

From general State-Space to VARMAX models

José Casals* Alfredo García-Hiernaux[†] Miguel Jerez[‡]

October 27, 2010

Abstract

Fixed coefficients State-Space and VARMAX models are equivalent, meaning that they are able to represent the same linear dynamics, being indistinguishable in terms of overall fit. However, each representation can be specifically adequate for certain uses, so it is relevant to be able to choose between them. To this end, we propose two algorithms to go from general State-Space models to VARMAX forms. The first one computes the coefficients of a standard VARMAX model under some assumptions while the second, which is more general, returns the coefficients of a VARMAX echelon. These procedures supplement the results already available in the literature allowing one to obtain the State-Space model matrices corresponding to any VARMAX. The paper also discusses some applications of these procedures by solving several theoretical and practical problems.

Keywords: State-Space, VARMAX models, Canonical forms, echelon

*Departamento de Fundamentos del Análisis Económico II. Facultad de Ciencias Económicas. Campus de Somosaguas. 28223 Madrid (SPAIN). Email: jcasalsc@cajamadrid.es

[†]Corresponding author. Departamento de Fundamentos del Análisis Económico II. Facultad de Ciencias Económicas. Campus de Somosaguas, 28223 Madrid (SPAIN). Email: agarciah@ccee.ucm.es, tel: (+34) 91 394 25 11, fax: (+34) 91 394 25 91.

[‡]Departamento de Fundamentos del Análisis Económico II. Facultad de Ciencias Económicas. Campus de Somosaguas. 28223 Madrid (SPAIN). Email: mjerez@ccee.ucm.es

1 Introduction

This paper describes two algorithms to compute the coefficients of a VARMAX model, in its *standard* (Quenouille, 1957) or *echelon* (Hannan and Deistler, 1988) form, from a general fixed coefficients State-Space model. To avoid cumbersome wordings, from now on we will refer to these models as “VARMAX” and “SS”, respectively.

The relationship between VARMAX and SS representations goes in both directions. First, it is well known (Aoki, 1990) that any fixed-coefficient linear stochastic process can be written in an equivalent SS form. Second, a not so-well known result states that a specific canonical SS model can be written in canonical VARMAX form, see Dickinson et al. (1974). In this paper we present two structured algorithms to obtain the coefficients of an invertible VARMAX model corresponding to a general SS structure, which extends and refines in several ways the results of Dickinson et al. (1974). The first procedure is simpler, but requires two conditions that can be summarized in the idea that every component of the endogenous variable must have the same dynamic order, *i.e.*, their, so-called, observability or Kronecker indices must be equal. The second method is more complex, but does not constrain the model dynamics. Further, we illustrate how these procedures help in model building.

The structure of the paper is as follows. Section 2 presents the different VARMAX and SS representations that will be used in the rest of the article and summarizes some previous results. Section 3 describes the general structure of the algorithms proposed and provides all the details required to implement them in practice. Section 4 discusses some practical applications including examples with real data. The utility and implications of these procedures are pointed out in Section 5, which also indicates how to obtain a free MATLAB toolbox that implements them.

2 Preliminaries

2.1 VARMAX models

Much work in applied time series analysis is based on the linear dynamic model:

$$\bar{\mathbf{F}}(B)\mathbf{z}_t = \bar{\mathbf{G}}(B)\mathbf{u}_t + \bar{\mathbf{L}}(B)\mathbf{a}_t \quad (1)$$

where $\mathbf{z}_t \in \mathbb{R}^m$ is an observable output, $\mathbf{u}_t \in \mathbb{R}^r$ is an observable input, $\mathbf{a}_t \in \mathbb{R}^m$ is an innovation such that $\mathbf{a}_t \sim iid(\mathbf{0}, \boldsymbol{\Sigma}_a)$. Finally, B denotes the backshift operator, such that for any ω_t : $B^i\omega_t = \omega_{t-i}$, $i = 0, \pm 1, \pm 2, \dots, I$ and:

$$\bar{\mathbf{F}}(B) = \sum_{j=0}^p \bar{\mathbf{F}}_j B^j, \quad \bar{\mathbf{G}}(B) = \sum_{j=0}^s \bar{\mathbf{G}}_j B^j, \quad \bar{\mathbf{L}}(B) = \sum_{j=0}^q \bar{\mathbf{L}}_j B^j$$

An important characteristic of model (1) is the maximum dynamic order, defined as $p_{\max} = \max\{p, s, q\}$, which will be used throughout the paper. Model (1) is assumed to be left coprime but the roots of $\bar{\mathbf{F}}(B)$ and $\bar{\mathbf{L}}(B)$ are allowed to be greater or equal to unity. Note that even left coprimeness does not imply that the model is identified as there are still infinite parameter sets that realize \mathbf{z}_t . To achieve identification, one must impose some additional constraints over $\bar{\mathbf{F}}_0$ and $\bar{\mathbf{L}}_0$. For instance, $\bar{\mathbf{F}}_0 = \bar{\mathbf{L}}_0 = \mathbf{I}$ yields the *standard* VARMAX representation introduced, without exogenous inputs, by Quenouille (1957).

On the other hand, an interesting alternative, known as the VARMAX *echelon* form, is widely used. The system (1) is in echelon form if the triple $[\bar{\mathbf{F}}(B) : \bar{\mathbf{G}}(B) : \bar{\mathbf{L}}(B)]$ is in echelon canonical form, *i.e.*, denoting $\bar{F}_{kl}(B)$, the kl -th element of $\bar{\mathbf{F}}(B)$ and similarly $\bar{G}_{kl}(B)$ for $\bar{\mathbf{G}}(B)$ and $\bar{L}_{kl}(B)$ for $\bar{\mathbf{L}}(B)$, the polynomial operators may

be uniquely defined by:

$$\bar{F}_{kk}(B) = 1 + \sum_{i=1}^{p_k} \bar{F}_{kk}(i)B^i, \text{ for } k = 1, \dots, m. \quad (2a)$$

$$\bar{F}_{kl}(B) = \sum_{i=p_k-p_{kl}+1}^{p_k} \bar{F}_{kl}(i)B^i, \text{ for } k \neq l. \quad (2b)$$

$$\bar{G}_{kl}(B) = \sum_{i=0}^{p_k} \bar{G}_{kl}(i)B^i, \text{ for } k = 1, \dots, m. \quad (2c)$$

$$\bar{L}_{kl}(B) = \sum_{i=0}^{p_k} \bar{L}_{kl}(i)B^i, \text{ with } \bar{L}_{kl}(0) = \bar{F}_{kl}(0) \text{ for } k, l = 1, \dots, m. \quad (2d)$$

The integers p_k , $k = 1, \dots, m$ are called the Kronecker or observability indices and they determine the structure of ones/zeros in the echelon form. Equation (2b) uses the index p_{kl} defined as,

$$p_{kl} = \left\{ \begin{array}{l} \min(p_k + 1, p_l) \text{ for } k \geq l \\ \min(p_k, p_l) \text{ for } k < l \end{array} \right\} \quad k, l = 1, 2, \dots, m. \quad (3)$$

As an illustration, consider the standard restricted VARMA(2,2) model:

$$(\mathbf{I} + \bar{\mathbf{F}}_1 B + \bar{\mathbf{F}}_2 B^2) \mathbf{z}_t = (\mathbf{I} + \bar{\mathbf{L}}_1 B + \bar{\mathbf{L}}_2 B^2) \mathbf{a}_t \quad (4)$$

where,

$$\bar{\mathbf{F}}_1 = \begin{pmatrix} -0.70 & 0 & 0 \\ 0.48 & -0.50 & -0.90 \\ -0.02 & 0.30 & -0.20 \end{pmatrix}; \bar{\mathbf{F}}_2 = \begin{pmatrix} 0.30 & -0.20 & 0.50 \\ -0.12 & 0.08 & -0.20 \\ 0.18 & -0.12 & 0.30 \end{pmatrix};$$

$$\bar{\mathbf{L}}_1 = \begin{pmatrix} -0.20 & 0.40 & 0.70 \\ 0.68 & -0.46 & -0.68 \\ 0.18 & 1.24 & -0.38 \end{pmatrix}; \bar{\mathbf{L}}_2 = \begin{pmatrix} 0.30 & 0.50 & -0.80 \\ -0.12 & -0.20 & 0.32 \\ 0.18 & 0.30 & -0.48 \end{pmatrix};$$

Note that this representation has 34 non-zero parameters, excluding those of the covariance matrix, and $p_{\max} = 2$.

Consider now the following VARMA echelon form:

$$(\bar{\mathbf{F}}_0 + \bar{\mathbf{F}}_1 B + \bar{\mathbf{F}}_2 B^2) \mathbf{z}_t = (\bar{\mathbf{L}}_0 + \bar{\mathbf{L}}_1 B + \bar{\mathbf{L}}_2 B^2) \mathbf{a}_t, \quad (5)$$

with,

$$\bar{\mathbf{F}}_0 = \mathbf{Q}_0 = \begin{pmatrix} 1.0 & 0 & 0 \\ 0.4 & 1.0 & 0 \\ -0.6 & 0 & 1.0 \end{pmatrix}; \bar{\mathbf{F}}_1 = \begin{pmatrix} -0.7 & 0 & 0 \\ 0.2 & -0.5 & -0.9 \\ 0.4 & 0.3 & -0.2 \end{pmatrix};$$

$$\bar{\mathbf{F}}_2 = \begin{pmatrix} 0.3 & -0.2 & 0.5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \bar{\mathbf{L}}_1 = \begin{pmatrix} -0.2 & 0.4 & 0.7 \\ 0.6 & -0.3 & -0.4 \\ 0.3 & 1.0 & -0.8 \end{pmatrix}; \bar{\mathbf{L}}_2 = \begin{pmatrix} 0.3 & 0.5 & -0.8 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix};$$

One can find out that systems (4) and (5) are observationally equivalent by pre-multiplying (5) by $\bar{\mathbf{F}}_0^{-1}$. Kronecker indices in model (5) are $p_k = \{2, 1, 1\}$, corresponding to the maximum dynamic order of each component of \mathbf{z}_t . Obviously, p_{\max} must be the same in the standard (4) and echelon (5) representations. Finally, the VARMA echelon: (i) reduces the number of non-zero parameters from 34 to 24 and, (ii) is a canonical form, meaning that there are no alternative representations with the same (or less) number of parameters. These advantages have been pointed out by many authors, among others, Hannan and Deistler (1988), Lütkepohl and Poskitt (1996) or, more recently, Mélard et al. (2006).

2.2 SS models

The relationship between the variables \mathbf{z}_t and \mathbf{u}_t in (1) can also be described by the SS model:

$$\mathbf{x}_{t+1} = \mathbf{\Phi}\mathbf{x}_t + \mathbf{\Gamma}\mathbf{u}_t + \mathbf{E}\mathbf{a}_t \quad (6a)$$

$$\mathbf{z}_t = \mathbf{H}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t + \mathbf{a}_t \quad (6b)$$

where $\mathbf{x}_t \in \mathbb{R}^n$ is a vector of states variables or dynamic components. The innovations \mathbf{a}_t coincide with those in (1), if both representations are adequately normalized. This special SS structure is known as *innovations form*.

We use the innovations form (6a)-(6b) instead of a more common SS model because it has a single error term and, therefore, is closer to VARMAX models. However this representation is equally general, see Hannan and Deistler (1988) for a theoretical discussion and Casals et al. (1999) for a procedure to compute the parameters in (6a)-(6b) from any SS model.

The SS representation (even in innovations form) of a given dynamic system is not unique. To see this, note that for any nonsingular arbitrary matrix \mathbf{T} , applying the equivalence transformation $\mathbf{x}_t^* = \mathbf{T}^{-1}\mathbf{x}_t$, $\mathbf{\Phi}^* = \mathbf{T}^{-1}\mathbf{\Phi}\mathbf{T}$, $\mathbf{\Gamma}^* = \mathbf{T}^{-1}\mathbf{\Gamma}$, $\mathbf{E}^* = \mathbf{T}^{-1}\mathbf{E}$, $\mathbf{H}^* = \mathbf{H}\mathbf{T}$ to any SS form yields an alternative representation for the output.

Any canonical SS representation is characterized by two elements: 1) a certain structure of the transition matrix, *e.g.*, some specific rows or columns must be null or identity sub-matrices and, 2) a unique transformation matrix \mathbf{T} (the only matrix \mathbf{T} which keeps this ones/zeros structure is the identity matrix). The main interest of canonical representations lies in the fact that they realize the system output as a function of a unique parameter set and, therefore, are exactly identified and a must for many applications such as parameter estimation. In the context of this article, we will use the Observable Canonical Form, hereafter OCF, and the

Luenberger Canonical Form, from now on LCF, due to Luenberger (1967).

3 Main results

We propose two different procedures to derive the VARMAX coefficients corresponding to a given SS model. Algorithm #1 requires two conditions while Algorithm #2 is more general but also more complex. Note that they are mutually coherent as they lead to the same VARMAX model when the conditions required by Algorithm #1 hold. The following sub-sections detail how to compute these procedures.

3.1 Algorithm #1: From general SS model to the equivalent standard VARMAX representation

The first algorithm requires two conditions: 1) the system order, n , must be multiple integer of m (from now on C.1) and, 2) the observability matrix, $\mathbf{O}_{p_{\max}}$, for $p_{\max} = n/m$ must have full rank (hereafter C.2). It is straightforward to see that every single-output minimal system fulfills both conditions as, $m = 1$ and, consequently, $\mathbf{O}_{p_{\max}}$ becomes \mathbf{O}_n , which has full rank when the system is minimal. Analogously, every multivariate minimal system whose components $(z_{k,t})$ have identical Kronecker indices (p_k) also fits to C.1 and C.2, as $n = \sum_{k=1}^m p_k$. The algorithm can be computed as follows.

Step 1. Minimality: If the initial SS model is not minimal, reduce it to an equivalent minimal SS realization by applying the staircase algorithm (Rosenbrock, 1970). Note that minimality is a necessary and sufficient condition for the system to be observable and controllable.

Step 2. Innovations form: Transform the model obtained from Step 1) to the corresponding innovations form. Casals et al. (1999) provide an efficient procedure to do it. This transformation has a suitable property: if we choose the strong

solution to the Riccati equation, then the eigenvalues of $(\Phi - \mathbf{E}\mathbf{H})$ will lie in or within the unit circle and there will be no moving average roots outside the unit circle in the resulting VARMAX model.

Step 3. Transformation to the OCF: To do so, we operate to find the transformation matrix, \mathbf{T} , such that $\Phi^* = \mathbf{T}^{-1}\Phi\mathbf{T}$ and $\mathbf{H}^* = \mathbf{H}\mathbf{T}$ present the OCF defined as:

$$\Phi^* = \begin{pmatrix} -\bar{\mathbf{F}}_1 & \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ -\bar{\mathbf{F}}_2 & \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\bar{\mathbf{F}}_{p_{\max}-1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} \\ -\bar{\mathbf{F}}_{p_{\max}} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{pmatrix}, \quad \mathbf{H}^* = [\mathbf{I} \quad \mathbf{0} \quad \dots \quad \mathbf{0}]. \quad (7)$$

Appendix A shows all the details of how to compute the matrix \mathbf{T} .

Step 4. Obtaining polynomial matrices $\bar{\mathbf{G}}(B)$ and $\bar{\mathbf{L}}(B)$:

$$\bar{\mathbf{G}}_0 = \mathbf{D}, \quad \begin{pmatrix} \bar{\mathbf{G}}_1 \\ \bar{\mathbf{G}}_2 \\ \vdots \\ \bar{\mathbf{G}}_{p_{\max}} \end{pmatrix} = \mathbf{T}^{-1}\mathbf{\Gamma} + \begin{pmatrix} \bar{\mathbf{F}}_1\bar{\mathbf{G}}_0 \\ \bar{\mathbf{F}}_2\bar{\mathbf{G}}_0 \\ \vdots \\ \bar{\mathbf{F}}_{p_{\max}}\bar{\mathbf{G}}_0 \end{pmatrix}, \quad \begin{pmatrix} \bar{\mathbf{L}}_1 \\ \bar{\mathbf{L}}_2 \\ \vdots \\ \bar{\mathbf{L}}_{p_{\max}} \end{pmatrix} = \mathbf{T}^{-1}\mathbf{E} + \begin{pmatrix} \bar{\mathbf{F}}_1 \\ \bar{\mathbf{F}}_2 \\ \vdots \\ \bar{\mathbf{F}}_{p_{\max}} \end{pmatrix} \quad (8)$$

3.2 Algorithm #2: From general SS models to the equivalent VARMAX echelon representation

This second algorithm is more general than the previous one, as it does not require any particular condition. The downside is that it is more complex. For example, it requires to identify the Kronecker indices which are directly specified in Algorithm #1. Algorithm #2 can be broken into two stages: 1) obtaining the Luenberger Canonical Form, and 2) deriving the VARMAX echelon coefficients.

Stage 1: Computing the LCF.

Steps 1, enforcing minimality, and 2, obtaining the innovations form, are identical to those in Algorithm #1.

Step 3. Identifying the Kronecker indices in the original SS representation: To this end, consider the observability matrix of a SS model such as (6a-6b):

$$\mathbf{O}_n = \left(\mathbf{H}' \quad (\mathbf{H}\Phi)' \quad (\mathbf{H}\Phi^2)' \quad \dots \quad (\mathbf{H}\Phi^{n-1})' \right)' \quad (9)$$

As the model is minimal, then this matrix has n linearly independent rows. If these rows are chosen in descending order we can build a base which, after re-ordering, can be written as:

$$\mathbf{M} = \left(\mathbf{h}_1 \quad (\mathbf{h}'_1 \Phi)' \quad \dots \quad (\mathbf{h}'_1 \Phi^{p_1-1})' \quad \mathbf{h}_2 \quad \dots \quad (\mathbf{h}'_2 \Phi^{p_2-1})' \quad \dots \quad \mathbf{h}_m \quad \dots \quad (\mathbf{h}'_m \Phi^{p_m-1})' \right)' \quad (10)$$

where \mathbf{h}_k is the k -th row of \mathbf{H} , p_k ($k = 1, \dots, m$) are the Kronecker indices and, therefore, $\sum_{k=1}^m p_k = n$.

Step 4. Transformation to the LCF: Again, this is done through a similar transformation. Appendix A shows all the details of how to compute the transformation matrix \mathbf{T} .

Stage 2: Identifying the VARMAX echelon coefficients.

The LCF is defined by the matrices Φ^* and H^* , being:

$$\Phi^* = \begin{pmatrix} \mathbf{F}_1 & \mathbf{Q}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{F}_2 & \mathbf{0} & \mathbf{Q}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{F}_{p_{\max}-1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{Q}_{p_{\max}-1} \\ \mathbf{F}_{p_{\max}} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{pmatrix} \quad \text{and} \quad H^* = [\mathbf{F}_0 \quad \mathbf{0} \quad \mathbf{0}]. \quad (11)$$

Φ^* is a companion matrix, where each \mathbf{F}_j block ($j = 1, \dots, p_{\max}$) has a number of rows equal to the number of Kronecker indices greater or equal to k , $\bar{m} = \sum_{k=1}^m \min\{p_k, 1\}$ columns and some null elements. In fact, the (k, l) -th element of \mathbf{F}_j will be nonzero only if $j \in [p_k - p_{kl} + 1, p_k]$, where p_{kl} was defined in (3). Each \mathbf{Q}_k block is a zeros/ones matrix, with as many columns as the number of observability indices which are greater or equal to k . If the endogenous variables are sorted according to their corresponding observability indices, the structure of \mathbf{Q}_k will be $\mathbf{Q}_k = [\mathbf{I}_{k+1} \quad \mathbf{0}]'$, where \mathbf{I}_{k+1} is an identity matrix with the same number of rows as \mathbf{F}_{k+1} . With respect to H^* , \mathbf{F}_0 is an $m \times \bar{m}$ matrix, such that the rows corresponding to components with nonzero observability indices can be organized in an $\bar{m} \times \bar{m}$ lower triangular matrix with ones in the main diagonal.

Transforming model (6a)-(6b) to the Luenberger canonical structure yields:

$$\mathbf{x}_{t+1}^* = \Phi^* \mathbf{x}_t^* + \Gamma^* \mathbf{u}_t + \mathbf{E}^* \mathbf{a}_t \quad (12a)$$

$$\mathbf{z}_t = H^* \mathbf{x}_t^* + D \mathbf{u}_t + \mathbf{a}_t \quad (12b)$$

where matrices Φ^* and H^* are as in (11). To write this model in an equivalent polynomial form it is convenient to increase the system dimension up to $m \cdot p_{\max}$ by adding as many non-excited states as needed. Then, the structure of Φ^* will be as in (11) but with: i) the identity matrix instead of \mathbf{Q}_j , and ii) an augmented dimension of matrices \mathbf{F}_j , now $m \times m$. Note that the constraints about potentially nonzero parameters also affect these augmented matrices. Consequently, the new non-excited states require adding null columns to H^* except for the endogenous

variables with a null observability matrix, so that the augmented \mathbf{F}_j is a $m \times m$ lower triangular matrix with ones in its main diagonal. This particular structure allows writing the observation equation as:

$$\mathbf{F}_0^{-1} \mathbf{z}_t = \mathbf{x}_{1:m,t}^* + \mathbf{F}_0^{-1} \mathbf{D} \mathbf{u}_t + \mathbf{F}_0^{-1} \mathbf{a}_t \quad (13)$$

where $\mathbf{x}_{1:m,t}^*$ denotes the first m elements of the state vector \mathbf{x}_t^* . According to (13), one can isolate $\mathbf{x}_{1:m,t}^*$ and substitute them in the state equation. Finally, taking into account the companion structure of $\mathbf{\Phi}^*$, we obtain the coefficients in the VARMAX echelon form (1)-(3) as:

$$\bar{\mathbf{F}}_0 = \mathbf{F}_0^{-1}, \quad \bar{\mathbf{F}}_j = -\mathbf{F}_j \mathbf{F}_0^{-1}; \quad (14a)$$

$$\bar{\mathbf{G}}_0 = \mathbf{F}_0^{-1} \mathbf{D}, \quad \bar{\mathbf{G}}_j = \mathbf{\Gamma}^* + \bar{\mathbf{F}}_j \mathbf{D}; \quad (14b)$$

$$\bar{\mathbf{L}}_0 = \bar{\mathbf{F}}_0, \quad \bar{\mathbf{L}}_j = \mathbf{E}^* + \bar{\mathbf{F}}_j, \quad j = 1, 2, \dots, p_{\max}; \quad (14c)$$

$$\bar{\mathbf{F}}(B) = \sum_{j=0}^{p_{\max}} \bar{\mathbf{F}}_j B^j, \quad \bar{\mathbf{G}}(B) = \sum_{j=0}^{p_{\max}} \bar{\mathbf{G}}_j B^j, \quad \bar{\mathbf{L}}(B) = \sum_{j=0}^{p_{\max}} \bar{\mathbf{L}}_j B^j. \quad (14d)$$

This representation has the characteristic structure of a canonical VARMAX echelon model, see Dickinson et al. (1974).

4 Examples

4.1 VARMAX representation of some common SS specifications

Structural SS models are adequate for many uses, such as displaying the structural components of a time series or dealing with nonstandard samples (*e.g.*, those with missing values, aggregated data or observation errors). On the other hand, VARMAX representations are more adequate for other purposes such as diagnostic checking, if only because they have a single source of errors. To get the best of both worlds one needs then the ability to obtain the VARMAX reduced form corresponding to SS model, bearing in mind that the converse transformation has

been solved by Aoki (1990).

Table 1 illustrates the results of the methods described in Section 3 by showing the ARMAX structures corresponding to some common SS models.

[TABLE 1 SHOULD BE AROUND HERE]

All the SS models in Table 1 are univariate, so these results could have been obtained by other approaches such as, *e.g.*, by identifying the autocorrelation function of the endogenous variable (Harvey, 1989, Chapter 2) or using the pseudo-spectrum implied by the unobserved components and reduced form models (Bujosa et al., 2007).

Table 2 shows that our method can also be applied to multivariate models. Note that, even in the simplest cases, it would be very difficult to obtain the corresponding VARMAX form by the previously mentioned autocorrelation or spectral approaches.

[TABLE 2 SHOULD BE AROUND HERE]

4.2 Identifiability and conditioning

The methods described in Section 3 are also useful to analyze two important (but often ignored) issues: model identifiability and conditioning.

A parametric model is said to be *identifiable* if no two parameter settings yield the same distribution of observations. By definition, canonical VARMAX models are always identified, while there may be infinite SS models realizing the same reduced-form VARMAX. In this case, the SS models would be unidentifiable.

Our methods can be applied to analyze identification of the structural SS model by the following procedure:

Step 1: Compute the response of VARMAX coefficients to small perturbations in the SS parameters. These values would be finite-difference approximations to the corresponding partial derivatives.

Step 2: Organize these derivatives into a Jacobian matrix, \mathbf{J} , with as many columns as the number of free parameters in the SS model (P_{SS}) and as many rows as the number of parameters in the VARMAX form (P_V).

Step 3: Compute the rank of \mathbf{J} , denoted as $\text{Rk}(\mathbf{J})$.

Step 4: Characterize the identifiability of the system as described in Table 3.

[TABLE 3 SHOULD BE AROUND HERE]

An efficient way to perform Step 3 above would consist of computing the singular-value decomposition (SVD) of \mathbf{J} in Step 2. Note that a null singular value corresponds to a linear combination of the VARMAX parameters that is not affected by perturbations on the SS model parameters and, therefore, points out to a specific source of non-identifiability. Accordingly, the rank of \mathbf{J} is the number of non-zero singular values. Moreover, by defining the transformation as the function $P_V = f(P_{SS})$, the Jacobian, \mathbf{J} , can also be used to compute its condition number as $c(f, P_{SS}) = \|\mathbf{J}\| \cdot \|P_{SS}\| / \|P_V\|$, which informs about the robustness of the transformation against numerical errors.

Consider, as an example, the following SS model:

$$x_{t+1} = -\phi x_t + w_t, \quad w_t \sim iid(0, .1) \tag{15}$$

$$z_t = x_t + v_t, \quad v_t \sim iid(0, 1) \tag{16}$$

with $E(w_t, v_t) = 0$, which has an ARMA(1,1) reduced form:

$$(1 + \phi B)z_t = (1 + \theta B)a_t, \quad a_t \sim iid(0, \sigma_a^2) \quad (17)$$

where the AR parameter in (17) coincides with the opposite of the transition scalar in (15). Figure 1 depicts the values of the θ and σ_a^2 parameters associated to different values of ϕ , as well as the smallest singular value of the Jacobian defined above in each case. Observe that when $\phi = 0$ the corresponding singular value is null. In this case the structural model degenerates to the sum of two white noise processes and is, accordingly, unidentifiable.

[FIGURE 1 SHOULD BE AROUND HERE]

4.3 Fitting an errors-in-variables model to Wolf's sunspot series

This example illustrates the use of our procedures to perform the diagnostic checking for a previously estimated SS model. To this end, consider the annual series of Wolf's Sunspot Numbers 1700-1988 taken from Tong (1990). This dataset draws on records compiled by human observers using optical devices of varying quality, so it seems natural to assume that the recorded values are affected by observation errors. On the other hand, many previous analyses have found that this series has a harmonic cycle with an 11 years period. Building on these two ideas, we fitted and estimated by gaussian maximum-likelihood an AR(2) plus white noise errors model to the square root of the original data. The resulting estimates are:

$$\left(\begin{array}{c} 1 - 1.444B + .743B^2 \\ (.048) \quad (.047) \end{array} \right) \hat{z}_t^* = \begin{array}{c} 1.476 + \hat{w}_t, \\ (.145) \end{array} \quad \hat{\sigma}_w^2 = 2.205; \quad (18)$$

$$z_t = \hat{z}_t^* + \hat{v}_t, \quad \hat{\sigma}_v^2 = .147; \quad (19)$$

where z_t and z_t^* are, respectively, the square root of the Wolf number at year t and the underlying "error free" figure. Note that the primary AR(2) structure has complex roots, which implies that the data follows a damped cycle with a period

of 10.87 years.

Using the algorithm described in Section 3, the SS representation of (18)-(19) can be written as the ARMA(2,2) model:

$$\begin{aligned} (1 + 1.444B + .743B^2)z_t &= 1.476 + (1 - .133B + .041B^2)\hat{a}_t, & \sigma_{\hat{a}}^2 &= 2.689, \\ Q(8) &= 10.59, & \ell^*(\hat{\theta}) &= 554.246; \end{aligned} \quad (20)$$

where $Q(8)$ is the portmanteau Q statistic computed with 8 lags and ℓ^* denotes the minus log-likelihood corresponding to the estimates.

Model (20) has six parameters while (18)-(19) has only five. Therefore, the latter is an overidentified structural form. It is immediate to check the empirical consistency of the structural model constraint by estimating freely the parameters in (20):

$$\begin{aligned} (1 + \underset{(.069)}{1.428}B + \underset{(.055)}{.733}B^2)z_t &= \underset{(.181)}{1.509} + (1 - \underset{(.090)}{.112}B + \underset{(.078)}{.064}B^2)\hat{a}_t, \\ \sigma_{\hat{a}}^2 &= 2.688, & \ell^*(\hat{\theta}) &= 554.169; \end{aligned} \quad (21)$$

so models (20) and (21) are almost identical. Their equivalence can be formally assessed by computing an LR statistic which value, .154, confirms that the structural constraint is consistent with the data.

4.4 “Bottom-up” modeling of quarterly US GDP trend

The model-building sequence followed in Section 4.3 can be described as “top-down”, meaning that we first fitted a structural (“top”) model and then obtained the corresponding VARMAX (“bottom”) reduced form. In this example we will show that our methods can also be applied to implement a “bottom-up” modeling strategy.

By “bottom-up” we refer to the situation when one fits a reduced-form VAR-

MAX model to the data and then computes the structural model parameters that realize, exactly or approximately, this reduced form. This approach, originally proposed by Nerlove et al. (1995), is justified if one wants to combine the advantages of a structural SS model with the ability of reduced form models to capture the data sample properties. Note also that this idea has a close relationship with the notion of ARIMA-based time series decomposition, originally suggested by Hillmer and Tiao (1982).

Consider now the quarterly and seasonally adjusted series of US Gross Domestic Product (GDP_t), from 1947 1st quarter to 2008 3rd quarter, in constant 2000 US Dollars. The trend of GDP series is often extracted using the filter proposed by Hodrick and Prescott (1997) which, as it is well known (see, *e.g.*, Harvey and Trimbur, 2008) is equivalent to the smoothed trend obtained from an integrated random-walk trend model:

$$\begin{aligned}
\mu_{t+1} &= \mu_t + \beta_t \\
\beta_{t+1} &= \beta_t + \zeta_t \\
z_t &= \mu_t + \varepsilon_t \\
\begin{pmatrix} \zeta_t \\ \varepsilon_t \end{pmatrix} &\sim iid \left\{ \begin{pmatrix} \zeta_t \\ \varepsilon_t \end{pmatrix}, \begin{pmatrix} \sigma_\zeta^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix} \right\}
\end{aligned} \tag{22}$$

with a signal-to-noise variance ratio such that $\sigma_\zeta^2/\sigma_\varepsilon^2 = 1/1600$.

While the Hodrick-Prescott filter is a simple and effective tool to extract a smooth long-term trend component, it does not capture well the data dynamics. In this case, if we fit model (22) to the series $z_t = \log(GDP_t) \times 100$, maximum-likelihood variance estimates would be $\sigma_\varepsilon^2 = 1.359$ and $\sigma_\zeta^2 = 1.359/1600$. Applying our method to this model yields the reduced form ARIMA model:

$$\begin{aligned}
(1 - B)^2 \log(GDP_t) \times 100 &= (1 - 1.777B + .799B^2)\hat{a}_t, \quad \sigma_a^2 = 1.699 \\
Q(15) &= 239.82; \quad \ell^*(\hat{\theta}) = 174.8
\end{aligned} \tag{23}$$

where the large value of the residual Q-statistic indicates that a strict Hodrick-Prescott specification does not capture all the autocorrelation of this series. Therefore, we may want to adjust a trend model with the dynamic structure of (22) so that it realizes a previously fitted ARIMA model. This modeling strategy can be implemented with the following process:

Step 1. Fit a VARMAX form to the dataset.

Step 2. Compute the SS model parameters that realize more closely the model previously fitted. This requires a non-linear iterative procedure to minimize a given loss function. In this example we specified this loss function as the squared root of the approximation error, computed as the difference between the parameters of: (a) the Step 1 model, and (b) those of the reduced-form corresponding to the SS model.

Note that there are many valid specifications for the loss function employed in Step 2. For example, one could minimize the squared sum of the difference between: (a) the log-likelihood of both models, or (b) the residual series generated by both models. These alternative functions would be particularly useful if the SS model cannot realize exactly the reduced form model.

Table 4 summarizes the results of the bottom-up sequence applied to the GDP data. In Step 1 we fitted an ARIMA model to $z_t = \log(GDP_t) \times 100$. Note that its parameters are very different from those of model (23).

In Step (2.a) we estimated the two variances of an integrated random-walk model by minimizing the loss function defined above and the corresponding reduced-form model. Note that the latter is similar but not identical to the model in Step 1, so an exact equivalence between both models could not be achieved.

On the other hand, comparing the models in Steps (1) and (2.a) it is immediate

to see that the latter is overidentified, as it only has two free parameters. In Step (2.b) we freed the null constraint imposed on the model covariance, to improve the fit between it and the reduced-form. The results indicate clearly that both are now equivalent.

Therefore one can conclude that, without the overidentifying constraints, the dynamic structure underlying the HP filter model could be flexible enough to capture most of the data autocorrelation.

[TABLE 4 SHOULD BE AROUND HERE]

5 Concluding remarks

The method described in this paper has several practical uses and some theoretical implications that can be summarized in the following items.

First, it transforms a structural SS form into an equivalent canonical reduced form, which identifiability is assured. Therefore, it provides the necessary conditions for the SS structure to be identified. Moreover, our method allows one to compute the derivatives of the VARMAX model parameters corresponding to any structural SS specification, providing: i) a natural and easy method to detect identifiability issues, and ii) the condition number of the transformation.

Second, obtaining the VARMAX form corresponding to a given SS specification is useful for diagnostic checking in two specific ways. On one hand, if the SS model is empirically adequate, its reduced form representation should be able to filter the data to white noise residuals. On the other hand, if the structural model is overidentified, unconstrained estimation of the reduced form provides an easy way to test the overidentifying constraints through a LR test.

Third, for some applications (*e.g.*, ARIMA-based seasonal adjustment or time

series disaggregation) one wants to obtain the structural model that more closely realizes a given reduced form. As shown in the example 4.4, our method provides the basic functionality required to do this by computing the numerical solution of a simple optimization problem.

Fourth, the method avoids strictly non-invertible representations of the VAR-MAX model, so the resulting models may be adequate for some specific uses requiring this property such as, *e.g.*, computing forecasts or performing the structural decomposition proposed by Casals et al. (2002).

Last, if a general linear stochastic process can be written either in SS or in VARMAX form, just assuming weak assumptions, this means that both representations are equally general in their ability to represent the data and, therefore, choosing any of these representations is just a matter of convenience.

The procedures described in the paper are implemented in a MATLAB toolbox for time series modeling called E4, which can be downloaded at www.ucm.es/info/icae/e4. The source code for all the functions in the toolbox is freely provided under the terms of the GNU General Public License. This site also includes a complete user manual and other materials.

Acknowledgements

The authors gratefully acknowledge financial support from Ministerio de Educacion y Ciencia, ref. ECO2008-02588/ECON.

References

Aoki, M. (1990). *State Space Modelling of Time Series*. Springer Verlag, New York.

- Bujosa, M., García-Ferrer, A., and Young, P. C. (2007). Linear dynamic harmonic regression. *Computational Statistics and Data Analysis*, 52(2):999–1024.
- Casals, J., Jerez, M., and Sotoca, S. (2002). An exact multivariate model-based structural decomposition. *Journal of the American Statistical Association*, 97(458):553–564.
- Casals, J., Sotoca, S., and Jerez, M. (1999). A fast and stable method to compute the likelihood of time invariant state space models. *Economics Letters*, 65(3):329–337.
- Dickinson, B., Morf, M., and Kailath, T. (1974). Canonical matrix fraction and state space descriptions for deterministic and stochastic linear systems. *IEEE Transactions on Automatic Control*, AC-19:656–667.
- Hannan, E. J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. John Wiley, New York.
- Harvey, A. and Trimbur, T. (2008). Trend estimation and the hodrick-prescott filter. *Journal of the Japan Statistical Society*, 38:41–49.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman Filter*. Cambridge University Press.
- Hillmer, S. and Tiao, G. (1982). An arima-model-based approach to seasonal adjustment. *Journal of the American Statistical Association*, 77:63–70.
- Luenberger, D. G. (1967). Canonical forms for linear multivariate systems. *IEEE Transactions on Automatic Control*, AC-12:290–293.
- Lütkepohl, H. and Poskitt, D. S. (1996). Specification of echelon form VARMA models. *Journal of Business and Economic Statistics*, 14(1):69–79.
- Mélard, G., Roy, R., and Saidi, A. (2006). Exact maximum likelihood estimation of structured or unit root multivariate time series models. *Computational Statistics and Data Analysis*, 50,11:2958–2986.

- Nerlove, M., Grether, D. M., and Carvalho, J. L. (1995). *Analysis of Economic Time Series: A Synthesis*. Academic Press, New York.
- Quenouille, M. H. (1957). *The Analysis of Multiple Time Series*. Griffin, London.
- Rosenbrock, H. H. (1970). *State Space and Multivariable Theory*. John Wiley, New York.
- Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford.

Appendix A

Algorithm #1. Step 3.

Matrix \mathbf{T} can be computed as follows:

1. Given the structure of Φ^* and using partitioned matrices, one can see that $\mathbf{T}_{1:n-m} = \Phi \mathbf{T}_{m+1:n}$, where $\mathbf{T}_{i:j}$ is the matrix made up of the columns $i, i+1, \dots, j$ of \mathbf{T} . From this and C.1, we can write \mathbf{T} as a function of its last m columns:

$$\mathbf{T} = \begin{bmatrix} \Phi^{\frac{n}{m}-1} \mathbf{T}_{n-m+1:n} & \Phi^{\frac{n}{m}-2} \mathbf{T}_{n-m+1:n} & \dots & \Phi \mathbf{T}_{n-m+1:n} & \mathbf{T}_{n-m+1:n} \end{bmatrix} \quad (24)$$

2. Premultiplying \mathbf{T} by matrix \mathbf{H} , we obtain the system of linear equations:

$$\begin{pmatrix} \mathbf{I} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{H} \Phi^{\frac{n}{m}-1} \\ \mathbf{H} \Phi^{\frac{n}{m}-2} \\ \vdots \\ \mathbf{H} \Phi \\ \mathbf{H} \end{pmatrix} \mathbf{T}_{n-m+1:n} \quad (25)$$

The left side of this equation corresponds to \mathbf{H}^* in (7) whereas the right side is \mathbf{HT} . The matrix of coefficients in this system of equations is the observability matrix $\mathbf{O}_{p_{\max}}$, and as C.1 holds, then system (25) has a single unique solution. Further, the product $\mathbf{O}_{p_{\max}} \mathbf{T}$ returns an inferior triangular matrix with ones in its main diagonal, so \mathbf{T} is necessarily nonsingular.

Algorithm #2. Step 4.

Matrix \mathbf{T} can be computed with the following procedure:

1. Invert matrix \mathbf{M} , defined in (10), and select, for each component with a nonzero observability index, the i_k -th column of \mathbf{M}^{-1} , denoted as $\boldsymbol{\mu}_k$, with $i_k = \sum_{l=1}^k p_k$.

2. For each component with a nonzero observability index, build the matrix
$$\mathbf{T}_k = [\Phi^{p_k-1}\boldsymbol{\mu}_k \quad \Phi^{p_k-2}\boldsymbol{\mu}_k \quad \dots \quad \boldsymbol{\mu}_k]'$$
3. Obtain \mathbf{T} by sorting the rows in $\mathbf{T}_k = [\mathbf{t}_{p_k,k} \quad \mathbf{t}_{p_{k-1},k} \quad \mathbf{t}_{1,k}]'$, with $\mathbf{t}_{p_{k-l},k} = \Phi^{p_k-l}\boldsymbol{\mu}_k$, so that $\mathbf{t}_{p_k-h,k}$ precedes $\mathbf{t}_{p_l-i,k}$ if $p_k - h > p_l - i$, or if $k < l$ and $p_k - h = p_l - i$.

Figures and Tables

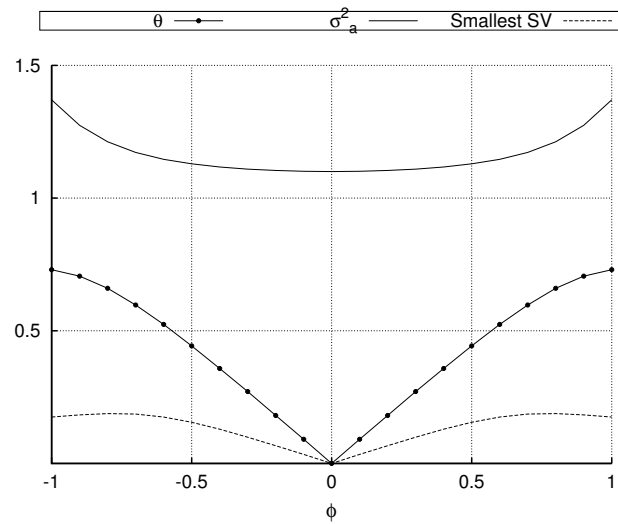


Figure 1: Parameters of the ARMA model (17) and smallest singular value for several values of ϕ in the AR(1)+error model.

Table 1: Univariate examples.

Model	Structural representation	ARIMAX representation
(i) Dynamic regression with errors in variables	$x_{t+1} = .5x_t + .7u_t + w_t, \quad w_t \sim iid(0, 1.5)$ $z_t = x_t + v_t, \quad v_t \sim iid(0, 1)$	$(1 + .5B)z_t = .7u_{t-1} + (1 - .188B)a_t,$ $\sigma_a^2 = 2.660$
(ii) Random walk plus noise model	$\mu_{t+1} = \mu_t + \zeta_t, \quad \zeta_t \sim iid(0, 1/100)$ $z_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim iid(0, 1)$	$(1 - B)z_t = (1 - .905B)a_t,$ $\sigma_a^2 = 1.105$
(iii) Integrated random walk plus noise model	$\mu_{t+1} = \mu_t + \beta_t$ $\beta_{t+1} = \beta_t + \zeta_t, \quad \zeta_t \sim iid(0, 1/100)$ $z_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim iid(0, 1)$	$(1 - B)^2 z_t = (1 - 1.558B + .638B^2)a_t,$ $\sigma_a^2 = 1.567$
(iv) Model (iii) with inputs	$\mu_{t+1} = \mu_t + \beta_t$ $\beta_{t+1} = \beta_t + \zeta_t, \quad \zeta_t \sim iid(0, 1/100)$ $z_t = \mu_t + .5x_t + \varepsilon_t, \quad \varepsilon_t \sim iid(0, 1)$	$(1 - B)^2 z_t = .5(1 - B)^2 x_t (1 - 1.558B + .638B^2)a_t,$ $\sigma_a^2 = 1.567$
(v) Model (iii) with quarterly dummy variable seasonality	$\mu_{t+1} = \mu_t + \beta_t$ $\beta_{t+1} = \beta_t + \zeta_t, \quad \zeta_t \sim iid(0, 1/100)$ $\gamma_{t+1} = \gamma_t - \gamma_{t-1} - \gamma_{t-2} + w_t, \quad w_t \sim iid(0, 1/10)$ $z_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim iid(0, 1)$	$(1 - B)(1 - B^4)z_t =$ $(1 - .714B + .114B^2 - .010B^3 - .563B^4 + .438B^5)a_t,$ $\sigma_a^2 = 2.283$

Note: In all cases the errors affecting the state and observation equations are assumed to be independent.

Table 2: Bivariate examples.

Model	Representations: Structural and ARIMAX
(vi) Seemingly unrelated time series equations with inputs	$\begin{pmatrix} \mu_{t+1}^1 \\ \mu_{t+1}^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_t^1 \\ \mu_t^2 \end{pmatrix} + \begin{pmatrix} \zeta_t^1 \\ \zeta_t^2 \end{pmatrix}, \quad \begin{pmatrix} \zeta_t^1 \\ \zeta_t^2 \end{pmatrix} \sim iid \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1/100 & 5/1000 \\ = & 2/100 \end{pmatrix} \right];$ $\begin{pmatrix} z_t^1 \\ z_t^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_t^1 \\ \mu_t^2 \end{pmatrix} + \begin{pmatrix} .5 \\ .7 \end{pmatrix} x_t + \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{pmatrix} \sim iid \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} .1 & .2 \\ = & .5 \end{pmatrix} \right].$ $\begin{pmatrix} 1-B & 0 \\ 0 & 1-B \end{pmatrix} \begin{pmatrix} z_t^1 \\ z_t^2 \end{pmatrix} = \begin{pmatrix} .5 \\ .7 \end{pmatrix} (1-B)x_t + \begin{pmatrix} 1-.910B & -.012B \\ -.009B & 1-.816B \end{pmatrix} \begin{pmatrix} a_t^1 \\ a_t^2 \end{pmatrix},$ $\begin{pmatrix} a_t^1 \\ a_t^2 \end{pmatrix} \sim iid \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1.104 & .232 \\ = & .610 \end{pmatrix} \right].$
(vii) Two series with a common random walk trend	$\mu_{t+1} = \mu_t + \zeta_t, \quad \zeta_t \sim iid(0, 1/100);$ $\begin{pmatrix} z_t^1 \\ z_t^2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \mu_t + \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{pmatrix} \sim iid \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} .1 & .2 \\ = & .5 \end{pmatrix} \right].$ $\begin{pmatrix} 1-B & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} z_t^1 \\ z_t^2 \end{pmatrix} = \begin{pmatrix} 1-.961B & .104B \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_t^1 \\ a_t^2 \end{pmatrix}, \quad \begin{pmatrix} a_t^1 \\ a_t^2 \end{pmatrix} \sim iid \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1.070 & .270 \\ = & .570 \end{pmatrix} \right].$
(viii) Two series with a common trend and an autonomous random walk component	$\begin{pmatrix} \mu_{t+1}^1 \\ \beta_{t+1} \\ \mu_{t+1}^2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_t^1 \\ \beta_t \\ \mu_t^2 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \zeta_t^1 \\ \zeta_t^2 \end{pmatrix}, \quad \begin{pmatrix} \zeta_t^1 \\ \zeta_t^2 \end{pmatrix} \sim iid \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1/100 & 5/1000 \\ = & 2/100 \end{pmatrix} \right];$ $\begin{pmatrix} z_t^1 \\ z_t^2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \mu_t^1 \\ \beta_t \\ \mu_t^2 \end{pmatrix} + \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{pmatrix} \sim iid \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} .1 & .2 \\ = & .5 \end{pmatrix} \right].$ $\begin{pmatrix} (1-B)^2 & 0 \\ 0 & 1-B \end{pmatrix} \begin{pmatrix} z_t^1 \\ z_t^2 \end{pmatrix} = \begin{pmatrix} 1-1.582B+.650B^2 & .60B-.021B^2 \\ .036B & 1-.777B \end{pmatrix} \begin{pmatrix} a_t^1 \\ a_t^2 \end{pmatrix},$ $\begin{pmatrix} a_t^1 \\ a_t^2 \end{pmatrix} \sim iid \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1.549 & .329 \\ = & .659 \end{pmatrix} \right].$

Note: In all cases the errors affecting the state and observation equations are assumed to be independent.

Table 3: Characterization of the identifiability of the SS model.

$\text{Rk}(\mathbf{J}) = P_{SS}$	
$P_{SS} < P_V$	The SS model parameters are overidentified
$P_{SS} = P_V$	The SS model parameters are exactly identified
$P_{SS} > P_V$	This combination is not possible

P_{SS} and P_V are, respectively, the number of parameters in the SS and VARMAX representation.
 When $\text{Rk}(\mathbf{J}) < P_{SS}$, the SS model parameters are underidentified (or not identified) for $P_{SS} \not\leq P_V$.

Table 4: Results from bottom-up modeling of the US GDP series.

Step	Structural model	ARIMAX model
(1) Fit a reduced-form model to the series	NONE	$(1 - B)^2 \log(GDP_t) \times 100 = (1 - \underset{(.075)}{.661}B + .000^*B^2)\hat{a}_t;$ $\sigma_a^2 = .250, \quad Q(15) = 16.54, \quad \ell^*(\hat{\theta}) = 75.371$
(2.a) Estimate the integrated random walk constraining the covariance to zero	$\text{cov} \begin{pmatrix} \zeta_t \\ \varepsilon_t \end{pmatrix} = \begin{pmatrix} .083 & 0^* \\ = & .044 \end{pmatrix}$	$(1 - B)^2 \log(GDP_t) \times 100 = (1 - .600B + .176B^2)\hat{a}_t;$ $\sigma_a^2 = .250, \quad Q(15) = 22.64, \quad \ell^*(\hat{\theta}) = 77.173$
(2.b) Estimate the integrated random walk freeing the covariance constraint	$\text{cov} \begin{pmatrix} \zeta_t \\ \varepsilon_t \end{pmatrix} = \begin{pmatrix} .029 & -.083 \\ = & .083 \end{pmatrix}$	$(1 - B)^2 \log(GDP_t) \times 100 = (1 - .661B + .000B^2)\hat{a}_t;$ $\sigma_a^2 = .250, \quad Q(15) = 16.54, \quad \ell^*(\hat{\theta}) = 75.371$

Notes: The figure in parenthesis is the standard error of the estimate. The values $Q(15)$ and $\ell^*(\hat{\theta})$ represent, respectively, the Ljung-Box Q statistic computed with 15 lags of the residual autocorrelation function and the minus log (gaussian) likelihood on convergence. The figures marked with an asterisk correspond to constrained parameters.