Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions [*]

Norman R. Swanson and Clive W.J. Granger

Pennsylvania State University, University Park, PA 16802, USA and

University of California, San Diego, La Jolla, CA 92093-0508, USA, respectively

this revision April 1996

ABSTRACT

A data-determined method for testing structural models of the errors in vector autoregressions is discussed. The method can easily be combined with prior economic knowledge, and a subjective analysis of data characteristics to yield valuable information concerning model selection and specification. In one dimension, it turns out that standard t-statistics can be used to test the various overidentifying restrictions which are implied by a model. In another dimension, the method compares *a priori* knowledge of a structural model for the errors with the properties exhibited by the data. Thus, this method may help to ensure that orderings of the errors for impulse response and forecast error variance decomposition analyses are sensible, given the data. Two economic examples which draw on data used by Sims (1980b), and King, Plosser, Stock, and Watson (1991) are used to illustrate the method.

KEYWORDS: variance decomposition, overidentifying constraints, structural models

1. INTRODUCTION

The vector autoregression (VAR) methodology has been the focus of much interest since Sims (1980a). For instance, VAR models provide an interesting alternate style of identification for constructing dynamic systems of equations in macroeconometrics. In essence, Sims (1980a) provided compelling criticism of the identification procedures often used in the construction of large-scale structural macroeconomic models. Sims suggested that: "It should be feasible to estimate large-scale macromodels as unrestricted reduced forms, treating all variables as endogenous.".

VAR analyses often center around the calculation of impulse response functions (IRFs) and forecast error variance decompositions (FEVDs), which track the evolution of economic shocks through the system. In practice, though, it often turns out that the covariance matrix of the residuals in a VAR model is not diagonal, implying contemporaneous correlation among the errors. Then, analysis of the evolution of the system caused *just* by an innovation in one variable may not be appropriate, since this innovation may occur at the same time as another innovation in the system. One solution to this problem is to orthogonalize the covariance matrix of the residuals, Σ_ε, using the well known Choleski decomposition, so that the evolution of shocks through the system is uni-directional. A problem with this orthogonalization method is that P changes, if the ordering of the variables in Σ_ε is changed. Since the ordering of the variables is essential to the interpretation of IRFs and FEVDs, Sims (1981) suggests investigating the sensitivity of the conclusions to the ordering of the variables. Blanchard and Watson (1984) and Bernanke (1986) suggest alternatives to the Choleski decomposition. Bernanke considers just-identified models of the structural errors. He suggests estimating these models using the method of moments. In this way, models which are not recursive in nature can be estimated. As with the Choleski decomposition method, however, one ultimately assumes a structural model of the errors based on subjective grounds.

In this paper, a method is suggested for testing structural models of the errors. The testing procedure can further be used in conjunction with subjective data analysis and prior economic information to help specify and select a model of the errors. The method is meant to complement existing strategies for decomposing the errors of a VAR, and essentially involves the analysis of overidentifying restrictions. The overidentifying restrictions considered include partial correlations and, to a lesser extent, tetrad differences. It is shown that a particular class of structural models of the errors in VARs imply a finite set of overidentifying constraints which can be tested easily. Furthermore, the same tests yield valuable information about a much broader class of recursive and non-recursive structural models of the errors.

The tests simply involve least squares regressions using the residuals from a VAR estimation. Overall, the method is meant to help in the specification of a sensible, data-determined ordering of the errors.

The ideas in this paper are quite close to the 'SUR' method discussed in Zellner (1962, 1979), where it is pointed out that introducing prior information generally increases the precision with which regression coefficients can be estimated (see Zellner (1979)). One of the differences is that we test the restrictions that are implied by a given model. Some of the ideas in this paper also draw on recent work in the area of causal modeling by Glymour, Scheines, Spirtes, and Kelly (1987) and Glymour and Spirtes (1988). Furthermore, a number of more general model search algorithms are discussed in Spirtes, Glymour, and Scheines (1993), and the references therein. Out procedure is different, though, as we do not impose the so-called "Faithfulness Condition" which is needed in more general algorithms. Other related recent advances are discussed in Geiger and Heckerman (1994), who examine Gaussian "networks". Historically, Sewell Wright's (1921) work on correlation and causation, and Simon's (1954) examination of causal ordering and identifiability are also closely related to the method proposed here. Other relevant work is contained in McElroy (1978), Wermuth and Lauritzen (1983), Granger (1988), Simon and Iwasaki (1988), Lütkepohl (1990), and Gilli (1992), from which many other useful references can also be found.

As motivation for the methods discussed below, a possible explanation for the presence of contemporaneous correlation between the errors in VAR models is also given. The explanation relies on the observation that uni-directional (Granger) causality in the presence of temporal aggregation can result in contemporaneous correlation among errors in a system that has uncorrelated errors at the 'true' time interval (see Wei (1978) and Weiss (1984) for interesting discussions of temporal aggregation in time series). It should be emphasized that this paper does not attempt to offer any further conceptualization of instantaneous causality. Although non-instantaneous causality has in the past been equated with contemporaneous correlation (e.g. Lütkepohl (1991 p. 102)), no such claim is made here. Broadly speaking, contemporaneous correlation may or may not be a characteristic of so called instantaneous correlation, but, regardless, shouldn't necessarily be equated with instantaneous causality (see Granger (1988 pp.9-10)). In fact, any definition which relies on the above characteristic alone, is very simplistic, and probably somewhat uninformative. Various conceptualizations of instantaneous causality can be found in Granger (1988 pp.7-10) as well as Lütkepohl (1990).

The paper is organized as follows. In the second section a standard VAR model is assumed. The overidentifying constraints implied by structural models of the errors in the presence of contemporaneous correlation are examined. The third section discusses estimation, and proposes a test which assumes that the errors of the VAR are white noise with nonsingular covariance matrix. All proofs are gathered in Appendix B. The fourth section contains two examples. In the first example, data used by King, Plosser, Stock, and Watson (1991) is modeled. The second example considers the VAR models analyzed in Sims (1980b) and Todd (1990). Conclusions and recommendations are presented in section five.

2. STRUCTURAL MODELS OF THE ERRORS IN VARs

In what follows, assume that a K-dimensional multiple time series, y_t, is available, and is generated by a stationary, stable VAR(p) process.

Assumption 1:

$$ y_t = \mu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t = B_1 z_{t-1} + \cdots + B_p z_{t-p} + u_t , \qquad (2.1) $$

where $y_t = (y_{1t} , y_{2t} , \cdots , y_{Kt})'$, $z_t = (1 , y_{1t} , y_{2t} , \cdots , y_{Kt})'$, and $u_t = (u_{1t} , u_{2t} , \cdots , u_{Kt})'$. The B_i are Kx(K+1) coefficient matrices. The u_t are continuous random vectors satisfying $E(u_t) = 0$, $\Sigma_u = E(u_t u'_t)$ is nonsingular, u_t and u_s are independent for $s \neq t$ (so that dynamic misspecification is ruled out, but contemporaneous correlation among the elements of u_t is allowed, i.e. the u_t can be termed *white noise*), and $E(| u_{it} u_{jt} u_{kt} u_{mt} |) \leq c$ for c a finite constant and for i,j,k,m = 1,...,K, and for all t. Then $plim \, \hat{B}_i = B_i$, $i = 1,...,p$, and the usual least squares estimators, (the $\hat{B}_i$), are asymptotically normally distributed with the usual covariance matrix.

It should be noted that linear as well as fixed parameter VARs have been seen to forecast poorly in many cases. However, as the method introduced below relies only on assumptions about the errors of the VAR, the method extends naturally to nonlinear and time-varying generalizations of (2.1). Consider a recursive structural models of the errors in (2.1).

$$ u_{1t} = v_{1t} ; \; u_{2t} = \beta_{2,1} v_{1t} + v_{2t} ; \; \cdots \; ; \; u_{Kt} = \beta_{K,1} v_{1t} + \beta_{K,2} v_{2t} + \cdots + \beta_{K,K-1} v_{(K-1)t} + v_{Kt} \qquad (2.2) $$

The system, (2.2), is structurally equivalent to models of the errors in VARs which are sometimes assumed in the literature, where the v_t are orthogonal underlying shocks. (For example, see Sims (1980a), Runkle (1987), and Todd (1990).) It should be noted, though, that (2.2) results from various different error specifications. For instance, the following two models both have the reduced form given by (2.2), where the βs are correspondingly functions of the γs or the αs.

$$ u_{1t} = v_{1t} ; \; u_{2t} = \gamma_{2,1} u_{1t} + v_{2t} ; \; \cdots \; ; \; u_{Kt} = \gamma_{K,1} u_{1t} + \gamma_{K,2} u_{2t} + \cdots + \gamma_{K,K-1} u_{(K-1)t} + v_{Kt} \qquad (2.3) $$

$$u_{1t} = v_{1t} \; ; \; u_{2t} = \alpha_2 u_{1t} + v_{2t} \; ; \; \cdots \; ; \; u_{Kt} = \alpha_K u_{(K-1)t} + v_{Kt} \qquad (2.4)$$

Assume that all of the coefficients in (2.3) and (2.4) are non-zero. Then, (2.3) is exactly identified, while (2.4) is overidentified. For example, if the population model is (2.3), then the partial correlation between any two elements of u_t conditioned on any other element is not zero, while there are many zero partial correlation restrictions implied by (2.4). For ease of exposition, in what follows we assume that the errors can be represented by (2.4). This assumption is later relaxed, though, and in practice it will be obvious whenever models like (2.4) are not appropriate. For example, when no partial correlations of the type discussed below appear close to zero, then (2.3) is a possible candidate model, or, when all restrictions are very close to zero then (2.4) with all αs identically zero will be a reasonable candidate model (i.e. the case where there is no contemporaneous correlation). We initially assume that all of the αs in (2.4) are non-zero. Orcutt (1987) introduces a much more general class of models which nest (2.2). In particular, (2.2) can be considered as an example of Orcutt's model, where the conditional mean functions are defined to be linear, and depend only on a cummulation of orthogonal shocks (the v_t).

Assumption 2: Define $v_t = (v_{1t}, v_{2t}, ..., v_{Kt})$, where v_t are continuous random vectors satisfying $E(v_t) = 0$, $\Sigma_v = E(v_t v'_t)$ is nonsingular, v_t and v_s are independent for any s,t (ruling out contemporaneous correlation), and $E(|v_{it} v_{jt} v_{kt} v_{mt}|) \leq c$ for c a finite constant and for i,j,k,m = 1,...,K, and for all t. Also, assume that the u_t can be modeled as (2.4)

As discussed above, the v_t are orthogonal underlying shocks which affect the reduced form errors. This is the usual assumption made in the VAR methodology. More generally, the v_t may be allowed to exhibit some second moment dependence, so that some form of heteroskedasticity consistent covariance matrix estimators can be used in place of the usual covariance estimators in the testing stage (see below).

Another convenient way of representing the u_t from (2.4) is with the following linear causal model (*graph*).

(2.5)

$$
\begin{array}{ccccc}
u_{1t} \rightarrow u_{2t} \rightarrow & \cdots & \rightarrow u_{Kt} \\
\uparrow & \uparrow & \uparrow & \uparrow \\
v_{1t} & v_{2t} & \cdots & v_{Kt}
\end{array}
$$

A Simple Causal Graph

A detailed discussion of various aspects of *graph theory*, and a discussion which compares graph theory and statistics is in an earlier version of this paper, and in Geiger and Pearl (1990). Display (2.5) is referred to in *graph theory* as a 'directed acyclic graph' (DAG). It is 'directed' because the arrows lead from one variable into another variable. It is acyclic because one cannot return to any starting variable by

following arrows which lead away from the starting variable in question. The arrows are assumed to represent a simple form of causal directionality which can be equated to a structural model. Display (2.5) is a particularly convenient way to express (2.4), because it is very easy to *read* the implied partial correlation overidentifying restrictions for a given structural model, when the model is expressed as a graph. Along these lines, the graph is an aid to identifying the constraints implied by a causal model of the u_t. By convention, an arrow from, for instance, u_{1t} into u_{2t} means that u_{2t} may be expressed as a function of u_{1t}, so that, for Display (2.5): $u_{1t} = v_{1t}$, $u_{2t} = \alpha_2 u_{1t} + v_{2t}$, ..., and $u_{Kt} = \alpha_K u_{(K-1)t} + v_{Kt}$, which is identically (2.4).

Consider the partial correlation constraints which are implied by (2.4). A partial correlation is the conditional correlation between two variables, conditioned on one or more other variables (see Anderson (1985). The partial correlation of u_{1t} on u_{3t} after subtracting out the influence of u_{2t} is

$$\rho(u_{1t}, u_{3t} \mid u_{2t}) = \frac{\rho(u_{1t}, u_{3t}) - \rho(u_{1t}, u_{2t})\rho(u_{2t}, u_{3t})}{(1 - \rho^2(u_{1t}, u_{2t}))^{1/2}(1 - \rho^2(u_{2t}, u_{3t}))^{1/2}}$$

Assume throughout that $\rho(u_{1t}, u_{2t})$ and $\rho(u_{3t}, u_{2t})$ (and all other correlations considered) are not identically one in absolute value. Then,

$$\rho(u_{1t}, u_{3t}) - \rho(u_{1t}, u_{2t})\rho(u_{3t}, u_{2t}) = 0 \implies \rho(u_{1t}, u_{3t} \mid u_{2t}) = 0 \qquad (2.6)$$

If the αs are identically 0, then (2.6) holds trivially. The testing procedure discussed below is robust to this (as well as cases where *some* αs are 0), though, since *all* possible constraints will be found to hold when the αs are all 0, for example.

Proposition 2.1: Given Assumption 2,

$$\rho(u_{it}, u_{kt} \mid u_{jt}) = 0, \quad \text{where } k < j < i \, ; \, i, j, k \in I \, ; \, i \leq K. \qquad (2.7)$$

Further, all other partial correlations among the u_t (where the conditioning set is one variable) must be nonzero.

It should be noted that if some αs are identically zero, then (2.4) should be partitioned appropriately. So that, for instance, if all of the αs are zero, and the u_t are independent, then *all* possible partial correlations between the u_t are constrained to be 0. As another example, consider the case where one error is independent, and the others in the system may not be. Assume that the rest of the errors can be represented as a chain with 3 or more variables in it, so that Proposition 2.1 applies. Then, Proposition 2.1 can also be applied to the lone independent error, by defining two new *null* variables, which by definition are not correlated with the error, and constructing a hypothetical 3 variable chain which contains the two null variables and the independent error, in any order. Then, the only stipulation is that the null variables

are *not* correlated with any true variables. In general, using the same framework, Proposition 2.1 can be applied for *any* model, although various other higher order constraints may be implied by more complex models. Interestingly, we have found that parsimonious models of the form given by (2.4) (with zero and non-zero αs) seem to capture the properties exhibited by all of the economic examples which we have examined (including systems with 4 to 9 variables). Overall, Proposition 2.1 simply states that alternate structural models given by (2.4) imply particular overidentifying partial correlation constraints. For example, in a four variable VAR, the associated structural model of the errors has 12 different (or, $[\sum_{i=1}^{K-1}(K-i)(K-2)]$) partial correlations (when conditioning on one variable). Proposition 2.1 states that 4 (or, $[\sum_{i=2}^{K-1}(K-i)(i-1)]$) of these partial correlations should be identically 0, and the rest should be non-zero. Proposition 2.1 uses conditioning sets of only one variable. The simple nature of (2.4) leads to a similar proposition concerning conditioning sets of more than one variable.

Proposition 2.2: Given Assumption 2,

$$\rho(u_{it}, u_{jt} \mid u_{qt}, u_{(q+1)t}, \cdots, u_{(q+r)t}) = 0 \qquad (2.8)$$

if and only if

$$\rho(u_{it}, u_{jt} \mid u_{qt}) = 0 \quad \text{or} \quad \rho(u_{it}, u_{jt} \mid u_{(q+1)t}) = 0 \quad \text{or} \quad \cdots \quad \text{or} \quad \rho(u_{it}, u_{jt} \mid u_{(q+r)t}) = 0 \qquad (2.9)$$

where, $0 < i, j, q, q+r \le K$, $i, j, q, q+r \in I^+$, and I^+ is the set of positive integers.

Proposition 2.2 is important because it ensures that in the class of all models given by (2.4), the partial correlations which condition on one variable nest all implied partial correlations. Thus, all partial correlation constraints can be tested by looking at the finite subset of partial correlations which are conditioned on only one variable.

A further proposition can be stated which concerns the tetrad difference constraints implied by (2.4). Tetrad equations equate products of two correlations among sets of four measured variables. Of note is that the analysis of tetrad constraints is nested by the analysis of partial correlation constraints in (2.4). The following proposition is nevertheless stated, because when non-recursive models of the errors are suspected, the *only* if part of the proposition holds, while the if part does not. Thus, in many cases, it may be of interest to examine tetrad difference as well as partial correlation constraints.

Proposition 2.3: Given Assumption 2,

$$\rho(u_{it}, u_{jt})\rho(u_{kt}, u_{lt}) - \rho(u_{it}, u_{kt})\rho(u_{jt}, u_{lt}) = 0 \qquad (2.10)$$

iff the following 4 conditions hold for some s, and

$$\rho(u_{it}, u_{jt}) - \rho(u_{it}, u_{st})\rho(u_{jt}, u_{st}) = 0 \qquad (2.11)$$

$$\rho(u_{kt}, u_{lt}) - \rho(u_{kt}, u_{st})\rho(u_{lt}, u_{st}) = 0 \qquad (2.12)$$

$$\rho(u_{it}, u_{kt}) - \rho(u_{it}, u_{st})\rho(u_{kt}, u_{st}) = 0 \tag{2.13}$$

$$\rho(u_{jt}, u_{lt}) - \rho(u_{jt}, u_{st})\rho(u_{lt}, u_{st}) = 0, \tag{2.14}$$

for $1 \le i, j, k, l, v \le K$ and $i \ne j \ne k \ne l$.

Clearly, other overidentifying constraints may also be implied by models of the errors. For instance, a model with at least five measured variables may imply pentad constraints, which have the products of five correlations as factors. Thus, tests of models by analysis of partial correlation and tetrad difference constraints tend to be too generous, in general. This is especially true for non-recursive models, since partial correlations *do not* generally nest product of correlation constraints of higher order. In the restrictive setting of Assumption 2, though, it is conjectured that all statistics constructed from products of correlations are nested by partial correlation constraints.

One intractable artifact of the above analysis concerns the issue of 'reversibility', or model 'opposites'. The unique subset of partial correlation constraints implied by each member of the class of recursive models also applies to the 'reverse' or 'opposite' model. This characteristic is just an artifact of the contemporaneous nature of the correlation constraints which are tested. For example, the same constraints are implied by a model of the type given by (2.4) in which the Wold causal chain runs from u_{1t} to u_{2t} to u_{3t} to u_{4t}, and its reverse, which has arrows running from u_{4t} to u_{3t} to u_{2t} to u_{1t}.

So far, no mention has been made as to how contemporaneous correlation among the errors may arise in economic theory. One intuitively appealing explanation is summarized in Granger (1988 pp.7). The idea is that *apparent* instantaneous causality results from temporal aggregation, in which case the true finite time delay between cause and effect is small compared to the time interval, k, over which the data is collected. This notion can be used to account for the presence of contemporaneous correlation among the errors when the true DGP can be expressed by the reduced structural form given in (2.1). Swanson (1994) further shows that for a class of very simple first order recursive VAR models exhibiting uni-directional Granger causality (with no contemporaneous correlation among the errors), not only does contemporaneous correlation among the errors arise for $k \ne 0$, but for k large, the pattern of partial correlations implied is the same as that given in Proposition 2.1.

Discussion to this point has focussed primarily on a particular class of recursive structural models of the errors where all off diagonal coefficients are zero. There is no obvious reason, however, why structural models should be recursive. Bernanke (1986) addresses this issue, and proposes a method of moments technique for the estimation of IRFs and FEVDs associated with a broad class of non-recursive

structural models. In fact, Propositions 2.2 and 2.3 do not have analogs in the case of non-recursive models. As stated above, the *only if* part of Proposition 2.3 is not valid for non-recursive models (specifically, for cyclic graphs), so that a tetrad difference constraint can be implied by a non-recursive model without in turn implying the partial correlation constraints that imply it. In a broad sense, then, the subset of partial correlation constraints implied by (2.4) occupy a very small part of the space containing all unique constraints which may be implied by a given joint probability distribution. Thus, the method for testing models discussed above is not strictly applicable when more general non-recursive models describe the true data generating process. This limitation does not cause the method to break down, though. Non-recursive models may still imply overidentifying constraints which can be tested by looking at the correlation (covariance) matrix of the residuals. In this vein, checking all possible partial correlation and tetrad difference constraints can act as an *aid* to the specification of non-recursive structural models of the errors.

A number of interesting, and somewhat related references are: Geiger and Pearl (1990), who examine the role of directed acyclic graphs (DAGs) as a representation of conditional independence relationships, Glymour and Spirtes (1988) who use a result similar to Proposition 2.3 to develop a method for modeling latent variables, and Spirtes, Glymour, and Scheines (1993), who provide a corrected version of Pearl and Verma's (1990, 1991) algorithm for latent variables. Although these approachs to latent variable modeling are not examined here, they may help in the detection of misspecification due to omitted variables, for example. To summarize, Propositions 2.1-2.3 suggest that an approach to testing models of the errors simply involves testing the partial correlation (and tetrad difference) overidentifying constraints which are implied by (2.4). In this manner a particular Choleski (or other) decomposition of the u_t can be used after it is shown that the model is *in accord* with the data.

3. ESTIMATION, TESTING, AND MODEL SEARCH

In this section a residual based test with a null hypothesis of zero partial correlation is considered, and a model search procedure is outlined. The test statistic is the standard (or heteroskedasticity consistent) t-statistic from a LS regression of the model $u_{it} = \alpha_j^* u_{jt} + \alpha_k^* u_{kt} + \xi_t$, where it is assumed that (2.4) holds, $0 < i,j,k \leq K$, $k<j<i$. Substitution into (2.4) shows that $\alpha_j^* \neq 0$ and $\alpha_k^* = 0$ *iff* (2.4) holds. The idea of the test is that is $\alpha_k^* = 0 <=> \rho(u_{it}, u_{kt} \mid u_{jt}) = 0$ and $\alpha_j^* \neq 0 <=> \rho(u_{it}, u_{jt} \mid u_{kt}) \neq 0$, so that the implications of Proposition 2.1 can easily be tested. Also, recall that one easy way to construct a list of constraints is by drawing a graph as in Display (2.5), and noting that whenever a variable separates two

other variables in the graph, then the partial correlation of the two variables is zero, when conditioned on the separating variable.

Define

$$u_{it} = \alpha_j^* u_{jt} + \alpha_k^* u_{kt} + \xi_t \, , \tag{3.1}$$

where $0 < i,j,k \leq K$, $k<j<i$, the true u_t are known, and $\alpha^* = (\alpha_j^*, \alpha_k^*)$, $U_t = (u_{jt}, u_{kt})$, $D \equiv M_{UU}^{-1} V M_{UU}^{-1}$, $M_{UU} = E(U_t U_t')$, and $V = E(U_t \xi_t^2 U_t')$, for some sample size T>2. (Also, if the linear model is correctly specified for $E(u_{it} \mid U_t)$ and $E(\xi_t^2 \mid U_t) = \sigma_\xi^2$ yields that $D \equiv \sigma_\xi^2 M_{UU}^{-1}$.) Further, define $\hat{\alpha}^* = (U'U)^{-1} U' u_i$, where (3.1) has been written as $u_i = U\alpha + \xi$, and $\hat{\alpha}^* = (\hat{alpha}_j^*, \hat{\alpha}_k^*)'$ is the LS estimator which could only be calculated were the u_t known. In what follows, the u_t are unknown, and an estimator $\tilde{\alpha}^*$, which depends on the $\hat{u}_t$ is defined. The ξ_t are linear combinations of the v_t, so that the conditions placed on the v_t in Assumption 2 are sufficient to ensure that $plim \ \hat{\alpha}^* = \alpha^*$ for $k<j<i$, and the least squares estimator, $\hat{\alpha}^*$, is asymptotically normally distributed with $D^{-1/2} \sqrt{T} (\hat{\alpha}^* - \alpha^*) \overset{d}{\to} N(0, I_2)$

Theorem 3.1: Given Assumptions 1 and 2 define

$$\hat{\alpha}_k^* = \frac{\sum u_{jt} u_{it} \sum u_{kt}^2 - \sum u_{kt} u_{it} \sum u_{jt} u_{kt}}{\sum u_{jt}^2 \sum u_{kt}^2 - (\sum u_{jt} u_{kt})^2} \, ,$$

and

$$\tilde{\alpha}_k^* = \frac{\sum \hat{u}_{jt} \hat{u}_{it} \sum \hat{u}_{kt}^2 - \sum \hat{u}_{kt} \hat{u}_{it} \sum \hat{u}_{jt} \hat{u}_{kt}}{\sum \hat{u}_{jt}^2 \sum \hat{u}_{kt}^2 - (\sum \hat{u}_{jt} \hat{u}_{kt})^2} \, ,$$

where,

$$\hat{u}_{rt} = y_t - \sum_{i=1}^{p} \hat{B}_i z_{t-i} \, , \quad r=i,j,k.$$

Then,

(i) $\alpha_k^* = 0 \iff \rho(u_{it}, u_{kt} \mid u_{jt}) = 0, k<j<i$. (Or, equally, $\alpha_j^* \neq 0 \iff \rho(u_{it}, u_{jt} \mid u_{kt}) \neq 0, k<j<i$.)

(ii) $T^{1/2} (\tilde{\alpha}_s^* - \hat{\alpha}_s^*) = o_p(1)$ and $\tilde{\alpha}_s^*$ is asymptotically equivalent to $\hat{\alpha}_s^*$, $s=j,k$ and has the same limiting distribution.

(iii) $\rho(u_{jt}, u_{kt} \mid u_{it}) \neq 0$ for any $k<j<i$ if all of the equality and inequality constraints which can be tested using (i) hold.

In Theorem 3.1, $\hat{\alpha}_k^*$ ($\hat{\alpha}_j^*$) is the usual LS estimator of the coefficient on u_{kt} (u_{jt}) from the regression of of u_{it} on u_{jt} and u_{kt}. Regressions using $\hat{u}_t$ in place of u_t are valid because $\hat{\alpha}_k^*$ ($\hat{\alpha}_j^*$) is asymptotically equivalent to $\tilde{\alpha}_k^*$ ($\tilde{\alpha}_j^*$). Thus, the standard LS t-statistic associated with $\tilde{\alpha}_k^*$ can be used when testing the

null hypothesis that $\rho(u_{it}, u_{kt} \mid u_{jt}) = 0$ (and similarly for $\tilde{\alpha}_j^*$ and $\rho(u_{it}, u_{jt} \mid u_{kt}) = 0$). A heteroskedasticity consistent LS covariance matrix estimator may be used when the linear model is not correctly specified for $E(u_{it} \mid U_t)$, and/or if $E(\xi_t^2 \mid U_t) \neq \sigma_\xi^2$. Although part (iii) of the theorem essentially restates part of Proposition 2.1, it is still included, as it also states that all partial correlation constraints from (2.4) which *cannot* be tested due to standard LS inconsistency considerations when estimating the regressions, k<j<i, hold if all of the constraints which *can* be tested hold. For example, assume that K=4 in (2.5). Then $\rho(u_{1t}, u_{2t} \mid u_{3t}) \neq 0$, $\rho(u_{1t}, u_{2t} \mid u_{4t}) \neq 0$, $\rho(u_{1t}, u_{3t} \mid u_{4t}) \neq 0$, and $\rho(u_{2t}, u_{3t} \mid u_{4t}) \neq 0$ cannot be directly tested, but each is in turn implied by $\rho(u_{1t}, u_{3t} \mid u_{2t}) = 0$, $\rho(u_{1t}, u_{4t} \mid u_{2t}) = 0$, $\rho(u_{1t}, u_{4t} \mid u_{3t}) = 0$, and $\rho(u_{2t}, u_{4t} \mid u_{3t}) = 0$, respectively. The number of constraints which cannot be directly tested is $[\sum_{i=1}^{K-1} (K-i)(K-2)]/3$ (i.e 1/3 of the total number of partial correlations), and all of these are inequality constraints. Thus, one third of the partial correlations constraints implied by Proposition 2.1 must be tested indirectly. Although this clearly limits the number of constraints which can be directly tested, there are many constraints which *can* still be directly tested, particularly for large systems of equations. Also, we will see below that the first step of the recommended search procedure *still* involves looking at the values of the *all* partial correlations, including those that cannot be directly tested. In essence, then, we are left with a modeling framework which is akin to the Box-Jenkins philosophy. To summarize, tests for zero partial correlations between the various u_t in (2.1) can be constructed by regressing $\hat{u}_{it}$ on $\hat{u}_{jt}$ and $\hat{u}_{kt}$, k<j<i.

A joint testing procedure is currently unavailable, and is left to future work. Thus, results based on individual analysis of the constraints should be viewed with caution, as the usual size problems may arise. However, it turns out that in the examples given in this paper, the size issue is somewhat moot, as the results do not change, even when the significance level of the tests are adjusted significantly upwards. The following summarizes the recommended search procedure (*rule of thumb*) for structural models similar to that given by (2.4):

Step 1: Calculate all $[\sum_{i=1}^{K-1} (K-i)(K-2)]$ partial correlations and choose a model which is loosely in accord with a subjective examination of the correlation values, and which is also in accord with some set economic theory priors.

Step 2: Construct tests of H_0: $\rho(\cdot) = 0$ for all partial correlation constraints implied by Proposition 2.1, k<j<i, for a given model (2.4), and determine whether the model given by Step 1 is *in accord* with the

data.

For recursive and non-recursive models which are not in the class given by (2.4) a similar search procedure is recommended as a *heuristic* specification method. In particular, since the zero partial correlation constraints do not nest higher order constraints in these models, Steps 1 and 2 could be followed as above with the modification that tetrad and other higher order constraints also be examined in both steps, where feasible.

Generally, the specification of a recursive structural model of the errors makes the estimation of the impulse response functions (IRFs) and forecast error variance decompositions (FEVDs) very straightforward. Assume that the VAR process (2.1) has a canonical MA representation

$$y_t = \mu^* + \sum_{i=0}^{\infty} \phi_i u_{t-i} = \mu^* + \Phi(B)u_t , \quad \Phi_0 = I_K$$

Then construct the lower triangular matrix P such that $PP' = \hat{\Sigma}_{u_t}$ where the elements of u_t are ordered according to the recursive structural model chosen, where Σ denotes the covariance matrix. The Choleski decomposition yields the impulse response functions

$$y_t = \mu^* + \sum_{i=0}^{\infty} \Theta_i v_{t-i} , \tag{3.2}$$

where, $\Theta_i \equiv \Phi_i P$, $v_t \equiv P^{-1} u_t$, and $\Sigma_{v_t} = P^{-1} \hat{\Sigma}_{u_t} (P^{-1})' = I_K$. The v_t are referred to as the orthogonal residuals or innovations, and μ^* may be dropped because it is of no interest to the analysis of IRFs and FEVDs. Forecast error variance decompositions offer a further possibility to interpret (2.1). FEVDs are easily calculated as the optimal h-step forecast error accorded to each component of y_t using (3.2). For non-recursive models, IRFs and FEVDs can correspondingly be constructed using Bernanke's (1986) method of moments procedure. Briefly, a matrix C is specified so that (i) $Cu_t = v_t$, and (ii) C incorporates the particular structure of the errors which is found to hold in the data. Overall, the *rule of thumb* search procedure ensures that the specified structural model of the errors exhibits at least *some* features which are *in accord* with the data.

4. EXAMPLES USING MACROECONOMIC DATA

In this section two examples which utilize the method are discussed. The results should be viewed with caution, though, since specification issues such as unit roots, structural breaks, and omitted and latent variables are not emphasized. Thus, the structural models chosen for the errors are meant to be *illustrations*, and are based on a potentially incomplete analysis of the data. In both examples, recall that

we adopt a 2-step approach. In the first step, the partial correlations are examined to see if they may imply a model like (2.1). That this model is true becomes the null hypothesis, which is tested in step two by examining the relevant t-statistics, according to Theorem 3.1.

In the first example, a standard four variable VAR model using quarterly U.S. per capita data from 1949:1-1990:2 on real consumption expenditures, per capita gross private domestic fixed investment, per capita *private* gross national product, and real balances - the log of money minus the log of prices is constructed. All measures are the same as those used by King, Plosser, Stock and Watson (1991), with the exception that the Citibase measure of M1 rather than M2 is used in the construction of the real money supply series. (This was done because of problems *splicing* Citibase data with historical data.) Please refer to Appendix A for complete definitions of the data used. In accordance with the *three* variable model of King et al. (1991 pp.823-9) 8 quarterly lags are used. The model is estimated as a system of seemingly unrelated equations (see Zellner (1962)). All data is log levels.

Table 1 exhibits all possible partial correlations and correlation values, as well as test statistics. The correlation values are used in Step 1 of the search procedure, and the test statistics are used in Step 2 to test for zero partial correlation overidentifying restrictions. The variables C, I, Y, M correspond to the *error* series from the unrestricted VAR (the u_t in (2.1)), while the partial correlation constraints are tested using the corresponding *residual* series, the $\hat{u}_t$ (see above discussion). The partial correlation values are suggestive, as 3 values are less than 0.06, and the other 9 values are above 0.12 (with 6 of these being above 0.30). Recall that the simple 4 variable causal model given by (2.4) implies that exactly four partial correlations should be zero. It is therefore fairly natural to specify the following causal model, even without the aid of economic theory.

$$(4.1)$$

$$M_t \to C_t \to I_t \to Y_t$$
$$\uparrow \quad \uparrow \quad \uparrow \quad \uparrow$$
$$v_{1t} \quad v_{2t} \quad v_{3t} \quad v_{4t}$$

It should be noted that the technique discussed here allows for correlation amongst the v_t. For instance, if v_{1_t} and v_{2_t} are correlated, we can equivalently specify a new unmeasured common cause (latent variable) affecting both variables (C and M). In this case, the search procedure for non-recursive models is used. A more detailed discussion of these points is in an earlier version of this paper, where the following model is examined: $M_t = (v_{1t} + \beta_{11}L_t)$; $C_t = \beta_{21}v_{1t} + (v_{2t} + \beta_{22}L_t)$; $I_t = \beta_{31}v_{1t} + \beta_{32}v_{2t} + (v_{3t} + \beta_{33}L_t)$; $Y_t = \beta_{41}v_{1t} + \beta_{42}v_{2t} + \beta_{43}v_{3t} + (v_{4t} + \beta_{44}L_t)$. Note that our model of interest, Display (4.1), can be expressed as the following recursive system of the errors in the spirit of Sims (1980a pp.34), with restric-

tions given by Proposition 2.1:

$$M_t = v_{1t} \; ; \; C_t = \beta_{21}v_{1t} + v_{2t} \; ; \; I_t = \beta_{31}v_{1t} + \beta_{32}v_{2t} + v_{3t} \; ; \; Y_t = \beta_{41}v_{1t} + \beta_{42}v_{2t} + \beta_{43}v_{3t} + v_{4t} \quad (4.2)$$

In this example, and in a number of other four to nine variable VAR models analyzed by us, simple models of the errors which are all similar to (2.4) seem to explain the data best (i.e. parsimonious models appear to agree with the data in most cases). For instance, inspection of Table 1 shows that the 5 of 8 p-values are equal to 0.000, and 3 are greater than 0.424. Because of the large gap between p-values in these two groups, and noting that this is not a joint testing procedure, the significance level for the example can be set from 0.000 to 0.424, without changing the results. Model (4.2) implies all of the (zero) partial correlation constraints that were found to hold in the data (with p-values of at least 0.425!). However, it also implies one zero partial correlation constraint that was found not to hold in the data. That is, the partial correlation constraint $\rho(C_t, Y_t \mid I_t) = 0$ is implied by (4.2), but does not hold in the data. Conversely, the data does not imply *any* zero constraints which are not in turn implied by (4.2). Finally, for the 4 remaining correlations which cannot be tested (and for which no test statistics are reported in Table 1), we note that one, $\rho(C_t, M_t \mid I_t) \neq 0$ is not in turn implied by the results found for the "testable" restrictions. (In particular, since $\rho(I_t, M_t \mid C_t)$ is found to be different from zero, then we cannot infer that $\rho(C_t, M_t \mid I_t) \neq 0$ from the data. The logic of the argument here is that: *not* $\rho(I_t, M_t \mid C_t) = 0$ => *not* $\rho(C_t, M_t \mid I_t) \neq 0$.) Thus, Model (4.2) adheres to the data for 10 of 12 partial correlations.

Finally, we must consider the issue of reversibility discussed above. Using the method suggested above, a causal model which runs from Y_t to I_t to C_t to M_t is statistically equivalent to the specified model which runs from M_t to C_t to I_t to Y_t. The latter specification has been chosen on the basis of a *conjecture* that either of money, consumption, or investment is a leading indicator for output. Thus, in this illustration, for example, we may surmise that shocks to money are a leading indicator to shocks to output. In general, when specifying a model of the type given by (2.4), the economist must choose between two alternate models using a priori knowledge.

In the second example, a four variable VAR model which is similar in spirit to those in Sims (1980b) and Todd (1990) is estimated. Four monthly U.S. macroeconomic time series measuring money (monetary base), interest rates (T-bill rate), prices, and output are used. As in Sims (1980b), one year of lags is used. Todd (1990) found that there was evidence of a structural break in the data (1947-1978) that Sims used. As in Todd (1990) the sample period used is 1953:4 - 1979:9. All data is monthly log levels, and the model is estimated as a system seemingly unrelated equations.

Table 2 exhibits the test results for all possible partial correlations. First, notice that the sample partial correlations are either greater than 0.105 (correlations 1-4, or less than 0.050 (correlations 5-12). The plethora of small correlation values immediately suggests that a causal model of the type given in (2.4) is not going to do well in the second step of the search procedure. However, noting that relatively large correlations remain in 2 of 3 cases where M is conditioned on, we specify the following causal model:

$$M_t \qquad P_t \to R_t \to Y_t$$
$$\uparrow \qquad \uparrow \qquad \uparrow \qquad \uparrow$$
$$v_{1t} \qquad v_{2t} \qquad v_{3t} \qquad v_{4t}$$

This causal ordering can be written as:

$$M_t = v_{1t} \ ; \ P_t = v_{2t} \ ; \ R_t = \beta_{11}v_{2t} + v_{3t} \ ; \ Y_t = \beta_{41}v_{2t} + \beta_{42}v_{3t} + v_{4t} \qquad (4.3)$$

Moving to Step 2 of the search procedure, note that (4.3) implies all of the zero correlation constraints found using a test size of anywhere from 0.06 to 0.38. Furthermore, using a test size of 0.06, the tests do not accept the zero null hypothesis for any of the correlations that are not implied to be zero, with the exception only of constraint 8, where the p-value is 0.898. Finally, correlation 1 not equal to zero (in accord with equation (4.3)) is implied by the fact that correlation 7 is not statistically different from zero. Thus, (4.3) corresponds to the data for 11 of 12 correlations.

A number of special features of (4.3) should be pointed out. First, the only test statistic which is not reported is the one for correlation 1. This is different from our usual case (where models like (2.4) are specified) because M is assumed exogenous. Thus, regressions which have M as dependent variable still yield consistent estimates, and the t-test can be carried out. In other words, in this case there is only 1 of 12 correlation constraints that cannot be directly tested. This special feature, which arises because of the structure of (4.3), serves to illustrate the need to carefully determine which regressions yield consistent estimates on a case by case basis, whenever the posited model is other than one given by (2.4). Also, (4.3) is interesting since it implies that the errors associated with the money equation are not contemporaneously correlated with the other errors in the system. Thus, this model imposes a restriction which is quite different from the recursive model considered by Sims (1980b pp.253) and Todd (1990 pp.35-7), where the money error is allowed to feed into all of the other error equations. In general, it should be interesting to see whether incorporation of overidentifying constraints into VAR analyses has a significant impact on the observed IRFs and FEVDs. The opposite recursive causal model for (4.3) still has the money innovation as independent, but has the prices to interest to output ordering reversed.

5. CONCLUSIONS AND RECOMMENDATIONS

Forecast error variance decompositions and impulse response functions are an important part of the VAR methodology proposed by Sims (1980a). For example, Sims(1980b) uses FEVDs to conclude that shocks to money do not have a large impact on the evolution of output, contrary to standard predictions from the monetarist paradigm. Contemporaneous correlation among the residuals of a VAR has often been noted in practice, though. In the presence of such contemporaneous correlation, IRFs and FEVDs are easily interpretable *only* after the residuals of the VAR have been orthogonalized. One argument against the robustness of VAR results has been that this orthogonalization involves the subjective specification of a structural model of the errors.

In this paper it is reiterated that many structural causal models of the errors imply overidentifying constraints which can be tested. Hence, analysis of potential overidentifying constraints is useful at the specification stage of the VAR analysis, and may help to substantially reduce the subjective nature of error orthogonalization in the VAR methodology. For a simple class of structural causal models of the errors, a small finite set of partial correlation constraints spans the space of all constraints which are implied. A large sample test is proposed under fairly weak conditions which allows for the constraints to be tested using standard t-statistics from regressions involving the residuals from a VAR estimation, given a particular structural model of the errors. A two step search procedure which is meant to comple- ment existing specification methods is outlined. The procedure uses economic information as well as a subjective examination of correlation values in the first step, and uses the test in the second step. The search procedure is also extended to cases where more complex recursive and nonrecursive models may be appropriate, and extends readily to the case of error correction.

Two examples are used to illustrate the search procedure. The first example supports the specification of a simple restricted recursive structural model when consumption (c), investment (i), money (m), and output (y) are used. The Wold causal chain in this case is shown to run from m to c to i to y. The second example uses interest rates, money, prices, and output, as in Sims (1980b) and Todd (1990). A restricted recursive structural model of the errors is again shown to fit the data, although, interestingly, the money error does not enter into the equations for any of the other errors.

Some directions for further research are: adapting the latent variable search procedures discussed by Glymour and Spirtes (1988) to check for misspecification in VAR and EC models; checking the effects of aggregation on Granger causal relationships vis a vis the contemporaneous causal framework employed here; calculating the exact distribution of partial correlations as opposed to using the

asymptotic results discussed above; constructing a *joint* test of the restrictions; gauging how generous the search procedure recommended for non-recursive structural models is; and further investigating the empirical implications of the proposed strategy.

APPENDIX A: THE DATA

U.S. data from Citibase are used throughout.

Consumption: In the first example, the consumption measure used is per capita real consumption expenditures. The Citibase mnemonic is GC82. *Interest Rates:* In the second example, the interest rate is measured as the three-month new auction U.S. Treasury bill rate. The Citibase mnemonic is FYGN3 (nsa). *Investment:* In the first example, investment is per capita real gross private domestic fixed investment. The Citibase mnemonic is GIF82. *Money:* In the first example, the Citibase M1 series (FM1) is used. However, as this series begins in 1959:1, M1 data as reported in *Banking and Monetary Statistics, 1941-1970* is spliced onto the Citibase M1 series. In order to smooth apparent measurement differences between the Citibase and the spliced data, the earlier data is multiplied by a factor of 0.977. Further, the monthly observations are averaged to obtain the quarterly observations. Real balances is then constructed as $log(M1) - log(P)$ where, $P = [(GNP - GGE)/(GNP82 - GGE82)]$. In the second example, money is measured as the monetary base adjusted for reserve requirement changes (FMBASE, sa). *Output:* In the first example, per capita real "private" gross national product is defined as total gross national product less real total government purchases of goods and services. The Citibase mnemonics are GNP82-GGE82. In the second example, output is defined as the total index of industrial production (IP, 1987=100). *Prices:* In the second example, prices are measured as the producer price index for all commodities (1982=100, U.S. Labor Department). The Citibase mnemonic is PW.

APPENDIX B: PROOFS OF PROPOSITIONS 2.1-2.3 and THEOREM 3.1

Proof of Proposition 2.1:
It is assumed that the u_t can be expressed as (2.4). It is also assumed that the v_t are distributed as independent continuous random variables (see Assumption 2). Thus, a regression of

$$u_{it} = \alpha_j^* u_{jt} + \alpha_k^* u_{kt} + \xi_t \ , \ \ k < j < i \tag{B.1}$$

must have either $\alpha_j^* = 0$, $\alpha_k^* = 0$, or $\alpha_j^* = \alpha_k^* = 0$. Also, from Assumptions 2 and 3, notice that ξ_t is a linear function of the v_t, by construction. Now, given (2.4) with all non-zero coefficients, inspection shows that for i=(j+1) and j=(k+1), the zero coefficient will be α_k^*. (This is easily seen by substituting i=3, j=2, and k=1, for example.) More generally, consider k<j<i. Without loss of generality, consider

$$u_{5t} = \alpha_3^* u_{3t} + \alpha_1^* u_{1t} + \xi_t \ ,$$

Note from (2.4) that

$$u_{5t} = \alpha_5 \alpha_4 u_{3t} + \alpha_5 v_{4t} + v_{5t} \ ,$$

in which case $\alpha_3^* = \alpha_5\alpha_4$, $\alpha_1^* = 0$, and $\xi_t = \alpha_5\nu_{4t} + \nu_{5t}$ is consistent with the regression model. Thus, $\alpha_k^* = 0$ in (B.1) for all k<j<i, i,j,k=1,...,K and $i \neq j \neq k$. Furthermore, the non-zero αs in (2.4) ensures that $\alpha_j^* = \alpha_k^* = 0$ never holds in (B.2). Exactly the same argument ensures that for j<k<i, $\alpha_k^* \neq 0$, since $\alpha_j^* = 0$, and the nonzero condition on the αs ensures that only one of α_j^* and α_k^* can be equal to zero.

For the zero restriction statement given by (2.7) it remains only to show that $\alpha_k^* = 0 \Rightarrow \rho(u_{it} , u_{kt} \mid u_{jt}) = 0$. This is easily seen since

$$\rho(u_{it} , u_{kt} \mid u_{jt}) \ = \ \frac{cov\,[(u_{it} - E(u_{it} \mid u_{jt}))(u_{kt} - E(u_{kt} \mid u_{jt}))]}{\sqrt{E((u_{it} - E(u_{it} \mid u_{jt}))^2)}\,\sqrt{E((u_{kt} - E(u_{kt} \mid u_{jt}))^2)}} \tag{B.2}$$

and α_k^* can be written as

$$\alpha_k^* \ = \ \frac{cov\,[(u_{it} - E(u_{it} \mid u_{jt}))(u_{kt} - E(u_{kt} \mid u_{jt}))]}{E[(u_{it} - E(u_{it} \mid u_{jt}))^2]} \tag{B.3}$$

Assume that $u_{it} \neq E(u_{it} \mid u_{jt})$ and $u_{kt} \neq E(u_{kt} \mid u_{jt})$ in (B.2) and (B.3). (This is equivalent to assuming that none of the αs in (2.4) are zero. For special cases which break this condition, the models as in (2.4) are partitioned and Proposition 2.1 is applied to the partitioned "sub" models. In the extreme case, where the u_t are fully independent, Proposition 2.1 states that *all* possible partial correlation constraints are 0.) Then, it follows that

$$\alpha_k^* \ = \ 0 \ \Rightarrow \ \rho(u_{it} , u_{kt} \mid u_{jt}) \ = \ 0$$
$$\alpha_k^* \ \neq \ 0 \ \Rightarrow \ \rho(u_{it} , u_{kt} \mid u_{jt}) \ \neq \ 0$$

Finally, notice that one third of all possible partial correlations which can be calculated, and which are all non-zero according to Proposition 2.1, are of the type $\rho(u_{jt} , u_{kt} \mid u_{it}) \neq 0$, k<j<i. However, these correlations cannot be shown to be non-zero using the above proof, since the regression coefficients cannot be consistently estimated (i.e. since the dependent variable in the regression cannot be decomposed to yield useful restrictions on the α^*s.) Using proof by contradiction, and without loss of generality, it is straightforward to show that the remaining partial correlations must be non-zero. Assume that

$$\rho(u_{1t},u_{3t} \mid u_{5t}) \ = \ \frac{\rho(u_{1t},u_{3t}) - \rho(u_{1t},u_{5t})\rho(u_{3t},u_{5t})}{(1 - \rho^2(u_{1t},u_{5t}))^{1/2}(1 - \rho^2(u_{3t},u_{5t}))^{1/2}} \ = \ 0.$$ It follows that

$$\rho(u_{1t},u_{3t}) \ = \ \rho(u_{1t},u_{5t})\rho(u_{3t},u_{5t}), \tag{B.4}$$

as we have assumed throughout that $\rho(u_{1t},u_{5t})$ and $\rho(u_{3t},u_{5t})$ (and any other correlations) are not identically one in absolute value (otherwise some partial correlations would be infinity). However, we have shown above that for the model given by (2.4), $\rho(u_{1t},u_{5t} \mid u_{3t}) \ = \ 0$, which in turn implies that

$$\rho(u_{1t},u_{5t}) \ = \ \rho(u_{1t},u_{3t})\rho(u_{3t},u_{5t}) \tag{B.5}$$

Equating like terms in (B.4) and (B.5) immediately yields the result that $|\rho(u_{3t},u_{5t})| = 1$, which is a contradiction. □

Proof of Proposition 2.2: The proof is easy, given the simple recursive structure of (2.4). That (2.9) implies (2.8) follows from the definition of a partial correlation (see Anderson (1985)). (Clearly, if the conditional correlation between two random variables is 0 when conditioned on some third random variable, then the conditional correlation between those two random variables must be 0 when conditioned on the same third random variable, and any other group of random variables.) That (2.8) implies (2.9) follows from Proposition 2.1. It is easy to see by inspection of (2.4) (and using Proposition 2.1) that the partial correlation constraints implied (with conditioning sets of more than one variable) are

$$\rho(u_{it},u_{jt} \mid u_{qt},u_{(q+1)t}, \cdots ,u_{(q+r)t}) = 0 \tag{B.4}$$

for,

$$1 \le i \le K-3 \text{ and } i+3 \ge j \le K \text{ and } i < q,q+1,...,q+r < j$$

But, Proposition 2.1 ensures that all (and no more) partial correlations of the form given by (2.9) with i, j, q, r as in (B.4) are implied by (2.4). Thus, (2.8) implies (2.9). □

Proof of Proposition 2.3: The proof follows directly from the simple structure of (2.4). That (2.11)-(2.14) implies (2.10) can be seen by multiplying together equations (2.11) and (2.12) and setting them equal to the product of (2.13) and (2.14), which then gives (2.10). Generally, though, a model can imply a vanishing tetrad difference without implying vanishing partial correlations that, in turn, imply the tetrad equation. In the case of a linear causal model given by (2.4), this is not so, however. It is easy to see by inspection of (2.4) or (2.5) that the tetrad difference constraints implied are

$$\rho(u_{it},u_{jt})\rho(u_{kt},u_{lt}) - \rho(u_{it},u_{kt})\rho(u_{jt},u_{lt}) = 0$$

for,

$$1 \le i \le K-3 \text{ and } i+2 \le j \le K-1 \text{ and } i \le l \le j \text{ and } j \le k \le K \tag{B.5}$$

Thus, to prove that (2.10) implies (2.11)-(2.14) it remains to show that for any i, j, k, l that satisfy (B.5) there exists some v such that (2.11)-(2.14) hold. Choose v=j. Then, (2.11) and (2.14) hold by the definition of a partial correlation. (Here, again, to omit the trivial case, assume that the simple correlation between any two elements of u_t is not identically 0 in (2.4) i.e that the αs in (2.4) are *all* $\ne 0$.) Further, since k>l and i<l<j, Proposition 2.1 ensures that (2.12) holds. Finally, since j>i and k>j (2.13) also holds

by Proposition 2.1. □

Proof of Theorem 3.1:

First consider part (i) of the theorem. In Proposition 2.1 it was shown that $\alpha_k^* = 0 \Rightarrow \rho(u_{it}, u_{kt} \mid u_{jt}) = 0$. Clearly, though, given the αs in (2.4) are non-zero coefficients, we also have that $\rho(u_{it}, u_{kt} \mid u_{jt}) = 0 \Rightarrow \alpha_k^* = 0$, since $\rho(u_{it}, u_{kt} \mid u_{jt}) = 0$ only when its numerator is zero, and since the numerator of α_k^* is the same as the numerator of $\rho(u_{it}, u_{kt} \mid u_{jt})$ for any k<j<i.

The proof of part (ii) is slightly more involved, but still draws on straightforward asymptotic results. Without loss of generality, and given Assumption 1, define

$$y_t = B_1 z_{t-1} + u_t ,$$

where $y_t = (y_{1t}, y_{2t}, \cdots, y_{Kt})'$, $z_t = (1, y_{1t}, y_{2t}, \cdots, y_{Kt})'$, and $u_t = (u_{1t}, u_{2t}, \cdots, u_{Kt})'$. B_1 is a Kx(K+1) coefficient matrix. The rest of Assumption 1 is assumed as above. Write $B_1 = (B_{11} \ B_{21} \ \cdots \ B_{K1})'$, so that $u_{it} = y_{it} - B_{i1}z_{t-1}$, $i=1,...,K$ (all i subscripts will be assumed to run from i=1,...,K unless otherwise stated), $\hat{u}_{it} = y_{it} - \hat{B}_{i1}z_{t-1}$, and

$$\hat{u}_{it} = u_{it} + (B'_{i1} - \hat{B}'_{i1})'z_{t-1} \tag{B.6}$$

Define $A_i^* = (B'_{i1} - \hat{B}'_{i1})'$. Again without loss of generality, consider the following:

$$\tilde{\alpha}_1^* = \frac{T^{-1}\sum\hat{u}_{2t}\hat{u}_{1t}T^{-1}\sum\hat{u}_{3t}^2 - T^{-1}\sum\hat{u}_{1t}\hat{u}_{3t}T^{-1}\sum\hat{u}_{2t}\hat{u}_{3t}}{T^{-1}\sum\hat{u}_{2t}^2 T^{-1}\sum\hat{u}_{3t}^2 - (T^{-1}\sum\hat{u}_{2t}\hat{u}_{3t})^2} \tag{B.7}$$

Plugging in (B.6) yields the following for the numerator of (B.7):

$$T^{-1}[\sum u_{2t}u_{1t} + \sum u_{2t}A_1^* z_{t-1} + \sum A_2^* z_{t-1}u_{1t} + \sum A_2^* z_{t-1}A_1^* z_{t-1}]* \tag{B.8}$$
$$*T^{-1}[\sum u_{3t}^2 + 2\sum u_{3t}A_3^* z_{t-1} + \sum(A_3^* z_{t-1})^2] -$$
$$- T^{-1}[\sum u_{3t}u_{1t} + \sum u_{3t}A_1^* z_{t-1} + \sum A_3^* z_{t-1}u_{1t} + \sum A_3^* z_{t-1}A_1^* z_{t-1}]*$$
$$*T^{-1}[\sum u_{2t}u_{3t} + \sum u_{2t}A_3^* z_{t-1} + \sum A_2^* z_{t-1}u_{3t} + \sum A_2^* z_{t-1}A_3^* z_{t-1}]$$

and the denominator of $\tilde{\alpha}_1$ can be written as

$$T^{-1}[\sum u_{2t}^2 + 2\sum u_{2t}A_2^* z_{t-1} + \sum(A_2^* z_{t-1})^2]T^{-1}[\sum u_{3t}^2 + 2\sum u_{3t}A_3^* z_{t-1} + \sum(A_3^* z_{t-1})^2] - \tag{B.9}$$
$$- (T^{-1}[\sum u_{2t}u_{3t} + \sum u_{2t}A_3^* z_{t-1} + \sum A_2^* z_{t-1}u_{3t} + \sum A_2^* z_{t-1}A_3^* z_{t-1}])^2$$

Given the expression for $\hat{\alpha}_k^*$ (below), and upon forming the difference, $T^{1/2}(\tilde{\alpha}_k^* - \hat{\alpha}_k^*)$, inspection of (B.8) and (B.9) shows that all terms in the A_i^*, $i=1,2,3$ are of the form

(i) $T^{-1/2}\sum u_{it}A_i^* z_{t-1}$, $i=1,2,3$

(ii) $T^{-1/2}\sum z'_{t-1}A'^*_i A_j^* z_{t-1}$, $i,j=1,2,3$

But (i) can be written as

$$vec[T^{-1/2}\sum u_{it}A_i^* z_{t-1}] = T^{-1/2}\sum(z'_{t-1}\otimes u_{it})vecA_i^* = (T^{-1}\sum z'_{t-1}\otimes u_{it})[T^{1/2}(B'_{i1} - \hat{B}'_{i1})] = o_p(1) ,$$

since the LS estimator is consistent by Assumption 1. Similarly, (ii) can be written as

$$vec[T^{-1/2}\sum z'_{t-1}A'^*_iA^*_jz_{t-1}] = T^{-3/2}\sum(z'_{t-1}\otimes z'_{t-1})vec(T^{1/2}A'^*_i)(T^{1/2}A^*_j) = o_p(1)$$

since the LS estimator is consistent, and $T^{-1}\sum(z'_{t-1}\otimes z'_{t-1})$ is $O_p(1)$. As the LS estimator is

$$\hat{\alpha}^*_1 = \frac{\sum u_{2t}u_{1t}\sum u^2_{3t} - \sum u_{1t}u_{3t}\sum u_{2t}u_{3t}}{\sum u^2_{2t}\sum u^2_{3t} - (\sum u_{2t}u_{3t})^2} \quad ,$$

and is a consistent estimator of α^*_1, it follows directly that

$$T^{1/2}(\tilde{\alpha}^*_1 - \hat{\alpha}^*_1) = o_p(1).$$

Also, $\tilde{\alpha}^*_1$ is asymptotically equivalent to $\hat{\alpha}^*_1$ and has the same limiting distribution (by the asymptotic equivalence theorem). □

The proof of part (iii) is given above (as the final part of the proof of Proposition 2.1), as it is only required to show that $\rho(u_{jt}, u_{kt} \mid u_{it}) \neq 0$ whenever $\rho(u_{kt}, u_{it} \mid u_{jt}) \neq 0$, k<j<i. □

REFERENCES

Anderson, T.W. (1985), *An Introduction to Multivariate Statistical Analysis*, (2nd ed.), Wiley, New York.

Bernanke, B.S. (1986), "Alternative Explanations of the Money Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25, 49-100.

Blanchard, O. and Watson, M. (1984), "Are Business Cycles All Alike?," National Bureau of Economic Research Working Paper No. 1392.

Board of Governors of the Federal Reserve System (1976), *Banking and Monetary Statistics 1941-1970*, Board of Governors of the Federal Reserve System, Washington D.C..

Engle, R.F. and Granger, C.W.J. (1987), "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251-276.

Geiger, D. and Heckerman, D. (1994), "Learning Gaussian Networks," in: *Uncertainty in Artificial Intellegence, Proceedings of the Tenth Conference*, Morgan Kauffman Publishers, San Fransisco.

Geiger, D. and Pearl, J. (1990), "On the Logic of Causal Models," *Uncertainty in Artificial Intelligence*, 4, 3-14.

Gilli, M. (1992), "Causal Ordering and Beyond," *International Economic Review*, 4, 957-971.

Glymour, C., Scheines, R., Spirtes, P., and Kelly K. (1987), *Discovering Causal Structure*, Academic Press, San Diego.

Glymour, C. and Spirtes, P. (1988), "Latent Variables, Causal Models and Overidentifying Constraints," *Journal of Econometrics*, 39, 175-198.

Granger, C.W.J. (1988), "Some Recent Developments in a Concept of Causality," *Journal of Econometrics*, 39, 199-211.

King, R.G., Plosser, C.I., Stock, J.H., and Watson, M.W. (1991), "Stochastic Trends and Economic Fluctuations," *The American Economic Review*, 81, 819-840.

Lütkepohl, H. (1990), "Testing for Causation Between two Variables in Higher Dimensional VAR Models," paper presented at the World Congress of the Econometric Society, Barcelona 1990.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

McElroy, F.W. (1978), "A Simple Method of Causal Ordering," *International Economic Review*, 19, 1-24.

Orcutt, G.H. (1987), "Joint Conditional Probability Functions for Modeling National Economies," in:M.L. King and D.E.A. Giles, eds., *Specification Analysis in Linear Models*, Routledge & Kegan Paul,

London and New York, pp. 133-145.

Pearl, J. and Verma, T.S. (1990), "A Formal Theory of Inductive Causation," Technical Report R-155, Cognitive Systems Laboratory, University of California, Los Angeles.

Pearl, J. and Verma, T.S. (1991), "A Theory of Inferred Causation," Technical Report R-156, Cognitive Systems Laboratory, University of California, Los Angeles.

Runkle, D.E. (1987), "Vector Autoregressions and Reality," *Journal of Business and Economic Statistics*, 5, 437-451.

Simon, H. A. (1954), "Causal Ordering and Identifiability," in: *Models of Discovery*, Dordrecht, Holland, pp. 53-80.

Simon, H.A. and Iwasaki, Y. (1988), "Causal Ordering, Comparative Statistics, and Near Decomposability," *Journal of Econometrics*, 39, 149-173.

Sims, C.A. (1980a), "Macroeconomics and Reality," *Econometrica*, 48, 1-48.

Sims, C.A. (1980b), "Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered," *American Economic Review*, 70, 250-257.

Sims, C.A. (1981), "An Autoregressive Index Model for the U.S. 1948-1975," in: J. Kmenta & J.B. Ramsey, eds., *Large-Scale Macro-Econometric Models*, North-Holland, Amsterdam pp. 283-327.

Spirtes, P., Glymour, C., and Scheines, R. (1993), *Causation, Prediction, and Search*, Springer Verlag, Ney York.

Swanson, N.R. (1994), Aggregation and Granger Causality in Vector Autoregressions, unpublished Ph.D. dissertation, University of California, San Diego.

Todd, R.M. (1990), "Vector Autoregression Evidence on Monetarism: Another Look at the Robustness Debate," in: *Federal Reserve Bank of Minneapolis Quarterly Review*, Spring 1990, pp. 19-37.

Wei, W.W.S. (1978), "The Effect of Temporal Aggregation on Parameter Estimation in Distributed Lag Model," *Journal of Econometrics*, 8, 237-246.

Weiss, A.A. (1984), "Systematic Sampling and Temporal Aggregation in Time Series Models," *Journal of Econometrics*, 26, 271-281.

Wermuth, N. and Lauritzen, S.L. (1983), "On Substantive Research Hypotheses, Conditional Independence Graphs and Graphical Chain Models," *Journal of the Royal Statistical Society*, 52, 21-72.

Wright, S. (1921), "Correlation and Causation," *Journal of Agricultural Research*, 20, 557-585.

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 57, 348-368.

Zellner, A. (1979), "An Error-Components Procedure (ECP) for Introducing Prior Information About Covariance Matrices and Analysis of Multivariate Regression Models," *International Economic Review*, 20, 679-692.

Table 1: Partial Correlations, Correlation Values, and Constraint Test Statistics.
Example 1: Consumption, Investment, Output, and Real Balances

Partial Correlation	Partial Correlation Value	Test Statistic	$p-Value$
1. $\rho(C_t, I_t \mid Y_t)$	0.384	---	---
2. $\rho(C_t, I_t \mid M_t)$	0.427	6.030	0.000
3. $\rho(C_t, Y_t \mid I_t)$	0.303	4.055	0.000
4. $\rho(C_t, Y_t \mid M_t)$	0.491	7.189	0.000
5. $\rho(C_t, M_t \mid I_t)$	0.119	---	---
6. $\rho(C_t, M_t \mid Y_t)$	0.163	---	---
7. $\rho(I_t, Y_t \mid C_t)$	0.544	8.302	0.000
8. $\rho(I_t, Y_t \mid M_t)$	0.642	10.71	0.000
9. $\rho(I_t, M_t \mid C_t)$	0.062	0.780	0.425
10. $\rho(I_t, M_t \mid Y_t)$	0.133	---	---
11. $\rho(Y_t, M_t \mid C_t)$	-0.047	-0.601	0.548
12. $\rho(Y_t, M_t \mid I_t)$	-0.055	-0.713	0.477

NOTES: C_t, I_t, Y_t, and M_t correspond to the errors from the consumption, investment, output, and real balance equations, respectively. The null hypotheses that the correlations are zero are tested by running regressions of the form $u_{it} = \alpha_j^* u_{jt} + \alpha_k^* u_{kt} + \xi_t$, for i,j,k =C,I,Y,M, k<j<i, and using the estimated errors, $\hat{u}_t$ from (2.1), in place of the actual errors. The test statistic reported in the third column is the standard t-statistic from the relevant coefficient in the above regression. For example, for $\rho(C_t, I_t \mid M_t) = 0$ (restriction 2.) the regression above has i=I, j=C, and k=M; and the t-statistic reported is for the null hypothesis that $\alpha_j^* = 0$. Test statistic p-values under the null hypothesis are tabulated in the fourth column.

Table 2: Partial Correlations, Correlation Values, and Constraint Test Statistics.
Example 2: Interest Rates, Prices, Output, and Money

Partial Correlation	Partial Correlation Value	Partial Correlation Test Statistic	$p-Value$
1. $\rho(R_t, P_t \mid Y_t)$	0.1056	--	--
2. $\rho(R_t, P_t \mid M_t)$	0.1056	1.889	0.060
3. $\rho(R_t, Y_t \mid P_t)$	0.1396	2.507	0.013
4. $\rho(R_t, Y_t \mid M_t)$	0.1399	2.510	0.013
5. $\rho(R_t, M_t \mid P_t)$	-0.0408	-0.726	0.469
6. $\rho(R_t, M_t \mid Y_t)$	-0.0415	-0.739	0.461
7. $\rho(P_t, Y_t \mid R_t)$	-0.0202	-0.359	0.720
8. $\rho(P_t, Y_t \mid M_t)$	-0.0073	-0.129	0.898
9. $\rho(P_t, M_t \mid R_t)$	0.0490	0.871	0.384
10. $\rho(P_t, M_t \mid Y_t)$	0.0452	0.803	0.422
11. $\rho(Y_t, M_t \mid R_t)$	0.0431	0.767	0.444
12. $\rho(Y_t, M_t \mid P_t)$	0.0380	0.676	0.499

NOTES: See notes to Table 1. R_t, P_t, Y_t, and M_t correspond to the errors from the interest rate, prices, money, and output equations, respectively. Note that in the causal model used here, M_t is exogenous. Thus, only one partial correlation zero restriction cannot be tested directly, as opposed to the usual four restrictions for causal models given by (2.4) with four variables.