Decomposition of a State-Space Model with Inputs

José Casals†

Miguel Jerez††

Sonia Sotoca†††

Universidad Complutense de Madrid

Abstract: This paper shows how to compute the in-sample effect of exogenous inputs on the endogenous variables in any linear model written in state-space form. Estimating this component may be, either interesting by itself, or a previous step before decomposing a time series into trend, cycle, seasonal and error components. The practical application and usefulness of this method is illustrated by estimating the effect of advertising on the monthly sales of Lydia Pinkham's vegetable compound.

Keywords: Time series decomposition, Signal extraction, State-space, Lydia Pinkham, Seasonal adjustment

---

† 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: mjerez@ccee.ucm.es, tel: (+34) 91 394 23 61, fax: (+34) 91 394 25 91.

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

1. Introduction.

This paper shows how to compute the in-sample effect of exogenous inputs on the endogenous variables, given a general linear dynamic model. Estimating this component may be, either interesting by itself, or a previous step before decomposing a time series into trend, cycle, seasonal and error components.

About the latter application, many works define the additive structural decomposition of a time series, z_t, as:

$$z_t = t_t + c_t + s_t + \varepsilon_t \tag{1.1}$$

where t_t is the *trend component*, representing the long-term behavior of the series, c_t is the cyclical component, describing autocorrelated transient fluctuations, s_t is the seasonal component, associated with persistent repetitive patterns over seasons, and ε_t is an unpredictable irregular component.

Note that the decomposition (1.1) may be incomplete in some situations, as it does not consider that z_t may be affected by exogenous variables. The effects of such variables can be viewed from three different perspectives:

(1) the effect of the inputs is a nuisance that must be removed from z_t before computing the decomposition or,

(2) these affects constitute a new structural component or, finally,

(3) each component in (1.1) may be generated both, by the model disturbances and the exogenous variables; adopting this point of view we could speak of an "input-driven trend" and an "error-driven trend".

These three interpretations are equally valid and the methods described in this paper can support any one of them.

On the other hand, estimating the effect of exogenous variables may be interesting by itself in different ways. For example if the inputs are control variables, their effect on the output provides a measure of the effectiveness of past controls. Also, if they are leading indicators for z_t, their individual influence would show which ones have larger effect on forecasts. Finally, if they were intervention variables (Box and Tiao, 1975), such decomposition would distinguish between the deterministic-driven part of the time series and the purely stochastic component.

Depending on the model dynamics, this decomposition may be either trivial or really hard to compute. Consider, e.g., the following single-input-single-output (SISO) models:

$$z_t = \hat{\omega}_0 u_t + \hat{a}_t \tag{1.2}$$

$$z_t = \frac{\hat{\omega}_0}{1-\hat{\delta}_1 B} u_t + \hat{a}_t \tag{1.3}$$

$$z_t = \frac{\hat{\omega}_0}{1-\hat{\delta}_1 B} u_t + \frac{\hat{a}_t}{1-\hat{\phi} B} \tag{1.4}$$

where the "hat" symbol "^" denotes that the corresponding term is either a parameter estimate or a residual, $a_t \sim iid(0,\sigma_a^2)$ and B is the backshift operator, such that for any w_t: $B^k w_t = w_{t-k}$, $k = 0, \pm 1, \pm 2, \ldots$

In the static regression (1.2) the contribution of the input, u_t, to the output, z_t, is trivially $\hat{z}_t = \hat{\omega}_0 u_t$ or $\hat{z}_t = z_t - \hat{a}_t$. The latter formula can be also applied to (1.3), while the direct calculation of $\hat{z}_t$ as a function of u_t would require solving a first-order difference equation. Finally, model (1.4) has autocorrelated errors, so computing $\hat{z}_t$ requires solving difference equations, either to calculate $\hat{z}_t$ as a function of u_t, or to compute the autocorrelated residuals as a function of $\hat{a}_t$. On the basis of these simple examples it is easy to see that computing this decomposition “by hand” may be very hard in models with seasonal structure, multiple inputs or, even worst, with multiple inputs and multiple outputs.

In the remaining Sections we discuss the computation of this input-related component in a general linear state-space (SS) framework. Setting the problem in SS has two main advantages. First, the resulting procedures will be able to support any model having an equivalent SS representation, therefore including univariate transfer functions, VARMAX or unobserved components models. Second, the SS literature provides many computational algorithms useful to perform this decomposition.

The structure of the paper is as follows. Section 2 defines the notation and describes the decomposition of a SS model in two sub-systems, one exclusively related to the inputs and another one exclusively related to the errors. Section 3 presents the signal-extraction procedure that we propose to estimate the effect of inputs. After doing so, the remainder component can be further decomposed according to (1.1). Section 4 discusses the case of models with several inputs and provides the results required to estimate the individual effects of each one of them. Section 5 illustrates the application of these methods using the famous Lydia Pinkham monthly series of sales and advertising expenditures. In this framework, the part of sales related to advertising can be interpreted as an estimate of the return of investment (ROI) in advertising. Finally, Section 6 provides some concluding remarks and indicates how to obtain a free MATLAB toolbox which implements the methods described.

2. Model decomposition.

2.1 Basic notation.

Consider the $m \times 1$ random vector $\boldsymbol{z}_t$, which is the output of a steady-state innovations SS model (hereafter "innovations model") defined by:

$$\boldsymbol{x}_{t+1} = \boldsymbol{\Phi}\,\boldsymbol{x}_t + \boldsymbol{\Gamma}\boldsymbol{u}_t + \mathbf{E}\,\boldsymbol{a}_t \tag{2.1}$$

$$\boldsymbol{z}_t = \mathbf{H}\,\boldsymbol{x}_t + \mathbf{D}\,\boldsymbol{u}_t + \boldsymbol{a}_t \tag{2.2}$$

where:

$\boldsymbol{x}_t$ is an $n \times 1$ vector of *state variables* or *dynamic components*,

$\boldsymbol{u}_t$ is an $r \times 1$ vector of exogenous variables or *inputs*, and

$\boldsymbol{a}_t$ is an $m \times 1$ vector of errors, such that $\boldsymbol{a}_t \sim iid\,(\,\mathbf{0}\,, \boldsymbol{Q}\,)$.

The transition equation (2.1) characterizes the system dynamics while the state observer (2.2) describes how $\boldsymbol{z}_t$ is realized as the sum of: (a) a linear combination of the dynamic components, given by $\mathbf{H}\,\boldsymbol{x}_t$, (b) the instantaneous effect of the exogenous variables, given by $\mathbf{D}\,\boldsymbol{u}_t$, and (c) the error $\boldsymbol{a}_t$.

Without loss of generality we will assume that model (2.1)-(2.2) is minimal, meaning that the number of states *n* is the minimum required to realize $\boldsymbol{z}_t$, and that the errors $\boldsymbol{a}_t$ are independent of the exogenous variables $\boldsymbol{u}_t$ and the initial state vector $\boldsymbol{x}_1$.

Many SS models have different errors in the transition and observation equations, as well as a coefficient matrix affecting the observation error. Model (2.1)-(2.2) does not conform to this common specification, but is general in the sense that any fixed-coefficients SS model can be written in this form (Casals, Sotoca and Jerez 1999, Theorem 1). We favor the use of innovations models for signal extraction because the sequence of filtered/smoothed state covariances converge to a null matrix and, therefore, the components obtained do not change as the sample increases, see Casals, Jerez and Sotoca (2002). Despite this preference, the theoretical results in the following sections could have been obtained with any other state-space representation.

2.2. The deterministic/stochastic decomposition.

Assume now that a time series model has been fitted to $\boldsymbol{z}_t$ and has been written in the innovations form (2.1)-(2.2). Consider also the additive decompositions:

$$\boldsymbol{z}_t = \boldsymbol{z}_t^d + \boldsymbol{z}_t^s \tag{2.3}$$

$$\boldsymbol{x}_t = \boldsymbol{x}_t^d + \boldsymbol{x}_t^s \tag{2.4}$$

where $\boldsymbol{z}_t^d$ denotes the part of the time series that exclusively depends on the value of the inputs, while $\boldsymbol{z}_t^s$ only depends on the errors $\boldsymbol{a}_t$. Accordingly, $\boldsymbol{x}_t^d$ and $\boldsymbol{x}_t^s$ are the parts of the state vector related to the inputs and the errors, respectively. Substituting (2.4) in (2.1)-(2.2) and (2.3) in (2.2) yields:

$$\boldsymbol{x}_{t+1}^d + \boldsymbol{x}_{t+1}^s = \mathbf{\Phi}\left(\boldsymbol{x}_t^d + \boldsymbol{x}_t^s\right) + \mathbf{\Gamma}\boldsymbol{u}_t + \mathbf{E}\,\boldsymbol{a}_t \tag{2.5}$$

$$\boldsymbol{z}_t^d + \boldsymbol{z}_t^s = \mathbf{H}\left(\boldsymbol{x}_t^d + \boldsymbol{x}_t^s\right) + \mathbf{D}\,\boldsymbol{u}_t + \boldsymbol{a}_t \tag{2.6}$$

and (2.5)-(2.6) can be immediately decomposed in two different SS models. The "deterministic sub-system", that realizes $\boldsymbol{z}_t^d$:

$$\boldsymbol{x}_{t+1}^{d} = \boldsymbol{\Phi}\, \boldsymbol{x}_{t}^{d} + \boldsymbol{\Gamma} \boldsymbol{u}_{t} \tag{2.7}$$

$$\boldsymbol{z}_{t}^{d} = \mathbf{H} \boldsymbol{x}_{t}^{d} + \mathbf{D} \boldsymbol{u}_{t} \tag{2.8}$$

and the “stochastic sub-system”, which output is $\boldsymbol{z}_t^s$:

$$\boldsymbol{x}_{t+1}^{s} = \boldsymbol{\Phi}\, \boldsymbol{x}_{t}^{s} + \mathbf{E}\, \boldsymbol{a}_{t} \tag{2.9}$$

$$\boldsymbol{z}_{t}^{s} = \mathbf{H} \boldsymbol{x}_{t}^{s} + \boldsymbol{a}_{t} \tag{2.10}$$

This deterministic/stochastic decomposition is a rather standard result of linear systems theory, see e.g., Van Overschee and De Moor (1994). The terms “deterministic” and “stochastic” may be confusing for readers with a background in statistics or econometrics, who would probably find the terms “input-driven” and “error-driven” to be more accurate. Taking into account this remark, we will maintain the original terminology.

2.3. Reducing the sub-systems to the minimal dimension.

We assumed (2.1)-(2.2) to be minimal, this condition being sufficient to assure observability of $\boldsymbol{z}_t^s$ and $\boldsymbol{z}_t^d$. On the other hand, decomposing this model into its deterministic and stochastic sub-systems replicates the state equation. To see this, note that (2.7) and (2.9) can be jointly written as:

$$\begin{bmatrix} \boldsymbol{x}_{t+1}^{d} \\ \boldsymbol{x}_{t+1}^{s} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{t}^{d} \\ \boldsymbol{x}_{t}^{s} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Gamma} \\ \mathbf{0} \end{bmatrix} \boldsymbol{u}_{t} + \begin{bmatrix} \mathbf{0} \\ \mathbf{E} \end{bmatrix} \boldsymbol{a}_{t} \tag{2.11}$$

and adding (2.8) and (2.10) yields, taking into account (2.3):

$$\boldsymbol{z}_t = \begin{bmatrix} \mathbf{H} & \mathbf{H} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_t^d \\ \boldsymbol{x}_t^s \end{bmatrix} + \mathbf{D}\boldsymbol{u}_t + \boldsymbol{a}_t \tag{2.12}$$

Note that this model realizes $\boldsymbol{z}_t$ using $2n$ states, while (2.1)-(2.2) only requires n dynamic components. Therefore (2.11)-(2.12) is a non-minimal system where the output is expressed as the sum of two components with identical dynamics. In general, these components would be impossible to distinguish. However, it is possible in this case because the input of one of these sub-systems, $\boldsymbol{u}_t$, is exactly known. As we will see in the next Section, this fact reduces the decomposition problem to computing adequate initial conditions for the deterministic sub-system.

On the other hand, even if the outputs $\boldsymbol{z}_t^s$ and $\boldsymbol{z}_t^d$ were observable, minimality of both sub-systems is not assured. This happens because controllability of the pair (**Φ** [**Γ** **E**]), corresponding to (2.1)-(2.2) does not assure controllability of the pairs (**Φ** **Γ**) and (**Φ** **E**), corresponding to (2.7)-(2.8) and (2.9)-(2.10), respectively. For example, if part of the model dynamics is uniquely associated to the inputs, as happens in a transfer function, the states describing these dynamics will not be excited in the stochastic sub-system and, as a consequence, this sub-system will be uncontrollable.

These uncontrollable modes are redundant and may wreak havoc on standard signal extraction algorithms. Fortunately, it is easy to detect and eliminate them by applying, e.g., the Staircase algorithm (Rosenbrock 1970) to each sub-system. The following example illustrates the previous discussion and the elimination of uncontrollable modes.

Example. Consider the following SISO model:

$$z_t = \frac{.5}{1-.6B} u_t + \frac{1}{1-B} a_t \tag{2.13}$$

which, following Casals, Jerez and Sotoca (2002, Result 2) can be written in block-diagonal innovations form as:

$$\begin{bmatrix} x_{t+1}^1 \\ x_{t+1}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & .6 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + \begin{bmatrix} 0 \\ -.103 \end{bmatrix} u_t + \begin{bmatrix} 1.166 \\ 0 \end{bmatrix} a_t \tag{2.14}$$

$$z_t = \begin{bmatrix} .858 & -2.916 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + .5u_t + a_t \tag{2.15}$$

According to (2.7)-(2.8) and (2.9)-(2.10) the deterministic and stochastic sub-systems are, respectively:

$$\begin{bmatrix} x_{t+1}^{d1} \\ x_{t+1}^{d2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & .6 \end{bmatrix} \begin{bmatrix} x_t^{d1} \\ x_t^{d2} \end{bmatrix} + \begin{bmatrix} 0 \\ -.103 \end{bmatrix} u_t \tag{2.16}$$

$$z_t^d = \begin{bmatrix} .858 & -2.916 \end{bmatrix} \begin{bmatrix} x_t^{d1} \\ x_t^{d2} \end{bmatrix} + .5u_t \tag{2.17}$$

$$\begin{bmatrix} x_{t+1}^{s1} \\ x_{t+1}^{s2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & .6 \end{bmatrix} \begin{bmatrix} x_t^{s1} \\ x_t^{s2} \end{bmatrix} + \begin{bmatrix} 1.166 \\ 0 \end{bmatrix} a_t \tag{2.18}$$

$$z_t^s = \begin{bmatrix} .858 & -2.916 \end{bmatrix} \begin{bmatrix} x_t^{s1} \\ x_t^{s2} \end{bmatrix} + a_t \tag{2.19}$$

Note that the first state in (2.16) and the second state in (2.18) will never be excited, as they do not depend on the values of the input and the error, respectively. Therefore, they are uncontrollable and the overall system is not minimal. Applying the Staircase algorithm to (2.16)-(2.17) and (2.18)-(2.19) yields two equivalent and minimal systems:

$$x_{t+1}^d = .6\, x_t^d - .103u_t \tag{2.20}$$

$$z_t^d = -2.916 x_t^d + .5u_t \tag{2.21}$$

$$x_{t+1}^s = x_t^s + 1.166\, a_t \tag{2.22}$$

$$z_t^s = .858 x_t^s + a_t \qquad (2.23)$$

3. Signal decomposition.

3.1 Notation and previous results.

Assuming that the minimal sub-system matrices are known, we will now address the problem of estimating the sequences, $\boldsymbol{z}_t^s$ and $\boldsymbol{z}_t^d$ (or $\boldsymbol{z}_t^{d,i}$, $i = 1, 2, \ldots, r$), for a given sample $\{\boldsymbol{z}_t, \boldsymbol{u}_t\}_{t=1}^N$. To this end, note that the solution of $\boldsymbol{x}_t^d$ in (2.7) is given by:

$$\boldsymbol{x}_t^d = \boldsymbol{\Phi}^{t\text{-}1} \boldsymbol{x}_1^d + \sum_{i=1}^{t-1} \boldsymbol{\Phi}^{t\text{-}1\text{-}i}\, \boldsymbol{\Gamma}\, \boldsymbol{u}_i \qquad (3.1)$$

If we had an adequate initial condition for the deterministic sub-system, $\hat{\boldsymbol{x}}_{1|N}^d$, computing the decomposition would reduce to: (a) propagating (3.1) to obtain the sequence $\left\{\hat{\boldsymbol{x}}_{t|N}^d\right\}_{t=2}^N$; (b) estimating the output of the deterministic sub-system which, by (2.8), would be:

$$\hat{\boldsymbol{z}}_{t|N}^d = \mathbf{H}\hat{\boldsymbol{x}}_{t|N}^d + \mathbf{D}\boldsymbol{u}_t \qquad (3.2)$$

and (c) obtaining the corresponding outputs of the stochastic sub-system using (2.3):

$$\hat{\boldsymbol{z}}_{t|N}^s = \boldsymbol{z}_t - \hat{\boldsymbol{z}}_{t|N}^d \qquad (3.3)$$

De Jong (1988) shows that the parameter matrices in (2.1)-(2.2) and the Kalman filter gain, $\mathbf{K}_t$, are not affected by the initial state, $\boldsymbol{x}_1$, and also that:

$$\tilde{\boldsymbol{z}}_t = \mathbf{H}\overline{\boldsymbol{\Phi}}_{t\text{-}1}\, \boldsymbol{x}_1 + \tilde{\boldsymbol{z}}_t^* \qquad (3.4)$$

where:

$\tilde{z}_t$ the sequence of innovations resulting from a Kalman filter arbitrarily initialized with a null state and a null covariance, hereafter KF(0,0),

$\tilde{z}_t^*$ the Kalman filter innovations that would result if the true initial conditions were null, and

$\bar{\mathbf{\Phi}}_t$ the matrix sequence resulting from: $\bar{\mathbf{\Phi}}_t = \left(\mathbf{\Phi} - \mathbf{K}_t \mathbf{H}\right)\bar{\mathbf{\Phi}}_{t-1}$, with $\bar{\mathbf{\Phi}}_1 = \mathbf{I}$

Expression (3.4) can be written in matrix form as:

$$\tilde{\boldsymbol{Z}} = \boldsymbol{X}\,\boldsymbol{x}_1 + \tilde{\boldsymbol{Z}}^* \tag{3.5}$$

where $\tilde{\boldsymbol{Z}}$ and $\tilde{\boldsymbol{Z}}^*$ are $(m\times N)\times 1$ vectors containing the innovations $\tilde{z}_t$ and $\tilde{z}_t^*$, respectively, and $\boldsymbol{X}$ is matrix which row-blocks contain the values $\mathbf{H}\bar{\mathbf{\Phi}}_{t\text{-}1}$.

Taking into account that: $\boldsymbol{x}_1 = \boldsymbol{x}_1^d + \boldsymbol{x}_1^s$, it would be natural to write (3.5) as $\tilde{\boldsymbol{Z}} = \boldsymbol{X}\left(\boldsymbol{x}_1^d + \boldsymbol{x}_1^s\right) + \tilde{\boldsymbol{Z}}^* = \boldsymbol{X}\boldsymbol{x}_1^d + \boldsymbol{X}\boldsymbol{x}_1^s + \tilde{\boldsymbol{Z}}^*$. However, the matrices affecting $\boldsymbol{x}_1^d$ and $\boldsymbol{x}_1^s$ in this expression can be different, because we concentrate on the controllable modes of the sub-systems, see Section 2.3. Therefore, distinguishing between the initial states of the deterministic and stochastic sub-systems yields the following version of (3.5):

$$\tilde{\boldsymbol{Z}} = \boldsymbol{X}^d\boldsymbol{x}_1^d + \boldsymbol{X}^s\boldsymbol{x}_1^s + \tilde{\boldsymbol{Z}}^* \tag{3.6}$$

so the problem reduces to determining the initial conditions for the deterministic sub-system, $\boldsymbol{x}_1^d$.

3.2 Initial conditions for the deterministic sub-system.

The initial conditions for the deterministic sub-system can be determined by two basic approaches, by assuming that they are diffuse or by computing the generalized least squares (GLS) estimate of a non-zero initial state, $\boldsymbol{x}_1^d$, with null covariance. As we will see, both of them are equivalent in the context of this paper.

Assuming a diffuse initial state implies that $\boldsymbol{x}_1^d \sim \left(\overline{\boldsymbol{x}}_1^d, \boldsymbol{P}_1^d\right)$, with $\left(\boldsymbol{P}_1^d\right)^{-1} = \boldsymbol{0}$. In this case (De Jong, 1988) the initial conditions are:

$$\hat{\boldsymbol{x}}_{1|N}^d = \left[\left(\boldsymbol{P}_1^d\right)^{-1} + \left(\boldsymbol{X}^d\right)^T \overline{\boldsymbol{B}}^{-1}\boldsymbol{X}^d\right]^{-1}\left[\left(\boldsymbol{P}_1^d\right)^{-1}\overline{\boldsymbol{x}}_1^d + \left(\boldsymbol{X}^d\right)^T \overline{\boldsymbol{B}}^{-1}\tilde{\boldsymbol{Z}}\right] \tag{3.7}$$

where $\overline{\boldsymbol{B}} = \boldsymbol{X}^s\boldsymbol{P}_1^s\left(\boldsymbol{X}^s\right)^T + \boldsymbol{B}$, and $\boldsymbol{B} = \operatorname{cov}\left(\tilde{\boldsymbol{Z}}^*\right) = \operatorname{diag}\left(\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_m\right)$, being $\left\{\boldsymbol{B}_t\right\}_{t=1,2,\ldots,n}$ the sequence of covariances resulting from a KF(0,0). Taking into account that $\left(\boldsymbol{P}_1^d\right)^{-1} = \boldsymbol{0}$, expression (3.7) simplifies to:

$$\hat{\boldsymbol{x}}_{1|N}^d = \left[\left(\boldsymbol{X}^d\right)^T \overline{\boldsymbol{B}}^{-1}\boldsymbol{X}^d\right]^{-1}\left[\left(\boldsymbol{X}^d\right)^T \overline{\boldsymbol{B}}^{-1}\tilde{\boldsymbol{Z}}\right] \tag{3.8}$$

which is the GLS estimate of the non-null initial state, $\boldsymbol{x}_1^d$, see De Jong (1988) and Casals, Jerez and Sotoca (2000).

Note that Exp. (3.8) depends on $\overline{\boldsymbol{B}}^{-1} = \left[\boldsymbol{X}^s\boldsymbol{P}_1^s\left(\boldsymbol{X}^s\right)^T + \boldsymbol{B}\right]^{-1}$ and, therefore, on the covariance $\boldsymbol{P}_1^s$. If the stochastic sub-system has unit roots, this matrix would have non-finite elements. Following De Jong and Chu-Chun-Lin (1994), this inconvenient can be solved by applying the matrix inversion lemma to $\overline{\boldsymbol{B}}^{-1}$, which yields:

$$\overline{\boldsymbol{B}}^{-1} = \left[\boldsymbol{X}^s\boldsymbol{P}_1^s\left(\boldsymbol{X}^s\right)^T + \boldsymbol{B}\right]^{-1} = \boldsymbol{B}^{-1} - \boldsymbol{B}^{-1}\boldsymbol{X}^s\left[\left(\boldsymbol{P}_1^s\right)^{-1} + \left(\boldsymbol{X}^s\right)^T \boldsymbol{B}^{-1}\boldsymbol{X}^s\right]^{-1}\left(\boldsymbol{X}^s\right)^T \boldsymbol{B}^{-1} \tag{3.9}$$

and, substituting (3.9) in (3.8) yields:

$$\begin{aligned}\hat{\boldsymbol{x}}_{1|N}^{d} &= \left[\left(\boldsymbol{X}^{d}\right)^{T}\left\{\boldsymbol{B}^{-1}-\boldsymbol{B}^{-1}\boldsymbol{X}^{s}\left[\left(\boldsymbol{P}_{1}^{s}\right)^{-1}+\left(\boldsymbol{X}^{s}\right)^{T}\boldsymbol{B}^{-1}\boldsymbol{X}^{s}\right]^{-1}\left(\boldsymbol{X}^{s}\right)^{T}\boldsymbol{B}^{-1}\right\}\boldsymbol{X}^{d}\right]^{-1} \\ &\quad \left[\left(\boldsymbol{X}^{d}\right)^{T}\left\{\boldsymbol{B}^{-1}-\boldsymbol{B}^{-1}\boldsymbol{X}^{s}\left[\left(\boldsymbol{P}_{1}^{s}\right)^{-1}+\left(\boldsymbol{X}^{s}\right)^{T}\boldsymbol{B}^{-1}\boldsymbol{X}^{s}\right]^{-1}\left(\boldsymbol{X}^{s}\right)^{T}\boldsymbol{B}^{-1}\right\}\tilde{\boldsymbol{Z}}\right] \\ &= \left\{\left(\boldsymbol{X}^{d}\right)^{T}\boldsymbol{B}^{-1}\boldsymbol{X}^{d}-\left(\boldsymbol{X}^{d}\right)^{T}\boldsymbol{B}^{-1}\boldsymbol{X}^{s}\left[\left(\boldsymbol{P}_{1}^{s}\right)^{-1}+\left(\boldsymbol{X}^{s}\right)^{T}\boldsymbol{B}^{-1}\boldsymbol{X}^{s}\right]^{-1}\left(\boldsymbol{X}^{s}\right)^{T}\boldsymbol{B}^{-1}\boldsymbol{X}^{d}\right\}^{-1} \\ &\quad \left\{\left(\boldsymbol{X}^{d}\right)^{T}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}-\left(\boldsymbol{X}^{d}\right)^{T}\boldsymbol{B}^{-1}\boldsymbol{X}^{s}\left[\left(\boldsymbol{P}_{1}^{s}\right)^{-1}+\left(\boldsymbol{X}^{s}\right)^{T}\boldsymbol{B}^{-1}\boldsymbol{X}^{s}\right]^{-1}\left(\boldsymbol{X}^{s}\right)^{T}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}\right\}\end{aligned} \tag{3.10}$$

The initial condition (3.10) allows for both, stationary and unit roots because it does not depend on $\boldsymbol{P}_1^s$, but on its inverse $\left(\boldsymbol{P}_1^s\right)^{-1}$. As De Jong and Chu-Chun-Lin showed, when the system has both types of roots, the matrix $\left(\boldsymbol{P}_1^s\right)^{-1}$ converges to a finite value with as many null eigenvalues as nonstationary states.

There are three important particular cases of (3.10) depending on the system dynamics.

Case 1. The stochastic sub-system is static, so $\boldsymbol{X}^s = \boldsymbol{0}$ and (3.10) simplifies to:

$$\hat{\boldsymbol{x}}_{1|N}^{d} = \left[\left(\boldsymbol{X}^{d}\right)^{T}\boldsymbol{B}^{-1}\boldsymbol{X}^{d}\right]^{-1}\left(\boldsymbol{X}^{d}\right)^{T}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}} \tag{3.11}$$

Case 2. The deterministic and stochastic sub-systems share the same dynamics, so $\boldsymbol{X}^d = \boldsymbol{X}^s$ and (3.10) collapses again to (3.11)

Case 3. The deterministic and stochastic sub-systems share at least one nonstationary mode. This is equivalent to defining two sets of diffuse initial conditions over these shared modes and creates a severe identifiability problem because the matrix

$$\left(\boldsymbol{X}^{d}\right)^{T}\boldsymbol{B}^{-1}\boldsymbol{X}^{d}-\left(\boldsymbol{X}^{d}\right)^{T}\boldsymbol{B}^{-1}\boldsymbol{X}^{s}\left[\left(\boldsymbol{P}_{1}^{s}\right)^{-1}+\left(\boldsymbol{X}^{s}\right)^{T}\boldsymbol{B}^{-1}\boldsymbol{X}^{s}\right]^{-1}\left(\boldsymbol{X}^{s}\right)^{T}\boldsymbol{B}^{-1}\boldsymbol{X}^{d} \tag{3.12}$$

whose inverse is required to compute (3.10), is singular. The only option in this case consists of using the pseudoinverse to compute the initial conditions.

3.3 Computational issues.

To compute (3.10) according to the results in previous Sub-sections, we need to calculate five terms: $\left(\boldsymbol{X}^{\boldsymbol{d}}\right)^T \boldsymbol{B}^{-1}\boldsymbol{X}^{\boldsymbol{d}}$, $\left(\boldsymbol{X}^{s}\right)^T \boldsymbol{B}^{-1}\boldsymbol{X}^{s}$, $\left(\boldsymbol{X}^{\boldsymbol{d}}\right)^T \boldsymbol{B}^{-1}\boldsymbol{X}^{s}$, $\left(\boldsymbol{X}^{\boldsymbol{d}}\right)^T \boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}$ and $\left(\boldsymbol{X}^{s}\right)^T \boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}$. To do this, we first compute the matrices $\left(\boldsymbol{X}\right)^T \boldsymbol{B}^{-1}\boldsymbol{X}$ and $\left(\boldsymbol{X}\right)^T \boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}$ as:

$$\left(\boldsymbol{X}\right)^T \boldsymbol{B}^{-1}\boldsymbol{X} = \sum_{t=1}^{N} \bar{\boldsymbol{\Phi}}_{\boldsymbol{t}}^{\boldsymbol{T}} \mathbf{H}^{\mathrm{T}} \boldsymbol{B}_{\boldsymbol{t}}^{-1} \mathbf{H} \bar{\boldsymbol{\Phi}}_{\boldsymbol{t}} \tag{3.13}$$

$$\left(\boldsymbol{X}\right)^T \boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}} = \sum_{t=1}^{N} \bar{\boldsymbol{\Phi}}_{\boldsymbol{t}}^{\boldsymbol{T}} \mathbf{H}^{\mathrm{T}} \boldsymbol{B}_{\boldsymbol{t}}^{-1} \tilde{\boldsymbol{z}}_{\boldsymbol{t}} \tag{3.14}$$

and, as $\boldsymbol{X}$ is known, it is easy to find the matrices $\boldsymbol{T_d}$ and $\boldsymbol{T_s}$, such that: $\boldsymbol{X}\,\boldsymbol{T_d} = \boldsymbol{X}^{\boldsymbol{d}}$ and $\boldsymbol{X}\,\boldsymbol{T_s} = \boldsymbol{X}^{s}$. Under these conditions the five terms needed to compute (3.10) are given by:

$$\left(\boldsymbol{X}^{\boldsymbol{d}}\right)^T \boldsymbol{B}^{-1}\boldsymbol{X}^{\boldsymbol{d}} = \boldsymbol{T_d^T}\,\boldsymbol{X^T}\boldsymbol{B}^{-1}\boldsymbol{X}\,\boldsymbol{T_d} \tag{3.15}$$

$$\left(\boldsymbol{X}^{\boldsymbol{d}}\right)^T \boldsymbol{B}^{-1}\boldsymbol{X}^{s} = \boldsymbol{T_d^T}\,\boldsymbol{X^T}\boldsymbol{B}^{-1}\boldsymbol{X}\,\boldsymbol{T_s} \tag{3.16}$$

$$\left(\boldsymbol{X}^{s}\right)^T \boldsymbol{B}^{-1}\boldsymbol{X}^{s} = \boldsymbol{T_s^T}\,\boldsymbol{X^T}\boldsymbol{B}^{-1}\boldsymbol{X}\,\boldsymbol{T_s} \tag{3.17}$$

$$\left(\boldsymbol{X}^{\boldsymbol{d}}\right)^T \boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}} = \boldsymbol{T_d^T}\,\boldsymbol{X^T}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}} \tag{3.18}$$

$$\left(\boldsymbol{X}^{s}\right)^T \boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}} = \boldsymbol{T_s^T}\,\boldsymbol{X^T}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}} \tag{3.19}$$

Under weak assumptions, the previous decomposition algorithm assures that the input-driven component will have minimal revisions when the sample increases. To see this, note that Equations (3.1) and (3.2) imply immediately that the only source of revisions is a change in the GLS estimate of the initial state. It can be shown that, if the system is stabilizable (roughly speaking, invertible) then the sequence $\bar{\mathbf{\Phi}}_t$ converges to zero at exponential rate (De Souza, Gevers and Goodwin, 1986) implying the convergence the matrices (3.13)-(3.14) and (3.15)-(3.19) to null values. Accordingly, the GLS estimate (3.10) will converge to a finite and stable value at the same speed.

4. Unscrambling the effect of individual inputs.

The decomposition described in Sections 2 and 3 is useful when one is interested in the total effect of all the exogenous variables. On the other hand, in models with several inputs one may want to estimate the individual effect of each one of them.

This separation is easier if we write the deterministic sub-system (2.7)-(2.8) in its equivalent Canonic Controllable Luenberger form (hereafter CCL form), denoted by:

$$\bar{\boldsymbol{x}}_{t+1}^{d} = \bar{\mathbf{\Phi}}\, \bar{\boldsymbol{x}}_{t}^{d} + \bar{\mathbf{\Gamma}} \boldsymbol{u}_t \tag{4.1}$$

$$\boldsymbol{z}_{t}^{d} = \bar{\mathbf{H}} \bar{\boldsymbol{x}}_{t}^{d} + \mathbf{D} \boldsymbol{u}_t \tag{4.2}$$

Model (4.1)-(4.2) can be obtained by a similar transformation of (2.7)-(2.8) characterized by the matrix $\mathbf{T}$, see Petkov, Christov and Konstantinov (1991), such that $\bar{\mathbf{\Phi}} = \mathbf{T}^{-1}\mathbf{\Phi}\mathbf{T}$; $\bar{\mathbf{H}} = \mathbf{H}\mathbf{T}$; $\bar{\mathbf{\Gamma}} = \mathbf{T}^{-1}\mathbf{\Gamma}$ and $\bar{\boldsymbol{x}}_{t}^{d} = \mathbf{T}^{-1}\boldsymbol{x}_{t}^{d}$. Accordingly, the initial state of the CCL system corresponding to the vector $\hat{\boldsymbol{x}}_{1|N}^{d}$ defined in (3.10) would be:

$$\hat{\bar{\boldsymbol{x}}}_{\mathbf{1}|\boldsymbol{N}}^{\boldsymbol{d}} = \mathbf{T}^{-1}\hat{\boldsymbol{x}}_{\mathbf{1}|\boldsymbol{N}}^{\boldsymbol{d}} \tag{4.3}$$

The controllability matrix of the CCL realization $\left(\bar{\mathbf{\Phi}}, \bar{\mathbf{H}}, \bar{\mathbf{\Gamma}}, \mathbf{D}\right)$ separates the dynamics associated to the individual inputs, therefore allowing to discern whether a given input affects (or not) each system state. To see this, note that propagating the CCL model according to (3.1) yields:

$$\bar{\boldsymbol{x}}_{\boldsymbol{t}}^{\boldsymbol{d}} = \bar{\mathbf{\Phi}}^{t-1}\bar{\boldsymbol{x}}_{\mathbf{1}}^{\boldsymbol{d}} + \left[\bar{\mathbf{\Phi}}^{t-2}\bar{\mathbf{\Gamma}} \quad \bar{\mathbf{\Phi}}^{t-3}\bar{\mathbf{\Gamma}} \quad \ldots \quad \bar{\mathbf{\Gamma}}\ \right]\begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_{t-1} \end{bmatrix} \tag{4.4}$$

where $\bar{\boldsymbol{C}} = \left[\bar{\mathbf{\Phi}}^{t-2}\bar{\mathbf{\Gamma}} \quad \bar{\mathbf{\Phi}}^{t-3}\bar{\mathbf{\Gamma}} \quad \ldots \quad \bar{\mathbf{\Gamma}}\ \right]$ is the controllability matrix of the CCL system. According to (4.4), its zero/nonzero structure characterizes which inputs affect any given state. Specifically, if the *i*-th state is not excited by the *j*-th input, then the elements $\bar{\boldsymbol{c}}\left(i, j\right)$, $\bar{\boldsymbol{c}}\left(i, j+r\right)$, $\bar{\boldsymbol{c}}\left(i, j+2r\right)$, …, $\bar{\boldsymbol{c}}\left(i, j+(n-1)\times r\right)$ should be null, being *r* the number of inputs.

Under these conditions, estimating the individual effect of each input requires:

(1) Decomposing the initial condition (4.3) into *r* vectors, such that each one of them is exclusively related to one of the *r* system inputs.

(2) Propagating the *r* systems given by $\bar{\mathbf{\Phi}}$, $\bar{\mathbf{H}}$, and the *j*-th columns of $\bar{\mathbf{\Gamma}}$ and $\mathbf{D}$ ($j = 1, 2, \ldots, r$), starting from the corresponding initial conditions.

Ideally, the initial vectors $\hat{\bar{\boldsymbol{x}}}_{\mathbf{1},\boldsymbol{j}}^{\boldsymbol{d}}$ would be mutually orthogonal and such that $\hat{\bar{\boldsymbol{x}}}_{\mathbf{1}|N}^{\boldsymbol{d}} = \sum_{j=1}^{r} \hat{\bar{\boldsymbol{x}}}_{\mathbf{1},\boldsymbol{j}}^{\boldsymbol{d}}$. Its orthogonality can be assured by distributing the initial state of the CCL system according to the structure of the controllability matrix $\bar{\boldsymbol{C}}$:

$$\tilde{x}^d_{1,j}(i) = \begin{cases} 0 & \text{if } \bar{\boldsymbol{c}}(i,j) = \bar{\boldsymbol{c}}(i,j+r) = \bar{\boldsymbol{c}}(i,j+2r) = \ldots = \bar{\boldsymbol{c}}(i,j+(n-1)\times r) = 0 \\ \hat{\hat{x}}^d_{1|N}(i) & \text{otherwise} \end{cases}$$

$$j = 1, 2, \ldots, r \; ; \; i = 1, 2, \ldots, n \tag{4.5}$$

where $\hat{\hat{x}}^d_{1|N}(i)$ is the i-th element of $\hat{\hat{\boldsymbol{x}}}^d_{1|N}$. Last, the vectors resulting from (4.5), $\tilde{\boldsymbol{x}}^d_{1,j}$, should be transformed according to:

$$\hat{\hat{\boldsymbol{x}}}^d_{1,j} = \boldsymbol{\Pi}_j \, \tilde{\boldsymbol{x}}^d_{1,j} \; ; \; j = 1, 2, \ldots, r \tag{4.6}$$

where $\boldsymbol{\Pi}_j$ is the orthogonal projector in the null space of the matrix $\left[\hat{\hat{\boldsymbol{x}}}^d_{1,1} \quad \hat{\hat{\boldsymbol{x}}}^d_{1,2} \quad \ldots \quad \hat{\hat{\boldsymbol{x}}}^d_{1,j-1} \quad \hat{\hat{\boldsymbol{x}}}^d_{1,j+1} \quad \ldots \quad \hat{\hat{\boldsymbol{x}}}^d_{1,r} \right]$. The initial conditions (4.6) can then be interpreted, either as the part of the CCL initial state that is orthogonal to the rest of the inputs, or as the marginal contribution of each input to the initial condition of the CCL system.

In many models, such as transfer functions, each mode is affected only by a single input, so the orthogonal projector $\boldsymbol{\Pi}_j$ is the identity and the condition $\hat{\hat{\boldsymbol{x}}}^d_{1|N} = \sum_{j=1}^{r} \hat{\hat{\boldsymbol{x}}}^d_{1,j}$ is assured. On the other hand, if a given input affects more than one mode, the decomposition does not assure a complete split of (4.3) so, in general, there will be a remainder $\hat{\boldsymbol{x}}^d_{1|N} - \sum_{j=1}^{r} \hat{\boldsymbol{x}}^d_{1,j} \neq \boldsymbol{0}$. This remainder can be propagated with the pair ($\bar{\boldsymbol{\Phi}}, \bar{\mathbf{H}}$) and the resulting component could be interpreted as the common effect of the system inputs.

5. Empirical example.

Now we will show how the decomposition proposed can be used to gain insight on the relationship between different time series. To this end, we will use the famous monthly series of sales and advertising of the Lydia Pinkham vegetable compound, see Figure 1.

[Insert Figure 1]

This product was introduced in 1873. It was an herbal extract in a 18-20 percent alcoholic solution, and was considered to be effective against "all those painful Complaints and Weaknesses so common to our best female population". Additional medical claims followed the commercial success of the compound. The company gained strong publicity because of controversies around the product ingredients and a large court case, which made public this dataset. The company was finally sold in 1968, but some medicinal products with the generic "Lydia Pinkham" brand can be acquired today by direct order.

These series are important in empirical research about the effects of advertising for several reasons: (a) the product is a frequently purchased, low-cost consumer nondurable, being this class of products specially interesting to marketing researchers; (b) advertising, primarily in newspapers, was the only marketing instrument used by the company; (c) price changes were small and infrequent; whereas (d) the distribution, mainly through drug wholesalers, remained fairly stable; furthermore, (e) there were no direct competitors for this product, so the market under study can be considered a closed sales-advertising system. Because of these convenient features, this dataset was used by early researchers such as Palda (1964) or Bhattacharyya (1982), and in very recent works such as Kim (2005) or Smith, Naik and Tsai (2006).

Simple inspection of Figure 1 shows that both series have seasonal fluctuations and show a downward drift. We: (a) fitted an IMA(1,1)x(1,1)$_{12}$ model to log advertising; (b) prewhitened the logs of both series using this model; and (c) computed the corresponding sample cross-correlations. These cross-correlations show a rough sinusoidal pattern between current sales and lagged advertising, with no substantial feedback. This sinusoidal pattern leads us to specify the following family of models:

$$\ln S_t \times 100 = \frac{\omega_0 + \omega_1 B + \omega_2 B^2}{1 + \delta_1 B + \delta_2 B^2} \ln A_t \times 100 + \frac{(1+\theta B)(1+\Theta B^{12})}{(1-B)(1-B^{12})} a_t \qquad (5.1)$$

where S_t and A_t denote, respectively, sales and advertising in month *t* and $a_t \sim iid(0,\sigma_a^2)$. Table 1 summarizes the estimation results and some residual statistics for five particular cases of (5.1). On the basis of the small AIC and *Q*-statistic values, we choose model #5 as a valid specification for the illustrating purpose of this exercise.

[Insert Table 1]

Note that:

(1) The roots of the polynomials in the numerator and denominator of the transfer function are, respectively, -.186±1.040*i* and .475±1.052*i* so, even though this model may be somewhat overparametrized, there are no redundant dynamic factors.

(2) On the other hand, the impulse-response function implied by this specification has both, positive and negative values. In a sales-advertising system this can happen when the product has a loyal customer base and the advertising accelerates the consumption, therefore changing the distribution of re-stocking purchases over time.

(3) Note that the values of $\hat{\theta}$ are quite large in comparison with their standard errors. In fact, standard hypothesis testing would not reject $\theta = 1$. However one must take into account that, under $\theta = 1$, Gaussian maximum-likelihood (ML) estimates display a so-called "pile-up" effect, in which the probability of $\hat{\theta}_{ML} = 1$ converges to a large positive value. The intuition behind "pile-up" is that the Gaussian likelihood is the same for (θ, σ^2) and $(1/\theta, \theta^2\sigma^2)$, so $\theta = 1$ is always a critical point of this function. These deviations from usual ML behavior bias standard testing towards non-rejection of $\theta = 1$. Taking into account these distortions, the testing procedure proposed by Davis, Chen and Dunsmuir (1994, Section 3) safely rejects $\theta = 1$, in favour of $\theta < 1$, with a 5% significance.

Using the decomposition described in previous Sections, we estimated the additive contribution of log advertising to log sales, obtained the corresponding multiplicative estimate of this effect and discounted it from the original sales series. The results are shown in Figure 2. Figure 3 plots the resulting estimate of the value added by advertising, computed as the difference between the contribution of advertising to sales in each month and the corresponding investment. Note that the return in many months is negative. In fact, the sum of the estimates in Figure 3 is -.3 million dollars, so advertising in this case did not create value for the firm.

[Insert Figures 2 and 3]

Further decomposition of the stochastic components reveals a possible cause for this lack of performance. Figure 4 plots the multiplicative components of sales corresponding to seasonality and the estimated effect of advertising. Note that both series have a clear negative correlation, meaning that advertising was systematically increased in the lower phases of the seasonal cycle and decreased in the higher phases. Taking into account that both effects are multiplicative, it is clear that this anti-cyclic management of advertising does not maximize the sales.

[Insert Figure 4]

It is easy to illustrate the inefficiency of this distribution through a simple contrafactual experiment, consisting of: (a) re-distributing the yearly expenditures as a direct proportion of the multiplicative seasonal component, (b) backcasting the corresponding sales, and (c) predicting the ROI of advertising in this new scenario. This alternative allocation increases the estimate of total sales generated by advertising by 13%, from 4.5 to 5.1 million USD. As the total amount invested in both scenarios is 4.8 million USD, previous estimates imply that investing against seasonality generated a negative ROI (4.5-4.8=-.3 million USD) while the expected ROI of a pro-seasonal investment policy would have been positive (5.1 - 4.8 = .3 million USD).

6. Concluding remarks.

This paper describes a method to compute the effect of exogenous inputs on the outputs of a model. The problem is solved in a general SS framework, using standard signal-extraction techniques. Once the effect of model inputs has been estimated and extracted, the remaining “stochastic component” can be further decomposed by a standard method into trend, cycle, seasonal and irregular components.

With some work, the methods discussed here can be applied to VARMA-type models without exogenous inputs, by obtaining the causal transfer functions implied. Consider e.g., the example given by Wei (1994, pp. 338), which is based in the VAR model:

$$\begin{bmatrix} 1-\phi_{11}B & 0 \\ -\phi_{21}B & 1-\phi_{22}B \end{bmatrix} \begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix} = \begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix} \; ; \; \text{cov}(\boldsymbol{a_t}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ = & \sigma_{22} \end{bmatrix} a_t \tag{6.1}$$

Model (6.1) can be easily written in transfer function form as:

$$\begin{aligned} z_t^1 &= \frac{1}{1-\phi_{11}B} a_t^1 \\ z_t^2 &= \frac{\phi_{21}B}{1-\phi_{22}B} z_t^1 + \frac{1}{1-\phi_{22}B} a_t^2 \end{aligned} \tag{6.2}$$

however, this representation does not account for nonzero covariances between the disturbances in both equations. Assume that:

$$a_t^1 = \varepsilon_t^1 \;\; ; \;\; a_t^2 = \omega a_t^1 + \varepsilon_t^2 \tag{6.3}$$

where ε_t^1 and ε_t^2 are orthogonal errors. Under these conditions, it is easy to see that the causal transfer function relating z_t^2 and z_t^1 is:

$$z_t^2 = \frac{\omega + \left(\phi_{21} - \omega\phi_{11}\right)B}{1-\phi_{22}B} z_t^1 + \frac{1}{1-\phi_{22}B}\varepsilon_t^2 \qquad (6.4)$$

and the algorithm discussed in previous sections can be immediately applied to (6.4).

The procedures described in this article 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.

References:

AKAIKE, H. (1973). "Information Theory and an Extension of the Maximum Likelihood Principle," in B.N. Petrov and F. Csaki, Proc. 2nd International Symposium on Information Theory, (eds.), 267-281, Akademia Kiado, Budapest.

BHATTACHARYYA, M.N. (1982). "Lydia Pinkham Data Remodelled," *Journal of Time Series Analysis*, 3, 81-102.

BOX, G.E.P. and G. TIAO (1975). "Intervention Analysis with Applications to Economic and Environmental Problems," *Journal of the American Statistical Association*, 70, 70-79.

CASALS, J., S. SOTOCA and M. JEREZ (1999). "A Fast and Stable Method to Compute the Likelihood of Time Invariant State-Space Models," *Economics Letters*, 65, 329-337.

CASALS, J., M. JEREZ and S. SOTOCA (2000). "Exact Smoothing for Stationary and Nonstationary Time Series," *International Journal of Forecasting*, 16, 1, 59-69.

CASALS, J., M. JEREZ and S. SOTOCA (2002), "An Exact Multivariate Model-based Structural Decomposition," *Journal of the American Statistical Association*, 97, 458, 553-564.

DAVIS, R.A., M. CHEN and W.T.M. DUNSMUIR (1994). "Inference for MA(1) processes with a unit root on or near the unit circle," *Probability and Mathematical Statistics*, 15, 227–242.

DE JONG, P. (1988), "The Likelihood for a State Space Model," *Biometrika*, 75, 165-169.

DE JONG, P. and S. CHU-CHUN-LIN (1994), "Stationary and Non-Stationary State Space Models," *Journal of Time Series Analysis*, 15, 151-166.

DE SOUZA, C.E., M.R. GEVERS and G.C. GOODWIN (1986), "Riccati Equations in Optimal Filtering of Nonstabilizable Systems having Singular State Transition Matrices," *IEEE Transactions on Automatic Control*, AC-31, 9, 831-838.

KIM, JAE H. (2005). "Investigating the advertising-sales relationship in the Lydia Pinkham data: a bootstrap approach." *Applied Economics*, 37, 3, 347 – 354.

LJUNG, G. and G.E.P. BOX (1978). "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 67, 297-303.

PALDA, K. (1964). *The Measurement of Cumulative Advertising Effects*, Prentice-Hall, Englewood Cliffs (NJ).

PETKOV, P. HR., N.D. CHRISTOV and M.M. KONSTANTINOV (1991). *Computational Methods for Linear Control Systems*, Prentice-Hall, Englewood Cliffs, New Jersey.

ROSENBROCK, H.H. (1970). *State Space and Multivariable Theory*, John Wiley, New York.

SCHWARZ, G. (1978). "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461-464.

SMITH, A. NAIK, P. A. and C. TSAI (2006). “Markov-switching model selection using Kullback–Leibler divergence,” *Journal of Econometrics*, 134, 2, 553-577.

VAN OVERSCHEE, P. and B. DE MOOR (1994). “N4SID*: Subspace algorithms for the identification of combined deterministic-stochastic systems,” *Automatica*, 30, 75-93.

WEI, W.S. (1994). *Time Series Analysis: Univariate and Multivariate Methods*, Addison-Wesley, New York.

Acknowledgements:

This research received financial support from Ministerio de Ciencia e Innovación, ref. ECO2008-02588/ECON. Feedback from an anonymous reviewer is gratefully acknowledged.

Figure 1: Monthly series of sales and advertising of the Lydia Pinkham vegetable compound from January 1954 to June 1960 (78 monthly values). Source: Palda (1964).

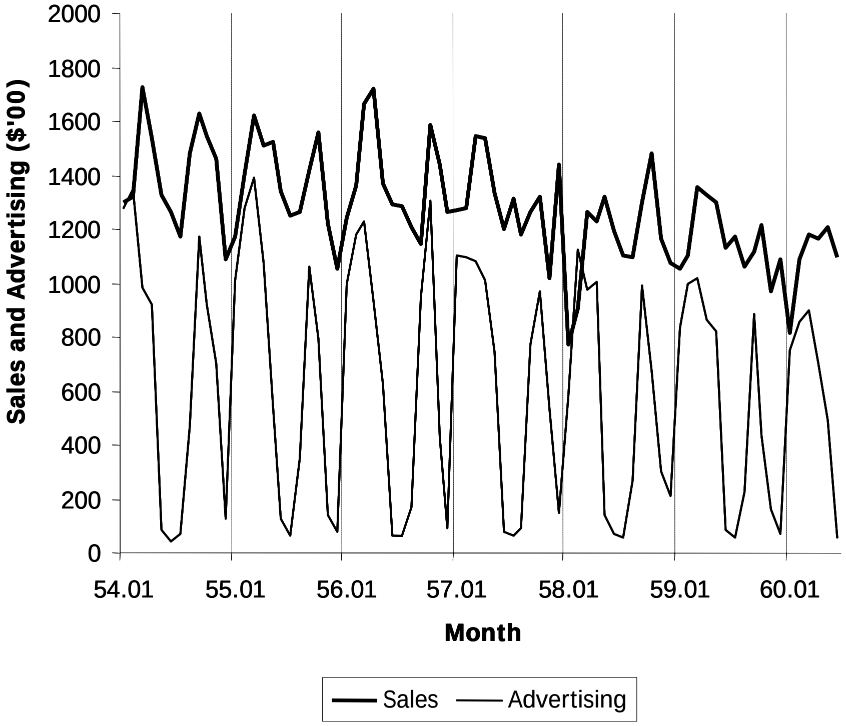

Figure 2: Sales versus the exponential of the stochastic component. The distance between both series (gray area) can be interpreted as an estimate of the effect of advertising over sales.

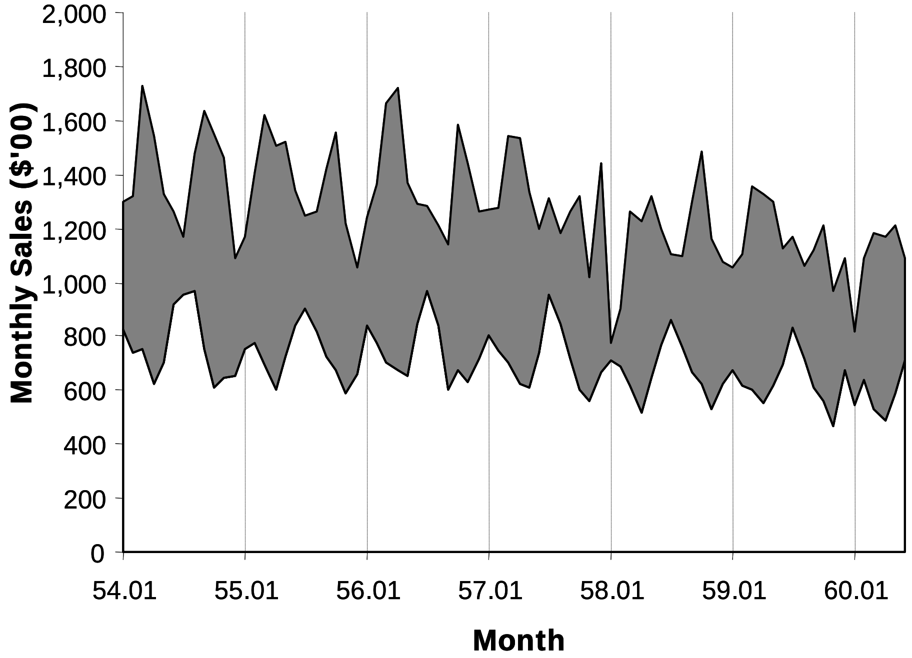

Figure 3: Estimated value added by advertising, computed in each period as the estimate of sales generated by advertising minus the advertising investment.

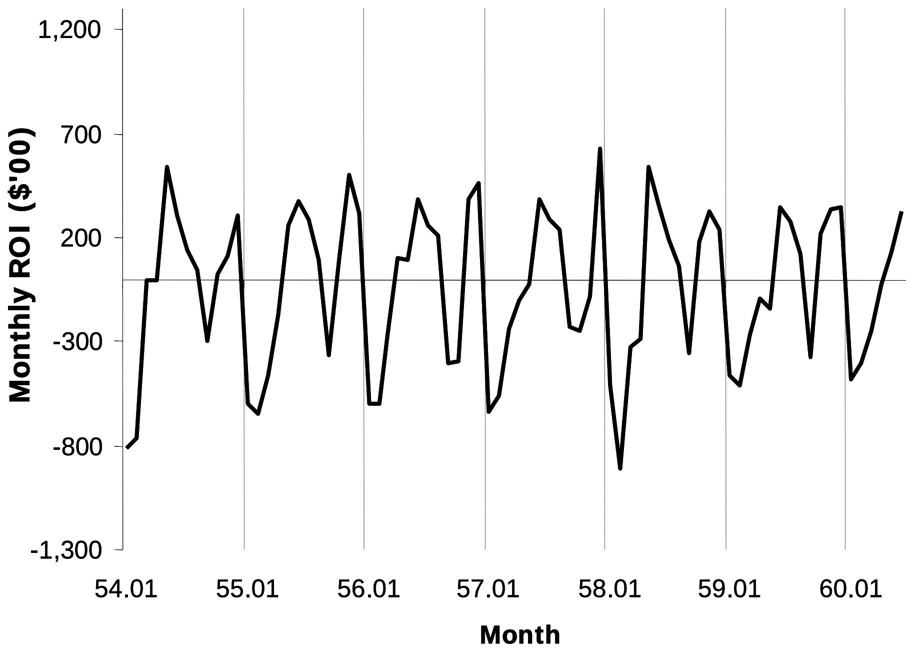

Figure 4: Multiplicative effect of advertising (thick line) versus multiplicative seasonality of sales (thin line). The values between November 1957 and February 1958 are fixed-interval smoothing interpolations. Note that Seasonal and Advertising factors are negatively correlated (-.87) so advertising was systematically increased in the seasonal cycle lows and decreased in the highs.

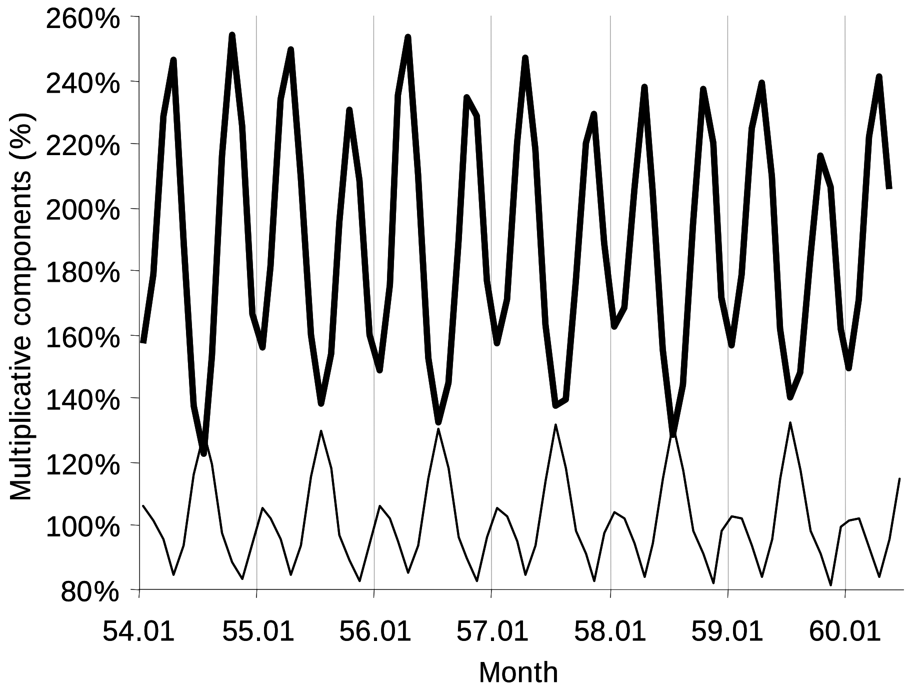

Table 1: Results from model estimation. Parameter estimates were obtained by maximizing the gaussian likelihood function, computed as described in Casals, Sotoca and Jerez (1999). The figures in parentheses are standard errors. Observations between November 1957 and February 1958 have been treated as missing values, because they were outliers.

Model #	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\theta}$	$\hat{\Theta}$	$\hat{\sigma}_a$	$\hat{g}$ [1]	Likelihood[2]	AIC[3]	SBC[4]	$Q(10)$[5]
1	.061 (.017)	--	--	-.391 (.165)	--	-.910 (.049)	-.687 (.186)	8.321	.101	-234.557	6.144	6.294	$\begin{bmatrix} 7.56 & 25.99 \\ 12.77 & 15.14 \end{bmatrix}$
2	.050 (.022)	.037 (.022)	.034 (.021)	--	--	-.913 (.058)	-.651 (.204)	8.174	.121	-229.941	6.050	6.231	$\begin{bmatrix} 5.68 & 23.91 \\ 9.95 & 15.14 \end{bmatrix}$
3	.042 (.012)	--	--	-1.143 (.122)	.721 (.084)	-.905 (.048)	-.523 (.120)	7.855	.073	-227.833	5.996	6.177	$\begin{bmatrix} 9.95 & 14.15 \\ 11.40 & 15.14 \end{bmatrix}$
4	.027 (.015)	.042 (.015)	--	-.817 (.081)	.697 (.083)	-.899 (.051)	-.572 (.117)	7.746	.078	-226.981	6.000	6.211	$\begin{bmatrix} 8.87 & 6.81 \\ 10.98 & 15.14 \end{bmatrix}$
5	.048 (.018)	.016 (.016)	.043 (.022)	-.713 (.118)	.751 (.088)	-.899 (.055)	-.628 (.162)	7.536	.103	-225.807	5.995	6.237	$\begin{bmatrix} 9.69 & 6.94 \\ 11.99 & 15.14 \end{bmatrix}$

---

[1] Steady-state gain, computed as: $\hat{g} = \dfrac{\hat{\omega}_0 + \hat{\omega}_1 + \hat{\omega}_2}{1 + \hat{\delta}_1 + \hat{\delta}_2}$

[2] Value of the Gaussian likelihood on convergence.

[3] Akaike (1973) information criterion.

[4] Schwartz (1978) information criterion.

[5] Matrix of Ljung-Box (1978) *Q* statistics for the sample auto and cross-correlation functions of the model residuals and the prewhitened advertising series.