NBER TECHNICAL PAPER SERIES

IMPLEMENTING CAUSALITY TESTS WITH PANEL DATA, WITH AN EXAMPLE FROM LOCAL PUBLIC FINANCE

Douglas Holtz-Eakin

Whitney Newey

Harvey Rosen

Technical Working Paper No. 48

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1985

We are grateful to Maria San Segundo for assistance with the computations and to Research Services at Princeton University for help in acquiring the data. We thank Gary Chamberlain, James Poterba, and members of Princeton's Industrial Realtions Seminar for useful comments. This research was supported in part by NSF Grants SES-8419238 and SES-8410249. The research reported here is part of the NBER's research program in Taxation and project in Government Budget. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Paper #48
June 1985

Implementing Causality Tests with Panel Data, with an Example from Local Public Finance

ABSTRACT

This paper considers estimation and testing of vector autoregression coefficients in panel data, and applies the techniques to analyze the dynamic properties of revenues, expenditures, and grants in a sample of United States municipalities. The model allows for nonstationary individual effects, and is estimated by applying instrumental variables to the quasi-differenced autoregressive equations. Particular attention is paid to specifying lag lengths and forming convenient test statistics. The empirical results suggest that intertemporal linkages are important to the understanding of state and local behavior. Such linkages are ignored in conventional cross sectional regressions. Also, we present evidence that past grant revenues help to predict current expenditures, but that past expenditures do not help to predict current revenues.

Douglas Holtz-Eakin
Department of Economics
Columbia University
New York, NY 10027
212 280-3681

Whitney Newey
Department of Economics
Princeton University
Princeton, NJ 08544
609 452-4000

Harvey Rosen
Department of Economics
Princeton University
Princeton, NJ 08544
609 452-4022

I. Introduction

Vector autoregressions are now a standard part of the applied econometrician's tool kit. To be sure, the interpretation of vector autoregressions remains controversial. Some argue that the results reveal causal relationships between variables. (See, e.g., Granger [1969] or Sims [1972].) Others such as Leamer [1984] deny that causality is revealed in any economically meaningful sense. Nonetheless, even critics of the causal interpretation admit that vector autoregressions are a parsimonious and useful means to summarize the time series "facts."

To date, vector autoregressive techniques have been used mostly to analyze macroeconomic time series where there are dozens of observations. (See, e.g., Taylor [1980] or Ashenfelter and Card [1982].) In principle, these techniques should apply equally well to disaggregate data. Does an individual's hours of work "cause" his or her wage? (See Lundberg [1984].) Does a community's tax structure "cause" its level of local public expenditure? Unlike macro applications, however, the available time series on micro units are typically quite short. Many of the popular panel data sets, for example, have no more than ten or twelve years of observations for each unit.[1]

Because the number of time periods observed is relatively small, the standard causality tests cannot be applied directly to panel data. In an important paper, Chamberlain [1983] developed a technique for estimating

---

[1] Nevertheless, our techniques are appropriate for more "traditional" macroeconomic applications. For example, Taylor [1980] examined and compared the time series properties of several key macroeconomic variables for a number of European countries. Our methods could be used to execute formal tests of similarity between them.

vector autoregressions using panel data. However, he does not explore the problems surrounding identification and hypothesis testing that are of major importance to practitioners.

The purpose of this paper is to explain carefully how to execute tests of causality on panel data so that this powerful tool becomes available to microeconomists. In doing so, we stress that it is not our intention to take sides in the debate over whether the results reveal anything about "causality." Each user must decide this for him or herself. However, for the sake of readability, we will henceforth refrain from putting quotation marks around that word.

Section II presents the basic model, discusses the relationship to conventional panel data estimators, and highlights the critical issues in a simple context. Section III presents a more general treatment of the statistical theory and computations. Section IV applies the methods to an example from local public finance. Specifically, we investigate the dynamic relationships between community expenditures, revenues from local sources, and grants. Section V provides a brief summary.

II. Basic Setup

A. What's the Problem?

We begin by considering causality tests in their usual time series context. The issue is to determine the causal relationship between the detrended variables x and y, on which the investigator has a large number of observations. The variable x is said to not (Granger) cause the variable y if:

$$(2.1)\quad E\{y_t | y_{t-1}, y_{t-2}, \ldots, y_1, x_{t-1}, x_{t-2}, \ldots, x_1\} = E\{y_t | y_{t-1}, y_{t-2}, \ldots, y_1\}$$

where $E\{\cdot|\cdot\}$ denotes a linear projection. Intuitively, if one's prediction of y_t, given the history of y, cannot be improved by including the history of x, then x does not cause y.

Essentially, the procedure is to estimate a regression of the form:

$$(2.2)\quad y_t = \alpha_0 + \sum_{\ell=1}^{m} \alpha_\ell y_{t-\ell} + \sum_{k=1}^{n} \delta_k x_{t-k} + u_t$$

where the α's and δ's are parameters and the lag lengths m and n are sufficient to ensure that u_t is a white noise error. While it is not essential that m equal n, we follow typical practice by assuming that they are identical. The test of whether x causes y is simply a test of the joint hypothesis that $\delta_1 = \delta_2 = \ldots = \delta_m$ are all equal to zero. This can be done using standard F-tests.

To perform the test, there must be enough observations on x and y to obtain consistent estimates of the parameters in equation (2.2). Panel data generally do not have the requisite number of observations. Instead, there often are a great number of cross-sectional units, but only a few years worth of data on each unit. To estimate any parameters, investigators typically pool data from different units, a procedure which imposes the constraint that the underlying structure is the same for each cross-sectional unit.

Given this, why not simply stack all the time series-cross section observations together and use them to estimate equation (2.2)? The main pitfall of such a procedure is that it ignores the possibility that each unit has an "individual effect"--which translates in practice to its own intercept. The individual effect summarizes the influence of unobserved variables which have a persistent effect on the dependent variable. For example, a worker's wage rate each period may be affected by his or her "ambition," or a community's expenditure each period might be affected by its "political make-up". To the extent that the other right hand side variables are correlated with the individual effect, its omission results in inconsistent estimates. Although there are standard methods for estimating individual effects, they are not appropriate for our problem. We now explain why.

B. The Standard Individual Effects Model

Assume that there are N cross-sectional units observed over T periods. Let i index the cross-sectional observations and t the time periods. The goal is to determine whether variable x causes variable y. In analogy to equation (2.2), the temptation is to specify the relationship as:

$$(2.3)\quad y_{it} = \alpha_0 + \sum_{\ell=1}^{m} \alpha_\ell y_{it-\ell} + \sum_{\ell=1}^{m} \delta_\ell x_{it-\ell} + u_{it} \qquad \begin{array}{l} i=1,\ldots,N \\ t=(m+1),\ldots,T \end{array}$$

If equation (2.3) is the correct model (and u_{it} is white noise), then the conventional techniques do indeed apply. If, however, one assumes the existence of an individual effect (f_i) for the i^{th} cross-sectional unit, the model becomes:[2]

$$(2.4)\quad y_{it} = \alpha_0 + f_i + \sum_{\ell=1}^{m} \alpha_\ell y_{it-\ell} + \sum_{\ell=1}^{m} \delta_\ell x_{it-\ell} + u_{it}$$

A standard method of estimating the individual effect is to first difference the data to eliminate f_i and then use ordinary or generalized least squares to estimate the differenced equation:[3]

$$(2.5)\quad y_{it} - y_{it-1} = \sum_{\ell=1}^{m} \alpha_\ell (y_{it-\ell} - y_{it-\ell-1}) + \sum_{\ell=1}^{m} \delta_\ell (x_{it-\ell} - x_{it-\ell-1})$$

$$+ (u_{it} - u_{it-1}) \qquad i=1,\ldots,N \qquad t=(m+2),\ldots,T$$

A quick examination of equation (2.5) indicates the flaw with this approach in the current context: because y_{it-1} depends on u_{it-1} (via equation (2.3)) the error term $(u_{it} - u_{it-1})$ is correlated with the regressor $(y_{it-1} - y_{it-2})$.

The fact that differencing can induce a simultaneity problem is well-known from the conventional literature on time series analysis and has

---

[2] If the x's and y's require detrending, equation (2.4) can be augmented with a time trend. Estimation and identification are not complicated by this addition; hence, for simplicity, it is suppressed below.

[3] Another common technique is to compute the difference between each variable and its time mean (by cross-sectional unit) to eliminate f_i. See, e.g., Lundberg [1984]. In the current context, this procedure will yield inconsistent estimates because the time means are correlated with all the error terms for each cross-sectional unit.

been explored in a panel data context. (See, e.g., Nickell [1981].) The usual solution is to employ an instrumental variables estimator. Here, too, this turns out to appropriate, but it is implemented in a different fashion than is typical. This is because, as we demonstrate below, the variables which are legitimate candidates for use as instrumental variables change over time. The technique is outlined in Section III.

Before leaving the standard model, we note that heteroskedasticity is likely to be a problem in the panel context--different units may be expected to have error terms with unequal variances. Efficient estimation and correct formulae for standard errors require that heteroskedasticity be taken into account.

C. Chamberlain's Solution

Equation (2.4) specifies that the parameters are constant not only across different units, but also over time. Similarly, each individual effect is time invariant. Both assumptions underlie virtually all applications using individual effects. In contrast, Chamberlain's [1983] investigation of the notion of causality (conditional upon the individual effect) allows both the parameters and the individual effect to vary over time. In this section we present a simple example of Chamberlain's approach, highlighting the features of interest prior to a more general treatment.

To keep the notation manageable, consider a panel extending over four periods and a model with a first order lag structure. Then the model is:

$$
\begin{aligned}
y_{i1} &= \alpha_{01} + \alpha_{11}y_{i0} + \delta_{11}x_{i0} + \psi_1 f_i + u_{i1} \\
y_{i2} &= \alpha_{02} + \alpha_{12}y_{i1} + \delta_{12}x_{i1} + \psi_2 f_i + u_{i2} \\
y_{i3} &= \alpha_{03} + \alpha_{13}y_{i2} + \delta_{13}x_{i2} + \psi_3 f_i + u_{i3} \\
y_{i4} &= \alpha_{04} + \alpha_{14}y_{i3} + \delta_{14}x_{i3} + \psi_4 f_i + u_{i4}
\end{aligned}
\tag{2.6}
$$

where ψ_t (t=1,2,3,4) is the coefficient multiplying the individual effect in period t. The standard model implicitly restricts the ψ_t's equal to one in each period. Notice that y_{i0} and x_{i0} are not observed by the econometrician, so the equation for y_{i1} cannot be estimated. We include it in the system (2.6) because of its implications for later observations.

To test for causality, one must estimate the equations (2.6) jointly and determine whether the data are consistent with the restriction:

$$(2.7) \quad \delta_{11} = \delta_{12} = \delta_{13} = \delta_{14} = 0$$

Given that the individual effects now have time-varying coefficients, how should the system be estimated? Simply differencing the data will no longer "work"--the individual effects will not disappear. Chamberlain suggests the following transformation: multiply the equation for time period t by (ψ_{t+1}/ψ_t) and then subtract it from the equation for period t+1. The result is:

$$(2.8) \quad \begin{aligned} y_{i2} &= (\alpha_{02} - r_2\alpha_{01}) + (\alpha_{12} + r_2)\, y_{i1} - r_2\alpha_{11}y_{i0} + \delta_{12}x_{i1} - r_2\delta_{11}x_{i0} \\ &\qquad + (u_{i2} - r_2u_{i1}) \\ y_{i3} &= (\alpha_{03} - r_3\alpha_{02}) + (\alpha_{13} + r_3)\, y_{i2} - r_3\alpha_{12}y_{i1} + \delta_{13}x_{i2} - r_3\delta_{12}x_{i1} \\ &\qquad + (u_{i3} - r_3u_{i2}) \\ y_{i4} &= (\alpha_{04} - r_4\alpha_{03}) + (\alpha_{14} + r_4)\, y_{i3} - r_4\alpha_{13}y_{i2} + \delta_{14}x_{i3} - r_4\delta_{13}x_{i2} \\ &\qquad + (u_{i4} - r_4u_{i3}) \end{aligned}$$

where:

$$r_2 = (\psi_2/\psi_1)$$
$$r_3 = (\psi_3/\psi_2)$$
$$r_4 = (\psi_4/\psi_3)$$

The procedure is to estimate (2.8) jointly and test the causality hypothesis. Again, some of the right hand side variables are correlated with the (transformed) error term, so an instrumental variables estimation technique is required. The natural choices for the instrumental variables are (appropriately) lagged values of x and y. However, not all of the equations in (2.8) are estimable. To see why, consider the equation for $t=4$. Recall the necessary condition for identification, that there be as many instrumental variables as right-hand side variables.[4] There are four right hand side variables (in addition to the constant) y_{i3}, y_{i2}, x_{i3} and x_{i2}. The lagged variables available as instruments, i.e., that are uncorrelated with $(u_{i4} - r_4u_{i3})$, are x_{i1}, x_{i2}, y_{i1} and y_{i2}. We thus have four instruments to match our four right hand side variables, and hence, no identification problem. But for $t=3$ and $t=2$ there are not enough instruments; the equations for these periods cannot be estimated.

Note that as more lags are added, a larger number of lagged observations are necessary for use as instrumental variables. Thus, in the case of a second order lag in (2.6), even the equation for period 4 would not be usable. In the absence of an additional year of data, no parameters could be identified. Since the investigator will not know the true lag structure a priori, all specifications are potentially open to this problem. We will refer to this phenomenon, which does not appear to have been discussed in previous treatments of this issue, as the lag truncation problem.

Importantly, the ψ's are not identified as only their ratios appear. Intuitively, since each ψ appears only interacting with the latent variable, disentangling its independent effect is impossible. As already suggested,

---

[4] For a discussion of this criterion, see Fisher [1966].

the identification of the other parameters is dependent upon the number of instruments available, and a full discussion is deferred to Section III.

We conclude this section by asking why one should depart from the standard model and examine this more general specification. Chamberlain does not address this question, but we can think of two possible reasons: 1) Time invariant parameters indicate that stationary behavior has been achieved. Chamberlain's approach explicitly assumes that the investigator has observed the entire time series for each unit, literally from its "birth" (thus circumventing the lag truncation problem). It is natural to allow for the possibility that the behavior of such a unit will change as it "matures." We do not assume that the panel data encompass the entire life of the unit. Behavior of "mature" units is more likely to be time invariant, but of course stationarity still may not obtain. 2) As long as it is possible to test whether the parameters are time invariant, why not do so? To the extent that investigators impose an unchanging structure when inappropriate, inconsistent estimates will result.

III. Statistical Inference

This section presents a discussion of identification, estimation and hypothesis testing. The criterion we use for identification is that there must be a sufficient number of instrumental variables to allow estimation of the equation in question. This leads directly to an instrumental variables estimator suggested by Chamberlain which has a generalized least squares (GLS) interpretation. In addition, we introduce a method for hypothesis testing, an issue which has not been discussed in this context before. Specifically, we show how to compute test statistics for some pertinent hypotheses--those of correct lag length, stationarity, and causality.

A. Identification

Consider the model (2.5), above, which results from assuming that the parameters are time invariant and differencing to eliminate the individual effect (see Section II). (The more general case is addressed below.) As usual, we assume that the error term, u_{it} , is uncorrelated with all past values of y and x , and the individual effect:

$$(3.1)\quad E\{y_{is}u_{it}\} = E\{x_{is}u_{it}\} = E\{f_i u_{it}\} = 0,\ s < t$$

The orthogonality conditions (3.1) can be used to identify the parameters of (2.5), since the disturbance term v_{it} $(= u_{it} - u_{it-1})$ will be uncorrelated with y_{it-s} and x_{it-s} for $s \geq 2$. The equation for each time period t has $2m$ right-hand side variables. To identify the parameters, there must be at least this many instrumental variables. The $2(t-2)$ variables $[y_{it-2},\ldots,y_{i1},x_{it-2},\ldots,x_{i1}]$ are available as instrumental variables to estimate the equation for time period t . Thus, to have at least as many instrumental variables as right-hand side variables, it must be true that $2(t-2) \geq 2m$, or $t \geq m+2$.[5]

Given our assumed lag structure, it is impossible to estimate the equations (2.5) for time periods before $t = m+2$. Thus, these equations are ignored. Clearly, the decision about which equations to "ignore" depends crucially on assumptions concerning lag length. If we make an incorrect assumption and truncate the lag distribution, the parameter estimates will be inconsistent. It is interesting to recall that Chamberlain circumvents this problem by assuming that the first observation in the panel coincides with the "birth" of the unit. Hence, he assumes that the entire lag structure is observed. Such an assumption seems untenable in most applications. An alternative assumption which is both more likely and in the spirit of causality testing in time series applications is the hypothesis that the birth of each unit occurred long before the first period of observation.

A difficulty associated with this more realistic assumption is that most panel data sets encompass a relatively small number of time periods. This creates a potential identification problem when the lag length is unknown. Causality testing will not be possible unless some a priori restriction is imposed on the lag distribution. We do not present a general treatment of the problem, and instead use only the straightforward restriction implicitly imposed above: that if the largest lag length is m, then the number of time periods T is greater than $m+2$.[6]

---

[5] A sufficient condition for identification is that in the limit the cross-product matrix between the instruments and the right-hand side variables be non-singular.

[6] A more general treatment might use methods similar to those of Pakes and Griliches [1984].

We turn now to issues that arise when the parameters are not constant across time periods. The simplest relaxation of stationarity is to introduce a time varying intercept; i.e. to allow α to depend on t. In this case, the differenced version (analogous to equation (2.5)) is:

$$(3.2)\quad y_{it} - y_{it-1} = a_t + \sum_{\ell=1}^{m} \alpha_\ell (y_{it-\ell} - y_{it-\ell-1}) + \sum_{\ell=1}^{m} \delta_\ell (x_{it-\ell} - x_{t-\ell-1}) + (u_{it} - u_{it-1})$$

where $a_t = \alpha_{0t} - \alpha_{0t-1}$. A different constant term now appears in the equation for each time period. The preceding identification discussion remains unchanged because a vector of 1's can be added to the instrumental variables for the equation for period t to identify the intercept. Operationally, all that the inclusion of a separate constant for each time period requires is the introduction of time dummy variables.

The most general specification is to allow all of the parameters to depend on the time period.[7] The general form for this case (refer to equation (2.6)) is:

$$(3.3)\quad y_{it} = \alpha_{0t} + \psi_t f_i + \sum_{\ell=1}^{m} \alpha_{\ell t} y_{it-\ell} + \sum_{\ell=1}^{m} \delta_{\ell t} x_{it-\ell} + u_{it} \quad [t=(m+1),\ldots,T]$$

This case, considered by Chamberlain, requires the transformation discussed in Section II to eliminate the individual effect. It leads to:

$$(3.4)\quad y_{it} = a_t + \sum_{\ell=1}^{m+1} c_{\ell t} y_{i,t-\ell} + \sum_{\ell=1}^{m+1} d_{\ell t} x_{i,t-\ell} + v_{it} \quad [t=(m+2),\ldots,T]$$

---

[7] A special case which may be of particular interest occurs when the α's and δ's are time invariant, but ψ_t is not.

where: $r_t = (\psi_t/\psi_{t-1})$

$$a_t = \alpha_{0t} - r_t\alpha_{0t-1}$$

$$c_{1t} = r_t + \alpha_{1t}$$

$$c_{\ell t} = \alpha_{\ell t} - r_t\alpha_{\ell-1,t-1} \qquad [\ell=2,\ldots,m] \qquad (3.4a)$$

$$c_{m+1,t} = -r_t\alpha_{m,t-1}$$

$$d_{1t} = \delta_{1t}$$

$$d_{\ell t} = \delta_{\ell t} - r_t\delta_{\ell-1,t-1} \qquad [\ell=2,\ldots,m]$$

$$d_{m+1,t} = -r_t\delta_{m,t-1}$$

$$v_{it} = u_{it} - r_t u_{i,t-1} .$$

Observe that in each of the equations in (3.4) there are 2(m+1) right hand side variables other than the constant, or a total of (2m+3). To identify the parameters of (3.4) an equal number of instrumental variables is required. Note that this is greater than the (2m+1) required in equation (3.2). Since the instrumental variables vector is

$$[1, y_{i\,t-2}, \ldots, y_{i1}, x_{i\,t-2}, \ldots, x_{i1}],$$

it is now required that $t \geq m + 3$ to have a sufficient number of instrumental variables for the equation for time period t .[8] Thus, as one would suspect, allowing for time varying parameters makes identification more difficult. To see this most clearly, note that if $T = (m+2)$, the parameters of the equation for the last year in the panel can be identified in the stationary case, but not when the parameters are time-varying. More generally, with time-varying parameters, the equation for one more year of data in the panel cannot be identified.

Not all of the parameters of equation (3.3) can be recovered from the estimates of equation (3.4). The parameters of (3.4) for the time period t depend upon the parameters for

---

[8] Hence, the equation for the earliest year estimated will be just exactly identified.

time period t-1. Since this is true of all periods and we do not observe the entire life of the unit, it is not in general possible to "solve back" to all of the original parameters. However, it is still possible to test for non-causality in the non-stationary case because, as the definition of $d_{\ell t}$ makes clear, all the δ's being zero implies that all the d's are zero. Moreover, under certain conditions some of the parameters of equation (3.3) can be recovered from the c's and d's estimated from equations (3.4). From (3.4a), in all cases, δ_{1t} is identified directly from d_{1t}. Further, in the special case of stationary individual effects ($r_t=1$), by solving equations (3.4a) recursively, one can obtain all the δ's and α's (except for the intercept), provided that the number of equations estimated is at least as large as the lag length. To see this, note that the number of estimated parameters is (T-m-2)(2m+2); this has to exceed the number of underlying parameters, (T-m-1)(2m). But $(T-m-2)(2m+2) \geq (T-m-1)(2m)$ implies that $(T-m-2) \geq m$.

Finally, if all parameters are stationary, one is back to the standard individual effects model, and all the parameters can be recovered.

B. Estimation

Unlike the discussion of identification, our presentation of the estimation procedure will proceed from the most general case (all parameters varying over time) to more restrictive specifications; showing how any restriction may be imposed. The spirit of the procedure is straightforward. For each time period, we have available a set of instrumental variables which may be used to obtain consistent estimates. However, the list of instrumental variables differs for each time period, so the procedure that is familiar from the typical simultaneous equations framework must be modified.

The presentation requires some additional notation. Let

$$Y_t = [y_{1t},\ldots,y_{Nt}]' \text{ and } X_t = [x_{1t},\ldots,x_{Nt}]'$$

be Nx1 vectors of observations on units for a given time period. Let

$$W_t = [e,Y_{t-1},\ldots,Y_{t-m-1},X_{t-1},\ldots,X_{t-m-1}]$$

be the Nx(2m+3) vector of right hand side variables for equations (3.4) where e is an Nx1 vector ones. Let

$$V_t = [v_{1t},\ldots,v_{Nt}]'$$

be the Nx1 vector of transformed disturbance terms, and let

$$B_t = [a_t,c_{1t},\ldots,c_{m+1,t},d_{1t},\ldots,d_{m+1,t}]'$$

be the (2m+3)x1 vector of coefficients for the equations. Then we can write equations (3.4) as:[9]

$$(3.5)\quad Y_t = W_tB_t + V_t \qquad [t=(m+3),\ldots,T]$$

To combine all the observations for each time period, we can "stack" equations (3.5). Let

$$Y = [Y'_{m+3},\ldots,Y'_T]'$$
$$[(T-m-2)N \times 1]$$

$$B = [B'_{m+3},\ldots,B'_T]'$$
$$[(T-m-2)(m+3) \times 1]$$

$$V = [V'_{m+3},\ldots,V'_T]'$$
$$[(T-m-2)N \times 1]$$

$$W = \text{diag}[W'_{m+3},\ldots,W'_T]$$
$$[(T-m-2)N \times (T-m-2)(m+3)]$$

---

[9] Observe that we exclude $t \le (m+2)$ because these equations are not identified. See the discussion above.

where diag[.] denotes a block diagonal matrix with the given entries along the diagonal. With this, the observations for equations (3.4) can be written:

(3.6) $Y = WB + V$

So far the discussion is quite similar to that of a classical simultaneous equations system where the equations are indexed by t and the observations by i. However, in the classical system the predetermined variables--which serve as instrumental variables--are the same for each equation. As we have stressed, this is no longer the case. The matrix of variables which qualify for use as instrumental variables in period t is:

$$Z_t = [e, Y_{t-2}, \ldots, Y_1, X_{t-2}, \ldots, X_1]$$

which changes with t. To allow the different instrumental variables for different equations, we choose the matrix of instrumental variables for the system in (3.6) to be block diagonal. Consider the matrix Z defined as:

$$Z = \text{diag}[Z_{m+3}, \ldots, Z_T]$$

The orthogonality conditions ensure that

$$\underset{N\to\infty}{\text{plim}} \ (Z'V)/N = \underset{N\to\infty}{\text{plim}} \begin{bmatrix} (Z'_{m+3} \ V_{m+3})/N \\ \\ (Z'_T \ V_T)/N \end{bmatrix} = 0$$

It follows directly that Z is the appropriate choice of instrumental variables for (3.6).[10]

[10] Limits are taken as $N\to\infty$, with T fixed.

To estimate B, premultiply (3.6) by Z' to obtain:

$$(3.7) \quad Z'Y = Z'WB + Z'V$$

We can then form a consistent instrumental variables estimator by applying GLS to this equation. As usual, such an estimator requires knowledge of the covariance matrix of the (transformed) disturbances, $Z'V$. This covariance matrix, Ω, is given by

$$\Omega = E\{Z'VV'Z\} .$$

Ω is not known and therefore must be estimated. To do so, let $\tilde{B}$ be the preliminary consistent estimator of B formed by estimating the coefficients of the equations for time periods t using two-stage least squares (2SLS) on the equation for each time period alone--using the correct list of instrumental variables.[11] Using these preliminary estimates, form the vector of residuals for period t: $\tilde{V}_t = Y_t - W_t\tilde{B}_t$. A consistent estimator of (Ω/N) is then formed by:[12]

$$(3.8) \quad (\tilde{\Omega}/N)_{rs} = \sum_{i=1}^{N} (v_{ir}v_{is}Z'_{ir}\, Z_{is})/N$$

where v_{it} (t=r,s) is the ith element of V_t and Z_{it} is the ith row of Z_t . Finally, $\tilde{\Omega}$ is used to form a GLS estimator of the entire parameter vector, $\hat{B}$, using all the available observations:

$$(3.9) \quad \hat{B} = [W'Z(\tilde{\Omega})^{-1}Z'W]^{-1}W'Z(\tilde{\Omega})^{-1}Z'Y \ .$$

---

11 That is:

$$\tilde{B}_t = [W_t'Z_t(Z_t'Z_t)^{-1}Z_t'W_t]^{-1}\, W_t'Z_t(Z_t'\, Z_t)^{-1}Z_t'Y_t \ .$$

12 This procedure is an extension of White's [1980] heteroskedasticity consistent covariance matrix estimator. It is appropriate if $E\{v_{ir}v_{js}\}=0$ for i,j,r,s such that $i\neq j$, that is, error terms for different units are uncorrelated. Note that common factors are controlled by inclusion of time dummy variables in the estimating equations.

To summarize, the estimator is formed in three steps: i) Estimate the equation for each time period using 2SLS. ii) Using the residuals and the matrix of instruments, estimate the joint covariance matrix Ω . iii) Estimate all parameters jointly using GLS on the stacked equations. Since there are no non-linearities in the parameters, our procedure is easily carried out using standard computer software for matrix manipulations.

Efficiency

Several comments concerning the efficiency of $\hat{B}$ are in order. $\hat{B}$ is efficient in the class of instrumental variable estimators which use linear combinations of the instrumental variables. This follows directly from the results of Hansen [1982]. (See also White [1982].) However, just as 3SLS on an entire system of equations may be more efficient than 3SLS on a subset of the equations, it may be possible to improve the efficiency by jointly estimating both the equation for y_{it} given past values of y and x and the equation for x_{it} given the history of x and y . Because such a procedure is much more complicated and requires us to make assumptions concerning the lag structure of x , we do not pursue it here.

Linear Constraints

Causality tests require estimating B subject to linear constraints. The most obvious is the constraint that all lagged values of x have zero coefficients. For the sake of illustration, consider the simple example in (2.6). After the transformation leading to (2.8), only the final year (t=4) of the data is identified. Using the general notation introduced in (3.4), the linear constraint that x not cause y is that $d_{14} = d_{24} = 0$ and that the constrained and unconstrained values of b_4, c_{14}, and c_{24} be identical. Using matrix notation, this may be written:

$$\begin{bmatrix} b_4 \\ c_{14} \\ c_{24} \\ d_{14} \\ d_{24} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} b_4 \\ c_{14} \\ c_{24} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

More generally, a simple way to formulate such constraints is to specify that

$$\text{(3.10)} \qquad B = H\gamma + G$$

where γ is a kx1 vector of parameters, H is a constant matrix with dimensions $[(T-m-2)(m+3) \times k]$ and G is a constant vector of the same dimension as B .[13] Since γ is the restricted parameter vector, it has dimension smaller than B .

Replacing B by $H\gamma + G$ and subtracting WG from both sides of (3.6) gives

$$\text{(3.11)} \qquad \tilde{Y} = Y - WG = WH\gamma + V = \tilde{W}\gamma + V$$

This equation has exactly the same form as (3.6). Thus, we can estimate γ as before--using the data matrices transformed by G and H .

Another constraint of interest is that of stationary coefficients. Consider again the simple example (2.6). Under the assumption of stationary coefficients, the equations for time periods 3 and 4 are both identified. To test for stationary coefficients (other than the time dummies), estimation involves using first-differences of the data. First differencing the equations (2.6) and re-arranging yields

---

[13] The rank of H must be k for the restrictions to be unique.

$$y_{i3} = b_3 + (1+\alpha_1)y_{i2} - \alpha_1 y_{i1} + \delta_1 x_{i2} - \delta_1 x_{i1} + (u_{i3} - u_{i2})$$

(2.6')

$$y_{i4} = b_4 + (1+ \alpha_1)y_{i3} - \alpha_1 y_{i2} + \delta_1 x_{i3} - \delta_1 x_{i2} + (u_{i4} - u_{i3}) .$$

The appropriate choice of matrices is:

$$\begin{bmatrix} b_3 \\ c_{13} \\ c_{23} \\ d_{13} \\ d_{23} \\ b_4 \\ c_{14} \\ c_{24} \\ d_{14} \\ d_{24} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} b_3 \\ b_4 \\ \alpha_1 \\ \delta_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Note that the choice of the matrix H and the vector G implicitly differences the data prior to estimation.

C. Hypothesis Testing

In this section we discuss the computation of statistics to test the hypothesis that x does not cause y , that the parameters are stationary, that m is the correct lag length, and other possible hypotheses. In each case, the test statistic revolves around the sum of squared residuals, resulting in tests with a chi-square distribution in large samples. The primary consideration is to transform the residuals to ensure that they have the requisite statistical properties. Further, we consider two additional topics: tests when parameters are not identified under the alternative hypothesis and sequences of tests.

We consider only tests of linear restrictions on the estimated parameters, B . Consider the null hypothesis:

$$(3.13) \quad H_0 : B = H\gamma + G$$

where the notation is as before. As we have shown, it is straightforward to impose this restriction during estimation. Let

$$(3.14) \quad \begin{aligned} Q &= (Y-W\hat{B})'Z(\tilde{\Omega})^{-1}Z'(Y-W\hat{B})/N \\ Q_R &= (Y-\tilde{W}\hat{\gamma})'Z(\tilde{\Omega})^{-1}Z'(Y-\tilde{W}\hat{\gamma})/N \end{aligned}$$

Q is the unrestricted sum of squared residuals and Q_R is the restricted sum of squared residuals. Q and Q_R each have a chi-square distribution as N grows.[14] By analogy with the F-statistic in the standard linear model, an appropriate test statistic is

$$(3.15) \quad L = Q_R - Q$$

L has the form of the numerator of the F statistic. By construction, the covariance matrix of the transformed disturbances is an identity matrix. As a result, L has a chi-squared distribution with degrees of freedom equal to the degrees of freedom of Q_R minus the degrees of freedom of Q. When all of the parameters are identified under both the null and the alternative hypotheses, the degrees of freedom of Q is equal to the number of

---

[14]To see this, let P be the matrix such that $P'P = \Omega^{-1}$. Then premultiplying (3.7) by P results in:

$$PZ'Y/\sqrt{N} = (PZ'W/\sqrt{N})B + (PZ'V/\sqrt{N})$$

Note that asymptotically, the disturbance $P'Z'V/\sqrt{N}$ is normally distributed with a covariance matrix equal to the identity matrix. As usual, sums of these squared residuals will have a chi-square distribution.

instrumental variables (the number of rows of $Z'V$ in (3.7)) minus the number of parameters, i.e. the dimension of B. Similarly, Q_R has degrees of freedom equal to the number of instrumental variables minus the number of restricted parameters--the dimension of γ. Thus, for (3.7) L has degrees of freedom equal to the dimension of B minus the dimension of γ.[15]

Unidentified Parameters

When executing causality tests with panel data, it is often the case that some parameters are not identified under the alternative hypothesis. For example, under the null hypothesis that x does not cause y, lagged x's can be used as instrumental variables for lagged y's. This is because lagged x's will be correlated with lagged y's via the individual effect. Use of these instruments permits us to identify the parameters in (3.6). Under the alternative hypothesis, the greater number of parameters means that not all of the parameters are identified. Nonetheless, a test of the null hypothesis is still possible. The method is analogous to conducting a Chow test with insufficient observations.[16] (See Fisher [1976].)

To address this issue, consider again the system (2.6) which leads to (2.8). Under the null hypothesis that x does not cause y, there are three parameters (b_3, c_{13}, c_{23}) in the equation for time period 3 and the same number of instrumental variables (e, y_1, x_1) available. Under

---

[15] L can be thought of as the extension of the Gallant and Jorgenson [1979] test statistic for 3SLS to this application. Of course, we could use other asymptotically equivalent test statistics to test the null hypothesis. In fact, the well known Wald test is numerically equivalent to our L. Newey and West [1985] discuss the relationship between L and other test statistics; including regularity conditions.

[16] The analogy is not exact because we consider more general hypotheses than simply hypotheses which impose equality across equations and because the joint covariance matrix across (3.7), above, and (3.16), below, is not diagonal.

the alternative hypothesis d_{13} and d_{23} must also be estimated; the number of parameters grows to five and the equation is underidentified. The equation for time period 4 is identified under either the null or alternative hypothesis.

In more general notation, suppose that the parameters of the equation

$$Y_s = W_s B_s + V_s$$

are not identified (in the absence of the restrictions imposed by the null hypothesis) for time period s, i.e. Z_s has fewer elements than B_s (and, hence fewer than W_s). That is, in the equation

$$(3.16) \quad Z_s'Y_s = Z_s'W_sB_s + Z_s'V_s$$

the number of rows in $Z_s'Y_s$ is fewer than the number of parameters in B_s.

The appropriate test statistic once again uses the difference between the restricted and unrestricted sum of squared residuals, but care must be taken in constructing the covariance matrix Ω. Since the same covariance matrix must be used when computing both the restricted and unrestricted sum of squared residuals, the following procedure is appropriate.

First obtain the restricted sum of squares, incorporating the fact that B_s is identified under the null hypothesis by adding equation (3.16) to the list of equations to be estimated. Let $B^* = [B', B_s']'$ be the coefficients for the equations for all time periods. The parameters B are identified under either hypothesis, but those for time period s are not. For the simple example

$$B = [b_4, c_{14}, c_{24}, d_{14}, d_{24}]'$$

$$B_s = [b_3, c_{13}, c_{23}] .$$

Consider the null hypothesis

(3.17) $H_0 : B^* = H\gamma + G$,

where the elements of γ are identified. Using similar notation, let

$$V^* = [V', V_s']' \quad Y^* = [Y', Y_s'],$$
$$W^* = \text{diag}[W, W_s] \; Z^* = \text{diag}[Z, Z_s] \; .$$

Under the null hypothesis, we may add equation (3.16) to equation (3.7) as:

(3.18) $Z^{*\prime}Y^* = Z^{*\prime}W^*\gamma + Z^{*\prime}v^*$

where $Y^* = Y^* - W^*G$ and $W^* = W^*H$.

Next estimate the parameters, B^*, and covariance matrix, Ω , using the procedure described above.

To obtain the unrestricted sum of squares and the appropriate test statistic, only those equations identified under the alternative hypothesis are employed. Accordingly, the appropriate estimate of the covariance matrix, Ω , is a submatrix of the covariance matrix estimated under the null hypothesis. The desired submatrix is that for equations identified under the alternative hypothesis.[17]

As before, $Q^*_R - Q$ will have a chi-square distribution in large samples. In this instance, the degrees of freedom is given by:

(3.19) $[\dim(Z^{*\prime}v^*) - \dim(\gamma)] - [\dim(Z'v) - \dim(B)]$

---

[17] Importantly, the submatrix must be obtained from the estimated covariance matrix, Ω , prior to inverting the matrix and constructing the unrestricted sum of squares.

In our example, $\dim(Z^{*\prime}, v^{*})$ equals 8 (three instrumental variables for time period 3 and five for time period 4), $\dim(\gamma)$ equals 6 (three parameters in the equation for each time period, $\dim(Z'v)$ equals five, and $\dim(B)$ equals 5 (the number of parameters in the equation for time period 4 under the alternative hypothesis. Thus, the degrees of freedom for the causality test is 2 in this example.

Sequences of Tests

Two important questions in this framework are whether the data are consistent with a lag of length m and whether x causes y. It seems natural to nest the hypothesis of non-causality within the hypothesis about the lag length. That is, it makes sense to think of testing for non-causality conditional upon the outcome of a test for the lag length. When hypotheses are nested in this manner, we can construct a sequence of test statistics which will be (asymptotically) statistically independent. This permits us to isolate the reason for the rejection of the joint hypothesis.

To see how such a sequence is constructed, consider the two hypotheses

$$H_1 : B = H\gamma + G ,$$

and the second hypothesis, nested within H_1

$$H_2 : \gamma = \overline{H\gamma} + \overline{G} .$$

Let Q be the unrestricted sum of squares, Q_{R1} the restricted sum of squares from imposing H_1, and Q_{R2} the sum of squares from imposing both H_1 and H_2, i.e. the restriction:

$$B = H\overline{H\gamma} + (H\overline{G} + G)$$

Then $Q_{R1} - Q$ is the appropriate test statistic for testing H_1 and $Q_{R2} - Q_{R1}$ is the appropriate statistic for testing H_2 conditional upon H_1 being true. Furthermore, it is the case that the two statistics are asymptotically independently distributed.[18]

The significance of a joint test of H_1 and H_2 may be determined. Suppose that the test consists of rejecting H_1 and H_2 if either statistic is too large. Let the first test have significance level a_1 and the second a_2 . The significance of the joint test is:

$$a_1 + a_2 - a_1 a_2 \, .$$ [19]

Notice that, if H_1 is accepted, we can infer the correctness of H_2 from whether or not the test statistic for H_2 is too large. However, if H_1 is rejected we can say nothing about H_2 because it is nested within H_1 .

[18] See Newey and West [1985].

[19] A similar procedure based upon Wald tests is discussed by Sargan [1980] in the closely related context of testing for dynamic specification of time series models.

IV. An Example

In this section, we demonstrate the techniques described above by investigating the dynamic interrelationships between local revenues, grant revenues, and local expenditures for a sample of 171 United States municipalities.

A. The Issues

Public finance economists have long investigated the determinants of the level of local government expenditure; particularly the impact of alternative federal grant policies. (For a review of this literature, see Gramlich [1977].) One empirical regularity is the "flypaper effect": a dollar increase in exogenous grant monies stimulates local spending more than a dollar increase in local income. One interpretation of this result turns on the hypothesis that local bureaucrats, for a variety of reasons, seek to increase the magnitude of public spending beyond the level desired by the representative voter. Assuming further that the bureaucrats have better information than voters on the magnitude of outside grants, the bureaucrats may "trick" the voters into supporting larger expenditures than they might otherwise permit.

Since local governments are considered the most likely (via some variant of the Tiebout mechanism) to supply an efficient level of public services, the implications for efficient resolution of the social choice problem are disheartening. (Less pessimistic interpretations are possible. See, e.g., Hamilton [1983].) As far as we know, all evidence on the flypaper effect comes from cross-sectional analysis of local governments. The dynamic and stochastic properties of local revenue, grant, and spending streams have not been explored.

A straightforward dynamic reinterpretation of the flypaper effect is that grant monies cause (in the sense discussed above) local expenditure.

A second hypothesis of interest is whether expenditures cause local revenues, since either the median voter model or the benign tyrant interpretation of government behavior imply that revenues are raised only due to the necessity of financing desired public provisions. Finally, a third hypothesis is suggested by the common belief that revenues cause expenditures. A typical expression of this view is that of a state senator from New Jersey. "It is axiomatic that government spending will rise to meet and eventually exceed available revenues."[20] Our framework allows a test of this "axiom."

Even if the investigator is unwilling to accept the causal interpretations of the results, investigation of the dynamic relationships will summarize the empirical facts to which theoretical studies must conform. For example, how long are the distributed lags which relate revenues and expenditures? In the absence of some empirical benchmark, it is not even clear what the theory has to explain.

B. Data Description

Our data are drawn from the annual Survey of Governments between 1973 and 1980 and the Census of Governments conducted in 1972 and 1977. A random sample of municipal government records was selected from the data tape for 1979 (the year with the least coverage) and these same government records were selected for the remaining eight years when possible.[21] There is usable information on 171 municipal governments over a period of nine (fiscal) years.

---

[20] New York Times, New Jersey Weekly, March 17, 1985, p. 22.

[21] To remain in the sample, communities had to report positive school expenditures.

In each year, the record for each government essentially presents the budget identity--including revenues from a variety of sources, expenditures by program and type (current, capital, etc.), debt transactions, grant receipts (by source), and grant transfers (of minimal importance at the municipal level). These dollar amounts were converted to per capita real dollars using a regional price index with December 1977 = 100. All variables are entered as natural logarithms.

The data contained virtually no information on the economic and demographic characteristics of the communities. Such variables typically play an important role in regression analyses of local government spending. However, to the extent that economic and demographic characteristics can be regarded as "individual effects," this absence of information will cause no problems. In essence, the statistical procedure eliminates these effects via differencing. In addition, the effects of any cyclical variables that might influence outcomes are captured by the time dummy variables.

For the estimations presented here we use total local current expenditures, total local revenues, and total grants received. Less aggregative work focusing on specific revenue and expenditure categories, while important and interesting, must await future research.

C. Estimation and Testing

Our focus is on the dynamic interrelationships between three variables: expenditures, revenues, and grants. For each variable, we estimate a model in which it appears on the left hand side. On the right hand side are its own lags and lags of the other two variables. The results are used to investigate issues of parameter stationarity, lag length, and causation. Below, we present the results for the expenditures, revenues, and grants equations in turn.

Expenditures. We begin by estimating an equation with 2 lags of each of the right hand side variables; in terms of our earlier notation, m=2.[22] The quasi-differenced version, then, has 3 lags. Given that in our data T=9, m=2 implies that we can estimate parameters for only the last 5 years in the data set; i.e., t=1976,...,1980. When the equations for these years are estimated jointly using the three-stage procedure described above, the minimized value of the χ^2 test statistic (Q in our earlier notation) is equal to 1.99, and has 30 degrees of freedom.[23] (For convenience, this result and others to follow are summarized in Table 1.) Now, inferences about causality will be incorrect if the lag distribution is incorrectly truncated and/or parameter stationarity is incorrectly imposed. In order to avoid these (type II) errors, we choose 10% significance levels for the tests on lag length and parameter stationarity, rather than the conventional 5% or 1% levels.

---

22 We began at m=2 in order to estimate the covariance matrix necessary to test for this and other (including larger) lag lengths. However, it is obvious from the test statistic that the restriction m=2 is consistent with the data. This also turns out to be true for the revenues and grants equations.

23 The calculation of degrees of freedom is as follows: For 1980, we have available 7 years of data for each variable (1972-1978). Adding a constant gives 22 instrumental variables. This number falls to 19 in 1979, 16 in 1978, and so forth. Thus, the total number of instrumental variables is 22 + 19 + 16 + 13 + 10 = 80. For each year we estimate 10 parameters; for a total of 50. The degrees of freedom is simply the difference: 80-50.

Table 1

Expenditures Equation

	Q	L	Degrees of Freedom	χ^2*
i) m=2	1.99	-	30	40.26
ii) stationary fixed effects (given i)	3.54	1.55	9	14.68
iii) all parameters stationary (given ii)	46.48	42.94	30	40.26
iv) m=1 (given ii)	19.03	15.49	18	25.99
v) m=0 (given ii)	108.44	89.41	18	25.99
vi) exclude revenue (given iv)	39.41	20.38	6	12.6
vii) exclude grants (given iv)	32.72	13.69	6	12.6

* For lines i through v, χ^2 is evaluated at the 0.10 significance level; for lines vi and vii, at the 0.05 significance level.

Because the value of the χ^2_{30} at the 10% level is 40.26, we can easily accept m=2.

When we examined the coefficients of this specification, we noticed that most of them were quite small relative to their standard errors. To see if we could sharpen the results by putting more structure on the model, we imposed the condition that the coefficients on the individual effect be stationary, i.e., that $r_t=1$ for all t. The value of Q from this restriction is 3.54. Therefore, from equation (3.15), the value of L is 3.54 ($=Q_R$) minus 1.99 (=Q), or 1.55. (There are 9 degrees of freedom because there are 4 restrictions with $r_t=1$, and 5 restrictions on the parameters for the third lag of each variable.) The associated critical value of χ^2_9 at a .10 significance level is 14.68; hence, the model with stationary individual effects passes the test by a wide margin. (See line ii of Table 1.)

Are <u>all</u> the parameters similarly stationary over time? Using the procedure outlined in Section III.B, we imposed this constraint. The associated value of Q is 46.48. In this case, then, L = 46.48 - 3.54, or 42.94, and has 30 degrees of freedom. (There are 30 degrees of freedom because the 6 lag parameters for each of 1975 through 1979 are constrained equal to their 1980 values.) The critical value of the χ^2_{30} distribution at the 0.10 level is 40.26. We therefore reject the hypothesis that all the coefficients are stationary across time.

We next investigate results relating to lag length (conditional on the assumption that $r_t=1$). The first question is whether the data will permit us to shorten the lag length from 2 to 1. When we impose m=1, the value of Q is 19.03. Comparing this to the value of Q in line ii of Table 1, we find that L = 15.49, and has 18 degrees of freedom. (There are 18 degrees of freedom because we are restricting 3 lags in each of 6 years.) The critical value of the χ^2_{18} distribution at the 0.10 level is 25.99. We can accept the restriction that one lag in each variable adequately characterizes the data.

The fact that m=1 passes the test gives rise to the thought that an even more parsimonious specification, m=0, might do so as well. When we estimate the expenditures equation with no lags at all, the value of Q jumps to 108.44; the associated value of L is 89.41 (= 108.44-19.03). The data clearly reject this hypothesis by a wide margin. (See line v of Table 1.)

Conditional on m=1 and stationary fixed effects, we next turn to causality issues. As noted above, to test whether revenues cause expenditures, we simply estimate the expenditures equation excluding revenues, and evaluate the increase in the minimum χ^2 test statistic. The value of Q when revenues are excluded is 39.41; the value of L is 39.41 minus 19.03, or 20.38, and it has 6 degrees of freedom. (There are 6 degrees of freedom because the coefficient on the lagged value of revenues in each year 1975-1980 is restricted equal to zero.) The critical value of the χ^2_6 distribution at the 0.05 level is 12.6; hence, the data reject by a wide margin the notion that revenues do not cause expenditures.

When we estimate the expenditures equation excluding grants, we find that Q = 32.72, L = 13.69, and the hypothesis of non-causality is again rejected, although by a smaller margin.

To summarize: We find that community expenditures can be described by a dynamic process which has only 1 year lags. The individual effects are stationary across time periods, but the other parameters (taken as a group) are not. Further, one can reject the hypothesis that revenues and grants do not cause expenditures.

So far we have not discussed the values of the parameters themselves. Because the expenditures, revenues, and grants equations form a system which determines the evolution of all three variables, it is most useful to analyze the parameters of all three equations jointly. Hence, we defer that discussion until after the revenues and grants equations have been discussed.

Revenues. The calculations for the revenues equations are very similar to those of the expenditures equation which were just described in detail. We therefore merely summarize the results which are reported in Table 2: (a) A lag length of 2 is at least sufficient to characterize the data (see line i). (b) Given m=2, one cannot reject the hypothesis that all the parameters are stationary across time periods (see lines ii and iii). (c) One can reject the hypothesis that m=1 (see line iv). (d) One cannot reject the hypothesis that expenditures do not cause revenues (line v), but one can reject the hypothesis that grants do not cause revenues (line vi).

Grants. The results for the grants equation are reported in Table 3. The main conclusions are: (a) A lag length of 2 is at least sufficient to characterize the data (see line i). (b) As before, the data are consistent with stationary fixed effects (see line ii). As in the case of the expenditures equation, however, the data are not consistent with all the parameters being

Table 2

Revenues Equation

	Q	L	Degrees of Freedom	χ^2 *
i) m=2	1.24	-	30	40.26
ii) stationary fixed effects (given i)	3.49	2.25	9	14.68
iii) all parameters stationary (given ii)	24.13	20.64	30	40.26
iv) m=1 (given iii)	47.23	23.10	3	6.25
v) exclude expenditures (given iii)	25.43	1.3	2	5.99
vi) exclude expenditures and grants	47.37	23.24	4	9.49

*For lines i through iv, χ^2 is evaluated at the 0.10 significance level; for lines v and vi at the 0.05 significance level.

Table 3

Grants Equation

	Q	L	Degrees of Freedom	χ^2 *
i) m=2	9.48	-	30	40.26
ii) stationary fixed effects (given i)	22.17	12.69	9	14.68
iii) all parameters stationary (given ii)	162.37	140.2	30	40.26
iv) m=1 (given ii)	50.36	28.19	18	25.99
v) exclude expenditures (given ii)	51.09	28.92	12	21.0
vi) exclude revenues (given ii)	56.60	34.43	12	21.0

* For lines i through iv, χ^2 is evaluated at the 0.10 level; for lines v and vi at the 0.05 level.

stationary (see line iii). (c) A lag length of one is not consistent with the data. (d) One can reject the hypotheses that expenditures or revenues do not cause grants (see lines v and vi).

This last result should be interpreted with caution given that some grants are allocated on the basis of expenditure and revenue behavior. Are the equations telling us something about causality, or simply reproducing the "law"? To the extent there is a random component in grant receipts, the causality test traces the effect of this innovation on revenues and expenditures. That there is such a random component seems extremely likely; indeed, local officials cite grant uncertainty as one of their most important budgetary problems. [See Advisory Council on Intergovernmental Relations (1977, p. 5).]

Parameter Estimates. We next turn to an examination of the parameter estimates. In Table 4 we generally report the lag coefficients of the most parsimonious specification of each equation that is consistent with the data, based on the discussions surrounding Tables 1, 2, and 3. (To conserve space, the coefficients of the unrestricted equations are not reported; these are available upon request.) Table 4 suggests the following thoughts:

1. While the processes generating expenditures, revenues, and grants share the important characteristic of a stationary fixed effect, they differ with respect to lag length and whether the lag parameters change over time. More coefficients are reported for expenditures and grants than for revenues, because only for the latter are all the parameters stationary over time.

2. In general, parameter stationarity can be rejected for one of two reasons. Either the estimates are qualitatively "close" but are very precisely

Table 4

Parameter Estimates *

	Expenditures		Revenues		Grants	
1980						
E_{t-1}	0.0981	(.786)	0.054	(.202)	-0.172	(.566)
E_{t-2}			0.054	(.051)	43.7	(36.7)
R_{t-1}	0.988	(.616)	0.543	(.179)	-1.40	(1.69)
R_{t-2}			-0.019	(.0245)	-34.0	(25.6)
G_{t-1}	-0.212	(.546)	-0.164	(.0640)	0.913	(.448)
G_{t-2}			-.101	(.0245)	-11.3	(9.36)
1979						
E_{t-1}	-0.160	.727			50.22	(35.0)
E_{t-2}					-7.24	(3.00)
R_{t-1}	1.13	.659			-36.8	(23.8)
R_{t-2}					2.35	(1.61)
G_{t-1}	-0.211	.175			-10.1	(8.80)
G_{t-2}					-.0207	(.361)
1978						
E_{t-1}	0.201	.220			2.30	(2.36)
E_{t-2}					2.02	(2.35)
R_{t-1}	0.642	.219			-1.78	(1.63)
R_{t-2}					1.60	(2.22)
G_{t-1}	-0.145	.0576			-0.433	(.599)
G_{t-2}					1.04	(1.12)

Table 4 (continued)

	Expenditures		Revenues	Grants	
1977					
E_{t-1}	0.170	(.185)		2.12	(1.60)
E_{t-2}				-0.228	(.680)
R_{t-1}	0.628	(.208)		2.15	(1.74)
R_{t-2}				-0.299	(.761)
G_{t-1}	-0.207	(.0809)		1.09	(1.00)
G_{t-2}				0.0753	(.446)
1976					
E_{t-1}	0.262	(.117)		0.311	(1.33)
E_{t-2}				1.09	(.653)
R_{t-1}	0.684	(.166)		3.61	(1.94)
R_{t-2}				-1.19	(.565)
G_{t-1}	-0.137	(.0443)		1.32	(.883)
G_{t-2}				0.0598	(.179)
1975					
E_{t-1}	0.151	(.335)		0.00754	(1.07)
E_{t-2}				3.05	(1.22)
R_{t-1}	.662	(.288)		1.20	(1.37)
R_{t-2}				0.457	(.275)
G_{t-1}	-0.176	(.141)		0.605	(.679)
G_{t-2}				-0.0733	(.341)

*Numbers in parentheses are standard errors.

estimated, or the parameters differ greatly in magnitude even if they are individually estimated without much precision. The former seems roughly to be the case in the expenditures equation. The latter appears to be the case in the grants equation. We conjecture that the behavior of the grants equations may be due to the instability over time of the process by which grants are administered. For example, the provision of special one year project grants could radically alter the relevant lag correlations.

3. In virtually all cases, causality hypotheses cannot be rejected. As noted earlier, the important exception is that expenditures do not cause revenues. For the sake of completeness, however, we have reported the coefficients of lagged expenditures in the revenues equation.[24] It is interesting to note that in addition to being statistically insignificant, they are small in magnitude compared to the other coefficients in the revenues equation. From either perspective, they are not an "important" determinant of revenues.

4. Analysis of the dynamic behavior of the system as a whole is complicated by the non-stationarity of the estimated coefficients. For example, examination of steady state multipliers is not meaningful when the coefficients are changing over time. Also, we feel that because of the unreasonable size of the estimated coefficients in the grants equation, it may be misspecified, and that this misspecification would adversely affect the analysis of the joint dynamic behavior of revenues, expenditures and grants.

---

[24] When expenditures are excluded from the equation, the other coefficients barely change. The coefficients on R_{t-1}, R_{t-2}, G_{t-1} and G_{t-2}, respectively, are 0.567 (.122), -0.00677 (.0216), -0.156 (.0575), and -0.0833 (.018).

D. Summary

What have we learned from all of this? The figures from the lag truncation tests suggest that one or two lags are sufficient to summarize the dynamic interrelationships. The results of the causality tests suggest that it is generally inappropriate to regard any of the members of the expenditures-revenues-grants nexus as exogenous.[25] A proper theory must take into account their joint determination. In addition, the results suggest the importance of intertemporal linkages that are omitted from conventional cross sectional regressions. Recognition of such linkages complicates the interpretations of such regressions. Finally, the results from the stationarity tests suggest that it is dangerous, as is common, to assume that all parameter estimates from panel data do not change over time.[26]

As we stressed at the beginning of this section, our numerical results must be regarded as tentative. For example, it would be useful to estimate a model in which relative prices (such as tax rates, as opposed to tax revenues) play a role. Nevertheless, we think that the results demonstrate the utility of our technique and its operational feasibility.

V. Conclusion

We have presented a simple method of executing causality tests using panel data. The key to its simplicity is the fact that estimation and testing have straightforward GLS interpretations--no non-linear optimization is required.

Our empirical example demonstrates the importance of testing for the appropriate lag length prior to causality tests; an issue of considerable importance in short panels. In the absence of such tests, no inferences concerning causal relationships may be drawn.

---

[25] But see Leamer [1984] for a careful discussion of alternative notions of exogeneity in this context.

[26] See, for example, Craig and Inman [1982].

REFERENCES

Advisory Council on Intergovernmental Relations, The Intergovernmental Grant System as Seen by Local, State, and Federal Officials, Washington: U.S. Government Printing Office, 1977.

Ashenfelter, Orley and Card, David, "Time Series Representations of Economic Variables and Alternative Models of the Labour Market," Review of Economic Studies, Volume 49, 1982, pp. 761-782.

Chamberlain, Gary, "Panel Data," Chapter 22 in The Handbook of Econometrics Volume II, Z. Griliches and M. Intrilligator (eds.), (Amsterdam: North-Holland Publishing Co.), 1983.

Craig, Steven and Inman, Robert, "Federal Aid and Public Education: An Empirical Look at the New Fiscal Federalism," The Review of Economics and Statistics, Volume 44, 1982, pp. 541-552.

Fisher, Franklin, The Identification Problem in Econometrics, (Huntington, N.Y.: Krieger Publishing Co.), 1966.

Gallant, Ronald and Jorgenson, Dale, "Statistical Inference for a System of Nonlinear, Implicit Equations in the Context of Instrumental Variables Estimation," Journal of Econometrics, Volume 11, 1979, pp. 275-302.

Gramlich, Edward, "Intergovernmental Grants: A Review of the Empirical Literature," in The Political Economy of Fiscal Federalism, W. Oates (ed.), (Lexington, MA: D.C. Heath Co.), 1977, pp. 219-240.

Granger, C. W. J., "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," Econometrica, Volume 37, 1969, pp. 424-438.

Hamilton, Bruce, "The Flypaper Effect and Other Anomalies," Journal of Public Economics, Volume 22, 1983, pp. 347-362.

Hansen, Lars, "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica, Volume 50, 1982, pp. 1029-1054.

Leamer, Edward, "Vector Autoregressions for Causal Inference?", mimeo, University of California-Los Angeles, 1984.

Lundberg, Shelly, "Tied Wage-Hours Offers and the Endogeneity of Wages," National Bureau of Economic Research Working Paper No. 1431, 1984.

Newey, Whitney and West, Kenneth, "Minimum Chi-Square Hypothesis Testing in Time Series Models," in preparation, Princeton University, 1985.

Nickell, Stephen, "Biases in Dynamic Models with Fixed Effects," Econometrica, Volume 49, 1981, pp. 1417-1426.

Pakes, Ariel and Zvi Griliches, "Estimating Distributed Lags in Short Panels with an Application to the Specification of Depreciation Patterns and Capital Stock Constructs," Review of Economic Studies, 1984, pp. 243-262.

Sargan, J. D., "Some Tests of Dynamic Specification for a Single Equation," Econometrica, Volume 48, 1980, pp. 879-898.

Sims, Christopher, "Money, Income, and Causality," The American Economic Review, Volume 62, 1972, pp. 540-552.

Taylor, John, "Output and Price Stability--An International Comparison," Journal of Economic Dynamics and Control, Volume 2, 1980, pp. 109-132.

White, Halbert, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," Econometrica, Volume 48, 1980, pp. 817-838.

White, Halbert, "Instrumental Variables Regression with Independent Observations," Econometrica, Volume 50, 1982, pp. 483-500.