©Journal of Business & Economic Statistics, Vol. 1, No. 4, October 1983

Lag Specification in Rational Distributed Lag Structural Models

Dominique M. Hanssens
Graduate School of Management, University of California, Los Angeles, CA 90024

Lon-Mu Liu
Department of Quantitative Methods, University of Illinois, Chicago, IL 60680

This article presents methods for the preliminary specification of distributed lags in structural models in the absence of sufficient a priori information. It is argued that an instrumental-variables procedure produces good results as long as the goodness of fit in the instrument equations is satisfactory. When this is not the case, simple OLS specification is preferable to more complex instrumental-variables methods, even though it may produce some spurious parameters. The procedures are illustrated with a simulated two-equation system and a model of annual supply and price of hogs.

1. INTRODUCTION

One of the most difficult and important problems in structural model building on time series data is the identification or specification of distributed lag relationships. This problem is approached from different angles in the econometric and the time series literature. [We are restricting the discussion to the time-domain literature (i.e., excluding spectral methods).] Econometricians typically suggest that distributed lag effects be specified from a priori considerations and estimated via clever data transformations (e.g., Koyck, Almon) or theoretically justified (and often complicated) estimation techniques. Time series analysts typically specify a lag structure from cross-correlation analysis, but offer little guidance in parameter estimation beyond what is known from econometrics. In both cases, however, the *specification* part is underdeveloped: on the one hand, a priori theory is seldom able to identify a lag structure (e.g., economists usually do not predict how long it takes for supply to react to a price change), and on the other hand, cross-correlation analysis is often unreliable in multiple-input models, especially those with feedback relationships. Since it is well known that the properties of estimates in a model are conditional upon the model being adequately specified, this problem merits further attention.

Several methods for specifying lag structures have been proposed in recent years. Restricting the discussion to those techniques that can handle multiple-equation models with feedback effects, two basic approaches have emerged, albeit with several variants: pairwise cross-correlation methods on prewhitened data (e.g. Haugh and Box 1976, Granger and Newbold 1977), and vector ARMA methods (e.g. Jenkins and Alavi 1981; Tiao and Box 1981). The first technique appears to be most valuable when the direction of causality between two variables is a priori unclear; however, it is cumbersome to derive lag structures from cross-correlations on prewhitened data, especially in k-variate models with $k > 2$. For example, Granger and Newbold's proposed specification technique is actually restricted to the bivariate case (Granger and Newbold 1977, pp. 244–254). On the other hand, the vector ARMA methods, which are extensions of the famous Box-Jenkins technique for univariate analysis, are specifically designed for the k-variate case. Initial experience with this approach (e.g. Jenkins and Alavi 1981, Tiao and Box 1981, Tiao and Tsay 1983) is encouraging in the sense that it produces dynamic *reduced*-form models and facilitates better understanding of the interrelationships among the variables. However, it is not directly applicable when structural-form modeling is desired, and it does not necessarily lead to parsimonious models.

The present approach is different from either double prewhitening or vector ARMA modeling and can be used in both reduced-form and structural-form models. We will make use of the frequently occurring case in applied econometric modeling in which the researcher has good a priori knowledge (usually from subject-matter theory) about the endogenous and the exogenous variables in each equation, but poor a priori understanding of the dynamics of the system. We propose that simpler and more robust lag specification can be done by using this a priori information. Practically speaking, this approach positions our method somewhere in the middle, between "pure" time series modeling (i.e., without any theory-driven lag specification) and "pure" econometric modeling (i.e., without any data-driven lag specification).

The starting point for our procedure is found in earlier work on least squares identification of multiple-input transfer function models (Liu and Hanssens 1982). After summarizing this method, we develop its extension in multiple-equation situations based on instrumental-variables techniques well known in econometrics, and discuss properties of the technique. Two illustrations are provided: one based on simulated data and one on actual data of annual hog price and supply.

2. STATISTICAL BACKGROUND

The model we advocate for the study of distributed lags is the "rational distributed lag structural form" (RSF), proposed by Wall (1976). Although not the most general in its class, it is a very useful and sufficiently flexible model, which combines the classical multiple regression model with Box-Jenkins transfer function and ARMA noise processes. Also, since the RSF model can include multiple equations, it is adequate for modeling feedback and bidirectional causality cases, which are fairly common in the social sciences. In its most general form the RSF is

$$\mathbf{B}(\mathrm{L})\mathbf{Y}_t = \mathbf{\Gamma}(\mathrm{L})\mathbf{X}_t + \mathbf{u}_t,$$

$$\mathbf{u} = \mathbf{\Lambda}(\mathrm{L})\mathbf{e}_t,$$

and

$$\mathbf{e}_t \sim N(\mathbf{0}, \mathbf{\Sigma}),$$

where $\mathbf{B}$, $\mathbf{\Gamma}$, and $\mathbf{\Lambda}$ are rational matrix operators for the endogenous, exogenous, and disturbance terms, respectively, $\mathbf{\Lambda}(\mathrm{L})$ is a diagonal matrix, and L is the lag operator. The model is assumed to satisfy the conditions of identifiability, stability, and invertibility (Wall 1975). The conditions of identifiability are discussed in Section 3.2.

Without loss of generality, we study the following dynamic simultaneous system with two endogenous variables Y_1 and Y_2 and two exogenous variables X_1 and X_2. We assume that the input variables for each equation are known and that the equations can be represented as

$$Y_{1t} = \beta_1(\mathrm{L})X_{1t} + \gamma_1(\mathrm{L})Y_{2t} + u_{1t},$$

$$Y_{2t} = \beta_2(\mathrm{L})X_{2t} + \gamma_2(\mathrm{L})Y_{1t} + u_{2t}, \qquad (2.1)$$

where $\beta_i(\mathrm{L})$ and $\gamma_i(\mathrm{L})(i = 1, 2)$ are polynomial or *rational* parameter functions in the lag operator L. The dynamic regression coefficients $\beta_i(\mathrm{L})$ and $\gamma_i(\mathrm{L})$ are also referred to as the transfer functions between the corresponding input and output variables. In general, the transfer function may be expressed as $\omega(\mathrm{L})/\delta(\mathrm{L})$ where

$$\omega(\mathrm{L}) = (\omega_1 + \omega_2\mathrm{L} + \ldots + \omega_s\mathrm{L}^{s-1})\mathrm{L}^b,$$

and

$$\delta(\mathrm{L}) = 1 - \delta_1\mathrm{L} - \delta_2\mathrm{L}^2 - \ldots - \delta_r\mathrm{L}^r,$$

where all roots of the δ polynomial lie outside the unit circle. The disturbances u_{1t} and u_{2t} each follow ARMA processes

$$\phi_i(\mathrm{L})u_{it} = \theta_i(\mathrm{L})e_{it}, \quad i = 1, 2,$$

where e_{1t} and e_{2t} are vector white noise with covariance matrix $\mathbf{\Sigma}$, and

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}.$$

Without loss of generality, we also assume that all the series in the system are stationary and have zero means, possibly after a suitable stationarity-inducing transformation.

Ignoring the correlation between u_{1t} and u_{2t}, we find that each equation in the dynamic simultaneous system is similar to a transfer function model except that the random error (u_{it}) may be correlated with the endogenous variable (Y_{jt}, $j \neq i$). Thus, we could apply techniques for the specification of transfer functions in each equation of the simultaneous system, for example the least squares identification procedure by Liu and Hanssens (1982). This procedure is based largely on the fact that each rational form transfer function $\omega(\mathrm{L})/\delta(\mathrm{L})$ can be expressed in linear form as

$$V(\mathrm{L}) = v_0 + v_1\mathrm{L} + v_2\mathrm{L}^2 + \ldots$$

by power series expansion where $V(\mathrm{L}) \equiv \omega(\mathrm{L})$ has a finite number of terms if $\delta(\mathrm{L}) \equiv 1$ and an infinite number of terms if $\delta(\mathrm{L}) \neq 1$. In the latter case, the function may be approximated by a finite number of terms since all roots of the δ polynomial lie outside the unit circle. Thus, the rational form model (2.1) can always be approximated in linear form as below:

$$Y_{1t} = V_{11}(\mathrm{L})X_{1t} + V_{12}(\mathrm{L})Y_{2t} + u_{1t},$$

$$Y_{2t} = V_{21}(\mathrm{L})X_{2t} + V_{22}(\mathrm{L})Y_{1t} + u_{2t}. \qquad (2.2)$$

Liu and Hanssens's procedure consists of two steps. The first step obtains the estimates of the transfer function weights $V(\mathrm{L})$ for each input variable by performing OLS or GLS estimation on a linear form of the model. After a linear transfer function is obtained, it can be expressed in rational form if necessary, by using an extension of the corner method (Beguin, Gourieroux, and Monfort 1980, Liu and Hanssens 1982).

The unique problem in the application of Liu and Hanssens's procedure to RSF models is that the disturbances u_{1t} and u_{2t} may not be independent of Y_{2t} and Y_{1t}, causing the OLS estimates of the transfer function weights to be inconsistent. In the estimation literature this problem is usually handled by instrumental-variables (IV) regression, such as two-stage least squares (2SLS). The procedure proposed in this article is to apply the instrumental-variables approach to the estimation of transfer function weights as follows:

1. We obtain an approximation of Y_{1t} and Y_{2t} via lag regression over X_{1t} and X_{2t}, using a sufficiently large number of lags:

$$Y_{1t} = \pi_{11}(L)X_{1t} + \pi_{12}(L)X_{2t} + n_{1t},$$
$$Y_{2t} = \pi_{21}(L)X_{1t} + \pi_{22}(L)X_{2t} + n_{2t}. \qquad (2.3)$$

The estimates of the endogenous variables, say $\hat{Y}_{1t}$ and $\hat{Y}_{2t}$, are independent of u_{1t} and u_{2t}.

2. Consequently, we can apply least squares estimation (OLS or GLS, see Liu and Hanssens 1982) of transfer function weights for (2.2) using the model

$$Y_{1t} = V_{11}(L)X_{1t} + V_{12}(L)\hat{Y}_{2t} + u'_{1t}$$
$$Y_{2t} = V_{21}(L)X_{2t} + V_{22}(L)\hat{Y}_{1t} + u'_{2t}. \qquad (2.4)$$

3. After the linear form transfer functions are appropriately estimated, the estimates of u_{1t} and u_{2t} in (2.2) may be obtained by

$$\hat{u}_{1t} = Y_{1t} - \hat{V}_{11}(L)X_{1t} - \hat{V}_{12}(L)Y_{2t},$$
$$\hat{u}_{2t} = Y_{2t} - \hat{V}_{21}(L)X_{2t} - \hat{V}_{22}(L)Y_{1t}. \qquad (2.5)$$

We can then identify ARMA models for u_{1t} and u_{2t} based on $\hat{u}_{1t}$ and $\hat{u}_{2t}$.

This instrumental-variables procedure provides initial information about the lag structure of the transfer functions for each input variable. As discussed in Section 3.1, the least squares estimates of the transfer function weights for model (2.2) are consistent but not necessarily efficient since we do not use the information that u_{1t} and u_{2t} may be correlated.

3. PROPERTIES OF THE METHOD

3.1 Consistency of the Transfer Function Weights

A careful use of the proposed method can be shown to produce consistent estimates for the linear transfer functions in model (2.2). To see this, one can view the proposed method as the two-stage least squares method since essentially no outside exogenous variables (i.e., variables absent from the structural model) are used to create the instruments. When the $\mathbf{u}_t$s are independent (not autocorrelated), the consistency property has been proved in various econometric textbooks (e.g., Maddala 1976, pp. 475–478). Cragg (1982) also shows that consistency holds when each individual u_{it} follows a moving-average model, which can be easily extended to an ARMA model.

In the two-stage least squares method, all exogenous variables (including their lags) are used to create instruments. However, the lags for the linear structural model (2.2) are yet to be identified. To ensure the consistency of the method and also to improve efficiency, it is important to include all statistically significant lagged exogenous variables in the first-stage estimation.

We limit our attention to model (2.2) in the previous discussion of consistency. Since model (2.2) is only an approximation of (2.1), it will contain truncation bias for the terms with rational transfer functions. Therefore, strictly speaking, the transfer function weights obtained by using two-stage least squares may not be consistent estimates for the "true" transfer function weights. However, this deficiency will not limit the application of the proposed method since we are more interested in the pattern of the transfer function weights than in the actual estimates.

3.2 Identifiability of the Model

Both the econometric and the time series literature recognize that models are not necessarily unique in generating the same likelihood of the data. This problem is referred to as "identification" in econometrics, and "model multiplicity" in time series analysis; Granger and Newbold (1977) conveniently use the terms "E-identification" and "TS-identification."

Neither E-identification nor TS-identification are necessarily satisfied a priori because the ultimate form of the model is determined posterior to data analysis. Therefore, some conditions need to be met *prior* to empirical analysis. For linear form models (e.g., model (2.2)), a complete set of identification conditions has been given by Hannan (1971) and Hatanaka (1975). However, conditions for the rational-form models (e.g., model (2.1)) are not well established. Assuming linear form models, we can use Hatanaka's (1975) definition of an *excluded variable* as one that does not appear in current or lagged form in an equation. We assume that the model contains a sufficient number of excluded variables in each equation so as to make it (over) identified. (The necessary condition is that there are $(m - 1)$ or more exclusions in each of the m equations. The necessary and sufficient condition is that the matrix formed by the columns of excluded variables in each equation is of rank $(m - 1)$. See Granger and Newbold (1977, pp. 219–224) and Hatanaka (1975).) In other words, we impose a dynamic version of E-identification on the model, which will also ensure TS-identification. Following Hatanaka's rule, it is easy to see that the linear form model (2.2) is identified as long as the original rational form model (2.1) is identified. For rational-form models, Hatanaka's identification conditions seem to be stronger than necessary, so that weaker conditions may be possible (see, e.g., Wall 1975, Kohn 1979).

Should the a priori model not be E-identified, then one may or may not obtain a unique empirical lag specification. In such cases, lag specification should be done with great care, and unique results are not guaranteed a priori.

3.3 Improving the Fit of Instruments

It is well known that the instrumental-variables method will not produce good second-stage estimates if the goodness of fit (R^2) for the instruments is poor. To remedy this deficiency, it would appear appropriate to model the noise in the first-stage estimation by ARMA processes. Such a procedure will be referred to as a "combined procedure," which is as follows:

1. Obtain a better approximation $\hat{Y}_{1t}$ and $\hat{Y}_{2t}$ of Y_{1t} and Y_{2t} via lag regression over X_{1t} and X_{2t} (i.e., model (2.3)), but with added noise represented by ARMA models,
2. Apply least squares estimation of transfer function weights for (2.2) using the model

$$Y_{1t} = V_{11}(\mathrm{L})X_{1t} + V_{12}(\mathrm{L})\tilde{Y}_{2t} + u'_{1t}$$
$$Y_{2t} = V_{21}(\mathrm{L})X_{2t} + V_{22}(\mathrm{L})\tilde{Y}_{1t} + u'_{2t} \qquad (3.1)$$

where $\tilde{Y}_{it} = \hat{Y}_{it}$ if the lag order (i.e., the power of L) is 0, and $\tilde{Y}_{it} = Y_{it}$ if the lag order is greater than 0.

Note that this procedure may not produce consistent estimates since $\tilde{Y}_{it}$ ($\hat{Y}_{it}$ and lagged Y_{it}s) may be correlated with u_{jt} ($j \neq i$) due to the presence of contemporaneous and serial correlations of u'_{1t} and u'_{2t}.

As will be explained later, we find that this combined procedure is inappropriate, even in cases where a substantial improvement in instrument fit over the strict instrumental-variables method is obtained. In such cases, it is better to accept the inconsistency of the OLS specification method.

The instrumental-variables procedure has been known for a long time, and the combined procedure and its variants have been suggested by Sargan (1961), Phillips (1978), Zellner and Palm (1974), Palm and Zellner (1980), and Perryman (1980). However, they are proposed for the purpose of model estimation rather than of model specification. In this article, we use these rather simple and robust methods to determine the algebraic form of each equation in a system. After the model for a system is completely specified, we can then use more efficient methods such as Full Information Maximum Likelihood (FIML), to estimate the parameters. Because of the recent development of time series computer programs (e.g., Liu et al. 1983), the computations in the proposed specification procedure can be performed easily.

4. SIMULATED EXAMPLE

The simulation model used to illustrate the proposed procedure is

$$Y_{1t} = \frac{5\mathrm{L}}{1 - 1.1\,\mathrm{L} + 0.3\mathrm{L}^2} X_{1t} - .5\,Y_{2t} + (1 - .6\mathrm{L})e_{1t},$$

$$Y_{2t} = (-.6 - .4\mathrm{L})X_{2t} + \frac{.7\mathrm{L}^2}{1 - .6\mathrm{L}} Y_{1t} + \frac{1}{1 - .5\mathrm{L}} e_{2t}.$$

120 observations were simulated, using the following noise series:

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} 1.21 & .55 \\ .55 & 1.00 \end{bmatrix}\right),$$

which implies that the contemporaneous correlation between e_{1t} and e_{2t} is .5. The exogenous variables were simulated as

$$X_{1t} = (1 - .5\mathrm{L})a_{1t},$$
$$(1 - .6\mathrm{L})X_{2t} = a_{2t},$$

with the noise process:

$$\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} 6.76 & 1.79 \\ 1.79 & 5.29 \end{pmatrix}\right),$$

so that the contemporaneous correlation between X_1 and X_2 is .3.

For the purpose of evaluating various model specification procedures, it is useful to compute the "true" transfer function weights of the model shown in Table 1. In the first step, the instruments $\hat{Y}_1$ and $\hat{Y}_2$ are obtained by simple OLS (see Table 2). The R^2's for these equations are .91 and .94, which is quite satisfactory. Many of the parameters are statistically significant(*), although one should realize that the reported parameter standard deviations could be underestimated, as there is no guarantee that the residuals are white noise.

In the second step, the instruments $\hat{Y}_1$ and $\hat{Y}_2$ are used in two polynomial lag regressions, still using OLS. The choice of number of lags was made with the true models in mind; in real-world situations, this choice is not critical as long as a *sufficient* number of lags is used (i.e., including all significant lags). The results are shown in Table 3. The goodness of fit is .92 and .95. More importantly, the OLS impulse response weights are uniformly close to the theoretical values. Limiting the analysis to the non-trivial transfer function specification, these weights produce the results shown in corner Tables 4 and 5. The tables suggest that the transfer functions are ($r = 2, s = 1, b = 1$) and ($r = 1, s = 1, b = 2$), respectively.

At this point, we would specify a "correct" distributed lag model, except for the added noise terms. Using (2.5), the residuals of this system of equations revealed an MA(1) process on u_{2t} and an AR(1) process on u_{2t}, as expected.

Using a computer program developed by Liu et al. (1983) the structural model has the following FIML

Table 1. Implied Transfer Function Weights of Simulated Models

Equation	Variable	v_0	v_1	v_2	v_3	v_4	v_5	v_6	v_7
1	X_1	0	.50	.55	.46	.34	.24	.16	.10
	Y_2	−.50	0	0	0	0	0	0	0
2	X_2	−.60	−.40	0	0	0	0	0	0
	Y_1	0	0	.70	.42	.25	.15	.09	.05

Table 2. Instrument Equations

Equation	Variable	$\hat{\pi}_0$	$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\pi}_3$	$\hat{\pi}_4$	$\hat{\pi}_5$	$\hat{\pi}_6$	$\hat{\pi}_7$	$\hat{\pi}_8$
1	X_1	.063*	.474*	.512*	.285*	.023	−.041	−.007	.002	.005
		(.028)	(.030)	(.030)	(.030)	(.030)	(.030)	(.030)	(.029)	(.025)
	X_2	.262*	.257*	−.092*	−.174*	−.044	.001	.034	.019	.003
		(.031)	(.038)	(.038)	(.039)	(.039)	(.039)	(.039)	(.040)	(.031)
	constant	.016	(.148)							
2	X_1	−.027	.033	.087*	.408*	.589*	.542*	.328*	.118*	.032
		(.032)	(.034)	(0.35)	(.035)	(.035)	(.034)	(.034)	(.034)	(.029)
	X_2	−.605*	−.401*	.162*	.309*	.093*	−.027	−.061	−.019	.029
		(.036)	(.044)	(.044)	(0.45)	(.045)	(.045)	(.046)	(.046)	(.036)
	constant	−.056	(.171)							

estimates:

$$Y_{1t} = \frac{.500L(.016)}{1 - \underset{(.033)}{1.130}L + \underset{(.030)}{.330}L^2} X_{1t} - \underset{(.010)}{.510} Y_{2t} + (1 - \underset{(.063)}{.686}L)a_{1t},$$

$$Y_{2t} = (\underset{(.018)}{-.596} - \underset{(.018)}{.412}L)X_{2t} + \frac{.700L^2(.019)}{1 - .615L(.016)} Y_{1t} + \frac{1}{1 - \underset{(.070)}{.609}L} a_{2t},$$

$$\hat{\Sigma} = \begin{bmatrix} 1.593 & .722 \\ .722 & 1.140 \end{bmatrix},$$

which confirms the originally simulated structure.

We also investigated the specification error if a straight OLS procedure is used. The estimates for the transfer function weights are shown in Table 6, which identifies the following model:

$$Y_{1t} = \frac{\omega_{111} + \omega_{112}L}{1 - \delta_{111}L - \delta_{112}L^2} X_{1t} + (\omega_{121} + \omega_{122}L + \omega_{123}L^2) Y_{2t} + (1 - \theta_{11}L)a_{1t},$$

$$Y_{2t} = (\omega_{211} + \omega_{212}L) X_{2t} + \frac{\omega_{221}L^2}{(1 - \delta_{221}L)} Y_{1t} + \frac{1}{1 - \phi_{21}L} a_{2t}.$$

The first equation has three spurious lags, but the second equation is correctly identified. The FIML estimates of this model are

$$Y_{1t} = \frac{.052(.018) + .430L(.029)}{1 - \underset{(.037)}{1.156}L + \underset{(.034)}{.351}L^2} X_{1t} + (\underset{(.024)}{-.472} - \underset{(.036)}{.074}L + \underset{(.022)}{.049}L^2) Y_{2t} + (1 - \underset{(.062)}{.686}L)a_{1t},$$

$$Y_{2t} = (\underset{(.018}{-.597} - \underset{.018)}{.409}L) X_{2t} + \frac{.708L^2(.020)}{1 - .603L(.018)} Y_{1t} + \frac{1}{(1 - \underset{(.071)}{.594}L)} a_{2t},$$

$$\hat{\Sigma} = \begin{bmatrix} 1.419 & .653 \\ .653 & 1.130 \end{bmatrix}$$

The three spurious parameters are only marginally significant and, thus, the specification bias caused by OLS is relatively minimal. In several other unreported simulations, we have found that OLS often provides useful initial information about lag structures, even though an equation with a contemporaneous endogenous variable is not identified correctly.

The key determinant of the quality of the specification results is the goodness of fit in the instrument equations. Several additional simulations were run with different variances for a_{1t} and a_{2t}, which imply different R^2 in the

Table 3. Lag Specification Model: IV Method

Equation	Variable	$\hat{v}_0$	$\hat{v}_1$	$\hat{v}_2$	$\hat{v}_3$	$\hat{v}_4$	$\hat{v}_5$	$\hat{v}_6$	$\hat{v}_7$	$\hat{v}_8$
1	X_1	.043	.488*	.554*	.470*	.329*	.217*	.161*	.051*	.009
		(.029)	(.031)	(.032)	(.039)	(.038)	(.038)	(.038)	(.036)	(.030)
	$\hat{Y}_2$	−.452*	−.106	.124	−.107	.106	−.034			
		(.052)	(.087)	(.106)	(.107)	(.085)	(.047)			
	constant	.013	(.154)							
2	X_2	−.650*	−.406*	−.001	.011	−.012	.070			
		(.040)	(.048)	(.053)	(.053)	(.054)	(.047)			
	$\hat{Y}_1$	.121	−.080	.787*	.405*	.204*	.157*	.060	−.024	.133*
		(.079)	(.105)	(.108)	(.106)	(.102)	(.095)	(.095)	(.083)	(.056)
	constant	−.069	(.173)							

Table 4. Corner Table: X_1 in Equation 1

$b + s - 1$	r = 1	2	3	4
0	.08	.01	.00	.00
1	.88	.70	.55	.43
2	1.00	.25	−.02	−.03
3	.85	.13	.01	.00
4	.59	.02	.00	.00

Table 5. Corner Table: Y_1 in Equation 2

$b + s - 1$	r = 1	2	3	4
0	.16	.02	.00	.00
1	−.10	−.14	.04	.02
2	1.00	1.05	1.11	1.17
3	.51	−.01	.07	.07
4	.26	−.04	.00	.01

first-stage regressions. Our findings can be summarized as follows:

1. As long as the exogenous variables explain a reasonable portion of the variance in the system (in our experience an R^2 of .6 or better in this simulated model), the IV method performs well; that is, it leads to a dynamic lag structure that is consistent with the theoretical structure;

2. When the instrument fit is poor, the specification results *cannot* be improved by the combined method. The combined method deteriorates rapidly as the variances in a_{1t} and a_{2t} are reduced, and thus it is an unreliable specification tool in such circumstances. On the other hand, when the instrument fit is good, the combined method provides results that are essentially the same as under the IV approach. Therefore, this method is always dominated by the simpler, more robust IV technique;

3. Instead of trying to improve the instrument fit by the combined method, one should compare specification findings under OLS and IV when the instrument R^2 is low. Unless the variances in a_{1t} and a_{2t} are very small, this strategy leads to mild overparameterization of the structure, which probably will be rectified in the ultimate estimation of the system. However, when the exogenous variables have no explanatory power, one cannot adequately assess the contemporaneous endogenous effects in the system under any method. Reduced-form modeling is probably the only viable approach in such situations.

5. RECOMMENDATIONS FOR THE PROCEDURE

The various simulation experiments revealed that the instrumental-variables method works very well. In general, this procedure will provide adequate results if the goodness of fit for the instruments is satisfactory. When the instrument-R^2 is low, the result under the combined procedure is no better than the instrumental-variables or the OLS procedure. This is because the mixture of instruments and lagged endogenous variables causes the latter variables to explain more information than they do if the instruments are poor. In such a situation, we suggest using *both* the instrumental-variables and the OLS procedures to reconcile an overparameterized model. If the system is only slightly overparameterized, the spurious parameters will probably be rectified in the estimation stage, as shown in the simulated example. When the goodness of fit is small for all the instruments, it implies that the system is not driven by the exogenous variables. Such a system behaves almost like a closed-loop time series model and any attempt to model contemporaneous relationships may not be fruitful.

When a system does not have any contemporaneous endogenous parameters, the rational structural-form model is an alternative parameterization of the vector ARMA model. The OLS procedure can be considered as a variant of the partial autoregression procedure discussed in Tiao and Box (1981). Tentative model specification for this class of models is quite straightforward.

In the previous discussions, we assume that all the variables are stationary. For nonstationary variables, we may use their differenced data for the analysis. When the entire system is near-nonstationary, the common filtering techniques proposed in Liu and Hanssens (1982) may be applied to all the variables in the system. This recommendation is illustrated in the following model of hog price and supply.

6. APPLICATION TO HOG SUPPLY AND PRICE DATA

In his important book on multiple time series, Quenouille (1957) published an interesting data set on annual

Table 6. Lag Specification Model: OLS Method

Equation	Variable	$\hat{v}_0$	$\hat{v}_1$	$\hat{v}_2$	$\hat{v}_3$	$\hat{v}_4$	$\hat{v}_5$	$\hat{v}_6$	$\hat{v}_7$	$\hat{v}_8$
1	X_1	.058*	.495*	.555*	.459*	.324*	.232*	.157*	.052	.006
		(.023)	(.025)	(.026)	(.029)	(.030)	(.031)	(.030)	(.029)	(.024)
	Y_2	−.404*	−.166*	.128*	−.035	.018	.019			
		(.036)	(.054)	(.060)	(.060)	(.051)	(.032)			
	constant	.025	(.127)							
2	X_2	−.594*	−.410*	−.023	.012	−.021	.066			
		(.026)	(.033)	(.035)	(.035)	(.035)	(.031)			
	Y_1	.018	−.005	.713*	.450*	.178*	.179*	.044	−.044	.076*
		(.042)	(.049)	(.049)	(.048)	(.046)	(.043)	(.042)	(.039)	(.031)
	constant	−.102	(.120)							

Table 7. IV Lag Specification Model: Hog Supply and Price

Equation	Variable	$\hat{v}_0$	$\hat{v}_1$	$\hat{v}_2$	$\hat{v}_3$	$\hat{v}_4$	$\hat{v}_5$
HS	HP	—	.246*	.045	−.023	.037	.022
			(.053)	(.055)	(.049)	(.048)	(.037)
	CS	.068	.350*	.029	−.057	.036	−.009
		(.060)	(.067)	(.069)	(.070)	(.070)	(.064)
	CP	.029	.056	−.217*	−.148*	−.009	−.001
		(.039)	(.042)	(.050)	(.053)	(.056)	(.058)
	constant	52.341	(50.215)				
HP	$\hat{HS}$	−.588*	−.903*	−.236	—	—	—
		(.294)	(.286)	(.275)			
	CP	.164*	.059	.169	.137	.002	−.077
		(.080)	(.105)	(.108)	(.109)	(.101)	(.096)
	FW	.823*	.402	−.339	.146	−.254	.225
		(.273)	(.292)	(.292)	(.288)	(.306)	(.229)
	constant	239.943	(100.557)				

hog supply, price, and the related variables corn supply, corn price, and farm wages. A fairly simple econometric model can be postulated to underlie movements in hog supply and hog price, but it is difficult to specify the dynamics of such a model a priori. In addition, since the data are annual it is reasonable to expect a contemporaneous relationship between price and supply. Therefore, Quenouille's hog data appear to be appropriate as a test case for our specification procedure. The following analysis is given mainly for illustrative purposes. We realize that there are several problems associated with this data set, such as possible outliers, adjustment of the original data, and other potential endogenous and exogenous variables in the model.

The endogenous variables hog supply (HS and hog price (HP) are postulated to be affected by corn supply (CS), corn price (CP) and farm wages (FW) as follows:

$$\mathrm{HS}_t = c_1 + \gamma_1(\mathrm{L})\mathrm{HP}_{t-1} + \beta_1(\mathrm{L})\mathrm{CS}_t + \beta_2(\mathrm{L})\mathrm{CP}_t + u_{1t},$$

$$\mathrm{HP}_t = c_2 + \gamma_2(\mathrm{L})\mathrm{HS}_t + \beta_3(\mathrm{L})\mathrm{CP}_t + \beta_4(\mathrm{L})\mathrm{FW}_t + u_{2t}.$$

The supply of hogs is expected to vary positively with the price of hogs and the supply of corn, but negatively with the price of corn. The price of hogs, in turn, is negatively affected by the hog supply and positively affected by the cost factors, corn price and farm wages. No exact a priori information is available on the dynamics of these relations, except that the current hog price can only affect next year's (and subsequent years') hog supply because of the one-year production cycle.

Quenouille's data cover the years 1867–1948, for a total of 82 observations; the variables are expressed in logarithms so that the various lag coefficients can be conveniently interpreted as elasticities. Since the entire system appears to be near-nonstationary, all five series are prefiltered by a common filter for the purpose of lag specification. Following Liu and Hanssens (1982), the choice of a filter is determined by the highest AR order that has roots close to 1 in the univariate series, in this case AR(1). Then, a filter parameter is chosen in the neighborhood of the highest AR parameter, but not too close to 1, for example $(1 - .7\mathrm{L})$ in this case. Furthermore, as the contemporaneous effect of HP on HS is ruled out a priori, the first equation can be specified by simple OLS methods. The second equation, on the other hand, will be specified by the instrumental-variables method. The instrument is obtained by regressing HS on up to 12 lags of CS, CP, and FW, yielding an instrument fit of 85 percent. Then, the lag structures shown in Table 7 are found (with standard errors between brackets). For the purpose of comparison, the OLS impulse response weights for the second equation are shown in Table 8. The OLS results are inferior in that they fail to indicate an expected contemporaneous effect of hog supply on hog price.

The empirically specified dynamics of the model are rather simple; the only relationship that may involve a rational transfer function is the effect of CP on HS. Its corner table, shown in Table 9, suggests a transfer function ($r = 1$, $s = 1$, $b = 2$), or perhaps a transfer function ($r = 0$, $s = 2$, $b = 2$) which, in final estimation, provides similar results. In conclusion, the proposed dynamic

Table 8. OLS Lag Specification Model: Hog Price

Equation	Variable	$\hat{v}_0$	$\hat{v}_1$	$\hat{v}_2$	$\hat{v}_3$	$\hat{v}_4$	$\hat{v}_5$
HP	HS	−.235	−1.059*	−.149	—	—	—
		(.206)	(.207)	(.206)			
	CP	.040	.123	.151	.099	−.067	−.163*
		(.064)	(.081)	(.083)	(.084)	(.082)	(.074)
	FW	.812*	.348	−.165	.031	.069	.070
		(.219)	(.249)	(.248)	(.249)	(.254)	(.193)
	constant	187.624	(64.203)				

Table 9. Corner Table: Hog Supply

$b+s-1$	r = 1	2	3
0	.13	.02	.00
1	.26	.20	.07
2	−1.00	1.18	−1.30
3	−.68	.42	−.27

structure of the model is

$$HS_t = c_1 + \omega_{111}HP_{t-1} + \omega_{121}CS_{t-1} + \frac{\omega_{131}}{1-\delta_{111}L}CP_{t-2} + u_{1t},$$

$$HP_t = c_2 + \omega_{211}HS_t + \omega_{212}HS_{t-1} + \omega_{221}CP_t + \omega_{231}FW_t + u_{2t}.$$

In FIML estimation (Liu et al. 1983) of this two-equation system it was found that μ_{1t} and μ_{2t} follow $(1-\theta L^2)(1-\varphi_{11}L)^{-1}a_{1t}$ and $(1-\varphi_{21}L)^{-1}a_{2t}$ processes. The reestimated model is:

$$HS_t = \underset{(58.170)}{423.397} + \underset{(.030)}{.306}\ HP_{t-1} + \underset{(.033)}{.278}\ CS_{t-1} + \frac{-.278(.026)}{\underset{(.074)}{1-.423\ L}}CP_{t-2} + \frac{1+.326L^2(.117)}{\underset{(.076)}{1-.802\ L}}a_{1t}$$

$$HP_t = \underset{(177.130)}{962.578} - (\underset{(.183)}{.638} + \underset{(.181)}{1.426}\ L)HS_t + \underset{(.058)}{.011}\ CP_t + \underset{(.101)}{1.388}\ FW_t + \frac{1}{\underset{(.072)}{1-.766\ L}}a_{2t}$$

$$\text{with}\quad \hat{\Sigma} = \begin{bmatrix} 523 & 97 \\ 97 & 3210 \end{bmatrix}.$$

The hypothesis that the residual series $\hat{a}_{1t}$ and $\hat{a}_{2t}$ are white noise cannot be rejected by the conventional Ljung-Box test.

All FIML coefficients have the expected sign and, with the exception of ω_{221}, are highly statistically significant. Thus, the model is consistent with prior expectations, although the effect of corn price on hog price is apparent only via hog supply. Considering that the original variance in HS is 5886 and that of HP is 61660, it is evident that the goodness of fit with the data is excellent.

We have placed considerable emphasis on the need to prefilter the data prior to lag specification when the series are near-nonstationary. As an illustration of its importance, we report in Table 10 the specification results obtained on the original data (i.e., without prefiltering).

Several findings are disturbing, such as the lack of significance of the HP effects on HS and the positive association between HS and HP in the second equation. However, a simple transformation by the filter $(1-.7L)$ removes these problems easily.

Quenouille's data have recently been analyzed by Tiao and Tsay (1983), using vector ARMA methods. A comparison of our findings with the hog supply and hog price equations, (4.2a) and (4.2b), in Tiao and Tsay reveals that the goodness of fit is similar and that the dominating causal factors are confirmed in both cases; thus, there is evidence of specification robustness between the methods, which is a positive result for the applied model builder. The main difference is that the vector ARMA method is limited to reduced- or final-form analysis because contemporaneous endogenous relations cannot be modeled. In this application, the strong negative contemporaneous effect of hog supply on hog price is not represented in the vector ARMA model. Writing our structural form in reduced form provides a basis for direct comparison with Tiao and Tsay's results. Many of their reported coefficients are implied by our results, although their counterintuitive positive lagged corn-supply-to-hog-price relation is not. The fact that the reported vector ARMA model contains significant coefficients for some a priori questionable economic relations suggests that vector ARMA modeling is not safe from spurious associations and should not be the sole source for the interpretation of a system's structure. However, the vector ARMA model is a convenient and useful tool for forecasting and control.

7. DISCUSSION AND CONCLUSION

When embarking upon an empirically oriented research project, the time series model builder may have different levels of a priori knowledge (Hanssens and Manegold 1980):

level 0: only the information set is known.
level 1: the distinction between endogenous and exogenous variables in the information set is known,
level 2: the functional forms and lag structures of the relationships are known.

The procedures advanced in this article are applicable to the level-1 case; thus they aim at moving the researcher to level-2 knowledge, where a wide array of powerful parameter estimation techniques is available. In contrast, some other systematic empirical specification procedures (such as in Tiao and Box 1981, Tiao and Tsay 1983) do not even require a priori level-1 knowledge, but move the user from level-0 to level-2 via an interesting set of cross-correlation and stepwise autoregression procedures.

We find Tiao and Box's method most insightful when ambiguities about endogenous versus exogenous variables occur—witness the current debate on money supply and interest rates in macroeconomics. However, once level-1 knowledge is obtained (by theory or empiricism), it is far more efficient and reliable in our opinion to use such knowledge in the tentative specification of a model.

Table 10. Lag Specification Without Prefiltering Hog Supply and Price

Equation	Variable	$\hat{v}_0$	$\hat{v}_1$	$\hat{v}_2$	$\hat{v}_3$	$\hat{v}_4$	$\hat{v}_5$
HS	HP	—	.109	.054	−.078	.022	.078
			(.064)	(.077)	(.076)	(.070)	(.047)
	CS	.065	.408*	.104	−.065	−.067	−.075
		(.074)	(.092)	(.095)	(.098)	(.091)	(.078)
	CP	.024	.074	−.138	−.071	−.051	−.011
		(.047)	(.070)	(.080)	(.078)	(.078)	(.068)
	constant	83.024	(84.785)				
HP (by IV)	HS	.177	−.739*	.124	—	—	—
		(.274)	(.320)	(.263)			
	CP	.086	.169	.241*	.154	−.083	−.106
		(.083)	(.105)	(.111)	(.106)	(.111)	(.098)
	FW	.551*	.313	−.199	.126	−.023	−.052
		(.272)	(.415)	(.416)	(.418)	(.424)	(.231)
	constant	149.744	(154.664)				
HP (by OLS)	HS	.581*	−1.128*	.371	—	—	—
		(.213)	(.267)	(.223			
	CP	−.048	.253*	.233*	.123	−.123	−.202*
		(.068)	(.084)	(.087)	(.086)	(.088)	(.072)
	FW	.673*	.078	−.009	−.061	.411	−.263
		(.226)	(.349)	(.361)	(.368)	(.355)	(.195)
	constant	13.505	(102.106)				

Such knowledge also leads us to reduce the parameter space drastically in the initial model estimation. For example, vector ARMA modeling would require some 50 parameters or more for an initial ARMA(1, 1) model in the hog case.

Summarizing our results, we find that the instrumental-variables procedure is very useful if the goodness of fit for the instruments in an equation is not too low. In the absence of contemporaneous endogenous relationships, the OLS procedure can be applied directly. Model identification in such cases is similar to the single-equation situation. Contrary to many statisticians' suggestions, the combined procedure and its variants do not seem to be effective when the goodness of fit for the instruments is low. We also find that slightly overparameterized models can be rectified during model estimation.

It is important to understand that our efforts are aimed at a preliminary model specification and that the ultimate parameter estimation will often involve more sophisticated procedures (e.g., joint equation methods such as FIML), for which an abundant literature is available.

[Received March 1982. Revised April 1983.]

REFERENCES

BEGUIN, J. M., GOURIEROUX, C., and MONFORT, A. (1980), "Identification of a Mixed Autoregressive-Moving Average Process: The Corner Method," in *Time Series*, ed. O. D. Anderson, Amsterdam: North-Holland.

CRAGG, J. G. (1982), "Estimation and Testing in Time-Series Regression Models with Heteroscedastic Disturbances," *Journal of Econometrics*, 20, 135–157.

GRANGER, C. W. J., and NEWBOLD, P. (1977), *Forecasting Economic Time Series*, New York: Academic Press.

HANNAN, E. J. (1971), "The Identification Problem for Multiple Equation System with Moving Average Errors," *Econometrica*, 39, 751–765.

HANSSENS, D. M., and MANEGOLD, J. G. (1980), "Descriptive Model Building on Multivariate Time-Series Data," UCLA Center for Marketing Studies, Working Paper No. 98, October.

HATANAKA (1975), "On the Global Identification of the Dynamic Simultaneous Equation Model with Stationary Disturbances," *International Economic Review*, 16, 545–554.

HAUGH, L. D., and BOX, G. E. P. (1976), "Identification of Dynamic Regression (Distributed Lag) Models Connecting Two Times Series," *Journal of the American Statistical Association*, 72, 121–130.

JENKINS, G. M., and ALAVI, A. S. (1981), "Some Aspects of Modelling and Forecasting Multivariate Time Series," *Journal of Time Series Analysis*, 2, 1–48.

KOHN, R. (1979), "Identification Results for ARMAX Structures," *Econometrica*, 47, 1295–1304.

LIU, L.-M., and HANSSENS, D. M. (1982), "Identification of Multiple-Input Transfer Function Models," *Communications in Statistics—Theory and Methods*, 11, 297–314.

LIU, L.-M., HUDAK, G. B., MULLER, M. E., BOX, G. E. P., and TIAO, G. C. (1983), *The SCA System for Univariate-Multivariate Time Series and General Statistical Analysis*, Scientific Computing Associates, P.O. Box 625, DeKalb, Illinois 60115.

MADDALA, G. S. (1976), *Econometrics*, New York: McGraw-Hill.

PALM, F., and ZELLNER, A. (1980), "Large Sample Estimation and Testing Procedures for Dynamic Equation Systems," *Journal of Econometrics*, 12, 251–283.

PERRYMAN, M. R. (1980), "Consistent Estimation of Large Dynamic Autoregressive Time Series Models: An Adaptable Multiple Stage Approach," paper presented at the Third International Time Series Meeting, Houston, Texas.

PHILLIPS, A. W. (1978), "Estimation of Systems of Difference Equations with Moving Average Disturbances," in *Stability and Inflation*, eds. A. R. Bergstrom et al., New York: John Wiley.

QUENOUILLE, M. H. (1957), *The Analysis of Multiple Time-Series*, New York: Hafner Publishing Co.

SARGAN, J. D. (1961), "The Maximum Likelihood Estimation of Economic Relationships with Autoregressive Residuals," *Econometrica*, 19, 414–426.

TIAO, G. C., and BOX, G. E. P. (1981), "Modeling Multiple Time Series with Application," *Journal of the American Statistical Association*, 76, 802–816.

TIAO, G. C. and TSAY, R. S. (1983), "Multiple Time Series Modeling and Extended Sample Cross-Correlations," *Journal of Business & Economic Statistics*, 1, 43–56.

WALL, K. D. (1975), "FIML Estimation of Rational Distributed Lag Structural Form Models," NBER Working Paper 77.

——— (1976), "FIML Estimation of Rational Distributed Lag Structural Form Models," *Annals of Economic and Social Measurement*, 5, 53–64.

ZELLNER, A., and PALM, F. (1974), "Time Series Analysis and Simultaneous Equation Econometric Models," *Journal of Econometrics*, 2, 1754.