GMM estimation with non-causal instruments

Lanne, Markku

2011

Lanne , M & Saikkonen , P 2011 , ' GMM estimation with non-causal instruments ' , Oxford Bulletin of Economics and Statistics , vol. 73 , pp. 581-592 . https://doi.org/10.1111/j.1468-0084.2010.00631.x

http://hdl.handle.net/10138/335581
https://doi.org/10.1111/j.1468-0084.2010.00631.x

Word count: 4779

GMM Estimation with Noncausal Instruments*

MARKKU LANNE† and PENTTI SAIKKONEN‡

†*Department of Political and Economic Studies, University of Helsinki, P.O.Box 17 (Arkadiankatu 7), 00014 University of Helsinki, Finland (e-mail: markku.lanne@helsinki.fi)*

‡*Department of Mathematics and Statistics, University of Helsinki, P.O.Box 68 (Gustaf Hällströmin katu 2b), 00014 University of Helsinki, Finland (e-mail: pentti.saikkonen@helsinki.fi)*

Abstract

This note provides a warning against careless use of the generalized method of moments (GMM) with time series data. We show that if time series follow noncausal autoregressive processes, their lags are not valid instruments, and the GMM estimator is inconsistent. Moreover, endogeneity of the instruments may not be revealed by the J-test of overidentifying restrictions that may be inconsistent and has, in general, low finite-sample power. Our explicit results pertain to a simple linear regression, but they can easily be generalized. Our empirical results indicate that noncausality is quite common among economic variables, making these problems highly relevant.

JEL Classification numbers: C12, C22, C51.

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

*We thank Joose Sauli for excellent research assistance and an autonomous referee and Anindya Banerjee (the editor) for useful comments The usual disclaimer applies. Financial support from the Academy of Finland and the OP-Pohjola Group Research Foundation is gratefully acknowledged.

I. 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 likelihood (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 the GMM to be applicable, a sufficiently large number of instrumental variables 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 II) and, as we argue in this paper, in that case lags are not, in general, valid instruments. A major difference between these two types of AR processes is that errors of a noncausal AR process can be predicted by past values of the process whereas this is not the case for a causal AR process.

It is worth emhasizing that the problems caused by noncausal instruments are quite distinct from those due to weak instruments. Even though instruments are noncausal, they need not be weak in the sense of having low explanatory power for the instrumented regressors. For instance, as discussed below, the (theoretical) R^2 of the first-stage in two-stage least squares (2SLS) estimation only depends on the autocovariance structure of the instrument and not on its causality or noncausality. Our empirical example in Section IV illustrates how these two problems may arise independently although they are likely to be encountered together in a large number of applications.

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. In this case, the J-test can even be inconsistent and, therefore, futile in checking the exogeneity of the instruments. According to our simulations, the finite-sample power of the J-test can also be very low in many other cases, indicating that, in practice, one cannot rely on the ability of the test to reveal the noncausality of the employed instruments. 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 instruments.

The plan of the paper is as follows. In Section II, the noncausal AR process is introduced and checking for its presence is discussed. We also present evidence that economic time series are quite often better described as noncausal than causal AR processes. Section III contains our main results concerning the asymptotic and finite-sample properties of the GMM estimator and the J-test. Section IV gives an empirical illustration of the effects of noncausal instruments. Finally, Section V concludes.

II. Noncausal Autoregression

In this section, we briefly discuss noncausal AR processes as a prelude to the results concerning the GMM estimation in Section III. 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.

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

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

These conditions 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 $|j| \to \infty$.

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 x_t depending on (present and) past but not future values of the error process ϵ_t. 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 (present and) future errors only. In the mixed AR(r, s) case where neither restriction holds, x_t depends on (present and) past as well as future errors. 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.

It is well-known that causal and noncausal AR processes cannot be distinguished by autocovariance functions. This particularly means that the autocovariance function of a noncausal AR(r,s) process is identical to the autocovariance function of a causal AR(p) (or purely noncausal AR(0,p)) process with $p = r+s$ (see Brockwell and Davis (1987, p. 124–125) and Lanne and Saikkonen, 2008). Furtheremore, the noncausal AR(r,s) process x_t defined in (1) also has a causal AR(p) representation with $p = r + s$ and the autoregressive polynomial given by $a(B) = 1 - a_1 B - \cdots - a_p B^p = \varphi(B)\,\phi(B)$. 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.

Checking for noncausality

As causal and noncausal AR processes cannot be distinguished by autocovariance functions 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.

Lanne and Saikkonen (2008) provide simulation evidence in favor of the model selection procedure based on maximizing the likelihood function among the different pth order AR(r, s) models. Their results, pertaining to noncausal second-order AR models indicate that a causal model is rarely selected when the true model is noncausal, even with as few as 100 observations. In addition, we simulated causal and purely noncausal first-order AR models with different degrees of persistence and found that the correct model is selected in at least 89% and 99% of the cases with 200 and 500 observations, respectively.

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 above. In particular, we considered 343 time series from the seven-country data set of Stock and Watson (2004).[1]

[1] This data set contains various asset prices, measures of activity (such as the real GDP, unem-

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 at 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, 242 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.

III. GMM with noncausal instruments

Model

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

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

ployment 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).

where the error term ε_{1t} is *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 and empirical application intercept terms are included, however.

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 briefly discuss the case where other 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 if noncausality of the instruments is erroneously ignored. They also make clear that problems caused by noncausal instruments are distinct from those due to weak instruments. When z_t is used as the vector of instruments, a low value of the R^2 (or equivalently the implied F-statistic) in a regression of x_t on z_t is conventionally interpreted to indicate weakness of the instruments. As x_t has the causal AR(p) representation (4) (with $\sigma^2 = \sigma_2^2$), the theoretical version of this measure can be written as $R^2 = 1 - \sigma_2^2 / Var(x_t)$. Our results show that the adverse effects of noncausal instruments can be severe even when the instruments cannot be deemed weak according to this measure.

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 OLS estimation of the parameters $a_1, ..., a_p$ in (4). 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 $\hat{a}$ the OLS estimator of the parameter vector $a = [a_1 \ \cdots \ a_p]'$

in a regression of x_t on z_{t-1} .

The inconsistency of $\tilde{\beta}$ in the noncausal case was already made clear. Using stationarity and standard arguments one can show that $\tilde{\beta} \xrightarrow{p} \beta + (a'E(z_{t-1}x_t))^{-1} a'E(z_{t-1}\varepsilon_{1t})$. For subsequent purposes we obtain a convenient expression for the probability limit of $\tilde{\beta}$. Let $\gamma_k = E(x_{t-k}x_t)$ be the autocovariance function of x_t and $\rho_k = \gamma_k/\gamma_0$ the corresponding autocorrelation function. Then, if $\rho = \begin{bmatrix}\rho_1 & \cdots & \rho_p\end{bmatrix}'$ and $\psi = \begin{bmatrix}\psi_{-1} & \cdots & \psi_{-p}\end{bmatrix}'$ we can write $E(z_{t-1}x_t) = \gamma_0\rho$ and, using (3), $E(z_{t-1}\varepsilon_{1t}) = \sigma_{12}\psi$. With this notation we have

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

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) is also zero when σ_{12} is nonzero and $\rho = c\psi$ for some nonzero real number c. This 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 because then $\rho = \psi$.[2] Arguments similar to those already used show that then we also have $J = O_p(1)$, implying that the J-test is inconsistent. This result is rather special, however, because it pertains to a case where $p - 1$ of the components of z_{t-1} are irrelevant instruments and could be omitted. For instance, if x_t follows an AR(0,1) process, its partial autocorrelations at lags larger than one are zero and the instruments $x_{t-2}, ..., x_{t-p}$ provide no expalanatory power over x_{t-1} for the instrumented regressors. Nevertheless, the preceding discussion suggests that the power of the J-test is expected be low when the instrument is purely noncausal and one of the coeffients $\varphi_1, ..., \varphi_p$ is 'dominating' or, more generally, when the limit in (9) is 'small'. From a practical point of view it is also of interest to note that, according to our simulations (see the next section), the power of the J-test can be very low even in cases where the right hand side of (9) is nonzero and there is no instrument that could be interpreted as 'irrelevant' in the sense discussed above.

It is also straightforward to check that in the aforementioned purely noncausal case with only one of the parameters $\varphi_1, ..., \varphi_p$ nonzero, the probability limit of the OLS estimator of β is $\beta + \sigma_{12}/\gamma_0$ which equals the probability limit of the 2SLS estimator (see (7)). 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.

[2] For instance, if $\varphi_1 \neq 0$ and $\varphi_2 = \cdots = \varphi_p = 0$ then $\varphi_i = a_i$ $(i = 1, ..., p)$ and from (3) and (4) it follows that $\psi_{-j} = \varphi_1^j = \rho_j$ $(j \geq 1)$. This shows that $\psi = \rho$ and that the right hand side of (9) is zero.

The preceding derivations can straightforwardly be modified to the case with lags of a general noncausal AR(r, s) process w_t instead of lags of x_t used as instruments. The condition where the right hand side of (9) becomes zero (or close to zero) is basically as before but giving concrete examples of this is more difficult than in the case where the instruments are lags of the regressor.

Simulation results

In this section, we report results of some simulation experiments to demonstrate the relevance of the asymptotic results derived above in finite samples. Specifically, we simulate 10,000 realizations from model (5)–(6) with $r + s = 2$. 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 Tables 1 and 2, 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 of Table 1, where the instruments follow a purely noncausal first-order 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, the differences between the biases do not seem to get smaller as the sample size increases. Similar findings emerge from Table 2 where the instruments follow an

AR(0,2) process. With all combinations of the autoregressive parameters considered, the rejection rates of the J-test remain small. Moreover, somewhat surprisingly, the bias even seems to be larger for the 2SLS estimator in a number of cases.

TABLE 1

Simulation results for a number of cases where the explanatory variable follows an AR(1,1) process

		$T = 200$			$T = 500$		
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	0.1	0.074	0.052	0.050	0.077	0.061	0.052
	0.5	0.297	0.292	0.049	0.298	0.297	0.050
	0.9	0.150	0.151	0.048	0.142	0.143	0.052
0.1	0.1	0.072	0.018	0.052	0.076	0.031	0.055
	0.5	0.280	0.241	0.047	0.282	0.246	0.048
	0.9	0.136	0.134	0.045	0.129	0.127	0.047
0.5	0.1	0.049	–0.006	0.056	0.054	0.005	0.053
	0.5	0.178	0.113	0.107	0.179	0.117	0.202
	0.7	0.161	0.131	0.128	0.074	0.070	0.100
0.9	0.1	0.002	–0.014	0.059	0.009	–0.005	0.055
	0.5	0.039	0.021	0.527	0.039	0.023	0.895
	0.9	0.018	0.017	0.717	0.016	0.015	0.992

Notes : The figures are based on 10,000 realizations of length T from model (5)–(6) where $r = s = 1$, 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 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.

As to the cases with the instruments following a mixed noncausal AR process in the lower panels of Table 1, the results are similar for small values of ϕ_1. Although the 2SLS estimator seems to produce somewhat less biased estimates, the bias, reducing

TABLE 2

Simulation results for a number of cases where the explanatory variable follows an AR(0,2) process

		T = 200			T = 500		
φ_1	φ_2	$Bias(\widehat{\beta}_{OLS})$	$Bias(\widehat{\beta}_{2SLS})$	$Rej(J)$	$Bias(\widehat{\beta}_{OLS})$	$Bias(\widehat{\beta}_{2SLS})$	$Rej(J)$
0.2	0.3	0.222	0.242	0.048	0.226	0.253	0.060
	0.5	0.273	0.272	0.034	0.273	0.270	0.037
	0.7	0.193	0.173	0.222	0.178	0.157	0.060
0.4	0.3	0.224	0.250	0.056	0.225	0.252	0.067
	0.5	0.161	0.151	0.058	0.150	0.140	0.081

Notes : The figures are based on 10,000 realizations of length T from model (5)–(6) where $r = 0$ and $s = 2$, 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 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.

as ϕ_1 increases, can still be substantial and in several cases an increase in the sample size has only a minor effect. 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. It is worth noting that the low power of the J-test cannot be explained by weak instruments. For instance, in the cases $(\phi_1, \varphi_1) = (0.1, 0.9)$ (or alternatively $(\phi_1, \varphi_1) = (0.9, 0.1)$) and $(\phi_1, \varphi_1) = (0.5, 0.7)$ the (theoretical) R^2 discussed in the previous section takes values as high as 0.843 and 0.816 with the corresponding values of the F-statistic being 83.46 and 80.78 when the sample size is 200. One may still argue that in the former case x_{t-2} is a weak instrument because the second partial autocorrelation of the x_t process is only 0.09. However, this argument does not apply to the latter case where the second partial autocorrelation is 0.35 which, especially for the sample sizes considered, is not small. Indeed, in practice the corresponding estimated partial autocorrelation would be deemed nonzero with very high probability. Thus, even in relatively realistic cases the J-test can be rather useless in detecting the endogeneity of the instruments.

IV. Empirical application

To illustrate the effects of noncausal instruments, we consider an empirical application to the estimation of the elasticity of intertemporal substitution (EIS) that is a parameter of great interest in macroecnomics and finance. The estimation of EIS is typically based on the regression equation

$$\Delta c_t = \tau + \psi r_t + \varepsilon_t, \tag{10}$$

where Δc_t is the consumption growth, r_t is the real return on an asset, such as the real interest rate, and τ is a constant. The error term ε_t is, in general, correlated with the regressor r_t, but given a set of valid instruments, the parameters in equation (10) can be estimated by two-stage least squares.

Model (10) is often written in reversed form as

$$r_t = \mu + \frac{1}{\psi}\Delta c_t + \eta_t, \tag{11}$$

where μ is a constant and η_t an error. Under power utility, the reciprocal of EIS, $1/\psi$, is the coefficient of relative risk aversion. Both equations have been used to estimate EIS (or, equivalently, the coefficient of relative risk aversion) in the previous literature. Of particular interest is the hypothesis $\psi = 1$ because (with Epstein-Zin preferences) it implies that an investor's optimal consumption choice is a constant fraction of wealth, whereas the optimal consumption-wealth ratio is decreasing (increasing) in expected returns when EIS is greater (less) than unity (see Campbell and Viceira, 1999). Concerning this hypothesis, results in the previous literature are contradictory in that estimates of ψ as well as its reciprocal are typically small. The problem of weak instruments has been suggested as an explanation to these findings. According to this view, equation (10) is preferable because asset returns such as the real interest rate are more predictable than consumption growth. Below we show that at least for the U.K., both equations yield similar conclusions when the set of instruments includes only lags of causal time series, while otherwise quite different results emerge.

We estimate by two-stage least squares models (10) and (11) with quarterly U.K. data (1970:III to 1999:I) taken from Yogo (2004), and following that paper, we use the twice lagged log dividend-price ratio (z_{1t}), nominal interest rate (z_{2t}), inflation (z_{3t}) and consumption growth (z_{4t}) as instruments. Yogo (2004) showed that this set of four instruments does not suffer from the problem of weak instruments, but he did not pay attention to their time series properties. We show that some of the instruments indeed are noncausal without this necessarily being revealed by the J-test. Moreover, very different estimates are obtained depending on which subset of them is used in estimation.

We start by checking each of the four instruments for noncausality. The model selection procedure discussed in Section II, indicates that inflation and the log dividend-price ratio are purely noncausal with AR(0,4) and AR(0,1) models with t-distributed errors providing the best fit, respectively. The other two instruments, on the other hand, turn out to be causal, with the dynamics of the nominal interest rate being adequately captured by an AR(1,0) model and consumption growth following an AR(3,0) model. Also in these cases the residuals of the Gaussian AR model turned out to be leptokurtic while the t-distribution seemed sufficient to capture the fat tails. In sum, there are two noncausal instruments that should not be used, inflation and the log dividend-price ratio.

TABLE 3

Estimation results of the U.K. EIS and coefficient of relative risk aversion with different sets of instruments

				$\widehat{\psi}$		$\widehat{1/\psi}$	
Instruments	$\widehat{\psi}$	$\widehat{1/\psi}$	$AR\left(\widehat{\psi}\right)$	J	F	J	F
All $(z_{1t}, z_{2t}, z_{3t}, z_{4t})$	$[-0.08, 0.41]$	$[0.17, 1.95]$	$[0.04, 0.28]$	0.03	17.04	0.00	2.52
Noncausal (z_{1t}, z_{3t})	$[0.15, 0.89]$	$[0.54, 3.21]$	$[0.10, 1.17]$	0.63	13.35	0.63	4.50
Causal (z_{2t}, z_{4t})	$[-5.53, 1.90]$	$[-1.66, 0.58]$	$[-0.04, 1.59]$	0.90	0.65	0.90	3.19

Notes : The instruments are twice lagged log dividend-price ratio (z_{1t}), nominal interest rate (z_{2t}), inflation (z_{3t}), and consumption growth (z_{4t}). On each row, the second and third columns report the 95% confidence intervals based on 2SLS estimation using the instruments in the first column. The fourth column contains the Anderson-Rubin confidence intervals for ψ. The p-values of the J-test of overidentifying restrictions and the values of the F-statistics of the first-stage regression in 2SLS estimation of equations (10) and (11) are reported in the fifth and sixth, and the last two columns, respectively.

Table 3 presents the 95% confidence intervals of ψ and $1/\psi$ based on equations (10) and (11), using three different sets of instruments: all four instruments, only the noncausal instruments (z_{1t} and z_{3t}), and only the causal instruments (z_{2t} and z_{4t}). In each case, we also report the value of the F-statistic of the first-stage regression and the p-value of the J-test. The value of the first-stage F-statistic is less than 10 (the rule-of-thumb value suggested by Staiger and Stock, 1997) in a number of cases, and, therefore, we also report the weak-instrument-robust Anderson-Rubin confidence intervals (see Anderson and Rubin, 1949) for ψ.[3] Note that the Anderson-Rubin test is invariant such that it rejects the null hypothesis $\psi = \psi_0$ based on equation (10) if and only if it rejects $1/\psi = 1/\psi_0$ based on equation (11).

[3] We thank Motohiro Yogo for providing the GAUSS code for computing the Anderson-Rubin confidence interval on his web page.

A number of interesting conclusions emerge. First, the estimates of ψ and $1/\psi$ seem to vary considerably depending on the instruments. Second, the J-test fails to reject at the 5% level only in the case of the full instrument set, but not when the noncausal instruments alone are used. Third, it is only in the case of the causal instruments that the confidence intervals of ψ and $1/\psi$ are in concert, suggesting EIS greater than and coefficient of relative risk aversion less than unity. In the other cases, the confidence interval of ψ includes values whose inverse is not included in the corresponding confidence interval of $1/\psi$, leading to a contradiction. Finally, these conclusions do not seem to depend on the problem of weak instruments since the weak-instrument-robust confidence interval yields similar results for ψ when only the causal instruments are used.

V. 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 predetermined 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. This is particularly the case when the potential instuments are Gaussian and, therefore, there is no way of detecting noncausality.

References

Anderson, T.W. and Rubin, H. (1949). 'Estimation of the parameters of a single equation in a complete system of stochastic equations', *Annals of Mathematical Statistics*, Vol. 20, pp. 46–63.

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

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

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

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

Campbell, J.Y. and Viceira, L.M. (1999). 'Consumption and portfolio decisions when expected returns are time varying', *Quarterly Journal of Economics*, Vol. 114, pp. 433–495.

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

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

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

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

Staiger, D., and Stock, J.H. (1997). 'Instrumental variables regression with weak instruments', *Econometrica*, Vol. 65, pp. 557–586.

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

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

Yogo, M. (2004). 'Estimating the elasticity of intertemporal substitution when instruments are weak', *Review of Economics and Statistics*, Vol. 86, pp. 797–810.