sometimes write AR(r) for AR$(r, 0)$. If $\varphi_1 = \cdots = \varphi_s = 0$, model (1) reduces to the conventional causal AR(r) process with y_t depending on its past but not future values. The more interesting cases from the viewpoint of this paper arise, when this restriction does not hold. If $\phi_1 = \cdots = \phi_r = 0$, we have the purely noncausal AR$(0, s)$ model with dependence on future values only. In the mixed AR(r, s) case where neither restriction holds, y_t depends on its past as well as future values. Our simulation results suggest that the problems due to the endogeneity of the instruments are severest when the instruments follow a purely noncausal AR process, but they can be substantial also in the case of a mixed process. However, to some extent these problems are mitigated as the causal part becomes more dominant.

The conditions in (2) imply that x_t has the two-sided moving average representation

$$x_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}, \tag{3}$$

where ψ_j is the coefficient of z^j in the Laurent series expansion of $\phi(z)^{-1} \varphi(z^{-1})^{-1} \overset{def}{=} \psi(z)$. This expansion exists in some annulus $b < |z| < b^{-1}$ with $0 < b < 1$ and with $\psi_{|j|}$ converging to zero exponentially fast as $|j| \to \infty$. It is well-known that x_t also has a causal AR(p) representation with $p = r + s$ and the autoregressive polynomial given by $a(B) = 1 - a_1 B - \cdots - a_p B^p = \varphi(B) \phi(B)$ (see Brockwell and Davis (1987, p. 124–125) and Lanne and Saikkonen (2009)). Thus, we can write

$$a(B) x_t = \xi_t, \tag{4}$$

where the (stationary) error term ξ_t is uncorrelated but, in general, not independent with mean zero and variance σ^2.

2.2 Checking for noncausality

It is well-known that causal and noncausal AR processes cannot be distinguished by autocorrelation functions. This means that they are not identified by Gaussian likelihood, so non-Gaussian distributions must be considered in ML estimation. Therefore, the first step in modeling a potentially noncausal time series is to search for signs of nonnormality. To this end, Lanne and Saikkonen (2008) suggest estimating an adequate Gaussian AR(p) model and checking its residuals for nonnormality. For economic and financial time series, the residuals are often leptokurtic, indicating that Student's t-distribution might be suitable. In their application to the U.S. inflation series as well as for a large number of series discussed below, this indeed seems to be the case.

Once nonnormality has been established, the next step is to select among the alternative AR(r, s) specifications. As the AR(p) model has been found to adequately capture the autocorrelation in the series, it seems reasonable to restrict oneself to models with $r + s = p$. Following Breidt et al. (1991), Lanne and Saikkonen (2008) suggested selecting among these the model that produces the greatest value of the likelihood function. Finally, the adequacy of the selected specification is checked diagnostically and the model is augmented if needed. In addition to examining the fit of the t-distribution, Lanne and Saikkonen (2008) checked the residuals for remaining autocorrelation and conditional heteroskedasticity.

The purpose of fitting a Gaussian AR model in the first step is only to help in determining the correct lag length and checking for nonnormality. Sometimes it may not be possible to come up with a satisfactory Gaussian AR model, in which case an adequate model might still be found among different non-Gaussian AR(r, s) specifications.

2.3 Prevalence of noncausality

In order to assess the significance of the problems caused by noncausal instruments in practice, we checked a large number of macroeconomic and financial variables for noncausality using the algorithm discussed in Subsection 2.2. In particular, we considered 343 time series from the seven-country data set of Stock and Watson (2004).[1]

Using a Gaussian likelihood, we were able to find a causal AR model adequate in the sense of capturing all autocorrelation for 260 of the considered series. In 202 cases, it is a noncausal specification that maximizes the likelihood function, and 136 of the selected models satisfy the diagnostic checks mentioned in Subsection 2.2 at

[1] This data set contains various asset prices, measures of activity (such as the real GDP, unemployment and consumer price index), wages, commodity prices, and money measures from Canada, France, Germany, Italy, Japan, UK and US. The data are monthly or quarterly and for the most part cover the years 1959–1999 although some series are available only for a shorter period. For most series we used various transformations, such as logs or differences, and we consider these as different time series in counting the total number of series. For details on the data, see Stock and Watson (2004).

the 5% level. In the remaining cases, there were signs of some unmodeled conditional heteroskedasticity and fat tails not satisfactorily captured by the t-distribution. Of the 83 series for which an adequate causal AR model could not be found as a starting point, in 40 cases a noncausal AR model turned out to be diagnostically satisfactory such that this model also maximizes the likelihood function among all AR specifications of the same order. All in all, we then have quite strong evidence in favor of noncausality in economic time series: of the 343 time series considered, 300 series show clear signs of noncausality and for 176 series an adequate noncausal AR model can be specified. These findings indicate that the possibility of noncausality should be kept in mind when using instrumental variables methods.

To take an example from the empirical literature employing the GMM estimator, we checked the instruments used by Campbell and Mankiw (1990) for noncausality. These authors tested the permanent income hypothesis by testing for the significance of the slope coefficient in a regression of the change in US aggregate consumption on the change in disposable income (quarterly data from 1953:1 to 1985:4). whether the slope coefficient equals zero. As instruments they used lagged differences of aggregate consumption, disposable income and interest rates, in various combinations. By the algorithm discussed in Section 2.2, all of these variables can be described as noncausal AR processes. Therefore, as will be explained in Section 3 below, it is not suprising that their two-stage least squares (2SLS) estimates do not differ much from the OLS estimate. Moreover, the test of overidentifying restrictions failed to reject at conventional significance levels for any combination of the instruments.

3 GMM with noncausal instruments

3.1 Model

In order to illustrate our main points we consider the simple time series regression model.

$$y_t = \beta x_t + \varepsilon_{1t}, \tag{5}$$

where the error term ε_{1t} is independent and identically distributed ($i.i.d.$) with zero mean. Despite its simplicity, this model serves to make our main points, and, as a matter of fact, even this simple regression model has been used quite frequently in empirical analysis. Typical examples include testing the permanent income hypothesis (e.g., Campbell and Mankiw, 1990) and consumption-based asset pricing models (see, Campbell et al. (1997, 311–313, and the references therein). The regressor x_t is supposed to follow the noncausal autoregression (1), rewritten here for convenience,

$$\varphi\left(B^{-1}\right)\phi\left(B\right)x_t = \varepsilon_{2t}, \tag{6}$$

where ε_{2t} is a zero mean $i.i.d.$ error term. Because we are interested in the case where the regressor and error term in (5) are correlated we let $\varepsilon_t = [\varepsilon_{1t} \;\; \varepsilon_{2t}]'$ be a general $i.i.d.$ error vector. Thus, defining the covariance matrix $\Sigma = [\sigma_{ij}]_{i,j=1,2}$ with $\sigma_{ii} = \sigma_i^2$ we assume $\varepsilon_t \sim i.i.d.\,(0, \Sigma)$ where, unless otherwise stated, σ_{12} is nonzero. For simplicity, we have omitted intercept terms from (5) and (6). Their inclusion would only mean using mean corrected data and, by standard arguments, it can be seen that mean correction has no effect on our asymptotic derivations. In our simulations intercept terms are included, however.

3.2 GMM estimation

When regressors are correlated with the error term, OLS estimation is inconsistent, and, therefore, GMM estimation is typically employed. That correlation between the regressor and error term results in (5) can be seen from (3) and the assumption $\sigma_{12} \neq 0$. In a case like this it is quite common to use lagged values of the regressor as instruments in GMM estimation. However, in the noncausal case these are not valid instruments. This is immediately seen by using (3) to obtain $Cor\left(x_{t-i}, \varepsilon_{1t}\right) = E\left(x_{t-i}, \varepsilon_{1t}\right) = \sigma_{12}\psi_{-i}$, $i > 0$, where ψ_{-i} is generally nonzero when the regressor is noncausal. One might think that in practice an application of the standard J-test (see Hansen, 1982) would reveal the problem. However, the J-test is known to have low power or even to be inconsistent against some alternatives (see Newey, 1985), and this can actually happen when noncausal instruments are employed.

Our subsequent derivations assume that the vector of instruments is given by $z_{t-1} = [x_{t-1} \ \cdots \ x_{t-p}]'$. At the end of this section we discus how to modify the results when other choices of instruments are employed. Note that using p lagged values of the regressor as instruments is appropriate because the regressor has a causal AR(p) representation (see (4)). Given this, and the fact that the errors in (5) are *i.i.d.*, means that our results indicate how badly things can go wrong even in a fairly favorable situation.

In our simple set up the GMM estimation boils down to classical 2SLS estimation. Suppose we have data for $t = -p+1, ..., 0, 1, ..., T$ with the first p observations of the regressor used as initial values in the LS estimation of the parameters $a_1, ..., a_p$ in (4). It will be convenient to introduce the parameter vector $a = [a_1 \ \cdots \ a_p]'$. The 2SLS estimator is defined as

$$\tilde{\beta} = \left(\sum_{t=1}^{T} \hat{x}_t x_t \right)^{-1} \sum_{t=1}^{T} \hat{x}_t y_t,$$

where $\hat{x}_t = \hat{a}' z_{t-1}$ with

$$\hat{a} = \left(\sum_{t=1}^{T} z_{t-1} z_{t-1}' \right)^{-1} \sum_{t=1}^{T} z_{t-1} x_t,$$

the OLS estimator of a. The inconsistency of $\tilde{\beta}$ in the noncausal case was already made clear but we nevertheless derive its probability limit, as the result is needed later.

Stationarity and standard arguments show that $\hat{a} \xrightarrow{p} \left(E\left(z_{t-1} z_{t-1}' \right) \right)^{-1} E\left(z_{t-1} x_t \right) = a$ and, furthermore,

$$\begin{aligned} \tilde{\beta} &= \beta + \left(\hat{a}' T^{-1} \sum_{t=1}^{T} z_{t-1} x_t \right)^{-1} \hat{a}' T^{-1} \sum_{t=1}^{T} z_{t-1} \varepsilon_{1t} \\ &\xrightarrow{p} \beta + \left(a' E\left(z_{t-1} x_t \right) \right)^{-1} a' E\left(z_{t-1} \varepsilon_{1t} \right). \end{aligned}$$

Let $\gamma_k = E\left(x_{t-k} x_t \right)$ be the autocovariance function of x_t and $\rho_k = \gamma_k / \gamma_0$ the corresponding autocorrelation function. Then, if $\rho = \left[\rho_1 \ \cdots \ \rho_p \right]'$ and $\psi = \left[\psi_{-1} \ \cdots \ \psi_{-p} \right]'$

we can write $E\left(z_{t-1}x_t\right) = \gamma_0\rho$ and, using (3), $E\left(z_{t-1}\varepsilon_{1t}\right) = \sigma_{12}\psi$. With this notation the preceding result reads as

$$\tilde{\beta} \xrightarrow{p} \beta + \frac{\sigma_{12}a'\psi}{\gamma_0 a'\rho}. \tag{7}$$

Thus, the 2SLS estimator is inconsistent when the numerator of the latter term on the right hand side is nonzero.

Now consider the J-test which is based on the covariance between the instruments and the 2SLS residual $\tilde{\varepsilon}_{1t} = y_t - \tilde{\beta}x_t$ $(t = 1, ..., T)$. The test statistic or in this case Sargan's statistic can be written as

$$J = \frac{T}{\tilde{\sigma}_1^2}\left(T^{-1}\sum_{t=1}^{T} z_{t-1}\tilde{\varepsilon}_{1t}\right)'\left(T^{-1}\sum_{t=1}^{T} z_{t-1}z_{t-1}'\right)^{-1}\left(T^{-1}\sum_{t=1}^{T} z_{t-1}\tilde{\varepsilon}_{1t}\right), \tag{8}$$

where $\tilde{\sigma}_1^2 = T^{-1}\sum_{t=1}^{T}\tilde{\varepsilon}_{1t}^2$. The test assumes that the number of instruments is larger than the number of regressors or, in our case, that $p > 1$. In practice one applies the test by comparing the observed value of J to quantiles of the χ^2_{p-1} distribution.

On the right hand side of (8) we have

$$T^{-1}\sum_{t=1}^{T} z_{t-1}\tilde{\varepsilon}_{1t} = T^{-1}\sum_{t=1}^{T} z_{t-1}\varepsilon_{1t} - (\tilde{\beta} - \beta)T^{-1}\sum_{t=1}^{T} z_{t-1}x_t,$$

and, furthermore, (see (7) and the derivations preceding it)

$$T^{-1}\sum_{t=1}^{T} z_{t-1}\tilde{\varepsilon}_{1t} \xrightarrow{p} \sigma_{12}\psi - \frac{\sigma_{12}a'\psi}{a'\rho}\rho. \tag{9}$$

The limit on the right hand side is zero when σ_{12} is zero. Then the regressor is strictly exogenous, which is not the case of our interest. However, the limit in (9) can also be zero when σ_{12} is nonzero. This happens when $\rho = c\psi$ for some nonzero real number c. This in turn happens in the purely noncausal case where $r = 0$ in (6) (and $s = p$) if only one of the parameters $\varphi_1, ..., \varphi_p$ is nonzero. To see this, suppose that, for example, $\varphi_1 \neq 0$ and $\varphi_2 = \cdots = \varphi_p = 0$. Then we also have $\varphi_i = a_i$ $(i = 1, ..., p)$ and from (3) and (4) it follows that $\psi_{-j} = \varphi_1^j = \rho_j$ $(j \geq 1)$. Thus, $\psi = \rho$, demonstrating the preceding statement. It is straightforward to check that in this case the probability limit of the OLS estimator of β is $\beta + \sigma_{12}/\gamma_0$, and a comparison with (7) reveals that this equals the probability limit of the 2SLS estimator. Thus, in this special case, the 2SLS estimator can be expected to be equally biased as the OLS estimator. Our simulation results confirm this and show that the bias of the 2SLS can be substantial also in other cases.

When the right hand side of (9) is zero, arguments similar to those already used show that $J = O_p(1)$, implying that the J-test is inconsistent. Thus, one might suspect that the power of the J-test is poor also when the limit in (9) is nonzero but 'small'. According to our simulations this indeed seems to be the case.

The preceding derivations can straightforwardly be modified to the case of general instruments. To illustrate this, define the noncausal AR(r, s) process w_t by substituting w_t for x_t in (6). Then (3) and (4) also hold with w_t in place of x_t and defining $z_{t-1} = [w_{t-1} \cdots w_{t-p}]'$ the previous expressions for the OLS estimator $\hat{a}$ and its probability limit a apply. Furthermore, (7) holds with the ith component of the vector $\gamma_0\rho$ given by $Cov\,(w_{t-i}, x_t)$ $(i = 1, ..., p)$ whereas ψ is as before except that its components are obtained from (3) with w_t in place of x_t. It follows that (9) holds with ψ as defined above and the ith component of the vector ρ given by $Cor\,(w_{t-i}, x_t)$, the cross correlation between w_{t-i} and x_t $(i = 1, ..., p)$. The condition where the right hand side of (9) becomes zero is as before but giving concrete examples of this is more difficult than in the case where the instruments are lags of the regressor.

3.3 Simulation results

In this section, we report results of some simulation experiments to demonstrate the relevance of the asymptotic results of Section 3.2 in finite samples. Specifically, we simulate 10,000 realizations from model (5)–(6) with $r + s = 2$ using a number of combinations of φ_1 and ϕ_1. In all experiments, $\beta = 1.0$ and also an intercept, whose true value equals zero, is estimated. The errors are drawn from a bivariate normal distribution with $\sigma_1^2 = \sigma_2^2 = 1.0$ and $\sigma_{12} = 0.8$. Qualitatively the conclusions are not affected by the values of these parameters. From each simulated bivariate time series, the parameters of the simple regression model are estimated by both OLS and 2SLS, and the value of the J-test statistic is computed. We consider two sample sizes, 200 and 500, but the results do not seem to be much affected by the length of the simulated realization.

In Table 1 we present a subset of our simulation results to highlight the main findings. The biases of the OLS and 2SLS estimates are reported as averages over all replications and the rejection rate of the J-test with nominal size 5%. Let us first consider the cases in the uppermost panel, where the instruments follow a purely noncausal AR process ($\phi_1 = 0$). It is seen that instrumental variables estimation does not correct for the bias, which for a given value of φ_1 is of the same magnitude for both estimators. In accordance with our theoretical results in Section 3.2, the differences between the biases get smaller as the sample size increases. The rejection rates of the J-test never exceed the nominal size of the test, reflecting the inconsistency of the test shown above.

As to the cases with the instruments following mixed noncausal AR process, the results are similar for small values of ϕ_1. Although the 2SLS estimator seems to produce a somewhat less biased estimates, the bias, reducing as ϕ_1 increases, can still be substantial. The rejection rates of the J-test are somewhat higher than in the purely noncausal case, but the test only has reasonable power when both ϕ_1 and φ_1 are large. This suggests that even in relatively realistic cases the J-test is rather useless in detecting the endogeneity of the instruments. As far as the bias is concerned, the effect of an increase in the sample size is minor also in the case of a mixed noncausal process.

4 Conclusion

In this paper, we have pointed out a potential pitfall in using lags of time series as instruments in GMM estimation. Lagged values are thought to be predermined by construction and, therefore, valid instruments. However, if the variable whose lags are used as instruments, is generated by a noncausal AR process, its lags may be endogenous and, hence, unsuitable as instruments, yielding an inconsistent GMM estimator. In a simple special case with lags of the explanatory variable used as instruments, we have shown that the OLS and 2SLS estimators even converge in probability to the same limit. Moreover, the J-test typically used to test for the exogeneity of the instruments, may be inconsistent, and, in general, has low power against endogenous instruments. In other words, the J-test cannot be relied on to reveal the endogeneity problem. Our finite-sample simulation experiments confirm these findings.

Although our results pertain to a relatively simple setup, it is not difficult to see that similar problems arise in more general contexts. As our empirical results indicate that noncausality is quite common among economic and financial time series, care should be taken when the GMM is employed. Based on our findings, we recommend that the candidate instrumental variables be checked for noncausality prior to using their lags as instruments and any instruments exhibiting noncausal dynamics be discarded. To that end, we have presented an algorithm, originally suggested in Lanne and Saikkonen (2008).

References

Breidt, J., R.A. Davis, K.S. Lii, and M. Rosenblatt (1991). Maximum likelihood estimation for noncausal autoregressive processes. Journal of Multivariate Analysis 36, 175–198.

Brockwell, P.J., and R.A. Davis (1987). Time Series: Theory and Methods. Springer-Verlag. New York.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay (1997). The Econometrics of Financial Markets. Princeton University Press. Princeton, New Jersey.

Campbell, J.Y., and N.G. Mankiw (1990). Permanent income, current income, and consumption. Journal of Business and Economic Statistics 8, 265–279.

Hansen, B.E., and K.D. West (2002). Generalized method of moments and macroeconomics. Journal of Business and Economic Statistics 20, 460–469.

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

Lanne, M., and P. Saikkonen (2008). Modeling expectations with noncausal autoregressions. HECER Discussion Paper No. 212.

Newey, W. (1985). Generalized method of moments specification testing. Journal of Econometrics 29, 229–256.

Stock, J.H., and M.W. Watson (2004). Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 23, 405–430.

Stock, J.H., J.H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. Journal of Business and Economic Statistics 20, 518–529.

Table 1: Simulation results.

		$T = 200$			$T = 500$		
ϕ_1	φ_1	Bias($\widehat{\beta}_{OLS}$)	Bias($\widehat{\beta}_{2SLS}$)	Rej(J)	Bias($\widehat{\beta}_{OLS}$)	Bias($\widehat{\beta}_{2SLS}$)	Rej(J)
0.0	0.2	0.149	0.115	0.049	0.152	0.135	0.051
	0.4	0.264	0.256	0.045	0.268	0.265	0.047
	0.6	0.306	0.305	0.047	0.307	0.307	0.047
	0.8	0.237	0.238	0.050	0.233	0.234	0.048
0.2	0.2	0.136	0.056	0.057	0.139	0.066	0.053
	0.4	0.234	0.163	0.049	0.237	0.170	0.053
	0.6	0.263	0.220	0.043	0.263	0.222	0.050
	0.8	0.197	0.183	0.041	0.193	0.180	0.043
0.4	0.2	0.112	0.030	0.056	0.116	0.039	0.061
	0.4	0.190	0.109	0.064	0.192	0.114	0.085
	0.6	0.208	0.157	0.075	0.208	0.159	0.130
	0.8	0.152	0.135	0.055	0.149	0.133	0.105
0.6	0.2	0.080	0.014	0.065	0.085	0.022	0.073
	0.4	0.135	0.068	0.118	0.137	0.072	0.209
	0.6	0.145	0.104	0.206	0.145	0.105	0.469
	0.8	0.104	0.091	0.181	0.104	0.102	0.461
0.8	0.2	0.040	2.0e–4	0.071	0.045	0.008	0.092
	0.4	0.071	0.031	0.216	0.072	0.034	0.465
	0.6	0.076	0.053	0.549	0.075	0.054	0.912
	0.8	0.054	0.048	0.633	0.052	0.046	0.973

The figures are based on 10,000 realizations of length T from model (5)–(6) where $r + s = 2$, the errors follow a bivariate normal distribution with $\beta = 1.0$, $\sigma_1^2 = \sigma_2^2 = 1.0$ and $\sigma_{12} = 0.8$. The first two lags of x_t are used as instruments in the 2SLS estimation. The reported biases are obtained as averages over all replications.The column 'Rej(J)' gives the fraction of replications where the J-test rejects at the 5% level of significance.