Econometrica, Vol. 56, No. 6 (November, 1988), 1371–1395

ESTIMATING VECTOR AUTOREGRESSIONS WITH PANEL DATA

BY DOUGLAS HOLTZ-EAKIN, WHITNEY NEWEY, AND HARVEY S. ROSEN[1]

This paper considers estimation and testing of vector autoregression coefficients in panel data, and applies the techniques to analyze the dynamic relationships between wages and hours worked in two samples of American males. 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, forming convenient test statistics, and testing for the presence of measurement error. The empirical results suggest the absence of lagged hours in the wage forecasting equation. Our results also show that lagged hours is important in the hours equation, which is consistent with alternatives to the simple labor supply model that allow for costly hours adjustment or preferences that are not time separable.

KEYWORDS: Vector autoregression, panel data, causality tests, labor supply.

1. INTRODUCTION

VECTOR AUTOREGRESSIONS are now a standard part of the applied econometrician's tool kit. Although their interpretation in terms of causal relationships is controversial, most researchers would agree that vector autoregressions are a parsimonious and useful means of summarizing 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. For example, a vector autoregression can be used to summarize the dynamic relationship between an individual's hours of work and wages (see below) or the dynamic relationship between a government's revenues and expenditures (see Holtz-Eakin, Newey, Rosen (forthcoming)). Unlike macroeconomic 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.[2] Also, it is possible that individual heterogeneity is an important feature of disaggregate data. For these reasons, it is inappropriate to apply standard techniques for estimating vector autoregressions to panel data.

The purpose of this paper is to formulate a coherent set of procedures for estimating and testing vector autoregressions in panel data. Section 2 presents the basic model, which builds upon Chamberlain (1983). Section 3 discusses identification and gives methods of parameter estimation and testing. The estimation

[1] This research was supported in part by NSF Grants SES-8419238 and SES-8410249. We are grateful to Joseph Altonji, the Editors, and three referees for useful comments. Joseph Altonji and David Card graciously provided us with the data used in the empirical analysis.

[2] 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.

method is similar in spirit to that of Anderson and Hsiao (1982). Section 4 applies the methods to an example from labor economics; we investigate the dynamic relationships between wages and hours worked. Section 5 provides a brief summary and conclusion.

2. THE MODEL

In the usual time series context, equations of a bivariate autoregression typically take the form

$$y_t = \alpha_0 + \sum_{l=1}^{m} \alpha_l y_{t-l} + \sum_{l=1}^{m} \delta_l x_{t-l} + u_t, \tag{2.1}$$

where the α's and δ's are the coefficients of the linear projection of y_t onto a constant and past values of y_t and x_t, and the lag length m is sufficiently large to ensure that u_t is a white noise error term. While it is not essential that the lag lengths for y and x are equal, we follow typical practice by assuming that they are identical.

Consistent estimation of the parameters of equation (2.1) requires many observations of x and y values. In time series applications these observations typically are obtained from a record of x and y over a long period of time. In contrast, panel data usually have a relatively small number of time series observations. Instead, there often are a great number of cross-sectional units, with only a few years of data on each unit. To estimate the parameters of equation (2.1) one must pool data from different units, a procedure which imposes the constraint that the underlying structure is the same for each cross-sectional unit.

The constraint that the time series relationship of x and y is the same for each cross-sectional unit is likely to be violated in practice, so that it is desirable to be able to relax this restriction. One way to relax the pooling constraint is to allow for an "individual effect," which translates in practice into an individual specific intercept in equation (2.1). Changes in the intercept of a stationary vector autoregression correspond to changes in the means of the variables, so that allowing for an individual effect allows for individual heterogeneity in the levels of x and y. A second way to allow for individual heterogeneity is to allow the variance of the innovation in equation (2.1) to vary with the cross-section unit. Changes in the innovation variance of a vector autoregression correspond to changes in the variance of the variables, so that allowing for cross-section heteroskedasticity in the innovation variance allows for individual heterogeneity in the variability of x and y. In what follows we allow for both an individual effect and cross-section heteroskedasticity in the variance of the innovation.

It is likely that the level and variability of the variables are important sources of individual heterogeneity, but it would also be nice to allow for individual heterogeneity in the time series correlation pattern of x and y. In this context, allowing for such heterogeneity is difficult because the variables on the right-hand side of the equation are lagged endogenous variables. Here it is impossible to interpret the α's and δ's as means of parameters that vary randomly across individual cross-section units, although this interpretation of regression parameters is possible when the right-hand side variables are exogenous. (See Pakes and Griliches (1984).)

On the other hand, pooling cross-sectional units does have certain advantages. First, the assumption of time stationarity can be relaxed. The presence of a large number of cross-sectional units makes it possible to allow for lag coefficients that vary over time. Second, the asymptotic distribution theory for a large number of cross-sectional units does not require the vector autoregression to satisfy the usual conditions that rule out unit and explosive roots. Of course, the presence of an explosive process may lead to difficulties in interpreting the model. Nevertheless, it is still possible, for example, to use standard asymptotic distribution theory to formulate valid tests for explosive behavior.[3]

A model with individual effects that relaxes the time stationarity assumption can be obtained by modifying a model presented by Chamberlain (1983). Assume that there are N cross-sectional units observed over T periods. Let i index the cross-sectional observations and t the time periods. A model that is analogous to equation (2.1), but allows for individual effects and nonstationarities across time is

$$(2.2) \qquad y_{it} = \alpha_{0t} + \sum_{l=1}^{m} \alpha_{lt} y_{it-l} + \sum_{l=1}^{m} \delta_{lt} x_{it-l} + \Psi_t f_i + u_{it} \qquad (i = 1, \ldots, N;\ t = 1, \ldots, T),$$

where f_i is an unobserved individual effect and the coefficients α_{0t}, $\alpha_{1t}, \ldots, \alpha_{mt}, \delta_{1t}, \ldots, \delta_{mt}, \Psi_t$ are the coefficients of the linear projection of y_{it} on a constant, past values of y_{it} and x_{it}, and the individual effect f_i.

The model of equation (2.2) is different than that of Chamberlain (1983, p. 1263) in that Chamberlain avoids restricting the lag length by assuming that the first period of observation corresponds to the first period of the life of the individual unit. This assumption implies that the projection of y_{it} on all the observed past values for y_{it} and x_{it} (i.e. $\{y_{it-1}, \ldots, y_{i1}, x_{it-1}, \ldots, x_{i1}\}$) is equal to the projection on the entire past. That is, the lag length m in equation (2.2) varies with t according to the relation $m(t) = t - 1$. In practice, the entire history of each economic unit is not usually observed and some assumptions must be imposed to identify the time series relationship of x and y using the observed data.[4] Our method takes this fact into account. The assumption embodied in equation (2.2) is that for each observed time period t the projection of y_{it} on the

[3] The asymptotic theory does require that various moments of the data exist. Existence of moments in models with unit or explosive roots requires an assumption concerning the initial conditions of the data such as the assumption that the first point in the life of the individual units is a constant.

[4] Pakes and Griliches (1984) have considered a similar identification issue in the context of distributed lag models with exogenous regressors.

entire past depends only on the past m observations. In the next section we will discuss identification, estimation, and inference under this assumption on lag length, as well as ways of testing restrictions on the parameters of equation (2.2).

3. STATISTICAL INFERENCE

A. Identification

The specification of equation (2.2) as a projection equation implies that the error term u_{it} satisfies the orthogonality condition

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

These orthogonality conditions imply that lagged values of x and y qualify as instrumental variables for equation (2.2). Our analysis of identification will be restricted to use of these orthogonality conditions.[5] Of course, if other restrictions are imposed on equation (2.2), such as absence of cross-section heteroskedasticity in the forecast error u_{it}, then it will be easier to identify the parameters.[6] Such extra restrictions will often take the form of imposing additional cross-section or time-series homogeneity in the relationship of x and y, so that restricting attention to the orthogonality conditions (3.1) is consistent with allowing as much heterogeneity as possible.

In order to use the orthogonality conditions (3.1) to identify the parameters of equation (2.2), the investigator must deal with the presence of the unobserved individual effect, f_i. It is well known that in models with lagged dependent variables it is inappropriate to treat individual effects as constants to be estimated.[7] Instead, we can transform equation (2.2) to eliminate the individual effect. Let $r_t = \Psi_t/\Psi_{t-1}$, and consider multiplying equation (2.2) for time period $t-1$ by r_t and subtracting the result from the equation for period t. Collecting all x and y terms dated $t-1$ or before on the right-hand side yields

$$y_{it} = a_t + \sum_{l=1}^{m+1} c_{lt} y_{it-l} + \sum_{l=1}^{m+1} d_{lt} x_{it-l} + v_{it} \qquad (t = (m+2), \ldots, T), \tag{3.2}$$

[5] Note that it would be valid to use nonlinear functions of lagged values of x and y as instruments only if equation (2.2) could be interpreted as a conditional expectation rather than a linear projection. We choose to work with a linear projection specification because specification of the form of the conditional expectation of y_{it} given lagged values of x and y is difficult. It seems likely that this conditional expectation would involve nonlinear functions of lagged values of x and y.

[6] If the first and second moments of the data are the same for different cross-section units, then the minimum distance methods of MaCurdy (1981a) and Chamberlain (1983) could be used to estimate the parameters from cross-section moments. In this case it will be easier to identify the parameters, because the orthogonality conditions (3.1) do not involve all the cross-section moments.

[7] One common technique is to compute the difference between each variable and its time mean (by cross-section unit) to eliminate the individual effect. See, e.g., Lundberg (1985). In the current context, this procedure will yield inconsistent estimates, even when the parameters are stationary, because of the presence of lagged endogenous variables. See Nickell (1981).

where

$$
\begin{aligned}
(3.3)\qquad & a_t = \alpha_{0t} - r_t\alpha_{0t-1}, \\
& c_{1t} = r_t + \alpha_{1t}, \\
& c_{lt} = \alpha_{lt} - r_t\alpha_{l-1,t-1} && (l = 2, \ldots, m), \\
& c_{m+1,t} = -r_t\alpha_{m,t-1}, \\
& d_{1t} = \delta_{1t}, \\
& d_{lt} = \delta_{lt} - r_t\delta_{l-1,t-1} && (l = 2, \ldots, m), \\
& d_{m+1,t} = -r_t\delta_{m,t-1}, \\
& v_{it} = u_{it} - r_t u_{i,t-1}.
\end{aligned}
$$

Note that in the special case of $r_t = 1$ for each t, then this transformation is simple differencing of equation (2.2). This has been suggested for use in estimation of univariate autoregressive models in panel data by Anderson and Hsiao (1982). More generally, this transformation is a quasi-differencing transformation that has been suggested by Chamberlain (1983). We will proceed by first discussing identification of the parameters of the transformed equation (3.2), and then discussing identification of the original parameters of equation (2.2) from (3.2).

The orthogonality conditions of equation (3.1) imply that the error term of the transformed equation (3.2) satisfies the orthogonality condition

$$(3.4)\qquad E[y_{is}v_{it}] = E[x_{is}v_{it}] = 0 \qquad (s < (t-1)).$$

Thus, the vector of instrumental variables that is available to identify the parameters of equation (3.2) is

$$Z_{it} = [1, y_{it-2}, \ldots, y_{i1}, x_{it-2}, \ldots, x_{i1}].$$

Using the orthogonality conditions (3.4), a necessary condition for identification is that there are at least as many instrumental variables as right-hand side variables.[8] Since there are a total of $2m+3$ right-hand variables in equation (3.2) and the dimension of Z_{it} is $2t-3$, this order condition reduces to $t \geqslant m+3$. Thus, we must have $T \geqslant m+3$ in order to estimate the parameters of equation (3.2) for any t.

Consider the identification of the original parameters. Note first that the parameters of equation (3.2) involve only the ratio r_t of the coefficients of the individual effect. This is to be expected. Since changes in the level of these coefficients correspond to changes in the scale of the individual effect, the level of these coefficients is not identified. The original coefficients that can be identified are therefore the lag coefficients and the ratios of the coefficients of the individual

[8] See Fisher (1966). A sufficient condition for identification is that in the limit, the cross-product matrix between the instruments and the right-hand side variables have rank equal to the number of right-hand side variables.

effect. We will ignore identification of the constant terms (i.e., the α_{0t}'s) since they are usually considered nuisance parameters.

To identify the original parameters there must be at least as many parameters in the transformed equation as in the original equation. Note that it is possible to estimate the parameters of (3.2) for a total of $T-m-2$ time periods. Ignoring the constant terms, there is a total of $(T-m-2)2(m+1)$ parameters in the transformed equation, which involve a total of $(T-m-2)+(T-m-1)2m$ original parameters. Thus, it will not be possible to identify the original parameters unless $T-m-2 \geqslant 2m$, i.e., the number of estimable time periods is at least as large as twice the lag length. Also, because many of the parameters of the transformed equation consist of nonlinear functions of the original parameters with complicated interactions across time periods, for some values of the parameters it may not be possible to recover the original parameters, even when $T \geqslant 3m+2$.

Importantly, there is no need to recover the original parameters to test certain interesting hypotheses. For example, the hypothesis that x does not (Granger) cause y conditional on the individual effect restricts $\delta_{lt}=0$ for each l and t. Since this further implies that $d_{lt}=0$ for each l and t, this noncausality hypothesis can be tested by testing for zero coefficients for the lagged x variables in the transformed equation.

It is also useful to note that additional restrictions on the original parameters can aid their identification. When the restriction $r_t=1$ for each t is imposed, the parameters of the estimable transformed equations involve only $(T-m-1)2m$ original parameters. There will be at least as many parameters in the estimable time periods for the transformed equation as original parameters when $T-m-2 \geqslant m$. Thus, when $r_t=1$ and the number of estimable time periods is at least as large as the lag length, recovery of the original parameters from the transformed parameters is possible and, as is apparent from equation (3.3), straightforward.

Identification of the original parameters is easiest when the individual effect coefficients and the lag coefficients are stationary. In this case the transformed equation (3.2) can be written:

$$(3.5) \qquad y_{it}-y_{it-1}=a_t+\sum_{l=1}^{m}\alpha_l(y_{it-l}-y_{it-l-1})+\sum_{l=1}^{m}\delta_l(x_{it-l}-x_{it-l-1})+v_{it}.$$

Here there are only $2m+1$ right-hand side variables, so that there are enough instruments to identify the parameters if $t \geqslant m+2$. In the stationary case it is possible to obtain estimates of the lag parameters when $T \geqslant m+2$.

A final case that is of interest occurs when measurement error is present. Suppose that x_{it} and y_{it} are unobserved, and that instead we observe

$$(3.6) \qquad \tilde{x}_{it}=x_{it}+e_{it}^{x}, \qquad \tilde{y}_{it}=y_{it}+e_{it}^{y},$$

where e_{it}^{x} and e_{it}^{y} are measurement errors that are uncorrelated with all x and y observations and are uncorrelated across time. For simplicity we consider the implications of such measurement error for the stationary case. Substitution of

x_{it} and y_{it} from equation (3.6) in equation (3.5) yields

$$(3.7) \qquad \tilde{y}_{it} - \tilde{y}_{it-1} = a_t + \sum_{l=1}^{m} \alpha_l (\tilde{y}_{it-l} - \tilde{y}_{it-l-1}) + \sum_{l=1}^{m} \delta_l (\tilde{x}_{it-l} - \tilde{x}_{it-l-1}) + \tilde{v}_{it},$$

where

$$(3.8) \qquad \tilde{v}_{it} = v_{it} + e_{it}^{y} - e_{it-1}^{y} - \sum_{l=1}^{m} \alpha_l (e_{it-l}^{y} - e_{it-l-1}^{y}) - \sum_{l=1}^{m} \delta_l (e_{it-l}^{x} - e_{it-l-1}^{x}).$$

By the assumption that the measurement errors are uncorrelated across time, the vector

$$\tilde{Z}_{it} = [1, \tilde{y}_{it-m-2}, \ldots, \tilde{y}_{i1}, \tilde{x}_{it-m-2}, \ldots, \tilde{x}_{i1}]$$

will be uncorrelated with $\tilde{v}_{it}$ and thus qualify as the vector of instrumental variables for equation (3.7).[9] Here there are only $2(t-m-2)+1$ instrumental variables, so that the requirement that there are at least as many instrumental variables as right-hand side variables becomes $t \geqslant 2m+2$. Thus, it will only be possible to estimate the lag parameters in the stationary case with uncorrelated measurement error when $T \geqslant 2m+2$.

B. *Estimation*

The presentation requires some additional notation. Let

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

be $N \times 1$ 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 $N \times (2m+3)$ vector of right-hand side variables for equations (3.2), where e is an $N \times 1$ vector of ones. Let

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

be the $N \times 1$ 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 $(2m+3) \times 1$ vector of coefficients for the equations. Then we can write equations (3.2) as[10]

$$(3.9) \qquad Y_t = W_t B_t + V_t \qquad (t = (m+3), \ldots, T).$$

[9] The autoregressive structure of equation (3.7) will result in $\tilde{Z}_{it}$ being correlated with the right-hand side variables. The use of instrumental variables to identify equation (3.7) under measurement error is similar to the methods for identification of panel data models with measurement error that have been suggested by Griliches and Hausman (1984).

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

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

$$
\begin{aligned}
Y &= \left[Y_{m+3}', \ldots, Y_t'\right]', \\
&\quad ((T-m-2)N \times 1) \\
B &= \left[B_{m+3}', \ldots, B_T'\right]', \\
&\quad ((T-m-2)(2m+3) \times 1) \\
V &= \left[V_{m+3}', \ldots, V_T'\right]', \\
&\quad ((T-m-2)N \times 1) \\
W &= \operatorname{diag}\left[W_{m+3}', \ldots, W_T'\right], \\
&\quad ((T-m-2)N \times (T-m-2)(2m+3))
\end{aligned}
$$

where diag[ ] denotes a block diagonal matrix with the given entries along the diagonal. With this, the observations for equations (3.2) can be written:

$$Y = WB + V. \tag{3.10}$$

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, here the instrumental variables are different for different equations. The matrix of variables which qualify for use as instrumental variables in period t is

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

which changes with T. Consider the matrix Z defined as

$$Z = \operatorname{diag}\left[Z_{m+3}, \ldots, Z_T\right].$$

The orthogonality conditions ensure that

$$\operatorname*{plim}_{N \to \infty} (Z'V)/N = \operatorname*{plim}_{N \to \infty} \begin{bmatrix} (Z_{m+3}'V_{m+3})/N \\ \vdots \\ (Z_T'V_T)/N \end{bmatrix} = 0.$$

It follows directly that Z is the appropriate choice of instrumental variables for (3.10).[11]

To estimate B, premultiply (3.10) by Z' to obtain:

$$Z'Y = Z'WB + Z'V. \tag{3.11}$$

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

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

given by

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

Ω is not known and therefore must be estimated. To do so, let $\tilde{B}_t$ be the preliminary consistent estimator of B_t formed by estimating the coefficients of the equation for time period t using two-stage least squares (2SLS) on the equation for each time period alone—using the correct list of instrumental variables.[12] 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[13]

$$(3.12) \quad (\tilde{\Omega}/N)_{rs} = \sum_{i=1}^{N} (v_{ir}v_{is}Z_{ir}'Z_{is})/N$$

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:

$$(3.13) \quad \hat{B} = \left[W'Z(\tilde{\Omega})^{-1}Z'W\right]^{-1} W'Z(\tilde{\Omega})^{-1}Z'Y.$$

(*i*) *Imposing Linear Constraints*

Stationarity of the individual effect and of the lag coefficients requires estimating B subject to linear constraints. The hypothesis that x does not cause y also imposes linear constraints on B. A simple way to formulate such constraints is to specify that

$$(3.14) \quad B = H\gamma + G$$

where γ is a $k \times 1$ vector of parameters, H is a constant matrix with dimensions $((T - m - 2)(2m + 3) \times k)$, and G is a constant vector of the same dimension as B.[14] 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 (3.10) gives

$$(3.15) \quad \tilde{Y} = Y - WG = WH\gamma + V = \tilde{W}\gamma + V.$$

This equation has exactly the same form as (3.10). Thus, we can estimate γ as before—using the data matrices transformed by G and H.

[12] That is:

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

[13] 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.

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

(ii) *Efficiency*

Several comments concerning the efficiency of $\hat{B}$ are in order. First, $\hat{B}$ 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.

Second, recall that our procedure involves dropping the equations for the first $m+2$ time periods. When the parameters are nonstationary this procedure involves no loss in efficiency. Although the equations that are dropped may be correlated with the remaining equations, there are no cross equation restrictions, and they are underidentified. When the parameters are stationary, dropping the first $m+2$ periods may involve some loss in efficiency. Because there are cross-equation restrictions, efficiency can be improved by adding back $t=m+2$ and $t=m+1$ period equations, both of which have observable lags. Also, if there is no heteroskedasticity (across time or individuals) in the innovation variance for y_{it} and x_{it}, then all of the parameters for the joint y_{it} and x_{it} process can be estimated without the earliest cross-section moments, so that it may be possible to further improve efficiency by using these moments. Cross-section moment based estimation of moving average (but not autoregressive) time series models in panel data has been considered by MaCurdy (1981a).

C. *Hypothesis Testing*

In this section we discuss the computation of statistics to test the hypotheses 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 chi-square distribution in large samples. 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:

$$H_0\colon\ B = H\gamma + G \tag{3.16}$$

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

$$\begin{aligned} Q &= (Y - W\hat{B})'Z(\tilde{\Omega})^{-1}Z'(Y - W\hat{B})/N, \\ Q_R &= \left(Y - \tilde{W}_{\hat{\gamma}}\right)'Z(\tilde{\Omega})^{-1}Z'\left(Y - \tilde{W}_{\hat{\gamma}}\right)/N. \end{aligned} \tag{3.17}$$

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.[15] By analogy with the F statistic in the standard linear model, an appropriate test statistic is

$$(3.18) \quad L = Q_R - Q.$$

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 instrumental variables (the number of rows of $Z'V$ in (3.11)) minus the number of parameters, i.e., the dimension of B. Similarly, Q_R has degrees of freedom equal to the dimension of B minus the dimension of γ.[16]

So far we have restricted our discussion to testing linear hypotheses concerning the coefficients of a single equation of a vector autoregression. In some contexts hypotheses that involve the coefficients of more than one equation and/or are nonlinear may be of interest. The case of linear hypotheses on the coefficients of more than one equation can be handled by simply stacking together the time periods for the several equations and proceeding in the manner we have discussed. The case of nonlinear restrictions can be handled by formulating the constraints as $B = H(\gamma)$, estimating γ by nonlinear GLS on equation (3.11), with $H(\gamma)$ in place of B, and forming the test statistic as before.

(*i*) *Unidentified Parameters*

When executing the tests, 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 (3.11). Under the alternative hypothesis, the greater number of parameters means that not all of the parameters are identified. Nonetheless, a test of the null

[15] To see this, let P be the matrix such that $PP = \Omega^{-1}$. Then premultiplying (3.11) 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.

[16] 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 (1987) discuss the relationship between L and other test statistics, including regularity conditions.

hypothesis is still possible. The method is analogous to conducting a Chow test with insufficient observations.[17] (See Fisher (1966).)

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

$$Z_s' Y_s = Z_s' W_s B_s + Z_s' V_s \tag{3.19}$$

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 (3.19) 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. Consider the null hypothesis

$$H_0\colon B^* = H\gamma + G \tag{3.20}$$

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

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

Under the null hypothesis, we may add equation (3.19) to equation (3.11) as:

$$Z^{*\prime}\tilde{Y}^* = Z^{*\prime}\tilde{W}^*\gamma + Z^{*\prime}v^*, \tag{3.21}$$

where $\tilde{Y}^* = Y^* - W^*G$ and $\tilde{W}^* = W^*H$.

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.[18]

[17] 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 (3.11), above, and (3.19), below, is not diagonal.

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

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

$$(3.22) \qquad [\dim(Z^{*\prime}v^{*}) - \dim(\gamma)] - [\dim(Z'v) - \dim(B)].$$

(*ii*) *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 noncausality within the hypothesis about the lag length. That is, it makes sense to think of testing for noncausality 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,

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

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

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

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.[19]

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:

$$a_1 + a_2 - a_1 a_2.$$[20]

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.

[19] This result is a simple extension of similar results for the likelihood ratio test for maximum likelihood.

[20] 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 in a dynamic analysis of the relationship between annual hours worked and hourly earnings.

A. *The Issues*

The conventional approach to analyzing the relationship between hours worked and the wage rate is to specify and estimate a model in which hours worked in a given period depend upon that period's wage rate. Implicitly, past hours and wages are assumed to have no impact on current hours. Similarly, the possibility that the past history of wages and hours affects the current wage is ruled out.

However, on theoretical grounds it is quite plausible to expect intertemporal relationships between wages and hours worked. For example, maximization of utility in some life cycle models leads to labor supply functions which depend on wages in other periods. (See, e.g., MaCurdy (1983).) Moreover, if there are costs to adjusting hours of work in response to changes in wages, one might expect that past hours of work would help predict current hours of work. At the same time, some human capital accumulation models suggest that present wage rates depend on past hours of work. (As hours increase, so does expertise on the job, leading to a higher subsequent wage.) Alternatively, one can imagine incentive schemes that link a worker's current wage rate to his past hours of work. (Hamilton (1986) argues that such schemes may help explain the behavior of medical interns, associates in law firms, and assistant professors.)

To fix ideas, consider the equation pair

$$h_{it} = \alpha_t^h + \beta w_{it} + \mu_i^h + \varepsilon_{it}^h, \tag{4.1a}$$

$$w_{it} = \alpha_t^w + \delta_1 w_{it-1} + \cdots + \delta_l w_{it-l} + \mu_i^w + \varepsilon_{it}^w, \tag{4.1b}$$

where

$$\begin{aligned} 0 &= E\left[\varepsilon_{it}^h \middle| \mu_i^h, \mu_i^w, h_{it-1}, h_{it-2}, \ldots, w_{it}, w_{it-1}, \ldots\right] \\ &= E\left[\varepsilon_{it}^w \middle| \mu_i^h, \mu_i^w, h_{it-1}, h_{it-2}, \ldots, w_{it-1}, w_{it-2}, \ldots\right] \end{aligned} \tag{4.2}$$

and h_{it} is the natural log of hours worked for individual i in period t, w_{it} is the natural log of the wage of individual i in period t, and μ_i^h and μ_i^w are unobserved individual effects. Equation (4.1a) is similar to a life cycle labor supply equation derived by MaCurdy (1981b) for a particular specification of preferences. In this equation μ_i^h represents the marginal utility of lifetime income (see also Heckman and MacCurdy (1980) and Browning, Deaton, and Irish (1985)), plus other individual specific variables. Unlike MaCurdy's (1981b) model, this equation imposes the strong restriction that ε_{it} is serially uncorrelated. Of course, variables which are a sum of a function of t and a function of i, such as experience, are allowed for, since they would be absorbed by the time specific term α_t^h and the individual specific term μ_i^h.

Equation (4.1b) is a wage forecasting equation. An important feature of this equation is that lagged hours are excluded. If lagged hours were of use in predicting wages, as might be the case in some of the scenarios previously discussed, then the individual would take into account the effect of today's choice of hours on tomorrow's wages, and the labor supply equation would take a different form.

Substituting w_{it} from equation (4.1b) into equation (4.1a) gives

$$(4.3) \qquad h_{it} = \left(\alpha_t^h + \beta\alpha_t^w\right) + \beta\delta_1 w_{it-1} + \cdots + \beta\delta_l w_{it-l} + \left(\mu_i^h + \beta\mu_i^w\right) + \left(\varepsilon_{it}^h + \beta\varepsilon_{it}^w\right),$$

which together with equation (4.1b) gives a VAR of the form considered in Section 2. Note that lagged hours is excluded from equation (4.1a) and (4.3) and that cross-equation restrictions are present. Evidence of the presence of lagged hours in either equation might therefore be interpreted as evidence against this specification. The presence of lagged hours in the wage equation might be interpreted as presence of the kind of human capital or incentive effects previously mentioned, while the presence of lagged hours in the hours equations might occur because of preferences that are not time separable or costs of adjusting hours. Of course, the presence of lagged hours in the hours equation could also be due to the omission of relevant variables or the violation of the assumption of no serial correlation in ε_{it}^h. Since serial correlation is often thought to be present in such models, this perhaps reduces the substantive implications of the finding that lagged hours appears in the hours equation.

The important point here is that this model and similar models imply the presence of dynamic interrelationships between wages and hours, and these interrelationships can be investigated using panel data on wages and hours. This fact, of course, has been recognized by earlier investigators. Lundberg (1985) used panel data to test whether hours Granger-cause wages. However, her estimation procedure involved taking deviations from means to account for the presence of individual effects. As we argued in Section II, such a procedure leads to inconsistent estimates. Abowd and Card (1986) analyzed the time series properties of the first differences in hours and earnings. Like MaCurdy (1981b) and Lundberg, they assumed that the individual effect was stationary. Moreover, although it would be possible to work backward from their estimates to learn about the time series properties of the *levels*, this would be extremely cumbersome, and they made no attempt to do so. In contrast, the procedures developed in Section 3 allow us to obtain consistent estimates of the time series properties of the *levels* of wages and hours without having to impose stationary individual effects, and without having to employ difficult nonlinear methods.

B. The Data

We estimate equations for wages and hours using a sample of 898 males from the Panel Study of Income Dynamics (PSID) covering the years 1968 to 1981.[21]

[21] Our data include the Survey of Economic Opportunity (SEO) subsample, which oversamples low income households. We performed all of the tests reported in the next section deleting the SEO subsample. The results, which are available upon request, do not in general differ substantially from those presented below. For the one exception, see footnote 22.

We study two variables for each individual. First is the log of the individual's annual hours of work ("hours", denoted h_t), and second is the log of his annual average hourly earnings ("wages", denoted w_t). (For a complete description of the data, see Altonji and Paxson (1986).) As discussed below, we also check some of our results using data from the National Longitudinal Survey of Men 45–59.

The wage variable in both data sets is constructed by dividing total earnings by hours worked. As a result, to the extent that measurement errors are present, they will be correlated across variables. While this presents a problem for full-information methods, the single equation techniques used here are unaffected by this correlation. Finally, to the extent that measurement error induces a serial correlation in the composite error terms of the autoregression, this problem will reveal itself as a correlation between instrumental variables and the transformed errors.

C. *Estimation and Testing*

Using the PSID data, we estimate two equations, one for hours and one for wages. On the right side of each are lags of both wages and hours. We conduct tests for parameter stationarity, minimum lag length, and causality or exclusion restrictions.

While in principle it is desirable to begin by specifying an arbitrarily long initial lag length, this poses a problem in practice. As additional lags are specified, the block structure of the matrix of instrumental variables causes the size of the weighting matrix (Ω in equation (3.12) above) to grow rapidly. For such large matrices, standard numerical procedures for inversion may yield unsatisfactory results. Therefore we initially assume a lag length $m = 3$, leading to four lags in the quasi-differenced reduced form. The wage and hours equations are estimated for the years 1977 to 1981.

(*i*) *Wage Equation*

Results for the wage equation are presented in Table I. The first step is to test for parameter stationarity; i.e., both the individual effects and the lag parameters in the equation are the same for each year. As the second line of the table shows, the chi-square statistic for this test (26.22) indicates that one cannot reject this hypothesis at any level of significance less than roughly 80%. Thus, the appropriate specification of the wage equation is a first-differenced form containing at least three lags each of wages and hours.

Column (1) of Table II shows estimates of the wage equation assuming parameter stationarity, but with no other constraints imposed. The only parameter that is statistically significant is that of the first lagged wage. Not only are the other coefficients insignificant, they are relatively small in absolute value as well. None of the other coefficients is more than about one-third the coefficient on the first lagged wage, in absolute value. While suggestive, these observations do not tell us which lag length is most consistent with the data. It is necessary to use the

TABLE I

WAGE EQUATION

	Q	L	DF	P
(i) $m = 3$	0.00	—	0	—
(ii) All Parameters Stationary	26.22	26.22	34	0.828
(iii) $m = 2$	27.14	0.92	2	0.631
(iv) $m = 1$	29.31	2.17	2	0.338
(v) $m = 0$	43.40	14.09	2	0.001
(vi) Exclude Hours $m = 1$	29.94	0.63	1	0.427

TABLE II

UNCONSTRAINED PARAMETER ESTIMATES[a]

	(1) Wage Equation	(2) Hours Equation
Δh_{t-1}	0.0623	0.145
	(0.0476)	(0.0262)
Δw_{t-1}	0.183	0.00116
	(0.0631)	(0.0385)
Δh_{t-2}	0.0189	−0.00489
	(0.0276)	(0.0158)
Δw_{t-2}	0.0359	−0.0455
	(0.0320)	(0.0190)
Δh_{t-3}	0.0200	−0.0158
	(0.0234)	(0.0202)
Δw_{t-3}	−0.00328	−0.00185
	(0.0245)	(0.0168)

[a] Figures in parentheses are standard errors.

methods of Section III.C to conduct the appropriate tests. The lag length results are recorded in lines (iii) through (v) of Table I, which show the results for the sequence $m = 2$, $m = 1$, and $m = 0$, respectively. The results provide no evidence that the wage equation contains more than a single lag of hours and wages.

Finally, we conduct a test of the hypothesis that hours do not cause wages. Line (vi) of Table I shows that one cannot reject the hypothesis that lagged hours may be excluded from the wage equation. The estimate of this single autoregressive parameter is shown in column (1) of Table III. Note that the exclusion of lagged hours from the wage equation rejects the notion that past hours of work affect the current wage. To the extent that workers face a market locus of hours and wages, it contains at most contemporaneous tradeoffs between hours and wages.

(ii) *Hours Equation*

To complete the investigation, we perform a symmetric set of tests for the specification of the hours equation. The results are reported in Table IV. As was

TABLE III

CONSTRAINED PARAMETER ESTIMATES[a]

	(1) Wage Eq.	(2) Hours Eq.	(3) Hours Eq.
		$(m=2)$	$(m=1)$
Δh_{t-1}	—	0.156	0.156
	—	(0.0206)	(0.0175)
Δh_{t-2}	—	0.002	—
	—	(0.0181)	—
Δw_{t-1}	0.135	−0.001	—
	(0.0368)	(0.05)	—
	—	−0.045	—
Δw_{t-2}	—	(0.0174)	—

[a] Figures in parentheses are standard errors.

TABLE IV

HOURS EQUATION

	Q	L	DF	P
(i) $m=3$	0.00	—	—	—
(ii) Parameters Stationary	26.69	26.69	34	0.810
(iii) $m=2$	27.33	0.64	2	0.726
(iv) $m=1$	34.09	6.76	2	0.034
(v) Exclude Wages, $m=2$	37.47	10.14	2	0.006
(vi) Exclude Wages, $m=1$	37.62	3.53	1	0.060

the case with the wage equation, we cannot reject the hypothesis that the appropriate specification is a first-differenced equation with constant lag parameters (see line (ii)). The unconstrained parameter estimates for this specification are presented in column (2) of Table II. As in the wage equation, the strongest effect both from the point of view of the absolute value of the coefficient and statistical significance, is the own first lag. However, while in the wage equation nothing else seems to "matter," in the hours equation, the second lag on wages is statistically significant.[22] Indeed, as in Table IV one cannot reject the hypothesis that $m=2$ (see line iii). Further restrictions depend, however, on the chosen level of significance. As line (iv) indicates, one cannot reject the hypothesis that $m=1$ at the 1% level. However, this conclusion is reversed using a 5% significance level.

Therefore, we test the hypothesis that lagged wages do not cause hours conditional on both $m=1$ and $m=2$. Using $m=2$ (and, thus, adopting a 5% significance level) one can reject the hypothesis that lagged wages may be excluded from the equation for hours (line (v)). The parameter estimates for this AR(2) model are shown in column (2) of Table III. Using the 1% level of

[22] When the SEO subsample was excluded, the data did not reject the hypothesis that lagged wages could be excluded from the hours equation.

significance leads to different results. As shown in line (vi) of Table IV, one cannot reject the hypothesis that lagged wages may be excluded from the AR(1) specification of the hours equation at a 5% significance level. The resulting parameter estimate is shown in column (3) of Table III.

Thus, using a 1% significance level, one is left with a parsimonious representation of the dynamic behavior of hours and wages. Both variables may be represented as autonomous autoregressive processes with a single lag. The robustness of this result is examined below.

(*iii*) *Deviations from Time Means*

It is interesting to determine whether an inappropriate statistical technique would lead to substantively different results. Recall that Lundberg tested for intertemporal relationships between wages and hours by conducting F-tests for exclusion of wages and hours from a VAR in which individual effects are removed by measuring all variables as deviations from individual time means. We applied this method to the PSID data. Specifically, we estimated a VAR with constant parameters and three lags of wages and hours, measuring all variables as deviations from time means. In direct contrast to the results presented above, this procedure indicates that hours Granger-cause wages and wages Granger-cause hours. The F statistic for the former test is 11.4 and for the latter 4.0. Both are significant at the 1% level. In short, in this context using an inappropriate estimation technique can lead to serious errors.

(*iv*) *Measurement Error*

The estimation procedure used in the wage and hours equations makes no special allowance for the possibility of measurement error. Altonji (1986) has estimated that a large part of the yearly variation in PSID data is due to measurement error. As noted in Section 3, the estimation procedure may be modified to accommodate measurement error by simply using a different set of instrumental variables.

We examine the effect of measurement error on our results by re-estimating the final form of both the wage and hours equations. In order to isolate the effect of measurement error, we focus on the correlation between the composite error term and the instrumental variables. For the AR(1) specification, measurement error will produce a correlation between instrumental variables dated $t-2$ and the composite error. In the absence of measurement error, no such correlation will be present. (See equation (3.8).) We estimate the equation for wages using two different sets of instrumental variables: (i) lagged wages dated both $t-2$ and $t-3$ and (ii) lagged wages dated $t-3$.[23] We estimate the hours equation using the corresponding sets of instrumental variables. One can formally test the null hypothesis of no correlation between the instrumental variables dated $t-2$ and

[23] Note that this set of instruments is more restrictive than that used in our previous estimations. This change is an attempt to increase the power of the test for measurement error.

TABLE V

MEASUREMENT ERROR CORRECTION[a]

	Wage Equation	Hours Equation
(i) Δh_{t-1}	—	0.169
		(0.0279)
Δw_{t-1}	0.179	—
	(0.050)	—
(ii) Δh_{t-1}		0.170
		(0.224)
Δw_{t-1}	0.460	—
	(0.172)	—
(iii) $\chi^2(5)$	6.23	1.427
	$(p = 0.284)$	$(p = 0.921)$

[a] Figures in parentheses are standard errors. Estimates in part (i) of the table use variables dated $t-2$ and $t-3$ as instrumental variables; the estimates in part (ii) use only variables dated $t-3$. Part (iii) contains the test statistic for a test of the null hypothesis of no correlation.

the error terms—and thus no measurement error—using a generalized method of moments test. (See Holtz-Eakin (forthcoming).)

The results are presented in Table V. Part (i) of the table shows the parameter estimates using instrumental variables lagged $t-2$ and $t-3$. Part (ii) gives the alternative parameter estimates using only variables dated $t-3$ as instrumental variables. Part (iii) contains the test statistic for a test of the null hypothesis of no correlation. The statistic is distributed as a chi-square with five degrees of freedom. The significance level of the test is shown in parentheses.

As the p-value indicates, the test of the null hypothesis of no correlation fails to reject at conventional significance levels.[24] It is important to note, however, that the estimated autoregressive parameter in the wage equation changes substantially with the measurement error correction. It increases from 0.179 to 0.460 when wages dated $t-2$ are excluded.

Turning to the hours equation, we find that the estimate of the autoregressive parameter for hours does not vary with the measurement error correction, although the precision of the estimate is affected. Not surprisingly, the null hypothesis of no measurement error in the hours equation cannot be rejected.

In sum, our attempt to correct for measurement error produces mixed signals. On one hand, changing the set of instrumental variables to allow for measurement error can have an important effect on the parameter estimates. On the other, the formal test suggests that this difference is not statistically significant. This suggests that the test may have low power in this particular application.

(v) *Evidence from the NLS*

As another check of the robustness of these results, we examine data on wages and hours from the National Longitudinal Survey of Men 45–59 (NLS). The

[24] A Hausman test on the difference between the two coefficients also fails to reject the hypothesis, although the p value of about 0.10 provides somewhat stronger evidence against the null hypothesis.

TABLE VI
NLS RESULTS

	Q	L	DF	P
	Hours Equation			
$m = 1$, All Parameters				
Stationary	21.21	—	16	0.170
Exclude wages	21.36	0.15	1	0.699
	Wage Equation			
$m = 1$, All Parameters				
Stationary	25.95	—	16	0.055
Exclude Hours	26.00	0.05	1	0.823

NLS sample consists of 1446 men who had positive earnings and hours in each of the survey years 1966, 1967, 1969, 1971, 1973, and 1975. (See Abowd and Card (1986) for details on the construction of the sample.) As in the PSID, we use two data series: the log of annual hours of work and the log of average hourly earnings. The latter is constructed as the difference between the log of annual earnings and the log of annual hours.

Use of the NLS is complicated by the above-noted fact that the data are not available for consecutive years. For the simple AR(1) model

$$y_{it} = f_i + \alpha_1 y_{it-1} + u_{it}, \tag{4.4}$$

the problem of missing years can be circumvented by successive substitution to yield a relationship between two-year differences:

$$y_{it} - y_{it-2} = \alpha_1^2 (y_{it-2} - y_{it-4}) + u_{it} + \alpha_1 u_{it-1} - u_{it-2} - \alpha_1 u_{it-3}. \tag{4.5}$$

Equation (4.5) is estimable using our methods. However, for an AR(2) (or longer lag), successive substitution gives an infinite order lag specification that is not estimable using our methods. Fortunately, the results from the PSID discussed above indicate (using the 1% significance level) that equation (4.4) may be an adequate representation of both wages and hours. For these reasons, we concentrate on the estimation of AR(1) models. We test the assumption of excludability and the overidentifying restrictions implied by the initial assumptions on stationarity and lag length.

The test results are contained in Table VI. For both wages and hours, we cannot reject the initial assumptions of stationarity and lag length; although the test for wages is borderline at the 5% level.[25] Under the identification restriction that both own lag coefficients are nonzero and have the same sign, we can test for causality by testing whether the wage is significant in the hours equation, and whether hours is significant in the wage equation. The result of these tests is that one cannot reject the hypotheses that hours do not cause wages and wages do not

[25] Note that we only identify the square of the matrix of autoregressive coefficients from the biannual data. These restrictions allow us to solve for the original matrix in terms of its square.

TABLE VII

NLS PARAMETER ESTIMATES[a]

	Wage Equation	Hours Equation
Δh_{t-2}	—	0.0263
	—	(0.0410)
Δw_{t-2}	0.0492	—
	(0.0259)	—

TRANSFORMED NLS PARAMETER ESTIMATES[a]

	Wage Equation	Hours Equation
Δh_{t-1}	—	0.162
	—	(0.127)
Δw_{t-1}	0.222	—
	(0.0584)	—

[a] Figures in parentheses are standard errors.

cause hours. The result in both cases is an equation of the form specified in (4.5). Thus, neither the PSID nor NLS data lend support to the notion that the current wage rate depends upon past hours of work.

The parameter estimates are shown in Table VII. Of course, because of the presence of the α_1^2 in equation (4.5), these parameters do not correspond directly to those we obtained using PSID data. To make the estimates comparable, we take square roots, imposing an identifying assumption that the underlying coefficients are positive. These transformed estimates (and their standard errors) are shown at the bottom of Table VII.

The correspondence between these point estimates and those in Table III is striking. This is particularly compelling when one considers that the two data sets cover different years and contain different types of individuals: the PSID has only hourly employees, while the NLS has only relatively old workers. With respect to the statistical significance of the estimates, in both data sets the lagged wage is significant at conventional levels. There is some disagreement with respect to lagged hours, however. For the PSID the t statistic for the coefficient of lagged hours in the hours equation is 8.90; for the NLS it is 1.28. While this evidence is mixed it does suggest that it is potentially dangerous to exclude lagged hours from the autoregressive hours equation. Whether this fact has consequences for the appropriate specification of the structural hours equation is unclear; it depends on whether the presence of lagged hours is due to serial correlation.[26]

[26] Using PSID data, we tested for the presence of first order serial correlation in ε_{it}^h of equation (4.1a) using a Wald test of the common factor restriction that the coefficient of H_{t-1} times the coefficient of w_{t-1} is equal to the negative of the coefficient of w_{t-2}. (This test is valid under the assumption that only w_{t-1} appears in the wage equation.) The test indicated that one can reject the hypothesis of first order autocorrelation at a 3 percent significance level. Of course, testing for other patterns of serial correlation might produce a different result. Hence, while this test is suggestive that the presence of lagged hours is indeed due to "structural" considerations, it is not conclusive.

TABLE VIII

NLS MEASUREMENT ERROR CORRECTION[a]

	Wage Equation	Hours Equation
(i) Δh_{t-2}	—	-0.092
	—	(0.0597)
Δw_{t-2}	0.063	—
	(0.0312)	—
(ii) Δh_{t-2}	—	-0.224
	—	(0.35)
Δw_{t-2}	0.420	—
	(0.154)	—
(iii) $\chi^2(3)$	10.913	2.295
	$(p = 0.017)$	$(p = 0.513)$

[a] Figures in parentheses are standard errors. Estimates in part (i) of the table use variables dated $t-2$ and $t-4$ as instrumental variables; the estimates in part (ii) use only variables dated $t-4$. Part (iii) contains the test statistic for a test of the null hypothesis of no correlation.

Finally, in the same fashion as for the PSID we check for the importance of measurement error by re-estimating the equations using alternative sets of instrumental variables. The results for the untransformed parameter estimates are shown in Table VIII. Comparing these results to those in the top of Table VII, we see that much like the results from the PSID, the most seriously affected coefficient is that in the wage equation. Here, however, the formal test for the wage equation supports the hypothesis of measurement error at close to the 1% level.[27] As before, for the hours equation, the hypothesis of measurement error is not supported.

5. CONCLUSION

We have presented a simple method of estimating vector autoregression equations using panel data. The key to its simplicity is the fact that estimation and testing have straightforward GLS interpretations—no nonlinear optimization is required.

We applied our estimation procedure to the study of dynamic relationships between wages and hours. Our empirical results are consistent with the absence of lagged hours in the wage forecasting equation, and thus with the absence of certain human capital or dynamic incentive effects. Our results also show that lagged hours is important in the hours equation, which is consistent with alternatives to the simple labor supply model that allow for costly hours adjustment or preferences that are not time separable. As usual, of course, these results might be due to serial correlation in the error term or to a functional form misspecification. However, we find it encouraging that broadly similar results are obtained from two different data sets.

[27] A Hausman test also supports the measurement error hypothesis for the wage equation; the p value is about 0.02.

More generally, our empirical example demonstrates the importance of testing for the appropriate lag length prior to causality testing, an issue of considerable importance in short panels. In the absence of such tests, no inferences concerning causal relationships can be drawn. The example also demonstrates that use of inappropriate methods to deal with individual effects in the VAR context can lead to highly misleading results.

Columbia University, New York City, New York, U.S.A.
and
Princeton University, Princeton, NJ, U.S.A.

Manuscript received June, 1985; final revision received October, 1987.

REFERENCES

ABOWD, JOHN M., AND DAVID CARD (1986): "On the Covariance Structure of Earnings and Hours Changes," mimeo, Princeton University, October, 1986.

ALTONJI, J. (1986): "Intertemporal Substitution in Labor Supply: Evidence from Micro Data," *Journal of Political Economy*, 94, S176–S215.

ALTONJI, J., AND C. PAXSON (1986): "Job Characteristics and Hours of Work," *Research in Labor Economics*.

ANDERSON, T. W., AND C. HSIAO (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

ASHENFELTER, ORLEY, AND D. CARD (1982): "Time Series Representations of Economic Variables and Alternative Models of the Labour Market," *Review of Economic Studies*, 49, 761–782.

BROWNING, M. J., A. S. DEATON, AND M. IRISH (1985): "A Profitable Approach to Labor Supply and Commodity Demands Over the Life-Cycle," *Econometrica*, 53, 503–544.

CHAMBERLAIN, GARY (1983): "Panel Data," Chapter 22 in *The Handbook of Econometrics Volume II*, ed. by Z. Griliches and M. Intrilligator. Amsterdam: North-Holland Publishing Company.

FISHER, FRANKLIN (1966): *The Identification Problem in Econometrics*. Huntington, N.Y.: Krieger Publishing Company.

GALLANT, RONALD, AND D. JORGENSON (1979): "Statistical Inference for a System of Nonlinear, Implicit Equations in the Context of Instrumental Variables Estimation," *Journal of Econometrics*, 11, 275–302.

GRILICHES, ZVI, AND J. HAUSMAN (1984): "Errors in Variables in Panel Data," National Bureau of Economic Research Technical Working Paper No. 37, May, 1984.

HAMILTON, BRUCE (1986): "Merit Pay Increases and the Supply of Labor," mimeo, Johns Hopkins University, November, 1986.

HANSEN, LARS (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HECKMAN, J. J., AND T. E. MACURDY (1980): "A Life Cycle Model of Female Labor Supply," *Review of Economic Studies*, 47, 47–74.

HOLTZ-EAKIN, D. (FORTHCOMING): "Testing for Individual Effects in Autoregressive Models," forthcoming in *Journal of Econometrics*.

HOLTZ-EAKIN, D., W. NEWEY, AND H. ROSEN (FORTHCOMING): "The Revenues-Expenditures Nexus: Evidence from Local Government Data," forthcoming in *International Economic Review*.

LUNDBERG, S. (1985): "Tied Wage-Hours Offers and the Endogeneity of Wages," *Review of Economics and Statistics*, 67, 405–410.

MACURDY, T. (1981a): "Time Series Models Applied to Panel Data," mimeo, Stanford University.

——— (1981b): "An Empirical Model of Labor Supply in a Life Cycle Setting," *Journal of Political Economy*, 89, 1059–1086.

——— (1983): "A Simple Scheme for Estimating an Intertemporal Model of Labor Supply and Consumption in the Presence of Taxes and Uncertainty," *International Economic Review*, 24, 265–289.

NEWEY, WHITNEY, AND K. WEST (1987): "Hypothesis Testing with Efficient Method of Moment Estimation," *International Economic Review*, 28, 777–787.

NICKELL, STEPHEN (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417–1426.

PAKES, ARIEL, AND ZVI GRILICHES (1984): "Estimating Distributed Lags in Short Panels with an Application to the Specification of Depreciation Patterns and Capital Stock Constructs," *Review of Economic Studies*, 51, 243–262.

SARGAN, J. D. (1980): "Some Tests of Dynamic Specification for a Single Equation," *Econometrica*, 48, 879–898.

TAYLOR, JOHN (1980): "Output and Price Stability—An International Comparison," *Journal of Economic Dynamics and Control*, 2, 109–132.

WHITE, HALBERT (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

——— (1982): "Instrumental Variables Regression with Independent Observations," *Econometrica*, 50, 483–500.