International Review of Business Research Papers
Vol. 8. No.6. September 2012. Pp. 149 – 157

Use of Instrumental Variables for the Granger-Causality Test

Chung-ki Min*

This study points out that the Granger-causality test can be biased due to both contemporaneous and lagged time-specific effects. To solve the bias problem this study employs estimation methods which use instrumental variables, particularly for a single time series. However, the success of these estimation methods depends critically on the relevance and exogeneity of the instrumental variables used. Using simulated data this study assesses the instrumental variables and compares the results by the instrumental variable estimation against the ones by the OLS estimation.

Field of Research: Econometrics

1. Introduction

Granger (2004) states that the causality has two components: (i) the cause occurs before the effect; and (ii) the cause contains information about the effect that is unique, and is in no other variable. Based on this definition of causality, Granger has developed a test method, called the Granger-causality test, which examines whether lagged values of a variable have significant effects on the current value of another variable.

This study focuses on the uniqueness of information contained in lagged values of a candidate cause. Since most economic variables are subject to time-specific changes which are common to all variables in the same period, the information associated with time-specific effects is not unique and should not be considered as a cause. Further, as the time-specific effects are independently and unexpectedly determined in each period, they are uncorrelated between periods and thus have nothing to do with any causal relation. Therefore, we have to control for the time-specific effects in the causality test.

The objective of this study is to illustrate how time-specific effects can distort the Granger-causality test and to suggest instrumental variables which can correct the problem. In particular, it points out that lagged time-specific effects make the right-hand-side lagged variables endogenous and therefore cause a bias in the causality test when the OLS estimation is employed. One widely-employed approach to control for time-specific effects is to use dummy variables. However, this dummy-variable approach can be applied only when there are multiple observations in each period, i.e., panel data.

Treating the lagged time-specific effects as measurement errors in explanatory variables, this study suggests use of instrumental variables from the literature on errors in variables. Griliches and Hausman (1984) use lagged and forward variables as instruments in their

* Department of Economics, Hankuk University of Foreign Studies, Seoul, Korea. email: cmin@hufs.ac.kr.
This work was supported by the Hankuk University of Foreign Studies Research Fund.

static panel data model. Since the model used for the Granger-causality test is a dynamic one, forward dependent variables are correlated with lagged time-specific effects and are not qualified for instruments. We therefore use only lagged variables as instruments. Another set of instrumental variables for measurement errors is constructed by higher moments of the regressors in a model (Dagenais and Dagenais ,1997; and Lewbel ,1997). These proposed instruments are useful in cross-sectional data. In time-series data for a causality test, however, it is relatively easy to find other instruments, e.g., lagged regressors. Use of the higher-moments instruments is not used in this study but is planned for future work.

There have been a fair number of applications of the Griliches and Hausman approach in the literature. One of these is Himmelberg and Petersen (1994) who estimate the effect of a permanent component in cash flow on R&D investment, with controlling for the confounding effect of a transitory component in cash flow. This decomposition of cash flow is based on an observation that R&D investment involves high adjustment costs and is therefore relatively unresponsive to transitory changes in cash flow. Since observed cash flow is a sum of the permanent and the transitory changes, they treat the transitory component as a measurement error and employ the instrumental variables estimation method suggested in Griliches and Hausman (1984). Another group of research extends the instrumental variables approach to more general cases of measurement errors, including measurement errors in nonlinear regressions, serially correlated measurement errors, etc. (Hausman, 2001; Wansbeek, 2001).

Instrumental variables are expected to solve the endogeneity problem by replacing the endogenous variables. They are required to be uncorrelated with the lagged time-specific effects which cause a bias (exogeneity). At the same time, they need to be highly correlated with the endogenous variables (relevance). Thus, success of these instrumental-variables estimation methods depends critically on the exogeneity and relevance of the instrumental variables used. This study evaluates the instrument exogeneity and relevance by applying the methods in Stock and Yogo (2005).

The following section explains how time-specific effects can bias the Granger-causality test. Section 3 discusses issues in the instrumental variable estimation and section 4 presents test methods for the issues. In section 5 we illustrate the estimation and test methods using simulated data. Conclusions are presented in section 6.

2. Issues in the Granger–causality Test: Time-specific Effects

The Granger-causality is defined as a relationship between a variable and the lagged values of other variables. Thus, the causality test uses VAR (vector autoregressive) models which are reduced-form equations of structural equations. For a case of two variables x^* and y^*, a VAR model is expressed as

$$\begin{aligned} x_t^* &= \alpha_0 + \alpha_1 x_{t-1}^* + \alpha_2 y_{t-1}^* + u_t^x \\ y_t^* &= \beta_0 + \beta_1 x_{t-1}^* + \beta_2 y_{t-1}^* + u_t^y \end{aligned} \qquad (1)$$

where only lagged-one variables are included for simplicity. A null hypothesis of $\alpha_2 = 0$ implies that y^* does not Granger-cause x^*, and a null hypothesis of $\beta_1 = 0$ implies that x^* does not Granger-cause y^*. While the OLS method is widely used in estimating the VAR models, it is well known that the OLS estimators become biased if any of the lagged variables on the right-hand side are correlated with the disturbances, called the endogeneity problem.

Every time period the economy experiences unexpected changes and almost all economic variables are influenced by the changes. The effects on economic variables of the changes in each period, called time-specific effects, are merged into the variables x^* and y^* in Eq.(1). A problem arises in estimating Eq.(1) since neither these time-specific effects nor the underlying variables x^* and y^* are observed. Letting δ_t^x and δ_t^y denote the time-specific effects for *x* and *y*, respectively, we can observe only the sum of the latent variables (x^* and y^*) and their corresponding time-specific effects.

$$\begin{aligned} x_t &= x_t^* + \delta_t^x \\ y_t &= y_t^* + \delta_t^y \end{aligned} \tag{2}$$

As changes in each period are unexpected, their time-specific effects are independent between periods. Therefore, they are not associated with the causal relationship. Instead, the causal relationship holds only for the latent variables. Since x^* and y^* are not observable, however, we need to transform Eq.(1) into equations expressed in observable variables. Substituting Eq.(2) for x^* and y^* in Eq.(1), we obtain

$$\begin{aligned} x_t &= \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 y_{t-1} + \theta_t^x + u_t^x \\ y_t &= \beta_0 + \beta_1 x_{t-1} + \beta_2 y_{t-1} + \theta_t^y + u_t^y \end{aligned} \tag{3}$$

where $\theta_t^x = \delta_t^x - \alpha_1 \delta_{t-1}^x - \alpha_2 \delta_{t-1}^y$ and $\theta_t^y = \delta_t^y - \beta_1^x \delta_{t-1}^x - \beta_2 \delta_{t-1}^y$; these are functions of current and lagged time-specific effects. Notice that the lagged time-specific effects are included not only in θ_t^x and θ_t^y but also in x_{t-1} and y_{t-1}. Thus, ignoring θ_t^x and θ_t^y (or δ_{t-1}^x and δ_{t-1}^y) biases the OLS estimators due to the endogenetiry. This resulting bias belongs to the problem of measurement errors in explanatory variables. It indicates that lagged time-specific effects need to be controlled for in the Granger-causality test.

If there are multiple observations in each time period, the time-specific effects can be eliminated by the within-period transformation. Consider a panel version of Eq.(3).

$$\begin{aligned} x_{i,t} &= \alpha_0 + \alpha_1 x_{i,t-1} + \alpha_2 y_{i,t-1} + \theta_t^x + u_{i,t}^x \\ y_{i,t} &= \beta_0 + \beta_1 x_{i,t-1} + \beta_2 y_{i,t-1} + \theta_t^y + u_{i,t}^y \end{aligned} \tag{4}$$

Under the assumption that the time-specific effects ($\delta_t^x, \delta_t^y, \delta_{t-1}^x, \delta_{t-1}^y$) are constant across observational units in each period, we can eliminate θ_t^x and θ_t^y by calculating the deviations from the means in each period. This belongs to the fixed effects model and includes dummy variables to control for the time-specific effects. In doing so, the lagged time effects (δ_{t-1}^x and δ_{t-1}^y) contained in the regressors $x_{i,t-1}$ and $y_{i,t-1}$ are also eliminated because they are unit-invariant.

However, we often need to apply the causality test for single time series. As one observation in each period does not allow us to estimate the unit-invariant time-specific effects, we employ alternative estimation methods which use instrumental variables (IVs). These IV methods can be applied to panel data as well if the time-specific effects are unit-variant.

3. Issues in the Instrumental Variable Estimation

For expositional simplicity we begin with the following model which includes one endogenous explanatory variable. For $t = 1,2,\cdots T$,

$$\begin{aligned} y_t &= \beta_0 + \beta_1 Y_t + \eta_t \\ Y_t &= \pi_0 + \pi_1 Z_{1t} + \pi_2 Z_{2t} + \cdots + \pi_K Z_{Kt} + v_t \end{aligned} \tag{5}$$

where Y_t is correlated with the disturbance η_t and can be expressed as a function of exogenous (thus instrumental) variables Z_{it}. In obtaining a 2SLS estimate, we use an estimate of π_i's from the first-stage regression, i.e., the second equation in Eq.(5), since their true values are unknown. Therefore, the 2SLS is biased for finite samples and its bias is approximately equal to, as derived in Hahn and Hausman (2002),

$$E\left[\hat{\beta}_1^{2SLS}\right] - \beta_1 \approx \frac{K \cdot Cov(\eta, v)}{R_f^2} \times \frac{1}{\sum_{t=1}^{T} Y_t^2} \tag{6}$$

where R_f^2 is the R^2 in the first stage regression to obtain $\hat{\pi}_i$'s. Eq.(6) indicates that the bias is monotonically increasing in $Cov(\eta, v)$, i.e., endogeneity of Y_t, and decreasing in R_f^2, i.e., correlation between the instrumental variables and the endogenous regressor in the first stage. It also shows that the bias becomes bigger as the number of instrumental variables (*K*) increases.[1]

For the 2SLS estimators to perform well, the instrumental variables should satisfy two conditions. The first condition requires that the IVs produce a large value of R_f^2 in the first stage regression; this is called the *instrument relevance*. If R_f^2 is too small and close to zero, the bias of the 2SLS will not disappear even with a very large sample size. Further, the sampling distribution of the 2SLS estimator is not normal and standard hypothesis tests are not valid. Such IVs are said to be weak.

The other condition is that the IVs should be exogenous, called the *instrument exogeneity*. To solve the endogeneity problem, the fitted values from the first stage have to be uncorrelated with the disturbances. As the fitted values are a linear combination of the IVs, the IVs themselves are required to be exogenous.

4. Tests for the Relevance and Exogeneity of IVs

Whether the 2SLS estimators are useful depends on whether the IVs are both relevant and exogenous. In this section we introduce test methods for the instrument relevance and exogeneity.

4.1Test for the Instrument Relevance

In developing tests for weak instruments, Stock and Yogo (2005) use two alternative definitions of weak instruments. The first definition is that a group of instruments are weak if the bias of the 2SLS estimator, relative to the bias of the OLS, could exceed a certain threshold, for example 10%. The second definition is that the IVs are weak if the conventional α-level Wald test based on the IV statistics has an actual size that could exceed a certain threshold, for example 15% when α =5%.

The idea underlying the test procedures is that in order to reject the null hypothesis of weak instruments, the first-stage partial R^2 contributed by the IVs (which are unique to the first-stage regression) has to be high enough. However, when there are multiple endogenous regressors, it is not enough to have large values of the first-stage partial R^2. It is because the fitted values of the endogenous regressors from the first stage regression can be highly correlated among them. To avoid this multicollinearity problem in the second stage regression, the IVs must be strong enough to overcome the two definitions of weak instruments; for example, the bias of the 2SLS estimator should not exceed 10% of the OLS bias and the actual size of the conventional 5%-level Wald test should not exceed 15%.

Stock and Yogo (2005) derive a test statistic for a regression model which includes n endogenous regressors (Y) and K_1 included exogenous regressors (X). In order to apply their test statistic for our Granger-causality test, we focus on the y_t equation in Eq.(3) and add two reduced-form equations which describe the relation between the endogenous lagged variables and instrumental variables. Using vector-matrix notations, the regression model is expressed as

$$\begin{aligned} y &= Y\beta + X\gamma + \eta \\ Y &= Z\Pi + X\Phi + V \end{aligned} \tag{7}$$

where y is a *T*x1 vector of observations on y_t; Y a *T*x2 matrix on x_{t-1} and y_{t-1} (*n*=2 here); β a 2x1 vector of coefficients (β_1, β_2); X a *T*xK_1 matrix of included exogenous variables (Since there is no included exogenous variable in Eq.(3) other than the intercept, X is just a *T*x1 column vector of 1's.); γ a K_1x1 vector of coefficients (K_1=1 in this application); η a (*T*x1) vector of disturbances $\eta_t = \theta_t^y + u_t^y$; Z a *T*xK_2 matrix of instrumental variables; Π a K_2x2 coefficient vector; and V a (*T*x2) matrix of the disturbances in the first-stage regression.

Stock and Yogo (2005) propose a test based on the eigenvalues of the following matrix.

$$G_T = \hat{\Sigma}_{VV}^{-1/2'} Y^{\perp'} P_{Z^{\perp}} Y^{\perp} \hat{\Sigma}_{VV}^{-1/2} / K_2 \qquad (8)$$

where the superscript "$\perp$" denotes the residuals from the projection on the included exogenous variables *X*; $\hat{\Sigma}_{VV} = (Y' M_{\underline{Z}} Y)/(T - K_1 - K_2)$; $\underline{Z} = [X\ Z]$; $M_{\underline{Z}} = I - \underline{Z}(\underline{Z}'\underline{Z})^{-1}\underline{Z}'$; and $P_{Z^{\perp}} = Z^{\perp}(Z^{\perp'}Z^{\perp})^{-1}Z^{\perp'}$. The test statistic is the minimum eigenvalue of G_T.

$$g_{\min} = \text{minimun eigenvalue}\,(G_T) \qquad (9)$$

The null hypothesis (H_0) is that the IVs are weak, and the alternative is that they are not (or strong). And the test procedure is

$$\text{Reject } H_0 \text{ if } g_{\min} \geq \text{critical value}(n, K_2, bias \text{ or } size) \qquad (10)$$

The critical values are a function of the number of included endogenous regressors (*n*) and the number of instrumental variables (K_2). The critical values also depend on which definition of weak instruments is used; the desired maximal bias of the 2SLS estimator relative to the OLS estimator ($bias$) or the desired maximal size of a 5% Wald test of $H_0 : \beta = \beta_0$ ($size$). Critical values are tabulated in Stock and Yogo (2005, pp100-101).

4.2Test for the Instrument Exogeneity

If the IVs are not exogenous, the 2SLS estimators are obviously inconsistent. Therefore, it is very important to make sure that the IVs used are exogenous. Since the structural disturbances are not observable, we test whether the IVs are uncorrelated with the residuals from the estimated 2SLS regression. This test is possible only when the model is over-identified, i.e., when the number of IVs exceeds the number of endogenous regressors, thus also called the over-identifying restrictions test.

The over-identifying restrictions test begins with a regression of the second-stage residuals on all exogenous variables (both included exogenous variables and instrumental variables). Under the null hypothesis that the IVs are exogenous, the coefficients on the IVs should be equal to zero. Thus, the test is based on R^2 from this regression. Formally, the test statistic TR^2 is known to follow a *Chi*-square distribution with $K_2 - n$ degrees of freedom, where T is the number of observations, K_2 is the number of IVs, and n is the number of endogenous regressors (Hansen, 1982).

5. Numerical Illustrations

For an empirical application of the above estimation and test methods, we generate time-series observations on (x_t, y_t) as follows. For t =1, 2, ... , 3000,

$$x_t^* = 0.5x_{t-1}^* + u_t^x, \qquad u_t^x \sim N(0, 0.5^2)$$
$$y_t^* = 0.3x_{t-1}^* + 0.5y_{t-1}^* + u_t^y, \quad u_t^y \sim N(0, 0.5^2)$$
$$x_t = x_t^* + \delta_t, \;\; y_t = y_t^* + \delta_t, \;\; \delta_t \sim N(0,1)$$

where common time-specific effects (δ_t) are added to both variables. According to this model, x Granger-causes y but y does not Granger-cause x. Using observable variables x_t and y_t, we estimate the following regression model:

$$y_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 y_{t-1} + \eta_t \tag{11}$$

where the lagged variables are endogenous due to the lagged time-specific effects, as explained in section 2.

Table 1: Estimation Results: Coefficient Estimates, and the Relevance and Exogeneity Tests of the Instrumental Variables (Number of observations = 3,000)

Instrumental Variables		x_{t-2} y_{t-2} x_{t-3} y_{t-3}		x_{t-2} y_{t-2} x_{t-3} y_{t-3} x_{t-4} y_{t-4}		x_{t-3} y_{t-3} x_{t-4} y_{t-4}	
Estimation method	OLS	2SLS	GMM	2SLS	GMM	2SLS	GMM
Coefficient for							
x_{t-1} (true value = 0.3)	-0.053*	0.473***	0.470***	0.477***	0.475***	0.431**	0.428
	(0.029)	(0.100)	(0.126)	(0.098)	(0.124)	(0.219)	(0.266)
y_{t-1} (true value = 0.5)	0.187***	0.430***	0.433***	0.435***	0.439***	0.362*	0.361
	(0.029)	(0.098)	(0.121)	(0.096)	(0.120)	(0.194)	(0.237)
Relevance test:							
Bias against OLS[a)]		< 5%		5~10%		5~10%	
Test size[b)]		10~15%		15~20%		15~20%	
Exogeneity test:							
p-value[c)]		0.638	0.609	0.878	0.901	0.861	0.840

a) This measures the relative bias of the 2SLS estimate against the bias of the OLS estimate at a significance level of 5%. For example, an entry of '5~10%' indicates that the bias of the 2SLS estimate is as much as 5~10% of the bias of the OLS estimate. The larger this number is, the weaker the IVs are for the purpose of estimating the coefficients.
b) This measures the actual size of the Wald test when the specified IVs are used at a 5% significance. For example, an entry of '10~15%' indicates that the Wald test based on the IVs would reject the null hypothesis 10~15% of the time even though the null is correct. The larger this number is, the weaker the IVs are for the purpose of testing the significance of coefficients.
c) This is to test whether the IVs are exogenous (i.e., orthogonal to the structural disturbances) and called an over-identifying restrictions test when the number of IVs exceeds the number of endogenous regressors. As the null hypothesis here is that all IVs are exogenous, large *p*-values cannot reject the null.
The symbols of *** and ** attached to the coefficient estimates indicate that its coefficient is significantly different from zero at the 1% and 5% levels, respectively.

The estimation results are summarized in Table 1. As expected, the OLS estimation yields biased estimates toward zero because it does not control for the time-specific effects. The coefficient estimate for is −0.053 and significant at the 10% level; this estimate is opposite in sign to the true value of 0.3 and its 95% confidence interval does not include the true value. In contrast, the 2SLS estimate for β_1 is much less biased and its 95% confidence interval includes the true value of 0.3. The GMM estimates are almost same as the 2SLS ones, although their standard errors are larger. Therefore, concerning the Granger-causality test, the 2SLS and GMM methods produce unbiased (correct) results but the OLS method produces biased (wrong) results.

Regarding the quality of the IVs used, the lagged values perform well. The relevance test shows that the bias of the 2SLS can be less than 5% of the OLS bias when lag 2 and 3 variables are used as IVs. With the same IVs, the actual test size is 10~15% when the significance level of the test is 5%. The IVs are exogenous as confirmed by the over-identifying restrictions; the p-values greater than 0.6 cannot reject the null hypothesis that the IV are exogenous.

6. Summary and Conclusions

This study points out a possible bias due to time-specific effects in the Granger-causality test. Treating the lagged time-specific effects as measurement errors in explanatory variables, this study suggests use of the instrumental variables from the literature on measurement errors. In this study we use appropriately lagged values as instrumental variables and assess their relevance and exogeneity.

The discussion in this study will be useful for testing causality of economic variables which are sensitive to time-specific effects, particularly variables related to the tourism research as tourism is highly sensitive to economic fluctuations; this application is planned for future work. In addition, regarding a search for instrumental variables, Dagenais and Dagenais (1997) and Lewbel (1997) develop alternative sets of instrumental variables using higher moments of the contemporaneous variables in a model. These higher moments are shown asymptotically uncorrelated with measurement errors; in future work, these higher moments will be supplemented as instrumental variables in addition to lagged regressors for possible improvement in efficiency.

Endnotes

[1] The denominator in Eq.(6) becomes large as the sample size T increases so that the bias of the 2SLS decreases. Thus, the 2SLS is consistent.

References

Dagenais, M and Dagenais, DL 1997, 'Higher moment estimators for linear regression models with errors in the variables', *Journal of Econometrics*, Vol. 76, pp. 193-221.

Granger, CWJ 2004, 'Time series analysis, cointegration, and applications', *American Economic Review*, Vol. 94, No. 3, pp. 421-425.

Griliches, Z and Hausman, JA 1986, 'Errors in variables in panel data', *Journal of Econometrics*, Vol. 31, pp. 93-118.

Hahn, J and Hausman, J 2002, 'Notes on bias in estimators for simultaneous equation models', *Economics Letters*, Vol. 75, pp. 237-241.

Hansen, LP 1982, 'Large sample properties of generalized method of moments estimators', *Econometrica*, Vol. 50, pp. 1029-1054.

Hausman, J 2001, 'Mismeasured variables in econometric analysis: problems from the right and problems from the left', *Journal of Economic Perspectives*, Vol. 15, No. 4, pp. 57-67.

Himmelgerg, CP and Petersen, BC 1994, 'R&D and internal finance: a panel study of small firms in high-tech industries', *The Review of Economics and Statistics*, Vol. 76(1), pp. 38-51.

Lewbel, A 1997, 'Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D', *Econometrica*, Vol. 65(5), pp. 1201-1213.

Stock, JH and Yogo, M 2005, 'Testing for weak instruments in linear IV regression', in DWK Andrews and JH Stock (ed.), *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg*, Cambridge: Cambridge University Press, Chapter 5.

Wansbeek, T 2001, 'GMM estimation in panel data models with measurement error', *Journal of Econometrics*, Vol. 104, pp. 259-268.