March 1993

State Space and ARMA Models: An Overview of the Equivalence

by
Paul D. Gilbert

Acknowledgments

I would like to thank Pierre Duguay, Stephen Poloz, and David Delchamps for helpful comments.

ISSN 1192-5434

ISBN 0-662-20494-8

Printed in Canada on recycled paper

Contents

Abstract

In this paper known results about the equivalence of state space and auto-regressive moving-average models with exogenous inputs (ARMAX) (including vector auto-regressive or VAR models) are reviewed. While most of these results are not new, no single reference appears to bring together several important related points. In addition, the original references have used a variety of different conventions for time-shift and parameter signs, have usually not included exogenous inputs (or have not distinguished them from noise), and have frequently not distinguished algebraic equivalence transformations from the statistical estimation. Furthermore, an overview of the subject area reveals some important gaps in the theory.

While the paper does not consider statistical estimation, it does suggest a subtle but important shift in methodology from the approach usually taken in econometrics. The traditional approach in econometrics has been to specify a representation based on economic theory, convert to a reduced form if necessary, impose identifiability constraints, and then estimate model parameters. If, on the one hand, forecasting is the final objective, this approach can work very well, though it is clear from the equivalence that it is not necessary to proceed in this order. On the other hand, if the eventual purpose is to use statistical hypothesis testing techniques to arrive at conclusions about the structure of the economy, then the equivalence can nullify many conclusions. In particular, it is often difficult or impossible to disentangle the statistical evidence from the implications imposed by the original economic theory used to specify the representation.

The alternative methodology suggested by the equivalence examined in this paper is first to estimate the model using whatever representation is convenient for estimation, then to impose economic constraints on the representation. This approach will disentangle structural conclusions that are supported by statistical evidence from those which are implied by economic theory. Statistical tests are applicable only to results that are invariant across all equivalent representations. Economic theory must be used to reach conclusions that do not correspond to invariants of the equivalence class. This latter observation has some implications for continuing debates in the econometrics literature.

Algorithms for model conversion and comparison are presented in an appendix.

Résumé

La présente étude contient une synthèse des résultats ayant trait à l'équivalence qui existe entre les modèles d'espace d'état et les modèles autorégressifs à moyennes mobiles dotés de variables exogènes ou modèles ARMAX (vecteur autorégressif compris). Même si la plupart de ces résultats ne sont pas nouveaux, aucun document ne semblait réunir jusqu'ici toute l'information à leur sujet. L'auteur de l'étude utilise dans différents modèles des conventions uniformes et il fait intervenir une variable exogène, contrairement à la plupart des auteurs des travaux publiés sur la question. Il s'efforce aussi de souligner la différence entre les équivalences algébriques et les estimations statistiques. Enfin, la théorie exposée dans ces travaux contient des lacunes importantes, que cette synthèse aidera sans doute à élucider.

Même si l'étude ne traite pas de l'estimation statistique des modèles, l'auteur y exploite une approche qui renferme de légères différences par rapport à l'approche économétrique habituelle. L'approche habituelle consiste à spécifier au départ une représentation du modèle, à en réduire au besoin la forme, à imposer des contraintes d'identification et enfin à en estimer les paramètres. Lorsque l'objectif final est de faire des prévisions, cette méthode se révèle efficace, bien que l'ordre de ces étapes puisse de toute évidence être modifié, étant donné l'équivalence entre les modèles. Cependant, s'il s'agit de tester des hypothèses sur la structure de l'économie, il se peut que l'on ne puisse conclure grand-chose des tests effectués à cause de cette relation d'équivalence. Par exemple, il est souvent difficile, voire impossible, de faire la distinction entre les résultats statistiques et les implications des hypothèses de base relatives à la représentation du modèle.

L'auteur propose plutôt d'estimer le modèle dont la représentation se prête le mieux à une estimation avant d'imposer les contraintes découlant de la théorie économique. Cette façon de procéder permet d'établir une distinction claire entre les conclusions qui s'appuient sur des données statistiques et celles qui découlent de la théorie économique. Comme les tests statistiques ne s'appliquent qu'aux invariants d'une même classe d'équivalences, il faut faire appel à la théorie économique pour tirer des conclusions qui ne se rapportent pas aux invariants. Cette constatation devrait avoir des conséquences d'ordre pratique pour l'analyse ultérieure de ces questions.

Les algorithmes utilisés pour la manipulation et la comparaison des modèles sont donnés en annexe.

1. Introduction

This paper reviews the input-output equivalence of various state space and ARMA (auto-regressive moving-average) model representations.[1] The equivalence is algebraic, not statistical, so it does not rely on any consideration of data nor on convergence in probability spaces. One model can be transformed into an equivalent model and the two models will give equivalent results, subject only to the accuracy of the numerical algorithms used for the transformation. This is typically many orders of magnitude better than the differences in statistical estimation techniques.[2]

Most of the results reviewed here are known but the implications do not appear to be widely appreciated. For example, while this paper does not specifically consider statistical estimation, it does suggest a subtle but important shift in methodology from the approach usually taken in econometrics. The traditional approach in econometrics has been to specify a representation based on economic theory, convert to a reduced form if necessary, impose identifiability constraints, and then estimate model parameters. If, on the one hand, forecasting is the final objective, this approach can work very well, though it is clear from the equivalence that it is not necessary to proceed in this order. On the other hand, if the eventual purpose is to use statistical hypothesis testing techniques to arrive at conclusions about the structure of the economy, then the equivalence can nullify many conclusions. In particular, it is often difficult or impossible to disentangle the statistical evidence from the implications imposed by the original economic theory used to specify the representation.

The alternative methodology, suggested by the equivalence, is first to estimate the model using whatever representation is convenient for estimation, then to impose economic constraints on the representation. This approach will disentangle structural conclusions that are supported by statistical evidence from those which are implied by economic theory. Statistical tests are applicable only to results that are invariant across all equivalent representations. Economic theory must be used to reach conclusions that do not correspond to invariants of the equivalence class.

In this paper the abbreviation ARMA includes models with exogenous inputs (ARMAX models), and vector auto-regressive (VAR) models are considered to be a subset. Univariate models can also be considered a special case, but do not receive much special attention. While most of these results are not new, no single reference appears to bring together several important related points. In addition, the original references have used a

1. The term "model" is used here to mean "statistical model." This is not an economic model (it is not based on economic theory). There are many equivalent parameterizations (representations) of the statistical model.

2. The distinction between the algebraic equivalence and the statistical estimation technique is not made very clear in the literature. The reason for this is that a given estimation technique usually works with a particular representation. For example, Aoki and Havenner (1991) in their concluding remarks make the statement "in finite samples all [equivalent representations] can be expected to result in different approximate models when the true model is unknown, despite their algebraic equivalence." However, these different approximate models are a result of different estimation techniques and are not in any way related to the algebraic equivalence, as one might think is implied by Aoki and Havenner.

variety of different conventions for time-shift and parameter signs, have usually not included exogenous inputs (or have not distinguished them from noise), and have frequently not distinguished algebraic equivalence transformations from the statistical estimation. Furthermore, an overview of the subject area reveals some important gaps in the theory.

The remainder of this first section of the paper defines the basic models. Section 2 outlines various representations and gives an indication of their potential usefulness. Section 3 gives an example. Section 4 summarizes many important observations which follow from the equivalences. Section 5 formalizes the equivalence relationships and leads to a precise technical description of the statistical parameter space. Algorithms for model conversion and comparison are presented in an appendix.

A linear time-invariant ARMA representation is given by

$$A(L)\, y_t = \mathrm{B}(L)\, \varepsilon_t + C(L)\, u_t \tag{1}$$

where y_t is a p dimensional vector of observed output variables, u_t is an m dimensional vector of input variables, ε_t is a p dimensional unobserved disturbance vector process and A, B and C are matrices of the appropriate dimension in the lag (back shift) operator L. VAR models will be given special attention but for many purposes can be thought of as a special case of ARMA models with $B(L)=I$.

In an ARMA model the dynamics of the processes are incorporated in the model by the lags in the polynomial arrays A, B and C. Using a different form of model it is possible to summarize the dynamics of the processes in a vector called the state vector. The state vector and the inputs at any point in time are all that is required in order to obtain the state at the next period. This makes the dynamics of the model into a Markov process given by the state transition equation. The outputs of the model are then obtained from the state by an equation called the output or measurement equation. The two equations are called a state space model.

A linear time-invariant state space representation in innovations form is given by[3]

$$z_t = F\, z_{t-1} + G\, u_t + K\, \varepsilon_{t-1} \tag{2}$$

$$y_t = H\, z_t + \varepsilon_t \tag{3}$$

where z_t is the unobserved underlying n dimensional state vector,[4] F is the state transition matrix, G, the input matrix, H, the output matrix, and K, the Kalman gain. This is a special case of the more general form

---

3. Other conventions are often used for the time subscript of u relative to z and y. This convention is chosen because in econometric models it is important for the control to be able to affect the output in the same period. Another way to achieve this is to have an additional term for the immediate effect of u_t in (3) and (5).

4. The objective of a Kalman filter is to estimate this unobserved underlying state. Sometimes an innovations form state space model is referred to as a Kalman filter model.

$$z_t = F\, z_{t-1} + G\, u_t + Q\, \nu_t \tag{4}$$

$$y_t = H\, z_t + R\, \varepsilon_t \tag{5}$$

where ν_t is the system noise, Q, the system noise matrix, and R the output (measurement) noise matrix. In the innovations representation both the system noise and the measurement noise are combined in the process ε_t, while in the general model these are separated into two distinct processes. Collectively (F,G,H,K) or (F,G,H,Q,R) are referred to as the system matrices and a particular parameterization of the system matrices is referred to as a realization. In the general state space model it is usually assumed that ε and ν are independent white noise processes, whereas in the innovations model, the covariance of ε_t is not fixed but it is assumed that there is no auto-correlation in the process.

For any general state space model defined by (4) and (5) it is possible to find an input-output equivalent innovations model of the form (2) and (3). The principal reason for using the general form is that some theoretical description of the system indicates a particular structure. In this case it can be possible to distinguish system noise, which affects the dynamics, from measurement noise. It should be observed that the state z_t is not observed and thus its scale and basis are not fixed a priori, so in both cases it is necessary to impose additional constraints in order to do statistical estimation. In this paper, attention will be focussed on the innovations form, since statistical information alone would usually not be sufficient to indicate the more general model. After the estimation is complete, a statistically determined innovations model can be converted to the more general form according to some knowledge or prescribed theory of the true system.

The models are not unique in the sense that there are many different parameterizations which will have the same input-output behaviour. An ARMA model can be pre- multiplied by any invertible matrix to give another equivalent model. In the state space case, the underlying state is not observed, so there is no statistical constraint on its ordering or scale, and any change of basis for the state will result in a new model representation. Different representations of the state give different parameterizations for the system matrices F, G, H and K (or F, G, H, Q and R). This non-uniqueness causes an estimation difficulty commonly referred to as the “identification problem.” If a sufficient number of constraints are imposed on allowable representations, then it is not possible to transform a given parameterization and find an equivalent form which satisfies the constraints. In the ARMA case the most obvious constraint is to restrict $A_0=B_0=I$ (which implies that the covariance of the disturbance process is not assumed to be orthonormal). In the state space case imposing constraints is referred to as selecting a “canonical form.” In both cases there is considerably more subtlety to this constraint process than at first appears. The last section of this paper considers those subtleties more carefully.

There is an equivalence between state space and ARMA models which has some similarity to the well-known fact that any nth order differential equation can be represented as n first order differential equations. As a simple example of equivalent state space and ARMA models, consider the univariate ($p=1$) AR(2) model with no input given by

$$(1+.5L+.3L^2)\, y_t = \varepsilon_t \tag{6}$$

An equivalent state space model is given by

$$z_t = \begin{bmatrix} 0 & -0.3 \\ 1 & -0.5 \end{bmatrix} z_{t-1} + \begin{bmatrix} -0.3 \\ -0.5 \end{bmatrix} \varepsilon_{t-1} \tag{7}$$

$$y_t = \begin{bmatrix} 0 & 1 \end{bmatrix} z_t + \varepsilon_t \tag{8}$$

where n, the dimension of the state, is 2. To see that this is equivalent, make the following substitutions:

$$\begin{aligned} y_t &= \begin{bmatrix} 0 & 1 \end{bmatrix} z_t + \varepsilon_t \\ &= \begin{bmatrix} 0 & 1 \end{bmatrix} \left(\begin{bmatrix} 0 & -0.3 \\ 1 & -0.5 \end{bmatrix} z_{t-1} + \begin{bmatrix} -0.3 \\ -0.5 \end{bmatrix} \varepsilon_{t-1} \right) + \varepsilon_t \\ &= \begin{bmatrix} 1 & -0.5 \end{bmatrix} z_{t-1} + \begin{bmatrix} -0.5 \end{bmatrix} \varepsilon_{t-1} + \varepsilon_t \\ &= \begin{bmatrix} 1 & -0.5 \end{bmatrix} \begin{bmatrix} -0.3\,(y_{t-2} - \varepsilon_{t-2}) - 0.3\varepsilon_{t-2} \\ y_{t-1} - \varepsilon_{t-1} \end{bmatrix} - 0.5\varepsilon_{t-1} + \varepsilon_t \\ &= -0.3y_{t-2} - 0.5y_{t-1} + \varepsilon_t \end{aligned}$$

The second line comes from substituting the state transition equation (7) into the output equation (8). The fourth line comes from using the state transition equation for the first element of the state at time $t-1$, $z_{1,\,t-1} = -0.3z_{2,\,t-2} - 0.3\,\varepsilon_{t-2}$, and the output equation at times $t-2$ and $t-1$, $z_{2,\,t-2} = y_{t-2} - \varepsilon_{t-2}$ and $z_{2,\,t-1} = y_{t-1} - \varepsilon_{t-1}$.

In the state space model the elements of the state vector summarize the dynamics of the system, and the dimension of the state vector largely determines the complexity of these dynamics. There is a minimum possible dimension for the state vector of any system with given dynamics. A set of system matrices with minimum state dimension is called a minimal realization of the system. The dimension of the state determines the maximum lag of any noise or output series which can affect the system. It is possible that no series affects the system at this maximum lag or it is possible that more than one series enters at this maximum lag. The state is analogous to principal components in multivariate statistics. All of the most relevant information up to the maximum lag is extracted from the variables and summarized in the state variables. In fact, the related technique of canonical correlation forms the basis for an estimation procedure developed by Akaike, which is one of the most widely used methods for estimating state space models.

Using (2), equation (3) can be expanded:

$$y_t = HF^r z_{t-r} + \varepsilon_t + \sum_{i=0}^{r-1} HF^i\,(Gu_{t-i} + K\varepsilon_{t-1-i}) \tag{9}$$

For a stable model, F^r will converge to zero as r becomes large. This illustrates that the state space model has a (possibly infinite) MA representation.

Using the substitution $F=F''+KH$ (2) and (3) give

$$y_t = HF^r z_{t-r} + \varepsilon_t + \sum_{i=0}^{r-1} HF''^{\,i} (Gu_{t-i} + Ky_{t-1-i}) \qquad (10)$$

This illustrates that the model has a (possibly infinite) VAR representation. By the Cayley-Hamilton theorem F^r can be expressed in terms of lower powers of F. Using this substitution in general gives an ARMA model (see, for example, Aoki and Havenner 1991).

It can be shown that any ARMA model has an equivalent state space model and vice versa.[5] Notice that the output matrix H is usually not square and cannot be assumed to be invertible. If it were, the relationship between state space and ARMA models would be trivial. However, typically, the state dimension exceeds the output dimension. Interest in the state space representation comes from at least three aspects. The first is that there can be some advantages for model estimation (in particular for model reduction as discussed in Aoki 1990; Aoki and Havenner 1991; Mittnik 1991; and Vaccaro and Vukina 1993). The second is that the model explicitly estimates an underlying state which summarizes the process. This underlying state may be of significant theoretical interest. The third is that the ideas of control theory can be applied.

The coefficient matrices of u_{t-i} and y_{t-1-i} implicitly defined by the matrix multiplications in equation (10) are sometimes called Markov parameters. These are important for an estimation technique proposed by Mittnik (1991).

A system with state dimension n is called observable[6] if the observability matrix, $O_n = [H' \mid (HF)' \mid (HF^2)' \mid \ldots \mid (HF^{n-1})']'$ has rank n or, equivalently, if the observability gramian, $O_n'O_n$ is positive definite. The system is called controllable if the controllability matrix, $C_n = [G \mid FG \mid F^2G \mid \ldots \mid F^{n-1}G]$ has rank n or, equivalently, if the controllability gramian, C_nC_n' is positive definite. Controllability is often called reachability in the literature. The idea behind these notions is that the state of a non-stochastic system (no noise terms) can be determined from a sufficient number of observations of the output, if the system is observable, and the system can be guided to any desired state if it is controllable.[7]

5. This will be addressed more precisely in the last section of this paper.

6. This term has had this precise technical meaning in the mathematical systems and engineering literature for more than a quarter of a century. The phrase “not observable” means the observability matrix has rank less than n, and thus it is impossible to estimate the state with any amount of data. In the recent economics literature the phrase “not observable” is often used to mean “not observed.” Caution is advised.

7. For linear time-invariant systems, “sufficient” is n. For stochastic systems, the definitions of controllable and observable are more complicated but the intuition is the same.

An ARMA model is stable if the roots of the determinant of $A(L)$ all have norm greater than one, and a state space system is stable if the eigenvalues of F are less than one. In fact, the roots of $A(L)$ are the inverse of the eigenvalues of F if the models are equivalent.

2. Some Typical Representations

This section briefly outlines some state space and ARMA representations of interest.

ARMA with restriction $A_0=B_0=I$

This restriction is almost universally applied to ARMA models since it seems also to be the most natural way to think of an ARMA model. Any ARMA model with non-singular A_0 can be converted to this form by pre-multiplying by A_0 inverse and adjusting the covariance of ε_t so that $B_0=I$.

VAR models

In theory any ARMA model can be converted to a VAR representation:

$$B(L)^{-1}A(L)\, y_t = \varepsilon_t + B(L)^{-1}C(L)\, u_t \tag{11}$$

where the $B(L)^{-1}$ can be thought of as the inverse in the field of matrices with rational polynomial entries, or as the operation of dividing one polynomial matrix by another. The result may have an infinite number of terms, but it is common practice to work with a truncated approximation. The zero lag term of the left side polynomial matrix is usually adjusted to be I by making the corresponding adjustment to the coefficient of u_t and the covariance of ε_t.

Canonical state space representations

Traditionally, state space models have been specified in some “canonical form,” with many of the entries in the system matrices set to one or zero. The most common of these forms are the controllability and the observability canonical forms, so named because the effect of the control on the state is clearly apparent in the first and the value of the state is clearly apparent from the observed outputs in the second. The control systems literature from the 1970s is replete with examples of different canonical forms and claims of superiority for purposes of estimation. The most insightful work in this regard is a procedure of van Overbeek and Ljung (1982). They proposed switching canonical forms during an iterative estimation process in order to avoid ill-conditioning problems which arise owing to the complicated nature of the underlying parameter space. (This underlying parameter space will be discussed further in Section 5.) Their procedure does not, however, provide a method to avoid the many other pitfalls of iterative estimation and it has not been widely adopted.

ARMA models by state elimination

The Cayley-Hamilton theorem can be invoked to eliminate states and convert state space models to ARMA models. (See, for example, Aoki and Havenner 1991.) Their presentation does not include exogenous variables but this is a straightforward (though tedious) extension. The method produces ARMA models with a diagonal *A(L)* matrix and a large number of lags in all parameters. This is usually not the most parsimonious (nor the most intuitive) representation of an ARMA model.

Nested realizations

Nested state space realizations provide an exceedingly simple approach to model reduction. The basic idea of model reduction is to find a smaller model which is a good approximation of a given model. For state space models this means finding a smaller state dimension and incorporating as much information as possible from the larger state into the smaller state. The idea of using a balanced representation to facilitate model reduction appears to be originally due to Moore (1981). It was further developed by Pernebo and Silverman (1982) and has been advocated by Aoki (1990).[8] Aoki and Havenner (1991) and Mittnik (1991) provide an excellent overview of this subject area and related estimation issues. The idea of nested realizations is that elements of the state should contain information in decreasing order of significance. The system matrices are arranged so that when the dimension is reduced the best[9] model for the smaller dimension system is obtained without re-estimation.

Mittnik suggests a “fully nested” approach to balanced realizations.[10] With a fully nested representation the system matrices of a model with smaller dimension state are calculated simply by eliminating the rows and columns corresponding to the last state element. To achieve the fully nested representation, it is necessary to convert an innovations model to a model with lagged outputs as inputs, employing the same transformation used to illustrate the VAR representation in (10). The model then has the form

$$z_t = F'' z_{t-1} + G u_t + K y_{t-1} \qquad (12)$$

$$y_t = H z_t + \varepsilon_t \qquad (13)$$

where $F'' = F - KH$.

---

8. While Aoki has advocated this approach, his algorithm for estimating these models contains an error. This has been noted in Heij, Kloek and Lucas (1992) and Vaccaro and Vukina (1993). Vaccaro and Vukina also propose a correction.

9. Mittnik (1991) points out a number of caveats with respect to the approximations involved. In reply Havenner (1991) makes the clarification that the nested realizations are a consistent estimate of the best model at each state dimension, where “best” is interpreted in the sense of the Hankel norm. It remains to be determined how good the reduced models are with respect to other criteria.

10. A realization is called balanced if the controllability and observability gramians are equal and diagonal. This happens as a result of the singular value decomposition procedure used to calculate these realizations, but for the purposes of model reduction the important feature is the nestedness property.

Aoki and Havenner work with the original innovations model but their nested realization is not *fully* nested in the sense that a more complicated calculation is required to get the system matrices for a reduced system.[11]

3. Example

This section describes a simple monthly model for monitoring the economy. The model uses 90-day interest rates as the input variable u_t and M1, GDP lagged two months[12] and CPI as the output variables y_t, in the order given.[13] M1, GDP and CPI were all seasonally adjusted data. The first difference of the log of all output variables was used but interest rates were not transformed.[14] The model is presented here to illustrate the equivalence of various representations. In particular the algebraic equivalence is emphasized and it is illustrated that differences among the representations are of the order of magnitude one would expect from numerical procedures. These are clearly smaller than the differences expected from different estimation techniques. The VAR model used as the initial model was estimated by a least squares technique.

A VAR representation of the model is given by

$$A(L)\ [\mathrm{M1}_t, \mathrm{GDP}_{t-2}, \mathrm{CPI}_t]' = B(L)\ \varepsilon_t + C(L)\ [\mathrm{R90}]_t$$

where

$$A(L) = \begin{bmatrix} 1\text{-}0.0582L^1\text{+}0.149L^2\text{-}0.0335L^3 & 0\text{-}0.0694L^1\text{-}0.0533L^2\text{+}0.125L^3 & 0\text{-}0.4L^1\text{-}0.0452L^2\text{-}0.657L^3 \\ 0\text{+}0.0211L^1\text{+}0.029L^2\text{-}0.0173L^3 & 1\text{+}0.204L^1\text{-}0.0339L^2\text{-}0.107L^3 & 0\text{+}0.0035L^1\text{+}0.171L^2\text{-}0.183L^3 \\ 0\text{-}0.0202L^1\text{-}0.0262L^2\text{-}0.0243L^3 & 0\text{-}0.0733L^1\text{-}0.0343L^2\text{+}0.081L^3 & 1\text{-}0.106L^1\text{-}0.238L^2\text{-}0.0876L^3 \end{bmatrix}$$

$$B(L) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad C(L) = \begin{bmatrix} \text{-}0.00142\text{-}0.00151L^1\text{+}0.00183L^2 \\ 7.09\text{e-}05\text{-}7.9\text{e-}05L^1\text{-}0.000393L^2 \\ 0.000506\text{+}9.18\text{e-}07L^1\text{-}7.68\text{e-}05L^2 \end{bmatrix}$$

11. Aoki and Havenner's approach is similar but they do not do this transformation. The result is that *F*,*G*, and *H* can be determined for a lower dimension system by truncation, but *K* must be determined by solving a matrix Riccati equation.

12. GDP is lagged because it is not available on as timely a basis. Lagging was necessary given the original intent of the model for timely monitoring. The most recent observation available was considered to be valuable information. The economic interpretation that interest rates can affect GDP two periods in advance of when they are set does not make much sense unless one allows for economic agents anticipating rate changes. In practice the estimated impact of interest rates on previous GDP appears to be very small.

13. The Statistics Canada series identifiers were: B14017, B1627, I37026, B820200.

14. The data was not stationary without differencing. The possibility of cointegration was not examined for this example.

This representation has a likelihood of -2567 with the sample used for the estimation.[15] The roots of the determinant of *A(L)* are 1.32, -1.34+0.975i, -1.34-0.975i, 2.16, -0.436+1.83i, -0.436-1.83i, 0.755-2.18i, 0.755+2.18i and -3.11. The inverses of these roots are 0.758, -0.488-0.355i, -0.488+0.355i, 0.463, -0.123-0.517i, -0.123+0.517i, 0.142+0.409i, 0.142-0.409i, and -0.321.

Figure1 shows the predicted CPI reconstituted from the differenced log prediction of the model and the actual CPI, both in terms of per cent change over 12 months. Figure 2 gives a more detailed view of the performance out of sample.

FIGURE 1. Predicted and actual CPI in terms of per cent change over 12 months

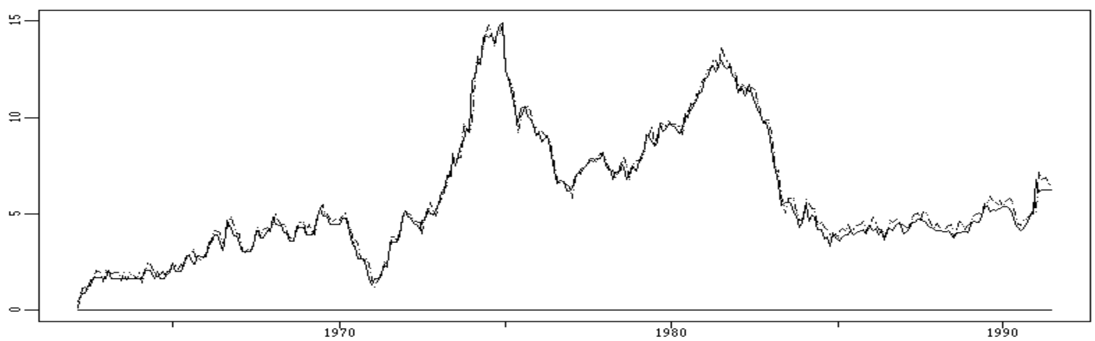
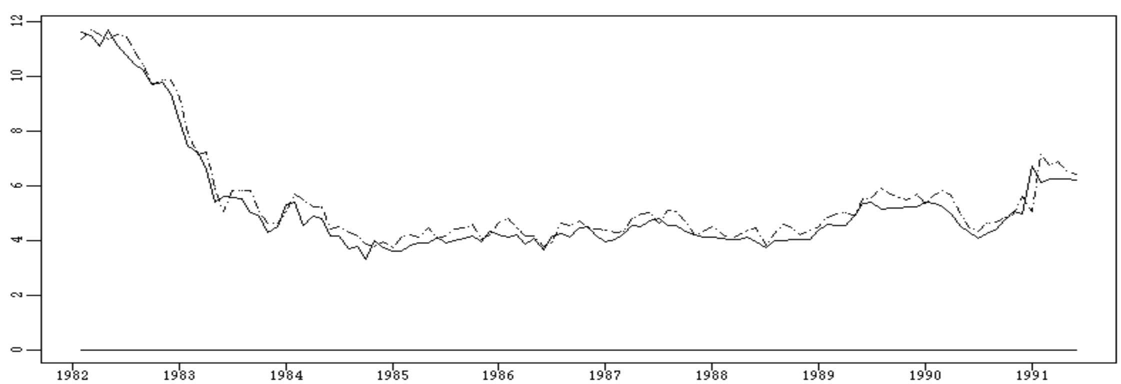


FIGURE 2. Predicted and actual CPI in terms of per cent change over 12 months - ex post period



Some equivalent representations will now be illustrated. The likelihood of all models was the same (-2567.32801321424) as that for the above VAR model up to 15 significant digits. The outputs from the different models are identical up to numerical precision. The roots of the determinant of *A(L)* for the ARMA model are almost identical to those of the VAR model above. The eigenvalues of the state transition matrix *F* are almost identical to the inverses of the roots of the determinant of *A(L)* for the VAR model. Table 1

15. Approximately 30 years of monthly data were available, starting in January 1961. The first 20 years were used for estimation. Parameters have been rounded for presentation. The likelihood and roots are based on parameters with more significant digits as estimated.

summarizes the differences.[16] To give an idea of the orders of magnitude involved, the standard deviations of the outputs from the above model are 0.00365, 0.00191 and 0.00262, for M1, GDP and the CPI (in first differences of logarithms). These comparisons can be summarized by saying that the models are the same to numerical precision; however, the additional detail seems necessary, given persistent references in the literature suggesting this is only a theoretical equivalence.

Table 1: Comparisons with VAR representation

VAR model	fully nested state space representation	state space representation of "gap" model	ARMA representation
inverse of roots of |A(L)|	**difference of eigenvalues of F or inverse of roots of |A(L)| from inverse of roots of |A(L)| for the VAR representation**		
0.758	-4.44e-16-1.87e-19i	1.15e-14-1.87e-19i	1.15e-09-9.63e-10i
-0.488-0.355i	2.78e-16-8.88e-16i	7.94e-15-3.05e-15i	3.25e-08+2.07e-08i
-0.488+0.355i	1.67e-16+7.77e-16i	-8.05e-15+2.94e-15i	-3.24e-06+1.20e-06i
0.463	-8.33e-16+7.19e-18i	0.00e+00+7.19e-18i	-4.87e-09-5.48e-08i
-0.123-0.517i	5.69e-16-5.55e-16i	1.57e-15-8.66e-15i	4.69e-07-4.20e-07i
-0.123+0.517i	5.69e-16+4.44e-16i	1.57e-15+8.55e-15i	-5.47e-08-2.44e-07i
0.142+0.409i	-1.11e-16+7.77e-16i	2.44e-15+1.39e-15i	2.38e-06+2.79e-06i
0.142-0.409i	-1.94e-16-6.66e-16i	2.36e-15-1.28e-15i	-4.78e-07+6.35e-07i
-0.321	-3.33e-16+4.11e-17i	-2.66e-15+4.11e-17i	-5.33e-15+7.20e-14i
	maximum absolute difference from VAR representation's outputs		
	3.46e-17	3.88e-16	2.94e-17
	4.68e-17	1.73e-16	5.03e-17
	8.50e-17	2.18e-16	8.15e-17

16. Parameter values are truncated to three significant digits for presentation. The comparisons are based on calculated parameters with considerably more significant digits.

Parameters for an equivalent state space representation which was balanced by Mittnik's technique are

$$F = \begin{bmatrix} 0.389 & 0.771 & 0.0204 & -0.104 & -0.0803 & -0.101 & 0.0853 & -0.0339 & 0.0116 \\ 0.565 & -0.275 & 0.695 & -0.0897 & 0.0128 & 0.0344 & 0.0167 & -0.00863 & 0.00144 \\ -0.0437 & -0.555 & 0.000239 & 0.154 & -0.134 & -0.189 & 0.0469 & -0.0189 & 0.00985 \\ 0.0574 & -0.179 & -0.151 & -0.133 & 0.546 & -0.168 & 0.112 & -0.0553 & 0.0348 \\ -0.0945 & -0.0388 & 0.0294 & -0.652 & -0.321 & -0.427 & -0.257 & 0.000204 & -0.00493 \\ -0.0876 & -0.0114 & -0.0385 & 0.100 & -0.434 & 0.387 & -0.110 & -0.0209 & 0.0553 \\ 0.111 & -0.0181 & -0.0520 & 0.0429 & -0.317 & -0.176 & 0.484 & 0.493 & 0.0394 \\ 0.0870 & -0.0205 & -0.0508 & -0.0206 & -0.0540 & 0.178 & -0.422 & -0.0465 & 0.354 \\ 0.00426 & -0.00218 & 0.00822 & -0.0142 & -0.0225 & 0.0587 & 0.0734 & -0.336 & -0.524 \end{bmatrix}$$

$$G = \begin{bmatrix} 0.000187 \\ 0.00139 \\ -0.00263 \\ 0.000103 \\ 0.00320 \\ 0.00143 \\ -0.000927 \\ 0.00444 \\ 0.00700 \end{bmatrix} \qquad K = \begin{bmatrix} -0.0591 & -0.00405 & -0.805 \\ 0.148 & -0.167 & 0.0692 \\ -0.00789 & 0.115 & -0.416 \\ -0.147 & -0.452 & -0.0350 \\ 0.0695 & -0.0774 & 0.0117 \\ 0.121 & -0.125 & -0.0316 \\ -0.126 & -0.0140 & 0.00880 \\ -0.0898 & 0.00418 & 0.00585 \\ -0.0138 & -0.00067 & 0.000645 \end{bmatrix}$$

$$H = \begin{bmatrix} -0.726 & 0.0955 & 0.466 & -0.0619 & -0.0326 & 0.00922 & 0.0377 & -0.0180 & 0.00341 \\ -0.0585 & 0.263 & 0.134 & 0.416 & -0.00783 & -0.0814 & -0.0915 & 0.0281 & -0.00600 \\ -0.181 & -0.246 & 0.0597 & -0.104 & 0.171 & 0.0928 & -0.117 & 0.0786 & -0.0124 \end{bmatrix}$$

This can be further transformed to another equivalent state space representation where the first and second elements of the state correspond to M1 and GDP, respectively, less the zero mean error term ε_t. This can be seen from the output matrix H. Thus the first two elements of the state could be given the interpretation of "underlying" or "fundamental" measures of M1 and GDP. It can also be seen from H that CPI is a multiple of the difference ("gap") between this fundamental GDP and the third element of the state, which might then be given the interpretation as a measure of unobserved underlying potential. The representation is given by

$$F = \begin{bmatrix} -1.05 & 2.37 & -0.497 & -0.153 & -0.912 & -0.292 & 0.784 & -0.425 & 0.0650 \\ -2.37 & 6.84 & -0.673 & -2.07 & -0.900 & -0.189 & 1.45 & -0.814 & 0.149 \\ -25.6 & 79.0 & -6.47 & -24.6 & -9.92 & -0.0850 & 16.9 & -7.19 & 1.06 \\ 0.335 & -1.92 & 0.107 & 0.528 & 0.728 & -0.219 & -0.193 & 0.0863 & 0.00936 \\ 0.231 & -0.495 & 0.0254 & -0.469 & -0.274 & -0.444 & -0.339 & 0.0375 & -0.0117 \\ 0.752 & -2.24 & 0.186 & 0.811 & -0.106 & 0.386 & -0.547 & 0.197 & 0.0170 \\ -0.300 & 0.501 & -0.0445 & -0.119 & -0.400 & -0.178 & 0.590 & 0.439 & 0.0488 \\ -0.178 & 0.176 & -0.0183 & -0.0782 & -0.0899 & 0.175 & -0.380 & -0.0686 & 0.358 \\ -0.0894 & 0.284 & -0.0251 & -0.101 & -0.0664 & 0.0574 & 0.130 & -0.365 & -0.519 \end{bmatrix}$$

$$G = \begin{bmatrix} -1.42e-03 \\ 7.09e-05 \\ -4.99e-03 \\ 1.03e-04 \\ 3.20e-03 \\ 1.44e-03 \\ -9.27e-04 \\ 4.45e-03 \\ 7.01e-03 \end{bmatrix} \qquad K = \begin{bmatrix} 0.0582 & 0.0694 & 0.399 \\ -0.0211 & -0.204 & -0.00349 \\ -0.222 & -0.937 & -1.06 \\ -0.147 & -0.452 & -0.0350 \\ 0.0695 & -0.0774 & 0.0117 \\ 0.121 & -0.125 & -0.0316 \\ -0.126 & -0.0140 & 0.00880 \\ -0.0898 & 0.00418 & 0.00585 \\ -0.0139 & -0.00067 & 0.000645 \end{bmatrix}$$

$$H = \begin{bmatrix} 1 & 0.0 & 0.0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.0 & 0.0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & -0.1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The underlying state variable is estimated by the Kalman filter and could be used as a measure of potential. If one thinks of “potential” as some fundamental underlying process which should only change gradually, then this representation is not entirely satisfactory. It is quite noisy, being simply defined as a multiple of the difference between the best estimate for GDP and the best estimate for CPI. If the model was further transformed into the more general state space form given by (4) and (5) and a smoother[17] was used on the state estimate, then the measure of potential would probably vary more gradually.

An equivalent ARMA model can be calculated by a substitution using the Cayley-Hamilton theorem as described in Aoki and Havenner (1991) but suitably modified to take account of an exogenous input. This is given by

17. A smoother is similar to a filter but uses future observations, in addition to past observations, to give an estimate of the state.

$$A(L) = \begin{bmatrix} 1+0.0397L^{1}-0.157L^{2}-0.238L^{3}-0.0677L^{4}-0.0417L^{5}+0.0157L^{6}-0.000295L^{7}+0.00366L^{8}+0.00218L^{9} & 0 & 0 \\ 0 & 1+0.0397L^{1}-0.157L^{2}-0.238L^{3}-0.0677L^{4}-0.0417L^{5}+0.0157L^{6}-0.000295L^{7}+0.00366L^{8}+0.00218L^{9} & 0 \\ 0 & 0 & 1+0.0397L^{1}-0.157L^{2}-0.238L^{3}-0.0677L^{4}-0.0417L^{5}+0.0157L^{6}-0.000295L^{7}+0.00366L^{8}+0.00218L^{9} \end{bmatrix}$$

$$B(L) = \begin{bmatrix} 1+0.0979L^{1}-0.294L^{2}-0.227L^{3}-0.00629L^{4}+0.00832L^{5}+0.0241L^{6} & 0+0.0694L^{1}+0.0752L^{2}-0.13L^{3}+0.0118L^{4}+0.0439L^{5}-0.0423L^{6} & 0+0.4L^{1}+0.127L^{2}+0.641L^{3}+0.0942L^{4}+0.0039L^{5}-0.0929L^{6} \\ 0-0.0211L^{1}-0.0269L^{2}+0.0219L^{3}+0.00606L^{4}-0.000935L^{5}+0.00292L^{6} & 1-0.165L^{1}-0.0916L^{2}-0.134L^{3}-0.0509L^{4}-0.0233L^{5}-0.013L^{6} & 0-0.0035L^{1}-0.179L^{2}+0.18L^{3}-0.0441L^{4}+0.0145L^{5}+0.00525L^{6} \\ 0+0.0202L^{1}+0.0287L^{2}+0.0261L^{3}+0.0039L^{4}-0.000669L^{5}-0.00399L^{6} & 0+0.0733L^{1}+0.0314L^{2}-0.0692L^{3}+0.00792L^{4}-0.0151L^{5}-0.000313L^{6} & 1+0.146L^{1}+0.104L^{2}-0.105L^{3}-0.00794L^{4}-0.0193L^{5}+0.00573L^{6} \end{bmatrix}$$

$$C(L) = \begin{bmatrix} -0.00142-0.00144L^{1}+0.00216L^{2}+0.0012L^{3}-0.000167L^{4}-0.000412L^{5}-0.000124L^{6}-3.55e\text{-}05L^{7}+6.8e\text{-}05L^{8} \\ 7.09e\text{-}05-6.26e\text{-}05L^{1}-0.000408L^{2}+0.000124L^{3}-5.61e\text{-}05L^{4}+8.1e\text{-}05L^{5}+3.54e\text{-}05L^{6}+2.99e\text{-}06L^{7}+1.01e\text{-}05L^{8} \\ 0.000506+5.15e\text{-}05L^{1}-9.86e\text{-}05L^{2}-0.000144L^{3}-1.07e\text{-}05L^{4}+6.66e\text{-}05L^{5}+1.53e\text{-}05L^{6}+1.23e\text{-}05L^{7}-7.62e\text{-}06L^{8} \end{bmatrix}$$

The procedure for calculating this representation is somewhat more numerically sensitive than those used for calculating the previous representations in the sense that it adds several additional roots very close to the correct roots. For this reason it is necessary to distinguish “distinct” roots, which have been reported in Table 1. This was done by arbitrarily picking one of the roots if there were several in a small neighbourhood. (The size of the neighbourhood used for this was 0.0001, but the procedure is not sensitive to this parameter.) A more judicious choice of the distinct roots might result in a smaller discrepancy between these roots and those of the VAR model, but they are already fairly close.

For these examples the VAR representation was taken as the starting point. If the original model were a state space or ARMA model with a moving-average component, then the VAR model might not provide as good an approximation as has been achieved by the conversions illustrated above.

4. Observations

The equivalence between different representations is at a very practical level. (The theoretical aspects will be given more substance in the next section.) A model can be converted to another representation and any loss in precision is minor compared to the errors involved in a statistical estimation with a finite sample. Certainly statistical techniques cannot distinguish equivalent representations, neither in theory nor in practice. This ability to convert between representations means that one can convert from one representation to another for the purpose at hand.

The equivalence also has an implication for the types of statistical inference one should attempt with time-series data, and the types of inferences which should be avoided. In order to be a valid statistical inference, a conclusion must be implied by all representations in an equivalence class: otherwise contradictory results would be implied by the same data, depending on the choice of representation. If a conclusion is not invariant within the equivalence class, then it is implied by the theory which suggested the representation, and is not a statistical inference. (It is not a function of the data.)[18]

Thus it is very useful to identify equivalence class invariants in order to help recognize valid inferences. Anything related to the input-output behaviour will be an invariant of the equivalence class. Thus the predicted outputs, the residuals, and any statistics based on the residuals, such as the likelihood and the auto-covariances, are clearly invariants. The Markov parameters and the eigenvalues of F or roots of the determinant of $A(L)$ are also all invariants though this is somewhat more difficult to see. Inference can be legitimately based on any of these. However, conclusions about the form or structure of the representation are much more difficult to justify statistically. In fact, it would seem those conclusions must almost always be based on economic theory.

Granger causality is an inference about the structure of a representation. As such, it is not invariant within an equivalence class.[19] As Granger (1969, 427) points out, some

---

18. This is a slight overstatement. It is possible for a result to be a function of both the theory and the data simultaneously. This is perhaps the proper interpretation of common practice. The difficulty is in interpreting what inferences should be attributed to the data and what inferences should be attributed to the theory. If an inference comes from an invariant of the equivalence class it is undeniably a statistical inference. The logic of the approach suggested here is to extract these inferences first. If one had an irrefutable economic theory, it could be applied first to limit the possible representations and remaining inferences posed as statistical tests (for example, in the manner in which causality testing is frequently practised). However, the conclusions would be the same as applying the statistical tests first and the irrefutable theory later. Doing the statistics first and applying the economic theory later makes the basis for the inference more obvious. Given the difficulty of finding irrefutable economic theories, it would seem preferable to follow the approach suggested here, which has the additional benefit of giving a clear indication of inferences which are not dependent on any economic theory. In any case, arbitrary constraints, such as identifiability constraints, are not irrefutable economic theories.

19. It is worth reiterating that this is not a question of estimation difficulties. Causality tests are commonly recognized to be sensitive to estimation techniques and data normalization. However, the problems described here would occur even if true parameter values were known. Truth can be transformed into a different representation which would have a different causal interpretation.

a priori information is required to provide sufficient constraints to avoid spurious causal inferences.[20] Identifiability constraints typically used by econometricians, such as $A_0 = B_0 = I$ in ARMA models, do not qualify as a priori information. As Basmann (1988) points out, Granger causality tests are being widely misapplied. The reader is referred to Basmann for a complete discourse on the problem. Despite the general problem, inferences about Granger causality do make some sense in the context of a model with a single input and a single output variable. In this context the transformations within an equivalence class do not admit a representation which would contradict the result.

To further illustrate the problems with causal inference, the last state space "gap" model of the previous section is considered. An examination of the state transition matrix F reveals that the first element of the state vector (M1) enters in the calculation of the second element of the state vector (GDP) with a coefficient of -2.37, which suggests that money influences output. The same conclusion follows from the original VAR representation, where the $A(L)$ polynomial has the term $0.0211L^1 + 0.029L^2 - 0.0173L^3$ for the influence of M1 on GDP. However, in the equivalent ARMA model given at the end of the previous section, there is no direct influence of M1 on GDP. In this equivalent ARMA representation, it turns out that the only connection between these two variables is that the same error terms enter both equations. The causal interpretation would be that both variables are influenced by the same disturbances - one does not cause the other. Moreover, the state space model can be further transformed to the following also equivalent state space model where the influence of M1 on GDP is 0.0 and "money does not matter."[21]

$$F = \begin{bmatrix} -0.876 & 2.370 & -0.4972 & -0.1532 & -0.9127 & -0.2928 & 0.784 & -0.4257 & 0.0650 \\ 0.0 & 6.845 & -0.6737 & -2.0712 & -0.9003 & -0.1892 & 1.453 & -0.8144 & 0.1499 \\ 2.63 & 79.030 & -6.4758 & -24.6266 & -9.9216 & -0.0850 & 16.965 & -7.1955 & 1.0672 \\ -1.28 & 0.802 & -0.4633 & 0.3529 & -0.3201 & -0.5558 & 0.707 & -0.4026 & 0.0840 \\ 0.770 & -0.495 & 0.0254 & -0.4694 & -0.2743 & -0.4443 & -0.339 & 0.0375 & -0.0117 \\ -0.179 & -2.249 & 0.1863 & 0.8110 & -0.1066 & 0.3860 & -0.547 & 0.1970 & 0.0171 \\ -0.164 & 0.501 & -0.0445 & -0.1194 & -0.4001 & -0.1782 & 0.590 & 0.4399 & 0.0488 \\ -0.0883 & 0.176 & -0.0183 & -0.0782 & -0.0899 & 0.1758 & -0.380 & -0.0686 & 0.3584 \\ 0.0271 & 0.284 & -0.0251 & -0.1014 & -0.0664 & 0.0574 & 0.130 & -0.3654 & -0.5194 \end{bmatrix}$$

---

20. The context which appears to be envisaged by Granger is one where economic theory imposes some structure on the form of the model representation, providing a natural selection of one representation from each equivalence class. (For example, if economic theory imposed a constraint like $A_0 = B_0 = I$, then this could be used as a starting point to test Granger causality.) Within this theory Granger causality can then be tested using sample data, under the assumption (non-testable) that the economic theory is correct. Unfortunately, it is usually very difficult to recognize the implications of the imposed constraints with respect to the hypothesis being tested.

21. The negative sign on the coefficient (-2.37) in the first "gap" model may not correspond to the way it is usually thought that "money matters." In general, the sign and magnitude of any single parameter can be changed to any value. Some economic theory must be used in order to judge whether it is possible to simultaneously keep other parameters fixed in a meaningful way.

$$G = \begin{bmatrix} -0.00142 \\ 0.0000709 \\ -0.00499 \\ -0.00153 \\ 0.00320 \\ 0.00144 \\ -0.000927 \\ 0.00445 \\ 0.00701 \end{bmatrix} \qquad K = \begin{bmatrix} 0.0582 & 0.06940 & 0.399622 \\ -0.0211 & -0.20436 & -0.003499 \\ -0.2226 & -0.93717 & -1.068027 \\ -0.0803 & -0.37272 & 0.423834 \\ 0.0695 & -0.07740 & 0.011797 \\ 0.1213 & -0.12531 & -0.031670 \\ -0.1263 & -0.01405 & 0.008805 \\ -0.0898 & 0.00418 & 0.005859 \\ -0.0139 & -0.00067 & 0.000645 \end{bmatrix}$$

$$H = \begin{bmatrix} 1.0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & -0.1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This example is not intended to resolve the economic dispute over the impact of monetary policy on output. It is only intended to illustrate how fragile estimation results may be. This will be the case even when the usual tests indicate that individual parameters are extremely significant. The fragility does not come from the data sample. It comes from the fact that equivalent representations may not have the same interpretation.

The general problem of causal inference can be further illustrated by means of the ARMA representation given by the substitution involving use of the Cayley-Hamilton theorem. Any model can be put in this form, illustrated by the final ARMA representation shown in the last section. This has a diagonal $A(L)$ polynomial. Interactions among the endogenous variables[22] are transformed into dependencies on noise terms. By the usual[23] interpretation this suggests that no endogenous variable causes any other endogenous variable. This would be true for any model with any set of variables, so that if the usual procedures are to be believed, it must be that no variable causes any other variable.

---

22. Selection of the input (exogenous) variable is completely arbitrary. A model can always be formulated with all variables as output (endogenous) variables. The terms “output” and “input” used in the engineering literature may be preferable to “endogenous” and “exogenous,” which are widely used in the econometrics literature. A dictionary definition of “exogenous” would suggest that the endogenous variables exert no influence on the exogenous variables. However, that does not appear to be the intended interpretation in most of the econometrics literature and certainly is not intended here. Rather, the distinction is that the model does not try to represent any influence of the output variables on the input variables. Thus, in the example used here, the model does not try to represent any reaction that the monetary authority may have to economic conditions. That does not mean that there is no reaction to economic conditions. Quite the contrary, one raison d’être for the monetary authority is to react to economic conditions such as inflation.

23. Even more usually, causality is tested in a VAR context. In that context a concise MA specification may result in a complicated VAR approximation that would equally lead to spurious causal interpretations.

With respect to another type of structural inference, Aoki and Havenner (1991) have discussed a rearrangement of the state of a model in terms of trend and cycle components, which is a transformation of one representation to another within the equivalence class. With trend arranged in one partition of the state and cycle in the other partition, they illustrate an error-correction model (cycle depends on trend but not vice versa) and a transformation of the state which gives a cointegration model (trend depends on cycles but not vice versa). They point out: "It is immediately clear that any conclusions regarding the separate components depend intimately on the restrictions embedded in [the transformation matrix], on which there is no broad consensus."[24] The fact that these models are in the same equivalence class leads to the conclusion that cointegration versus error correction is a question for economic theory which cannot be resolved by statistical analysis.[25] Another characterization of cointegration corresponds to questions about eigenvalues of a model and thus makes good sense as a statistical question. In this characterization, a model[26] would be called cointegrating of rank r if it contains r unit eigenvalues (r greater than or equal to one and strictly less than p), and the p individual models for the outputs each contain a single unit eigenvalue.[27] The implication of this is that the standard practice of pre-transforming the data by differencing will remove p unit roots when only r should be removed. This can be of considerable importance because the result may be that the joint dynamics are obscured. The reader is referred to Johansen (1988, 1991), Reinsel and Ahn (1992), Granger(1986) and Engle and Granger (1987).

Decomposition of underlying dynamics into cycle and trend (or permanent) components is often important for both theoretical and policy reasons. The non-uniqueness of the state representation, and thus different, statistically equivalent model representations, means that some economic theory or statistics unrelated to the input-output behaviour must be used in order to find a decomposition. Blanchard and Quah (1989) have attempted this by separating permanent and temporary effects based on the impulse response characteristics of the equivalence class of models. They claim these correspond to the economic ideas of permanent and cyclical dynamics, which they equate with demand and supply shocks. The reader is also referred to Cecchetti and Karras (1992), Gali (1992), Shapiro and Watson (1988), and Dea and Ng (1990).

The confusion caused by equivalent models is not new. Sargent (1976) points out that observational equivalence makes it necessary to bring to bear theoretical

---

24. Aoki and Havenner (1991. 47).

25. At least not with input-output data in the context of linear time-invariant models. This would also be the conclusion from the representation theorem in Engle and Granger (1987, 255).

26. Cointegration is usually considered as a question about the process (data) and it is being translated here to a question about the model, since that is the focus of this paper. It is not clear what the relationship between the p individual models and the joint model for the p outputs should be in this translation. In the statistical context this is resolved by using the "best" model in each case.

27. Some authors allow for the possibility that the individual models contain more than one eigenvalue of unit magnitude.

considerations in the debate on the natural rate theory.[28] A methodology suggested by the equivalence reviewed here is to do estimation first without imposing constraints from economic theory. This will be explored at much greater length in some future work. The unconstrained approach will provide a solid basis to test any statistically testable hypothesis, since the conclusion will be invariant within the equivalence class. Economic constraints are then imposed to specify a representative model. This will more clearly disentangle structural implications owing to the imposed economic theory from those which are statistically testable.

5. Equivalence Relationships and the Underlying Parameter Space

The results summarized in this section are scattered through a number of research papers produced over the past twenty years. The intention of this section is to outline briefly the importance of known results, give a flavour of the underlying mathematics and indicate the appropriate reference material. Where possible the intuition of the results is indicated. In most cases the original results treat either the deterministic case (there is an input u but no stochastic noise term) or the stochastic case with no exogenous input u. In either case the results summarized here usually follow directly, simply by some notational change combining the two processes.

When one model can be converted to another input-output equivalent model, then it is not possible to distinguish these models statistically. One is often inclined to think of ARMA or state space model estimation as an estimation over a large vector space with each unfixed entry of the various arrays being a dimension of the vector space. The equivalence relationships complicate this considerably. A simple example of the sort of complication caused by equivalence relationships is to equate the two ends of the real line. The resulting space is a circle. On a circle it is possible to move constantly in the same direction and arrive back at the same point. (That is not possible in a vector space.) A slightly more complicated example can be modelled by twisting a strip of paper (180 degrees) and taping together (equating) the two ends. The result is called a Möbius strip. This piece of paper has only one edge and one side, as can be verified by marking a point near the edge of the paper and tracing a line along the edge and back to the same spot. The point is that equivalence relationships on a set can result in a new set which has very different qualities than the original set. It may no longer be obvious how to define things like straight lines and distance between points. One should not expect the space of sets of equivalent models, that is, the quotient space under the equivalence relationship, necessarily to have any of the nice properties that one is accustomed to having in vector spaces. In fact, the quotient space could be extremely nasty (for example, quotient spaces are not always Hausdorff, so

28. He holds out some hope for statistical hypothesis testing but points out that it must be “empirical work of a kind considerably more subtle than that directed solely at estimating reduced forms.” Based on the equivalence reviewed in this paper, it must in fact be considerably more subtle than simple estimation with input-output data.

it may be impossible to separate different points in the space.) The space resulting from an accounting for the input-output equivalence relationship will be called the underlying parameter space. It is not a vector space and its dimension does not necessarily correspond to the number of unfixed entries in the arrays.[29] It turns out that the underlying parameter space is a smooth manifold[30] (Clark 1976a, 1976b; Hazewinkel 1977), which is about the next best thing to a vector space.

There is another possibility for dealing with the equivalence relationship. One could give pre-eminence to a special form for the representation. The difficulty with anointing one special canonical form is that it can seriously distort the view of the world. A familiar analogy is a two-dimensional map of the earth. Northern Canada and Antarctica look much larger than they are in reality. Even worse, places on the left side of the chart appear very far from places on the right side, when in reality they are quite close. Choosing any single canonical form will cause similar problems. Iterative estimation techniques will have a tendency to iterate toward the edge of the chart (diverge toward infinity) when, in fact, they are trying to move to a point which would be nearby on another chart. For example, consider the function on the surface of the world defined by distance from Ottawa. This function has a minimum at Ottawa. On a two-dimensional chart of the world, with Ottawa at the center, this will look like a well-behaved function which decreases everywhere toward Ottawa. The minimum can be easily found by many conventional optimization algorithms. However, on a chart which was centered elsewhere, say on Rome, there would be large regions where starting a search algorithm would result in iterations toward the edge of the chart. Starting in Tokyo, a steepest descent algorithm would lead toward the Aleutian Islands at the edge of the chart. This motion toward the edge appears as divergence if the chart is mapped onto all of R^2 rather than the closed and bounded representation given by a map of the world. If the problem is recognized, it may be possible to accommodate it, as by jumping to the opposite side of the chart, in this example. This amounts to explicitly recognizing the topology of the underlying space.

Another problem with choosing a particular chart as the canonical representation is that confidence region tests will be seriously distorted. There will be a tendency to count points on the opposite side of the chart as distant, even though they are close in another representation.

Using the quotient space as the underlying parameter space resolves these problems by making points close if they are close in any representation. In particular, if a sequence converges in any chart its image in the quotient space will converge. This means that many results which have been proven using a particular representation (such as

---

29. The topology of this underlying parameter space has been studied extensively. See, for example, Delchamps (1982, 1985), Delchamps and Byrnes (1982), Hazewinkel (1977, 1979), Hazewinkel and Kalman (1975), Kalman (1974), Helmke (1982).

30. A manifold is a space that can be approximated locally as a vector space. The dimension of a manifold is the same as the dimension of the vector space which approximates it locally. The vector space gives a local view and set of co-ordinates, called a chart. A manifold is called trivial if one chart will “work” for the whole manifold. Examples of non-trivial two-dimensional manifolds are the surface of a sphere and a torus (the surface of a donut).

asymptotic convergence of various estimators) will automatically be true on the quotient space. Furthermore, the vector space notions that one is accustomed to using are (usually good) local approximations using any representation. Thus, using the quotient space as the underlying parameter space not only is consistent with the way mathematicians would usually deal with an equivalence relationship, but also has important benefits in the current context.

In practice it is not possible to do calculations on the abstract quotient space, and it is necessary to choose a representation in order to do any computing. It should be kept in mind, however, that any chosen representation is just a representative and another representative could be used. This has at least three important consequences. The first is that it is possible to switch representatives at any time. That has been illustrated in this paper and can be used as a simple method to improve estimation (Gilbert 1993). It also has implications about the types of conclusions which can be drawn from a statistical estimation. The second consequence is that the topology of the parameter space is determined by the quotient space. The illusion of a vector space structure obtained by fixing entries in a chosen representative is only valid as a local approximation. The fact that the parameter space is not a vector space has global implications that cause serious difficulties with iterative estimation techniques. Delchamps (1982, 1985) established that convergence analysis such as developed in Ljung (1978) cannot be applied globally. (See also Helmke, 1982). In Gilbert (1988) it is established that non-global local optima are a generic problem to the parameter space. This is due to the nature of the parameter space and not due solely to the choice of a particular objective function. The third consequence is that some aspects of the geometry of the parameter space are determined by the quotient space. This has implications for hypothesis testing and any statistics based on second moments. These will be explored in future work.

In many situations a practitioner has prior knowledge about the physical or theoretical structure of a system and can choose a representation which reflects this structure. The estimation problem is then to establish certain unknown parameters. It should be noted, however, that choosing a structure and allowing only certain parameters to vary, even in the situation where “truth” is known, does not transform the parameter space into a vector space. The equivalence of different representations, even when they are not considered to be valid representations of truth or knowledge, implies the quotient space structure. This is a biproduct of the mathematical abstraction that must be acknowledged.

The remainder of this section is more technical. It is of importance for completing the description of the equivalence between state space and ARMA models and provides some background for more technically inclined readers, but it is not necessary for an intuitive understanding of the results. Unfortunately, it is impossible to give adequate coverage to the background areas of mathematics (topology, algebra and differential geometry), which are required for a full understanding of the material. It is intended for three different audiences. Readers with a good understanding of one portion of these related ideas may find the links help fill in the pieces of this complicated puzzle. Readers with a background in mathematics may find this to be a useful description of a very important applied problem. (These readers should ignore the obvious shortcomings of the abbreviated definitions given in the footnotes.) Finally, it may serve as an outline for young econometricians who are interested enough to pursue the required background material.

It is necessary to formalize the sets under consideration in order to discuss the problem more precisely. Let $E_n=\{(F,G,H,K)|$ *controllable and/or observable realizations with state dimension n*$\}$ be the total set of controllable and/or observable innovation state space models.[31] An ARMA model is specified by the triple (A,B,C) of polynomial matrices. Let E''_n be the total set of all ARMA models[32] subject to the constraint that the models have McMillan degree n.[33] That is, $E''_n = \{(A,B,C)|$ *(A,B,C) has McMillan degree n*$\}$. It is assumed throughout this paper that the input and output dimensions m and p are fixed. Each of these sets can be reduced to a set of equivalence classes of models. Models within an equivalence class give the same output for a given input (and thus cannot be distinguished on the basis of observed input and output data). The total sets E_n and E''_n are projected onto sets Θ_n and Θ''_n respectively by the equivalence relationship. Statistical techniques can be used to select among points in Θ_n or Θ''_n (among equivalence classes) and thus help support or reject economic theories that require a choice among equivalence classes. Economic theory alone can be used to choose among points within an equivalence class.

The usual identification problem amounts to choosing a set of representative models, one from each equivalence class. Roughly, a canonical form is a selection of one representative from each equivalence class.[34] For state space models, any matrix T in GL_n, the set of invertible nxn matrices, can define a change of basis for the state. This gives an input-output equivalence defined by mapping (F,G,H,K) *to* $(T^{-1}FT, T^{-1}G, HT, T^{-1}K)$. The set of realizations equivalent to (F,G,H,K) can be denoted $[F,G,H,K] = \{(T^{-1}FT, T^{-1}G, HT, T^{-1}K)|\ T \text{ in } GL_n\}$. Let $\Theta_n = \{[F,G,H,K]\}$ denote the set of equivalence classes, that is, Θ_n is the quotient space E_n / GL_n. Similarly let Θ''_n denote the quotient space $E''_n / GL_{p,k}$, where $GL_{p,k}$ is the general linear

31. "Controllable and/or observable" means that it is possible to consider the set of all controllable realizations, the set of all observable realizations, or the set of all realizations that are both controllable and observable. The theory does not hold if the set includes all realizations that are either controllable or observable. In that case the quotient space to be defined shortly is not Hausdorff. In the systems theory literature Kalman (1974) has used algebraic geometry techniques to develop a more general description which may not require this restriction. Unfortunately statistical methods to date are not capable of handling the resulting projective varieties.

32. It seems possible here to consider a larger set of models where the arrays *A*, *B* and *C* in (1) are allowed to be rational polynomial matrices (that is, matrices of the form $Q^{-1}P$, where *P* and *Q* are polynomial matrices). The definition of McMillan degree requires more care for this. Since any model in this larger set has an equivalent representation in the smaller set, the quotient space is unchanged. Thus "forward-looking" models are included without any additional complication. There is no specific consideration of this in the reemainder of this paper.

33. The McMillan degree is an indication of the number of poles in the transfer function representation (including poles at infinity). For ARMA models, when the degree of *A* exceeds the degree of *B* and *C*, this is the degree of the determinant of *A* (assuming that *A* has no left factor with a determinant of non-zero degree that is also a left factor of *B* and *C*).

34. This is sometimes suggested as a definition of the term "canonical form." The difficulty is that in practice most of the forms called canonical forms do not satisfy the definition because of the omission of some equivalence classes at the boundary of the space. Because of this, the term "pseudo-canonical forms" has sometimes been suggested for what have historically been called canonical forms, but the term has not been widely accepted.

group of pxp matrices over the field k of rational polynomials.[35] A particular representative model (realization) produces an output which is a point in the sample space that can be compared to data by some real valued estimation function such as least-squared error or maximum likelihood. Let Y_n be the image of E_n in the sample space. This subset of the sample space is sometimes called the model surface or expectation surface in the literature on statistical manifolds. Of course, any input-output equivalent representation will give the same point on the model surface. The estimation function is a map from E_n or E''_n to the real numbers, which can be factored through Θ_n or Θ''_n and Y_n. Figure 3 gives a diagrammatic summary of the relationships.

FIGURE 3. Summary of relationships

$$
\begin{tikzcd}
\mathbf{E}_n \arrow[rr, leftrightarrow] \arrow[d, "\pi_n"] & & \mathbf{E}''_n \arrow[d, "\pi''_n"] \\
\Theta_n \arrow[rr, leftrightarrow] \arrow[dr] & & \Theta''_n \arrow[dl] \\
& \mathbf{Y}_n \arrow[d] & \\
& \mathbf{R} &
\end{tikzcd}
$$

As suggested in the initial paragraphs of this section, it is important that the underlying parameter spaces Θ_n and Θ''_n have certain nice properties in order to be useful. One nice property is that they should be smooth manifolds. This would mean that they are locally like vector spaces and there is considerable theory that can be applied. The fact that the underlying parameter space is a smooth manifold is extremely important, because it ensures that known statistical techniques can be applied for estimation purposes. Locally, any chart can be used as if it were a vector space, and in fact this has been standard practice. Clarke (1976a, 1976b) and Hazewinkel (1977) have shown that the quotient space Θ_n is a smooth manifold of dimension $n(m+p)$ in the case of non-stochastic systems. There has been considerably less established in the case of Θ''_n for ARMA models, but see, for example, Deistler and Hannan (1981), Deistler (1983a, 1983b) and Hannan and Deistler (1988). To my knowledge it has not been shown that Θ''_n is a smooth manifold, though

35. The extension from polynomials to rational polynomials is primarily for the mathematical convenience of having the quotient space formed by a group action. Other than allowing for the inclusion of forward looking models, which is not explored here, this extension has little practical consequence.

papers by Deistler and Hannan have many related results based on the transfer function representations.[36]

Another nice property is that the underlying parameter spaces Θ_n and Θ''_n "resemble" the image Y_n in the sample space. Otherwise, statistics that are calculated in the sample space do not have an easily translatable meaning in the parameter space. Using the quotient topology on Θ_n and Θ''_n, some additional constraints insure that Θ_n, Θ''_n and Y_n all have the same topology. This is a very basic sense in which the spaces should resemble one another.

The relationship between state space and ARMA models can be stated formally as an equivalence between Θ_n and Θ''_n. Thus Θ_n or Θ''_n can be considered as an abstract representation of the underlying parameter space.

The first level of this equivalence is that these spaces are isomorphic and there is a bijection[37] between them.[38] Guidorzi (1981) has demonstrated the isomorphism by constructing a bijection between Θ and Θ''. Several authors have demonstrated other algorithms for converting between ARMA and state space representations (see, for example, Aoki and Havenner 1991). These correspond to functions between E_n and E''_n. The isomorphism result establishes the fact that it is always possible to make the conversion from one representation to the other and thus there is never a case when it is possible statistically to distinguish the representations on the basis of input-output data using first moment information. With some additional qualifications there will also be bijections from Θ_n to Y_n and from Θ''_n to Y_n.

The second level of the equivalence between Θ_n and Θ''_n would be that these spaces are homeomorphic.[39] This seems quite likely, but to my knowledge this has never been shown. E_n and E''_n are left G-spaces and thus the projections are open maps.[40] Thus the fact that Θ_n and Θ''_n are homeomorphic would follow from establishing homeomorphisms

---

36. They also use a pointwise topology on the space of transfer functions, which comes from the relative topology in the product space of power series coefficients. The reason for this is not intuitively obvious. It appears to be an attempt to avoid the complications caused by the non-existence of bundle sections. This amounts, however, to an arbitrary selection of one pre-eminent canonical form. It thus replaces one obstruction with what would seem a more serious problem, as indicated previously.

37. A bijection is a one-to-one onto function between spaces, also called an isomorphism (in the category of sets and functions). If an isomorphism exists, then the spaces are said to be isomorphic.

38. A similar result can be obtained for transfer function representations, which can also be identified with Hankel matrices of the transfer functions, but this will not be developed here.

39. A continuous function that has a continuous inverse is called a homeomorphism, and the two spaces are homeomorphic. If the function and its inverse are infinitely differentiable, then it is called a diffeomorphism and the spaces are diffeomorphic. (These terms are sometimes used to mean "once differentiable.")

40. A function is open if it takes open sets in its domain space to open sets in its range space. This is the converse of a continuous function. (A function is continuous if the inverse image of every open set is open.) A projection to a space with the quotient topology is automatically continuous by definition of the quotient topology. The fact that the projection of a G-space to the quotient space is an open map is shown in Husemoller (1975, 40) or with somewhat different terminology in Steenrod (1951) or Spanier (1966).

between complete sets of overlapping charts in the totals spaces E_n and E''_n.[41] This approach could also be used to establish that Θ''_n is a smooth manifold. A homeomorphism between the spaces would make any properties based on the first moment carry through from one space to the other. Thus results for ARMA estimation would automatically be true for state space estimation and vice versa. In particular, local results such as the asymptotic convergence of estimators need only be shown for one case, and they would be true for both.

The fact that the projections are open maps also helps guarantee that Θ_n and Θ''_n resemble the image Y_n in the sample space as mentioned above. Specifically, the functions from E_n and E''_n to Y_n, defined by evaluating a particular realization with data, are continuous functions, so the functions from Θ_n and Θ''_n to Y_n are also continuous. To complete the argument and show that Θ_n, Θ''_n and Y_n are homeomorphic requires that the inverse exists and is continuous. This requires some additional sample considerations to prevent the image from being degenerate.

The next level of the equivalence between Θ_n and Θ''_n would be that these spaces are diffeomorphic. This fact may be slightly more difficult to establish, though intuitively it seems likely to be true. A diffeomorphism would give a translation of second moment properties, such as the information matrix, from one space to the other. There is, however, a correction which should be applied to take consideration of the curvature of the manifold. This will be discussed further in future work. Given some additional conditions, it also seems likely that the bijections from Θ_n to Y_n and from Θ''_n to Y_n are diffeomorphisms, but the results here are also incomplete.

Some of the missing results may follow from an even more abstract formulation of the relationship between the total spaces E_n and E''_n and the quotient spaces Θ_n and Θ''_n. This requires an additional level of mathematical complication which can obscure the interpretation for the uninitiated. However, there are two very important advantages. One is that well-studied mathematical problems and known results may apply once the problem is formulated with the appropriate abstraction. The other is that the problem may be a special case of an outstanding mathematical problem for which there is not yet any known solution. In the direction of increasing the level of mathematical abstraction, Hazewinkel

41. The homeomorphisms between charts (sometimes called local homeomorphisms) can be established by examining algorithms for converting between state space and ARMA models. The algorithm for state elimination based on the Cayley-Hamilton theorem (see, for example, Aoki and Havenner 1991) is polynomial with coefficients determined by the characteristic polynomial of F, thus giving a continuous function in one direction (at least in the case of controllable *and* observable systems). In the other direction, the partial fraction expansion illustrated in Aoki and Havenner (1991) is defined by the ratio of a polynomial and the determinant of $A(L)$. This will be continuous given the restriction to systems of McMillan degree n (at least in the case where the McMillan degree is determined by the degree of the determinant of $A(L)$). These are in fact both differentiable functions. The shortcoming here is that a more precise definition of McMillan degree (that is, a more precise indication of the representations considered in the set E'') is required in order to insure the quotient is Hausdorff. Then it is necessary to be specific about whether E includes controllable, observable, or controllable and observable systems.

(1977) has shown that (E_n, π, Θ_n) is a principal GL_n bundle[42] and that the base space Θ_n is a differentiable manifold of dimension $n(m+p)$ in the non-stochastic case. (That would correspond to dimension $n(m+2p)$ in the current context.) To my knowledge it has not been established that $(E''_n, \pi'', \Theta''_n)$ is a principal $GL_{p,k}$ bundle or that Θ''_n is a differentiable manifold, though these seem quite likely.

There has also been considerable interest in questions related to the embedding of the sequence of spaces as n increases. This relates to questions of model order reduction and the approximation of high dimensional systems by lower dimensional systems. Work on this has been done by Hazewinkel (1977, 1979), Delchamps (1982, 1985) and Hannan and Deistler in the papers cited above. Although E''_n is not a subset of E''_{n+1}, it does form part of the boundary of E''_{n+1}, and similarly for E_n and E_{n+1}. This can be seen by showing that any point in E''_n is the limit of a sequence in E''_{n+1} as follows. Given a point in E''_n (an ARMA model of McMillan degree n), multiply the A polynomial by a matrix formed by replacing one diagonal element of a pxp identity matrix with $(1+xL)$. For non-zero, real valued x, this gives an ARMA model of McMillan degree $n+1$ (since the determinant of the product of matrices is the product of determinants). Any sequence of values for x converging to zero gives a sequence in E''_{n+1} converging to the given point in E''_n. The corresponding result for state space representations can be seen most easily by noticing that the set of non-controllable or non-observability realizations with state dimension n, which is eliminated in the definition of E_n, corresponds to deficient rank controllability and observability matrices. Thus the remainder, E_n, is open and dense in the set of all realizations with state dimension n, so its closure contains the lower dimension systems. With some attention to the details necessary to ensure that the quotient spaces are Hausdorff, these sequences should also project to give corresponding results on the quotient spaces. This means that the sequences Θ_n, Θ_{n+1}, ... and Θ''_n, Θ''_{n+1}, ..., corresponding to the abstract underlying parameter spaces, would increase in a way that allows model reduction to be done in a logical manner.

6. Conclusion

A practical demonstration of the equivalence between various state space and ARMA representations has been used to illustrate that it is possible to convert between representations with no practical loss of information. This also illustrates that statistical techniques could not distinguish among the various representations. A résumé of the theo-

---

42. See also Delchamps (1985). Roughly, a bundle might be thought of as a generalization of the vector space notion of orthogonality, though in general there is no inner product involved. The fibers above a point in Θ_n (that is, the set of equivalent points in E_n) are “orthogonal” to Θ_n. (More technically, this should be thought of as a generalization of product spaces.)

A continuous canonical form could be thought of as a continuous map from Θ_n to E_n (that is, a section of this bundle). Hazewinkel and Kalman (1975) have shown that continuous canonical forms in general do not exist. (See also Byrnes and Hurt 1979.) The implication of this is that any canonical form will be good only as a local representation of the underlying parameter space. One manifestation of this problem will be ill-conditioned numerical problems in estimation.

retical aspects indicates that there are still some important points which have not been completely resolved.

The equivalence relationships outlined in this paper and the resulting underlying parameter space give an interesting technical formulation of the complementary aspects of statistics and economic theory: statistics can help to choose among equivalence classes, but economic theory must be used to choose a representative model within the equivalence class.

One important question is: “What is invariant within an equivalence class, and thus statistically testable, and what is not?” Some indications have been given but there may be other important invariants. As a general rule, it is important to show that something that is to be tested statistically is an invariant of the equivalence class. Otherwise, the test is meaningless as a statistical test, since it is implied by some imposed theory rather than being a function of the data. One simple check of researchers’ conclusions is to try to transform their models into equivalent models with contradictory implications. If this can be done, then their conclusions do not come from a statistical analysis of the data, but rather from their assumptions. This is considerably more powerful than redoing the estimation with a different representation or different data, since the different results derived from redoing the estimation only suggest that the results are statistically ambiguous. However, a transformed model with contrary implications shows that the inference is completely unjustified. This approach of finding other equivalent representations with contradictory implications was illustrated by an example that suggests that the question of whether “money matters” is not statistically testable with time-series data. (Some economic theory must be imposed and the conclusions will be a result of this theory.)

The split between statistical estimation and the choice of a representation has a practical implication that will be exploited further in some future work. That is, the representations which are most convenient can be used for the purpose of estimation. The most powerful tools available can be applied to do the estimation, and the representation can be specified later.

Appendix: S Programs

This appendix contains some of the S code for the calculations done in this paper. Only a small part of the code, which may be of interest for understanding the calculations, has been included. The complete code is available on diskette from the author or over Internet from statlib@lib.stat.cmu.edu. The complete code provides a fairly general system for estimation of and conversion among linear time-invariant representations of ARMA and state space models.

```
to.SS.augment <- function # convert by augmentation - state dimension may not be minimal (model) {
# Assumes A[1,,] = B[1,,] = I
... error checking not printed ...
if ( “ARMA” == class(model)[2])
 {A <- model$A
 B <- model$B
 C <- model$C
 a <- dim(A)[1] - 1 # order of polynomial arrays
 if (is.na(a)) a <- 0
 b <- dim(B)[1] - 1
 if (is.na(b)) b <- 0
 cc <- dim(C)[1] - 1
 if (is.na(cc)) cc <- 0
 p <- dim(A)[3] # Dim of endoenous Variables.
 m <- dim(C)[3] # Dim of exogenous Variables.
 if(is.null(m)) m<-0
 if (b>a) return(“ The MA order cannot exceed the AR order to convert with state augmentation.”)
 if (cc>(a-1))
 {cat(“ The order of the input polynomial cannot exceed the AR order -1 to convert with state augmentation./n”)
 return(“ The order of the input polynomial cannot exceed the AR order -1 to convert with state augmentation.”)
 }
 #make three parameters A,B and C have convenient order by adding 0’s.
 k <- 1 + a
# if (b != 0)
 {BB <- array(0,c(k,dim(B)[2:3]))
 BB[1:(b+1),,] <- B
 }
 if (m!=0)
 {CC <- array(0,c(k,dim(C)[2:3]))
 CC[1:(cc+1),,] <- C
 }
 FF <- matrix(NA,a*p,p)
 for (i in 1:a) FF[(1+p*(i-1)):(p*i),] <- -A[a-i+2,,]
 if(a>1) FF<-cbind(rbind(matrix(0,p,(a-1)*p),diag(1,(a-1)*p)),FF)
 if (m == 0) G <-NULL
 else
 {G <- matrix(NA,a*p,m)
 for (i in 1:a) G[(1+p*(i-1)):(p*i),] <- CC[a-i+1,,]
 }
 H <- diag(1,p)
 if(a>1) H <- cbind( matrix(0,p,(p*(a-1))),H)
 K <- matrix(NA,a*p,p)
 for (i in 1:a) K[(1+p*(i-1)):(p*i),] <- -A[a-i+2,,]+BB[a-i+2,,]
 z0 <-NULL
 if(!is.null(model$TREND)) #add a constant state which feeds into the states
 {FF<-rbind(cbind(FF,0),0) # identified with outputs (through H).
 n <-dim(FF)[1]
```

```
 FF[n,n] <-1
 FF[n-p:1,n] <- model$TREND
 z0 <- rep(0,n)
 z0[n] <-1
 H<-cbind(H,0)
 if (m!=0) G <- rbind(G,0)
 K<- rbind(K,0)
 }
 descr<-c(model$description," Converted to state space by state augmentation.")
 model <- list(F=FF,G=G,H=H,K=K,z0=z0,description=descr)
 class(model) <- c("TSmodel","SS","innov")
 set.parameters(model)
 }
}


gmap <- function(g, model)
{# convert to an equivalent representation using a given matrix
... error checking not printed ...
if ( "SS" ==class(model)[2])# transform State space model by g in GL(n)
 {n <- dim(model$F)[1]
 if (!is.matrix(g)) g<-diag(g,n) # if g is not a matrix make it into a diagonal matrix.
 if ((n !=dim(g)[1]) | (n !=dim(g)[2]) )
 stop("g must be a square matrix of dimensions equal the model state (or a scalar).")
 inv.g <- solve(g)
 model$F <-inv.g%*%model$F%*% g
 if (!is.null(model$G)) model$G <-inv.g %*%model$G
 model$H <-model$H %*% g
 if (!is.null(model$z0)) model$z0 <-c(inv.g %*%model$z0)
 if (class(model)[3]=="innov")
 model$K <-inv.g %*% model$K
 else
 {model$Q <-inv.g %*% model$Q
 model$R <-model$R
 }
 }
 if ( "ARMA" ==class(model)[2])
 {if (! is.matrix(g)) g<- diag(g,dim(model$A)[2]) # if g is not a matrix make it into a diagonal matrix.
          for(l in 1:dim(model$A)[1]) model$A[l, , ] <- g %*% model$A[l, ,]
          for(l in 1:dim(model$B)[1]) model$B[l, , ] <- g %*% model$B[l, ,]
          for(l in 1:dim(model$C)[1]) model$C[l, , ] <- g %*% model$C[l, ,]
          if(!is.null(model$TREND)) model$TREND <- g %*% model$TREND
 }
 set.parameters(model)
}
to.ARMA <- function(model)
{ # convert to an ARMA representation by Cayley Hamilton (not very parsimonious)
#ref. Aoki and Havenner, Econometric Reviews v.10,No.1, 1991, p13.
... error checking not printed ...
if ("SS" ==class(model)[2]) # SS model
 {if (class(model)[3]=="non-innov") model <- to.SS.innov(model)
 FF<-model$F
 G <-model$G
 H <-model$H
 K <-model$K
 m <-dim(G)[2]
 if (is.null(m)) m <-0
 n <-dim(FF)[1]
```

```
 p <-dim(H)[1]
 poly <- - characteristic.poly(FF) # N.B. sign change in Aoki vs Kailath
 A <- array(0,c(1+length(poly),p,p))
 A[1,,] <-diag(1,p)
 for (i in 1:length(poly)) A[i+1,,] <- diag(-poly[i],p)
 Fn <- array(0,c(n-1,n,n))
 Fn[1,,] <- FF
 if (n > 2) for (i in 2:(n-1)) Fn[i,,] <- FF %*% Fn[i-1,,]
 HFnK <- array(0,c(n+1,p,p))
 HFnK[1,,] <- diag(1,p)
 HFnK[2,,] <- H %*% K
 for (i in 3:(n+1)) HFnK[i,,] <- H %*% Fn[i-2,,] %*% K
 B <- array(0,c(1+length(poly),p,p))
 B[1,,] <-diag(1,p)
 for (i in 1:length(poly))
 {B[i+1,,] <- HFnK[i+1,,]
 for (j in 1:i) B[i+1,,] <- B[i+1,,]-poly[j]*HFnK[i+1-j,,]
 }
 if (m == 0) C <- NULL
 else
 {C <- array(0,c(1+length(poly),p,m))
 HFnG <- array(0,c(n+1,p,m))
 HFnG[1,,] <- H %*% G
 for (i in 2:n) HFnG[i,,] <- H %*% Fn[i-1,,] %*% G
 C[1,,] <- HFnG[1,,]
 for (i in 2:length(poly))
 {C[i,,] <- HFnG[i,,]
 for (j in 1:(i-1)) C[i,,] <- C[i,,]-poly[j]*HFnG[i-j,,]
 } }
 newmodel <- list(A=A,B=B,C=C)
 class(newmodel) <- c("TSmodel","ARMA")
 newmodel <- set.parameters(newmodel)
 }
 else
 {newmodel <- model
 }
 newmodel
}


KF <- function(model, data, sampleT=NULL, predictT=NULL,
 return.state=F, return.track=F, result=NULL, fortran=T)
{# ref. B.D.O.Anderson & J.B.Moore "Optimal Filtering" p.44.
# sampleT is the length of data which should be used for estimation. y must be at least as
# long as sampleT. If predictT is large than sampleT then the model is simulated to
# predictT. y is used if it is long enough. u must be at least as long as predictT.
# The default result=0 returns a list of all the results. Otherwise only the
# indicated list element is return (eg. result=1 return the likelihood and
# result=3 returns the one step ahead predictions.
# Calculate the state, residuals, and likelihood value for the model:
#
# z(t) = Fz(t-1) + Gu(t) + Qe(t-1)
# y(t) = Hz(t) + Rw(t)
#
# or the innovations model:
# z(t) = Fz(t-1) + Gu(t) + Kw(t-1)
# y(t) = Hz(t) + w(t)
#
```

```
# FF (nxn) is the state transition matrix F.
# H (pxn)is the output matrix H.
# Q (nxn) is the input matrix of the system noise and the noise is assumed to be white. Some authors (eg. Harvey) modify
      this as rt*qt*rt' where rt is the matrix for the system noise and qt is the noise cov, but that is redundant.
# R (pxp) is the input matrix of the output (measurement) noise, which is assumed white. probably need R if p>n ??
# G (nxp)is the control (input) matrix.
# K (nxp)is the Kalman gain.
# y is the p dimensional output data.
# u is the m dimensional exogenous (input) data.
# z is the n dimensional (estimated) state at time t, E[z(t)|y(t-1), u(t)] denoted E[z(t)|t-1].
# state is the history of the state.
# Om is the estimated output cov matrix.
# vt is the prediction error.
# pred is the history of the one-step ahead predictions, E[y(t)|y(t-1),u(t)] denoted E[y(t)|t-1].
# The history of the prediction error is given by y-pred[1:predictT,]or y-pred[1:sampleT,]
# P is the estimate of the state tracking error matrix at each
# period. Cov{z(t)-E[z(t)|t-1]}
# trackError is the history of P.
# Tracking error pt can only be calculated if Q and R are provided ( gain FALSE).
# Using the Kalman gain K directly these are not necessary
# for the likelihood calculation,
# but the tracking error cannot be calculated.

... errror checking not printed ...

gain <- "innov" == class(model)[3]
if (gain & return.track)
 cat("Tracking error is zero for an innovations model. track will not be calculated.")

FF<- model$F
H <- model$H
n <- dim(FF)[2]
p <- dim(H)[1]
if (is.null(model$G))
 {m<-0
 G<-matrix(0,n,1) # can't call fortran with 0 length arrays
 u <- matrix(0,predictT,1)
 }
else
 {m <- dim(model$G)[2]
 G <-model$G
 u <- data$input
 }
if (gain) # K or Q,R can be NUll in model, which messes up fortran
 {K <- model$K
 Q <- matrix(0,1,1) #not used
 R <- matrix(0,1,1) #not used
 track <-array(0,c(1,1,1)) #not used
 }
else
 {cat("Non - innovations models have not been thoroughly tested. ")
 Q <- model$Q
 R <- model$R
 K <- matrix(0,n,p) # this is used
 if(return.track) track <-array(0,c(predictT,n,n))
 else track <-array(0,c(1,1,1)) #not used
 }
if (return.state) state <- matrix(0,predictT,n)
```

```
else state <- matrix(0,1,1) #not used
if(is.null(model$z0)) z <-rep(0,n) # initial state
else z <-model$z0

if (fortran)
 { }
else # S version
 {y <- data$output
 vt <- rep(0,p) # initial prediction error
 pred <- matrix(0,predictT,p)
 if ( ! gain )
 {P <- diag(1,n) # initial tracking error p0
 RR <- R %*% t(R)
 QQ <- Q %*% t(Q)
 # K <- matrix(0,n,p) # this might be initialized better.
 K<-Q%*% rbind(solve(R), matrix(0, (n - p), p))
 }

 for (Time in 1:sampleT) {
 if ( ! gain)
 {P <- (FF %*% P %*% t(FF) ) - ( K %*% H %*% P %*% t(FF) ) + QQ
 P <- (P + t(P))/2 # force symmetry (eliminate rounding error problems)
 if (return.track) track[Time,,] <- P
 PH <- P %*% t(H)
 ft <- ( H %*% PH ) + RR
 ft <- (ft + t(ft))/2 # force ft to be symmetric
 # ftinv <- solve(ft) # if possible solve rather than invert
 # K <- FF%*% PH %*% ftinv
 K <- t(solve(ft,t(FF %*% PH))) #if ftinv is not needed solve this instead
 }

 z<- c(FF%*%z) + c(K%*%vt) # E[z(t)| t-1 ]
 if (m !=0) z<- z + c(G%*%u[Time,])
 if (return.state) state[Time,]<- z
 pred[Time,] <- Ey <- c(H %*% z) # predicted output
 vt<- y[Time,] - Ey # prediction error
 }
 prederror <- y[1:sampleT,,drop=F]-pred[1:sampleT,,drop=F]
 like <- L(prederror)
 Om <-t(prederror) %*% prederror /sampleT

 # now simulate to predictT - requires u but not y (y is ignored if it is supplied)
 if (predictT > sampleT)
 {cat("Simulation has not been tested. \n")
 for (Time in (sampleT+1):predictT)
 {z <- c(FF%*% z)
 if (m !=0) z <- z + c(G%*%u[Time,])
 if (Time==sampleT+1) z <- z + c(K%*%vt)
 if (return.state) state[Time,] <- z
 pred[Time,] <- c(H %*% z) # predicted output
 }
 }
 r<- list(like=like,cov=Om,pred=pred,state=state,
 track=track,sampleT=sampleT, predictT=predictT)
 } # end of S version

 if ( is.null(result))
 {if (gain|(!return.track)) r$track <- NULL
```

```
 if (!return.state) r$state <- NULL
 r$pred <- ts(r$pred, start=start(data$output), frequency=frequency(data$output), names=dimnames(data$output)[[2]])
 r <-(list(estimates=r, data=data, model=model) )
 class(r) <- “TSestModel”
 return(r)
 }
 else
 {if (result ==”like”)
 {v <- svd(r$cov)$d
 # eigenvalues are not robust to degenerate density.
return((r$like[2]*sum(v!=0)/p+ 0.5*sampleT*log(prod(v[v!=0]))+r$like[4]))
 }
 else { return(r[[result]]) }
 }
}


ARMA <- function(model,data, sampleT=NULL, predictT=NULL,result=NULL, fortran=T)
{# calculate likelihood, residuals, prediction, etc. for ARMA model
 # N.B. The .f version is much preferred for speed.
 # sampleT is the length of data which should be used for estimation.
 # Calculate the one-step ahead predictions, and likelihood value for the model:
#
# A(L)y(t) = B(L)w(t) + C(L)u(t) + TREND
#
# A(L) (axpxp) is the auto-regressive polynomial array.
# B(L) (bxpxp) is the moving-average polynomial array.
# C(L) (cxpxm) is the input polynomial array.
# TREND is a constant vector added at each period.
# y is the p dimensional output data.
# u is the m dimensional control (input) data.
# Om is the estimated output cov matrix.

... error checking not printed...
u <- data$input
y <- data$output
A<- model$A
B <- model$B
C <- model$C
TREND <- model$TREND
m <- dim(C)[3]
if (is.null(m)) m <-0
p <- dim(A)[2]
a <-dim(A)[1]
b <-dim(B)[1]
if (fortran)
 { }
else # start S version
 {prederror <- matrix(0,predictT,p)
 invB0 <- solve(B[1,,])
 for (l in 1:a) A[l,,] <- invB0 %*% A[l,,] # set B(0) = I
 for (l in 1:b) B[l,,] <- invB0 %*% B[l,,]
 if (m!=0) for (l in 1:dim(C)[1]) C[l,,] <- invB0 %*% C[l,,]
 if(!is.null(TREND)) TREND <- invB0 %*% TREND
 #browser()
 for (Time in 1:sampleT)
 {if(!is.null(TREND)) vt <- -TREND
 else vt <- rep(0,p)
```

```
 for (l in 1:a)
 if (l<=Time) # this is cumbersome but drop=F leaves A,B,C as 3 dim. arrays.
 if (p==1) vt <- vt + c(A[l,,] * y[Time+1-l,])
 else vt <- vt + c(A[l,,] %*% y[Time+1-l,])
 if (b >= 2) for (l in 2:b)
 if (l<=Time)
 if (p==1) vt <- vt - c(B[l,,] * prederror[Time+1-l,])
 else vt <- vt - c(B[l,,] %*% prederror[Time+1-l,])
 if (m!=0) for (l in 1:dim(C)[1])
 if (l<=Time)
 if (m==1) vt <- vt - c(C[l,,] * u[Time+1-l,])
 else vt <- vt - c(C[l,,] %*% u[Time+1-l,])
 prederror[Time,] <- vt
 }
like <- L(prederror)
 Om <-t(prederror) %*% prederror /sampleT
 pred <- matrix(0,predictT,p)
 pred[1:sampleT,] <- y[1:sampleT,,drop=F] - prederror[1:sampleT,,drop=F]

r<-list(like=like,cov=Om,pred=pred,sampleT=sampleT, predictT=predictT)
} # end of S version
r$pred <- ts(r$pred, start=start(data$output), frequency=frequency(data$output), names=dimnames(data$output)[[2]])
if ( is.null(result) )
 {r <-(list(estimates=r, data=data, model=model) )
 class(r) <- “TSestModel”
 return(r)
 }
 else
 {if (result ==”like”)
 {v <- svd(r$cov)$d
 # eigenvalues are not robust to degenerate density.
return((r$like[2]*sum(v!=0)/p+ 0.5*sampleT*log(prod(v[v!=0]))+r$like[4]))
 }
 else { return(r[[result]]) }
 }
} # end of ARMA


information.tests.calculations <- function # return model selection criteria
 (lst, sample.start=1,sample.end=NULL){
 resid <- lst$estimates$pred-lst$data$output
 if (is.null(sample.end)) sample.end <- nrow(resid)
 resid <- resid[sample.start:sample.end,,drop=F]
 ml <- L(resid)[1] # neg. log( likelihood ).
 n <- length(lst$model$parms) # No. of parameters.
 # nt is theorical dimension of parameter space n(m+2p)
 if (class(lst$model)[2] == “ARMA”) nt <- NA
 if (class(lst$model)[2] == “SS”)
 if (is.null(lst$model$G)) nt <- nrow(lst$model$F)*2*nrow(lst$model$H)
 else nt <- nrow(lst$model$F)*(ncol(lst$model$G)+2*nrow(lst$model$H))
 r <- nrow(lst$estimates$pred)*ncol(lst$estimates$pred) #No. of residuals.
 port <-Portmanteau(resid)
 aic <- 2*ml + 2*n # AIC
 bic <- 2*ml + n*log(r) # BIC
 gvc <- 2*ml - 2*r*log(1-n/r) # GCV
 rice<- 2*ml - r*log(1-2*n/r) # RICE
 fpe <- 2*ml + r*(log(1+(n/r))-log(1-(n/r))) # FPE
 taic <- 2*ml + 2*nt # AIC
```

```
 tbic <- 2*ml + nt*log(r) # BIC
 tgvc <- 2*ml - 2*r*log(1-nt/r) # GCV
 trice<- 2*ml - r*log(1-2*nt/r) # RICE
 tfpe <- 2*ml + r*(log(1+(nt/r))-log(1-(nt/r))) # FPE
 z<- matrix(c(port,ml,aic,bic,gvc,rice,fpe,taic,tbic,tgvc,trice,tfpe),1,12)
 dimnames(z)_list(NULL,c("port","like","aic","bic","gvc","rice","fpe","taic","tbic","tgvc","trice","tfpe"))
 z
}
```

References

Aoki, M. 1990. *State Space Modelling of Time Series*. 2d ed. rev. and enl., Heidelberg, NY: Springer-Verlag.

Aoki, M. and A. Havenner. 1991. "State Space Modeling of Multiple Time Series." *Econometric Reviews* 10:1-59.

Basmann, R. L. 1988. "Causality Tests and Observationally Equivalent Representations of Econometric Models." *Journal of Econometrics* 39:69-104.

Blanchard, O. J. and D. Quah. 1989. "The Dynamic Effects of Aggregate Demand and Supply Disturbances." *The American Economic Review* 79:655-73.

Byrnes, C. and N. E. Hurt. 1979. "On the Moduli of Linear Dynamical Systems." In *Studies in Analysis: Advances in Mathematics,* vol. 4, edited by Gian Carlo Rota, 83-122. New York: Academic Press.

Cecchetti, S. G. and G. Karras. 1992. "Sources of Output Fluctuations During the Interwar Period: Further Evidence on the Causes of the Great Depression." Working Paper 4049, National Bureau of Economic Research.

Clark, J. M. C. 1976a. "The Consistent Selection of Local Coordinates in Linear System Identification." Photocopy.

Clark, J. M. C. 1976b. "The Consistent Selection of Parameterizations in System Identification." In *Proceedings of the Joint Automatic Control Conference,* 576-80. New York: American Society of Mechanical Engineers. [The result that the quotient space manifold dimension is n(m+p) in the non-stochastic case is often attributed to this paper but was actually demonstrated in Clarke, 1976a, which was widely circulated but never published. Clarke, 1976b, demonstrated the result that the manifold dimension is 2np for a stochastic system with no input.]

Dea, C. and S. Ng. 1990. "Sources of Business Cycles in Canada." Working Paper 90-4, Bank of Canada.

Deistler, M. 1983a. "The Properties of the Parameterization of ARMAX Systems and Their Relevance for Structural Estimation and Dynamic Specification." *Econometrica* 51:1187-1202.

Deistler, M. 1983b. "The Structure of ARMA Systems in Relation to Estimation." In *Geometry and Identification, Proceedings of APSM Workshop on System Geometry, System Identification, and Parameter Estimation*, Systems Information and Control, vol. 1, edited by P. E. Caines and R. Hermann, 49-61. Brookline, MS: Math Sci Press.

Deistler, M. and E. J. Hannan. 1981. “Some Properties of the Parameterization of ARMA Systems with Unknown Order.” *Journal of Multivariate Analysis* 11:474-84.

Delchamps, D. F. 1982.“The Geometry of Spaces of Linear Systems with an Application to the Identification Problem.” PhD thesis, Division of Applied Sciences, Harvard University.

Delchamps, D. F. 1985. “Global Structure of Families of Multivariable Linear Systems with an Application to Identification.” *Mathematical Systems Theory* 18:329-80.

Delchamps, D. F. and C. I. Byrnes. 1982. “Critical Point Behaviour of Objective Functions Defined on Spaces of Multivariable Systems.” In *Proceedings of the 21st IEEE Conference on Decision and Control*, vol. 2, 937-43. Orlando, FL: IEEE.

Engle, R. F. and C. W. J. Granger. 1987. “Co-integration and Error Correction: Representation, Estimation, and Testing.” *Econometrica* 55:251-76.

Gali, J. 1992. “How Well Does the IS-LM Model Fit Postwar U.S. Data?” *The Quarterly Journal of Economics* 107:709-38.

Gilbert, P. D. 1988. “Well Behaved Objective Functions and the Estimation of Discrete Time Linear Dynamic Systems.” PhD diss., University of Virginia.

Gilbert, P. D. 1993. “Estimation and Reduction of State Space Models: An Improvement on VAR Methodology.” Forthcoming.

Granger, C. W. J. 1969. “Investigating Causal Relations By Econometric Models and Cross-Spectral Methods.” *Econometrica* 37:424-38.

Granger, C. W. J. 1986. “Developments in the Study of Cointegrated Economic Variables.” *Oxford Bulletin of Economics and Statistics* 48:213-28.

Guidorzi, R. P. 1981. “Invariants and Canonical Forms for Systems Structural and Parametric Identification.” *Automatica* 17:117-33.

Hannan, E. J. and M. Deistler. 1988. *The Statistical Theory of Linear Systems*. New York: John Wiley and Son.

Hazewinkel, M. 1977. “Moduli and Canonical Forms for Linear Dynamical Systems II: The Topological Case.” *Mathematical Systems Theory* 10:363-85.

Hazewinkel, M. 1979. “On Identification and the Geometry of the Space of Linear Systems.” In *Proceedings* of the Bonn Workshop on *Stochastic Control Theory and Stochastic Differential Systems*, Lecture Notes in Control and Information Sciences, vol. 16, edited by M. Kohlmann and W. Vogel, 401-15. New York: Springer-Verlag.

Hazewinkel, M. and R. E. Kalman. 1975. “On Invariants, Canonical Forms and Moduli for Linear, Constant, Finite Dimensional, Dynamical Systems.” In *Proceedings of the International Symposium on Algebraic Systems Theory,* Lecture Notes in Economics and Mathematical Systems, vol. 131, 48-60. Udine, Italy: Springer-Verlag.

Heij, C., T. Kloek and A. Lucas. 1992. “Positivity Conditions for Stochastic State Space Modelling of Time Series.” *Econometric Reviews* 11:379-96.

Helmke, U. 1982. “The Topology of the Space of Linear Systems.” In *Proceedings of the 21st IEEE Conference on Decision and Control*, vol. 2, 948-49. Orlando, FL: IEEE.

Husemoller, D. [1975]. *Fibre Bundles*. 2nd ed. New York: Springer-Verlag.

Johansen, S. 1988. “Statistical Analysis of Cointegration Vectors.” *Journal of Economic Dynamics and Control* 12:231-54.

Johansen, S. 1991. “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models.” *Econometrica* 59:1551-80.

Kalman, R. E. 1974. “Algebraic Geometric Description of the Class of Linear Systems of Constant Dimension.” In *8th Annual Princeton Conference on Information Sciences and Systems,* 189-91. Princeton, NJ: Princeton University.

Ljung, L. 1978. “Convergence Analysis of Parametric Identification Methods.” *IEEE Transactions on Automatic Control* AC 23:770.

Mittnik, S. 1991. “State Space Modeling of Multiple Time Series: A Comment.” *Econometric Reviews* 10:75-90.

Moore, B. C. 1981.“Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction.” *IEEE Transactions on Automatic Control* AC 26:17-32.

Pernebo, L. and L. M. Silverman. 1982. “Model Reduction via Balanced State Space Representations.” *IEEE Transactions on Automatic Control* AC 27:382-87.

Reinsel, G. C. and S. K. Ahn. 1992. “Vector Autoregressive Models with Unit Roots and Reduced Rank Structure: Estimation, Likelihood Ratio Test, and Forecasting.” *Journal of Time Series Analysis* 13:353-75.

Sargent, T. J. 1976. “The Observational Equivalence of Natural and Unnatural Rate Theories of Macroeconomics.” *Journal of Political Economy* 84:631-40.

Shapiro, M. D. and M. W. Watson. 1988. “Sources of Business Cycle Fluctuations.” In *NBER Macroeconomics Annual*, 111-56. Cambridge: National Bureau of Economic Research.

Spanier, E. H. 1966. *Algebraic Topology.* New York: McGraw-Hill (reprinted by Springer-Verlag).

Steenrod, N. 1951. *The Topology of Fibre Bundles*. Princeton, NJ: Princeton University Press.

Vaccaro, R. J. and T. Vukina. 1993. “A Solution to the Positivity Problem in the State-Space Approach to Modeling Vector-Valued Time Series.” *Journal of Economic Dynamics and Control* 17:401-21.

van Overbeek, A. J. M. and L. Ljung. 1982. “Online Structure Selection for Multivariable State Space Models.” *Automatica* 18:529-43.