THE EXACT LIKELIHOOD FOR A STATE SPACE MODEL WITH STOCHASTIC INPUTS

José Casals
Sonia Sotoca[1]
Universidad Complutense de Madrid
(SPAIN)

Abstract: In this work we derive exact and approximate expressions for the conditional mean and variance of the initial state of a state space model, allowing for unit roots and stochastic inputs. These results provide adequate initial conditions to compute the exact likelihood using the Kalman filter. The exact conditional moments are the best choice when the stochastic structure of the inputs is known. If this is not the case, the approximate expressions are a good alternative, as some simulation results illustrate.

Keywords: Exact maximum likelihood, Initial conditions, Kalman filter, State space model, Unit roots.

---

[1] Correspondence should be addressed to: Sonia Sotoca, Departamento de Fundamentos del Análisis Económico II (Economía Cuantitativa), Facultad de C. Económicas y Empresariales, Universidad Complutense, Campus de Somosaguas, 28223 Madrid, Spain. E-mail: sotoca@ccee.ucm.es. Fax: 34-91-3942613.

1. Introduction

Recent literature shows an increasing interest in methods to compute maximum likelihood estimates for the parameters in State Space (SS) models. This may be due to three facts. First, the SS representation includes as particular cases many standard formulations, e.g., ARMA, VARMA and VARMAX. Second, the SS framework simplifies the analysis in many non-standard situations like errors in variables [1, Ch. 3], conditional heteroskedastic errors [2] or random coefficients [3]. Third, the exact likelihood function of a SS model, its first and second-order derivatives and its information matrix have been analyzed by the literature [4; 1] which hence provides all the elements required to implement robust and efficient estimation and testing procedures.

The joint log density function of the sample ($\boldsymbol{y}_t$, t=1,...,N), conditional on an input vector ($\boldsymbol{u}$), can be written as a function of the conditional mean and variance of the initial state ($\boldsymbol{x}_1$):

$$\ell(\boldsymbol{y}/\boldsymbol{u}) = \ell(\boldsymbol{x}_1/\boldsymbol{u}) + \ell(\boldsymbol{y}/\boldsymbol{u},\boldsymbol{x}_1) - \ell(\boldsymbol{x}_1/\boldsymbol{y},\boldsymbol{u}) \tag{1}$$

Under normality, the evaluation of (1) crucially depends on the values of $\mathrm{E}(\boldsymbol{x}_1/\boldsymbol{u})$ and $\mathrm{cov}(\boldsymbol{x}_1/\boldsymbol{u})$ because these moments characterize the first term in (1) and provide an adequate initialization of a Kalman filter (KF) required to calculate the second and third terms [5; 6]. As we will show in this paper, initialization is also important because final estimates are sensitive to the initial conditions of the filter, being this effect very important in models with high signal-to-noise ratios, seasonal models and when the sample is short.

The computation of adequate initial conditions depends on two issues, the stationarity of the SS model and the stochastic nature of its inputs.

When the model is stationary, the standard initialization consists of using the unconditional mean and variance of the initial state [7, Ch. 4]. In the non-stationary case these moments are not finite and the literature suggests several alternatives: a) setting $\mathrm{E}(\boldsymbol{x}_1) = \boldsymbol{0}$ and $\mathrm{cov}(\boldsymbol{x}_1) = k\boldsymbol{I}$, where k is an arbitrary large constant [6]; b) using the information filter [8]; c) initializing the filter with $\mathrm{E}(\boldsymbol{x}_1) = \boldsymbol{0}$ and $\mathrm{cov}(\boldsymbol{x}_1) = \boldsymbol{0}$ and adjusting the log-likelihood function to compensate for this initialization [4]; d) [9] suggested applying the transformation approach of [10] directly to the initial dataset and e) [11] presents a new exact solution for the initialization of the Kalman filter for state space models with diffuse initial conditions. [10] point out that method a) induces numerical errors and method b) cannot be applied to all cases. Method d) provides the appropriate initialization of the

Kalman filter for the remaining dataset but needs modification when missing observations occur in the initial dataset. Methods c), d) and e) does not allow for stochastic inputs.

[12] obtained the exact expressions of $\mathrm{E}(\boldsymbol{x}_1/\boldsymbol{u})$ and $\mathrm{cov}(\boldsymbol{x}_1/\boldsymbol{u})$ for a SS model with unit roots and deterministic inputs, thus providing adequate initial conditions when the inputs are constant terms or dummy variables. In most other cases, however, the exogenous variables in an econometric model should be regarded as stochastic. In this paper we generalize their results by deriving exact and approximate expressions for the conditional mean and variance of the initial state of a SS system, allowing for stochastic inputs. Previous literature, mostly belonging to the engineering field, did not pay attention to this issue, probably due to the fact that in most physical systems inputs can be safely assumed to be either deterministic or controllable. However, it is very relevant when modeling an economic time series.

The outline of the paper is as follows. Section 2 states the problem and defines the basic notation. In Section 3 we derive the exact expressions of $\mathrm{E}(\boldsymbol{x}_1/\boldsymbol{u})$ and $\mathrm{cov}(\boldsymbol{x}_1/\boldsymbol{u})$. In Section 4 we obtain approximations to these moments that both, avoid the need to know the stochastic structure of the inputs and reduce the computational cost of initialization. The performance of the exact and approximate criteria is tested in Section 5, by means of a simulation exercise. Appendices 1 and 2 contain the proof of the main results.

2. The likelihood of a SS model with stochastic inputs

A SS model in steady-state innovations form can be written as:

$$\boldsymbol{y}_t = \boldsymbol{H}\boldsymbol{x}_t + \boldsymbol{D}\boldsymbol{u}_t + \boldsymbol{\epsilon}_t \tag{2}$$

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{\Gamma}\boldsymbol{u}_t + \boldsymbol{E}\boldsymbol{\epsilon}_t \tag{3}$$

where the observation equation (2) generates the $(m \times 1)$ vector of measures $\boldsymbol{y}_t$, for all $t = 1, \ldots, N$, $\boldsymbol{u}_t$ is a $(r \times 1)$ vector of observable inputs and the state equation (3) describes the evolution of the $(n \times 1)$ state vector, $\boldsymbol{x}_t$. Model (2)-(3) can be stationary, non-stationary or partial non-stationary, depending on the eigenvalues of $\boldsymbol{A}$. Also, $\boldsymbol{u}_t$ may include deterministic and/or stochastic inputs. These two issues - stationarity and stochastic nature of the inputs - affect crucially the values of $\mathrm{E}(\boldsymbol{x}_1/\boldsymbol{u})$ and $\mathrm{cov}(\boldsymbol{x}_1/\boldsymbol{u})$ and, therefore, the evaluation of the likelihood function.

We make the following assumptions about (2)-(3):

i) The system is gaussian, *i.e.*, $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$.

ii) The initial state is independent of the sequence $\boldsymbol{\epsilon}_t$ and such that $\boldsymbol{x}_1 / \boldsymbol{U} \sim N(\boldsymbol{\mu}, \boldsymbol{P}_1)$, being $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N]$.

iii) The matrices $\boldsymbol{H}$, $\boldsymbol{D}$, $\boldsymbol{A}$, $\boldsymbol{\Gamma}$, $\boldsymbol{E}$, and $\boldsymbol{\Sigma}_\epsilon$ and the moments $\boldsymbol{\mu}$ and $\boldsymbol{P}_1$ are unknown.

Assuming i), ii) and iii) our purpose consists of computing the likelihood of a vector $\boldsymbol{\theta}$, containing all the parameters in $\boldsymbol{H}$, $\boldsymbol{D}$, $\boldsymbol{A}$, $\boldsymbol{\Gamma}$, $\boldsymbol{E}$ and $\boldsymbol{\Sigma}_\epsilon$, using the data in $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N]$ and $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N]$. Following [4], minus twice the log- likelihood is, apart from constants:

$$\begin{aligned} \ell(\boldsymbol{Z}/\boldsymbol{U}, \boldsymbol{\theta}) &= \log|\boldsymbol{P}_1| + \boldsymbol{\mu}^T \boldsymbol{P}_1^{-1} \boldsymbol{\mu} + \sum_{t=1}^{N} \log|\boldsymbol{B}_t| + \sum_{t=1}^{N} \boldsymbol{e}_t^T \boldsymbol{B}_t^{-1} \boldsymbol{e}_t + \\ &+ \log|\boldsymbol{P}_1^{-1} + \boldsymbol{W}_N| - (\boldsymbol{P}_1^{-1}\boldsymbol{\mu} + \boldsymbol{w}_N)^T (\boldsymbol{P}_1^{-1} + \boldsymbol{W}_N)^{-1} (\boldsymbol{P}_1^{-1}\boldsymbol{\mu} + \boldsymbol{w}_N) \end{aligned} \tag{4}$$

where $\boldsymbol{e}_t$ and $\boldsymbol{B}_t$ $(t = 1, \ldots, N)$ are the innovations and their covariance matrices resulting from a KF(**0**,**0**), *i.e.*, the Kalman filter applied to (2)-(3) with null values for the mean and covariance of the initial state. The values of $\boldsymbol{w}_N$ and $\boldsymbol{W}_N$ can be computed as follows:

$$\boldsymbol{w}_t = \boldsymbol{w}_{t-1} + \boldsymbol{Z}_{t-1}^T \boldsymbol{H}^T \boldsymbol{B}_t^{-1} \boldsymbol{e}_t \tag{5}$$

$$\boldsymbol{W}_t = \boldsymbol{W}_{t-1} + \boldsymbol{Z}_{t-1}^T \boldsymbol{H}^T \boldsymbol{B}_t^{-1} \boldsymbol{H} \boldsymbol{Z}_{t-1} \tag{6}$$

$$\boldsymbol{Z}_t = (\boldsymbol{A} - \boldsymbol{K}_t \boldsymbol{H}) \boldsymbol{Z}_{t-1} \tag{7}$$

where $\boldsymbol{w}_N$ and $\boldsymbol{W}_N$ should be initialized with null values, $\boldsymbol{Z}_0 = \boldsymbol{I}$, and $\boldsymbol{K}_t$ is the gain of the KF(**0**,**0**) applied to (2)-(3).

3. Exact initial conditions for a SS model with stochastic inputs

In this section we derive the exact expressions of the conditional mean and variance of the state $\boldsymbol{x}_1$ for a SS model with stochastic inputs. Assume that the dynamics of the inputs are given by:

$$\boldsymbol{x}_{t+1}^{u} = \boldsymbol{F}\boldsymbol{x}_{t}^{u} + \boldsymbol{G}\boldsymbol{a}_{t} \tag{8}$$

$$\boldsymbol{u}_{t} = \boldsymbol{J}\boldsymbol{x}_{t}^{u} + \boldsymbol{a}_{t} \tag{9}$$

where $\boldsymbol{a}_t$ is a vector of zero-mean random errors with covariance $\boldsymbol{\Sigma}_a$ and independent from $\boldsymbol{\epsilon}_t$ in model (2)-(3). Matrices $\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{J}$ and $\boldsymbol{\Sigma}_a$ are time-invariant and unknown. Following [1, Ch. 3] Eqs. (2)-(3) and (8)-(9) can be written in a single SS system:

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\Phi}\boldsymbol{\alpha}_{t} + \boldsymbol{M}\boldsymbol{\eta}_{t} \tag{10}$$

$$\boldsymbol{z}_{t} = \bar{\boldsymbol{H}}\boldsymbol{\alpha}_{t} + \boldsymbol{N}\boldsymbol{\eta}_{t} \tag{11}$$

where:

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{\Gamma}\boldsymbol{J} \\ \boldsymbol{0} & \boldsymbol{F} \end{bmatrix};\ \boldsymbol{M} = \begin{bmatrix} \boldsymbol{E} & \boldsymbol{\Gamma} \\ \boldsymbol{0} & \boldsymbol{G} \end{bmatrix};\ \bar{\boldsymbol{H}} = \begin{bmatrix} \boldsymbol{H} & \boldsymbol{D}\boldsymbol{J} \\ \boldsymbol{0} & \boldsymbol{J} \end{bmatrix};\ \boldsymbol{N} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}$$
$$\boldsymbol{\eta}_{t} = \begin{bmatrix} \boldsymbol{\epsilon}_{t} \\ \boldsymbol{a}_{t} \end{bmatrix};\ \boldsymbol{\alpha}_{t} = \begin{bmatrix} \boldsymbol{x}_{t} \\ \boldsymbol{x}_{t}^{u} \end{bmatrix};\ \boldsymbol{z}_{t} = \begin{bmatrix} \boldsymbol{y}_{t} \\ \boldsymbol{u}_{t} \end{bmatrix} \tag{12}$$

As $\boldsymbol{\Phi}$ is a square matrix, there exists Ω such that $\boldsymbol{\Phi} = \Omega\,\mathrm{diag}(\boldsymbol{\Phi}^{N}, \boldsymbol{\Phi}^{E})\Omega^{-1}$, or:

$$\boldsymbol{\Phi} = \begin{bmatrix} \Omega^{N} & \Omega^{E} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Phi}^{N} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Phi}^{E} \end{bmatrix} \begin{bmatrix} \boldsymbol{T}^{N} \\ \boldsymbol{T}^{E} \end{bmatrix} \tag{13}$$

where $\boldsymbol{\Phi}^{N}$ and $\boldsymbol{\Phi}^{E}$ are Jordan forms that contain, respectively, the non-stationary and stationary roots of $\boldsymbol{\Phi}$ [13, Ch. 1]. From (13), it is easy to decompose Eq. (10) in two subsystems:

$$\boldsymbol{\alpha}_{t+1}^{N} = \boldsymbol{\Phi}^{N}\boldsymbol{\alpha}_{t}^{N} + \boldsymbol{T}^{N}\boldsymbol{M}\boldsymbol{\eta}_{t} \tag{14}$$

$$\boldsymbol{\alpha}_{t+1}^{E} = \boldsymbol{\Phi}^{E}\boldsymbol{\alpha}_{t}^{E} + \boldsymbol{T}^{E}\boldsymbol{M}\boldsymbol{\eta}_{t} \tag{15}$$

where (14) and (15) represent the evolution of the non-stationary and stationary components of $\boldsymbol{\alpha}_t$, respectively. In these conditions, the following result holds:

Proposition 1: The exact first and second-order moments of the initial state $\boldsymbol{x}_1$ of (2)-(3), conditional on inputs, are:

$$E(\boldsymbol{x}_1/\boldsymbol{U}) \equiv \boldsymbol{\mu} = \boldsymbol{P}_{12}\boldsymbol{P}_{22}^{-1}(\boldsymbol{S} + \boldsymbol{P}_{22}^{-1})^{-1}\boldsymbol{s} \tag{16}$$

$$\text{cov}(\boldsymbol{x}_1/\boldsymbol{U}) \equiv \boldsymbol{P}_1 = \boldsymbol{P}_{11} - \boldsymbol{P}_{12}\boldsymbol{S}(\boldsymbol{S} + \boldsymbol{P}_{22}^{-1})^{-1}\boldsymbol{P}_{22}^{-1}\boldsymbol{P}_{21} \tag{17}$$

where

$$\boldsymbol{S} = \boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{X} \tag{18}$$

$$\boldsymbol{s} = \boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{d} \tag{19}$$

$\boldsymbol{X}$ is a matrix whose block-row is $\boldsymbol{J}\overline{\boldsymbol{F}}_{t-1}$, $\overline{\boldsymbol{F}}_t = (\boldsymbol{F} - \boldsymbol{K}_t\boldsymbol{J})\overline{\boldsymbol{F}}_{t-1}$, starting from $\overline{\boldsymbol{F}}_0 = \boldsymbol{I}$; and $\boldsymbol{K}_t$ is the Kalman filter gain resulting from the application of KF(**0**,**0**) to (8)-(9). We also denote $\boldsymbol{d}$ as the $(r\times N)\times 1$ innovations vector, $\boldsymbol{d}_t$, $\boldsymbol{D}$ is a block-diagonal matrix that contains the covariance matrices of each r innovations and $\boldsymbol{P}_{ij}$ denotes the (i,j) block of the covariance matrix $\boldsymbol{P}$, defined by:

$$\boldsymbol{P} = k\boldsymbol{\Omega}^N(\boldsymbol{\Omega}^N)^T + \boldsymbol{\Omega}^E\boldsymbol{P}^E(\boldsymbol{\Omega}^E)^T \tag{20}$$

being $\boldsymbol{P}^E$ the solution of the Lyapunov equation applied to (15) [7, Ch. 4] and k an arbitrary large constant. The proof of this theorem is given in Appendix 1.

Remarks:

1) Note that the variance of the initial state $(\boldsymbol{x}_1)$ in the augmented model (10)-(11), $\boldsymbol{P}_{11}$, does not coincide with (17), which is smaller. When the problem consists of estimating the unknown parameters in model (2)-(3) this difference is relevant, because if the uncertainty of the initial state is biased upwards, final estimates will be more imprecise than needed. However, when applied to compute forecasts or smoothed values of $\boldsymbol{y}_t$, the augmented model (10)-(11) can be used, yielding results identical to those of model (2)-(3).

2) Expressions (16)-(17) are general, as they can be applied in the stationary, non-stationary and partial non-stationary situations. Note that in the last two cases the direct computation of (16)-(17) may diverge to infinite. However, this is not important for likelihood computation because (4) does not depend on these conditional moments but on $\boldsymbol{P}_1^{-1}$, $\boldsymbol{P}_1^{-1}\boldsymbol{\mu}$ and $\boldsymbol{\mu}^T\boldsymbol{P}_1^{-1}\boldsymbol{\mu}$ and, as the following Proposition states, these terms always converge to finite and easy to compute values. Besides, $\boldsymbol{P}_1^{-1}$ exists but is rank-deficient [12].

3) It is well known that Jordan decomposition is numerically unstable, see *e.g.*, [14]. However, it is not necessary to compute $\boldsymbol{P}^{-1}$, as this only requires block-diagonalizing the transition matrix $\boldsymbol{\Phi}$. A stable and efficient procedure to do this consists of: a) applying a real Schur decomposition, to obtain a triangular form and b) obtaining the block-diagonal matrix by solving a set of Sylvester equations, see [13]. As $\boldsymbol{\Phi}_N^*$ and $\boldsymbol{\Phi}_E^*$ have no common eigenvalues, the solution of these equations exists.

Proposition 2: The terms $\boldsymbol{P}_1^{-1}$, $\boldsymbol{P}_1^{-1}\boldsymbol{\mu}$ and $\boldsymbol{\mu}^T\boldsymbol{P}_1^{-1}\boldsymbol{\mu}$ in (4) can be computed as:

$$\boldsymbol{P}_1^{-1} = \boldsymbol{V}_{11} - \boldsymbol{V}_{12}(\boldsymbol{S} + \boldsymbol{V}_{22})^{-1}\boldsymbol{V}_{21} \tag{21}$$

$$\boldsymbol{P}_1^{-1}\boldsymbol{\mu} = -\boldsymbol{V}_{12}(\boldsymbol{S} + \boldsymbol{V}_{22})^{-1}\boldsymbol{s} \tag{22}$$

$$\boldsymbol{\mu}^T\boldsymbol{P}_1^{-1}\boldsymbol{\mu} = \boldsymbol{s}^T(\boldsymbol{S} + \boldsymbol{V}_{22})^{-1}\boldsymbol{V}_{21}(\boldsymbol{P}_1^{-1})^{+}\boldsymbol{V}_{12}(\boldsymbol{S} + \boldsymbol{V}_{22})^{-1}\boldsymbol{s} \tag{23}$$

where $\boldsymbol{V}_{ij}$ is the (i,j) block of a matrix $\boldsymbol{V}$ such that $\boldsymbol{V} = \boldsymbol{P}^{-1}$ and $(\boldsymbol{P}_1^{-1})^{+}$ denotes the Moore-Penrose pseudo-inverse of $\boldsymbol{P}_1^{-1}$ and the terms $\boldsymbol{s}$ and $\boldsymbol{S}$ are defined in (18)-(19). Note that the left-hand-sides of (21)-(23) are finite even if (10)-(11) has unit roots. The proof of this result is in Appendix 2.

Building on Propositions 1 and 2, it is easy to derive specific expressions for the exact first and second-order moments conditional on inputs of stationary, non-stationary and partial non-stationary models:

1) In the stationary case, $\Omega^N = \boldsymbol{0}$ and $\Omega^E = \boldsymbol{I}$. Then: a) $\boldsymbol{P} = \boldsymbol{P}^E$, see Eq. (20), b) results (16)-(17) hold and c) the conditional mean and variance depend on the input model. For example, if the inputs follow a white noise process, it is easy to see that $\boldsymbol{\mu} = \boldsymbol{0}$ and $\boldsymbol{P}_1 = \boldsymbol{A}\boldsymbol{P}_1\boldsymbol{A}^T + \boldsymbol{E}\Sigma_\epsilon\boldsymbol{E}^T + \Gamma\Sigma_a\Gamma^T$.

2) In the pure non-stationary case, $\Omega^{E} = \mathbf{0}$, $\Omega^{N} = \boldsymbol{I}$ and $\boldsymbol{P} = k\boldsymbol{I}$ where k is a large constant, see Eq. (20). Then, expressions (21)-(23) collapse to zero. Since the likelihood function depends on (21)-(23), it is not necessary to choose an arbitrary value of k which could degrade the Kalman filter [4]. However, the term $\log|\boldsymbol{P}_1|$ in (4) cannot be computed. In this situation the log-likelihood function can be approximated by the limit when $\boldsymbol{P}_1$ tends to infinity of $\ell(\boldsymbol{Z}/\boldsymbol{U},\boldsymbol{\theta}) - \log|\boldsymbol{P}_1|$, which yields:

$$\ell(\boldsymbol{Z}/\boldsymbol{U},\theta) \simeq \sum_{t=1}^{N} \log|\boldsymbol{B}_t| + \sum_{t=1}^{N} \boldsymbol{e}_t^T \boldsymbol{B}_t^{-1} \boldsymbol{e}_t + \log|\boldsymbol{W}_N| - \boldsymbol{w}_N^T \boldsymbol{W}_N^{-1} \boldsymbol{w}_N$$

3) In the partial non-stationary case, *i.e.*, when some roots of $\boldsymbol{\Phi}$ are stationary and others are not, $\boldsymbol{\mu}$ and $\boldsymbol{P}_1^{-1}$ are finite and different from zero. This result coincides with that obtained by [12] in the deterministic input case. The expression of $\boldsymbol{P}_1^{-1}$ is given by (21) and the conditional mean is given by (16). About the determinant of $\boldsymbol{P}_1$ in (4), it is easy to compute the part corresponding to the stationary roots by setting $\Omega^{N} = \mathbf{0}$ in (20).

4. Approximate initial conditions of a SS model

Expressions (16)-(17) require knowledge of both, the model generating the inputs and its parameters. In this section we extend the previous framework by obtaining approximate initial conditions that can be applied when the model for the inputs is unknown. Besides, the computational cost of this initialization is smaller than that of the exact method. To do this, we reduce (2)-(3), into a deterministic subsystem:

$$\begin{aligned} \boldsymbol{x}_{t+1}^{d} &= \boldsymbol{A}\boldsymbol{x}_t^{d} + \Gamma \boldsymbol{u}_t \\ \boldsymbol{y}_t^{d} &= \boldsymbol{H}\boldsymbol{x}_t^{d} + \boldsymbol{D}\boldsymbol{u}_t \end{aligned} \tag{24}$$

and a stochastic subsystem:

$$\begin{aligned} \boldsymbol{x}_{t+1}^{s} &= \boldsymbol{A}\boldsymbol{x}_t^{s} + \boldsymbol{E}\epsilon_t \\ \boldsymbol{y}_t^{s} &= \boldsymbol{H}\boldsymbol{x}_t^{s} + \epsilon_t \end{aligned} \tag{25}$$

where $\boldsymbol{x}_{t+1} = \boldsymbol{x}_{t+1}^{d} + \boldsymbol{x}_{t+1}^{s}$, $\boldsymbol{y}_t = \boldsymbol{y}_t^{d} + \boldsymbol{y}_t^{s}$ and we assume that $\mathrm{cov}(\boldsymbol{x}_t^{s}\boldsymbol{x}_t^{d}) = \mathbf{0}$ for all t [15]. This hypothesis is not very restrictive, as the independence between $\boldsymbol{x}_1$ and ϵ_t implies that a sufficient condition for $\operatorname{cov}(\boldsymbol{x}_t^s \boldsymbol{x}_t^d) = \mathbf{0}$ to hold is $\operatorname{cov}(\boldsymbol{x}_1^s \boldsymbol{x}_1^d) = \mathbf{0}$.

From (24)-(25) it is immediate that:

$$\boldsymbol{\mu} = \boldsymbol{\mu}^d + \boldsymbol{\mu}^s \tag{26}$$

$$\boldsymbol{P}_1 = \boldsymbol{P}_1^d + \boldsymbol{P}_1^s \tag{27}$$

where $\boldsymbol{\mu}^d = \mathrm{E}(\boldsymbol{x}_1^d / \boldsymbol{U})$, $\boldsymbol{\mu}^s = \mathrm{E}(\boldsymbol{x}_1^s)$. Then, the problem reduces to computing estimates of $\boldsymbol{\mu}^s$, $\boldsymbol{\mu}^d$, $\boldsymbol{P}_1^s$ and $\boldsymbol{P}_1^d$.

As for the stochastic subsystem, the immemorial time argument assures that an adequate initialization is $\boldsymbol{\mu}^s = \mathbf{0}$ and the solution for $\boldsymbol{P}_1^s$ depends on the existence of unit roots in (25) [12, Theorem 3]. We use the definition of immemorial time given by [12, pp.154], *i.e.*, a SS model is said to have applied since time immemorial if the state transition equation (3) is assumed to hold for $t = 0, -1, -2, \ldots$, where $\boldsymbol{x}_r = \mathbf{0}$ and $r \to -\infty$.

In the most general case, when the matrix $\boldsymbol{A}$ in (25) has both stationary and non-stationary roots, $(\boldsymbol{P}_1^s)^{-1} = (\Psi^E)^T (\boldsymbol{P}^E)^{-1} \Psi^E$, where $\boldsymbol{P}^E$ satisfies $\boldsymbol{P}^E = \boldsymbol{A}^E \boldsymbol{P}^E (\boldsymbol{A}^E)^T + \Psi^E \boldsymbol{E} \Sigma_\epsilon \boldsymbol{E}^T (\Psi^E)^T$ and there exists a matrix $\boldsymbol{R}$ such that $\boldsymbol{A} = \boldsymbol{R}\,\mathrm{diag}(\boldsymbol{A}^N, \boldsymbol{A}^E)\boldsymbol{R}^{-1}$ or alternatively:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{R}^N & \boldsymbol{R}^E \end{bmatrix} \begin{bmatrix} \boldsymbol{A}^N & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A}^E \end{bmatrix} \begin{bmatrix} \psi^N \\ \psi^E \end{bmatrix} \tag{28}$$

where $\boldsymbol{A}^N$ and $\boldsymbol{A}^E$ are Jordan forms that contain the non-stationary and stationary roots of $\boldsymbol{A}$, respectively. In these conditions $(\boldsymbol{P}_1^s)^{-1}$ is always finite.

In the deterministic system (24) we assume that the initial state is a fixed and unknown value, $\boldsymbol{x}_1^d$, with $\boldsymbol{P}_1^d = \mathbf{0}$. Maximum likelihood estimates of $\boldsymbol{\mu}^d$ can be computed using the expressions $\boldsymbol{\mu}^d = \boldsymbol{W}_N^{-1} \boldsymbol{w}_N$, where $\boldsymbol{W}_N$ and $\boldsymbol{w}_N$ result from the propagation of (5)-(7) [4; 1].
In this case, the likelihood (4) can be concentrated with respect to $\boldsymbol{\mu}^d$ and $\boldsymbol{P}_1^d$. Substituting this expression back into (4) yields the concentrated likelihood:

$$\log|\boldsymbol{I} + \boldsymbol{P}_1^{d}\boldsymbol{W}| + \sum_{t=1}^{N}\log|\boldsymbol{B}_t| + \sum_{t=1}^{N}\tilde{\boldsymbol{z}}_t^{T}\boldsymbol{B}_t^{-1}\tilde{\boldsymbol{z}}_t - \boldsymbol{w}^{T}\boldsymbol{W}^{-1}\boldsymbol{w} \tag{29}$$

Therefore, the approximation consists of a) computing $\boldsymbol{\mu}^s$ using the immemorial time argument, b) computing $\boldsymbol{\mu}^d$ by maximum likelihood and c) using (26)-(27) to calculate the approximate initial conditions for the whole system. The resulting initial conditions are again different, according to the stationary or non-stationary nature of the model:

1) If (2)-(3) is stationary, both the deterministic and stochastic subsystems are stationary. Hence according to (28), $\boldsymbol{R}^N = \boldsymbol{0}$, $\boldsymbol{R}^E = \boldsymbol{I}$, $\boldsymbol{A}^E = \boldsymbol{A}$ and the approximation to the initial conditions is given by:

$$\boldsymbol{\mu} = \boldsymbol{W}_N^{-1}\boldsymbol{w}_N \tag{30}$$

$$\boldsymbol{P}_1 = \boldsymbol{A}\boldsymbol{P}_1\boldsymbol{A}^T + \boldsymbol{E}\boldsymbol{\Sigma}_\epsilon\boldsymbol{E}^T \tag{31}$$

Note that expressions (30)-(31) do not depend on the parameters of the model for the inputs (8)-(9).

2) When (2)-(3) is non-stationary, the conditional mean of the initial state is $\boldsymbol{\mu} = \boldsymbol{W}_N^{-1}\boldsymbol{w}_N$ and the inverse of the corresponding covariance is $\boldsymbol{P}_1^{-1} = \boldsymbol{0}$. This result coincides with that obtained in [12].

3) The starting conditions of a partial non-stationary model are $\boldsymbol{\mu} = \boldsymbol{W}_N^{-1}\boldsymbol{w}_N$ and $\boldsymbol{P}_1^{-1} = (\boldsymbol{\Psi}^{\boldsymbol{E}})^T(\boldsymbol{P}^E)^{-1}\boldsymbol{\Psi}^{\boldsymbol{E}}$, where $\boldsymbol{P}^E$ satisfies $\boldsymbol{P}^E = \boldsymbol{A}^E\boldsymbol{P}^E(\boldsymbol{A}^E)^T + \boldsymbol{\Psi}^{\boldsymbol{E}}\boldsymbol{E}\boldsymbol{\Sigma}_\epsilon\boldsymbol{E}^T(\boldsymbol{\Psi}^{\boldsymbol{E}})^T$, and $\boldsymbol{\Psi}^{\boldsymbol{E}}$ is defined in (28).

The idea to derive adequate initial conditions under conditions 1), 2) or 3), consists of reducing (2)-(3) into (24)-(25) and then applying the immemorial time argument. On the other hand, the correct initialization of a model with no stochastic errors, depends only of the deterministic or stochastic nature of the inputs. If these are stochastic, the solution consists of obtaining maximum likelihood estimates of the initial state for the subsystem.

5. Results with simulated data

In this section we test the initialization procedures proposed in Sections 3 and 4, using simulated data. The samples were obtained with the random number generator of MATLAB, initialized with the default seed, determined by the system clock, and discarding the first 100 values no matter the sample size.

Tables 1 and 2 show the average estimates obtained with 1000 replications, the sample standard errors and mean squared error (MSE) of the parameters of a stationary transfer function for different sample sizes. We assume that the input variable follows first an AR(1) process and afterwards a seasonal AR(1) process. We use three initialization methods of the Kalman filter: the exact moments given by (16)-(17); the standard initial conditions $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{P}_1 = \boldsymbol{A}\boldsymbol{P}_1\boldsymbol{A}^T + \boldsymbol{E}\Sigma_\epsilon\boldsymbol{E}^T$, which ignore the stochastic structure of the input [12] and the approximate initial conditions (30)-(31). The main conclusions this experiment are the following:

1) When the sample size is small (30 or 50 observations) the estimates computed with exact initial conditions are better than those corresponding to the alternative procedures yielding a lower MSE.
2) The estimates become less sensitive to starting conditions as the sample size increases.
3) The approximate initial conditions behave well for all sample sizes. These results suggest that the approximate criterion does not differ substantially from the exact initialization procedure.
4) When the model is seasonal (see Table 2) the importance of using exact initial conditions increases but the approximation proposed in Section 4 continues performing very well.

[Table 1]

[Table 2]

Acknowledgments. We acknowledge the helpful comments of an anonymous referee. Sonia Sotoca acknowledges financial support from the Fundación Caja de Madrid and the CICYT project PB95-0912/95 (nº 5290806).

Appendix 1: Proof of expressions (16)-(17).

From (14)-(15), [12] prove that:

$$\boldsymbol{\alpha}_1^* = \begin{bmatrix} \boldsymbol{\alpha}_1^N \\ \boldsymbol{\alpha}_1^E \end{bmatrix} \sim N\left\{ \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} k\boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{P}^E \end{bmatrix} \right\} \tag{A.1}$$

Therefore, the distribution of the initial state is $\Omega\boldsymbol{\alpha}_1^* \sim N(\mathbf{0}, \boldsymbol{P})$ where the blocks of $\boldsymbol{P}$ can be written as:

$$\boldsymbol{P}_{11} = k\boldsymbol{\Omega}_1\boldsymbol{\Omega}_1^T + \boldsymbol{\Omega}_3\boldsymbol{P}^E\boldsymbol{\Omega}_3^T \;\;;\;\; \boldsymbol{P}_{12} = \boldsymbol{P}_{21}^T = k\boldsymbol{\Omega}_1\boldsymbol{\Omega}_2^T + \boldsymbol{\Omega}_3\boldsymbol{P}^E\boldsymbol{\Omega}_4^T$$
$$\boldsymbol{P}_{22} = k\boldsymbol{\Omega}_2\boldsymbol{\Omega}_2^T + \boldsymbol{\Omega}_4\boldsymbol{P}^E\boldsymbol{\Omega}_4^T \tag{A.2}$$

where $(\Omega^N)^T = [\boldsymbol{\Omega}_1^T \; \boldsymbol{\Omega}_2^T]$ and $(\Omega^E)^T = [\boldsymbol{\Omega}_3^T \; \boldsymbol{\Omega}_4^T]$, see Eq. (20).

It is easy to prove [1] that:

$$\boldsymbol{d}_t = [\mathbf{0} \;\; \boldsymbol{X}] \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_1^u \end{bmatrix} + \tilde{\boldsymbol{d}}_t \tag{A.3}$$

being $\boldsymbol{d}_t$ the innovations resulting from a KF($\mathbf{0}$,$\mathbf{0}$) applied to (8)-(9) and $\tilde{\boldsymbol{d}}_t$ those corresponding to a process generated from a zero initial state and covariance matrix. The block-row of matrix $\boldsymbol{X}$ is $\boldsymbol{J}\bar{\boldsymbol{F}}_{t-1}$ and $\bar{\boldsymbol{F}}_t = (\boldsymbol{F} - \boldsymbol{K}_t\boldsymbol{J})\bar{\boldsymbol{F}}_{t-1}$ starting from $\bar{\boldsymbol{F}}_0 = \boldsymbol{I}$ and $\boldsymbol{K}_t$ is the KF($\mathbf{0}$,$\mathbf{0}$) gain. If we define $\boldsymbol{d}$ as the $(r \times N) \times 1$ vector of innovations $\boldsymbol{d}_t$ and $\boldsymbol{D}$ as the block-diagonal matrix that contains the covariance matrices of each r innovations, the whole system can be written as:

$$\begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ \boldsymbol{X}^* & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \tilde{\boldsymbol{d}} \end{bmatrix} \tag{A.4}$$

where $\boldsymbol{X}^* = [\mathbf{0} \;\; \boldsymbol{X}]$ and $\boldsymbol{\alpha}_1^T = [\boldsymbol{x}_1^T \;\; (\boldsymbol{x}_1^u)^T]^T$. Then, the joint distribution of $[\boldsymbol{\alpha}_1^T \;\; \boldsymbol{d}^T]^T$ is, again, normal with zero mean and covariance matrix:

$$\begin{bmatrix} \boldsymbol{P} & \boldsymbol{P}^{T}(\boldsymbol{X}^{*})^{T} \\ \boldsymbol{X}^{*}\boldsymbol{P} & \boldsymbol{X}^{*}\boldsymbol{P}(\boldsymbol{X}^{*})^{T}+\boldsymbol{D} \end{bmatrix} \tag{A.5}$$

Under these conditions, the conditional mean and covariance of $\boldsymbol{\alpha}_1$ are:

$$\mathrm{E}(\boldsymbol{\alpha}_1/\tilde{\boldsymbol{d}}) = \boldsymbol{P}^{T}(\boldsymbol{X}^{*})^{T}\{\boldsymbol{D} + \boldsymbol{X}^{*}\boldsymbol{P}(\boldsymbol{X}^{*})^{T}\}^{-1}\boldsymbol{d} \tag{A.6}$$

$$\mathrm{cov}(\boldsymbol{\alpha}_1/\tilde{\boldsymbol{d}}) = \boldsymbol{P} - \boldsymbol{P}(\boldsymbol{X}^{*})^{T}\{\boldsymbol{D} + \boldsymbol{X}^{*}\boldsymbol{P}(\boldsymbol{X}^{*})^{T}\}^{-1}\boldsymbol{X}^{*}\boldsymbol{P} \tag{A.7}$$

Taking into account that $\boldsymbol{\alpha}_1^{\boldsymbol{T}} = [\boldsymbol{x}_1^{\boldsymbol{T}} \quad (\boldsymbol{x}_1^{\boldsymbol{u}})^{T}]^{T}$, the exact expressions for the conditional mean and covariance would be:

$$\mathrm{E}(\boldsymbol{x}_1/\boldsymbol{U}) \equiv \boldsymbol{\mu} = \boldsymbol{P}_{12}\boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{D} + \boldsymbol{X}\boldsymbol{P}_{22}\boldsymbol{X}^{\boldsymbol{T}})^{-1}\boldsymbol{d} \tag{A.8}$$

$$\mathrm{cov}(\boldsymbol{x}_1/\boldsymbol{U}) \equiv \boldsymbol{P}_1 = \boldsymbol{P}_{11} - \boldsymbol{P}_{12}\boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{D} + \boldsymbol{X}\boldsymbol{P}_{22}\boldsymbol{X}^{\boldsymbol{T}})^{-1}\boldsymbol{X}\boldsymbol{P}_{21} \tag{A.9}$$

and the application of the matrix inversion lemma to (A.8)-(A.9), yields (16)-(17). ■

Appendix 2: Proof of expressions (21)-(23).

Let be Π a symmetric matrix such that:

$$\Pi = \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}_{11} & \boldsymbol{P}_{12} \\ \boldsymbol{P}_{21} & \boldsymbol{P}_{22} + \boldsymbol{S}^{-1} \end{bmatrix}^{-1} \tag{A.10}$$

where the blocks $\boldsymbol{P}_{11}$, $\boldsymbol{P}_{12}$, $\boldsymbol{P}_{21}$ and $\boldsymbol{P}_{22}$ are defined in (A.1) and $\boldsymbol{S}$ is defined in (18). Applying the inverse of partitioned matrices to (A.10) it holds that:

$$\Pi_{11} = \{\boldsymbol{P}_{11} - \boldsymbol{P}_{12}(\boldsymbol{S}^{-1} + \boldsymbol{P}_{22})^{-1}\boldsymbol{P}_{21}\}^{-1} \tag{A.11}$$

$$\Pi_{12} = -\Pi_{11}\boldsymbol{P}_{12}(\boldsymbol{S}^{-1} + \boldsymbol{P}_{22})^{-1} \tag{A.12}$$

and applying again the inverse of partitioned matrices, Eqs. (16)-(17) can be written as:

$$\boldsymbol{\mu} = \boldsymbol{P}_{12}(\boldsymbol{S}^{-1} + \boldsymbol{P}_{22})^{-1}\boldsymbol{S}^{-1}\boldsymbol{s} \tag{A.13}$$

$$\boldsymbol{P}_{1} = \boldsymbol{P}_{11} - \boldsymbol{P}_{12}(\boldsymbol{S}^{-1} + \boldsymbol{P}_{22})^{-1}\boldsymbol{P}_{21} \tag{A.14}$$

Comparing (A.11)-(A.12) with (A.13)-(A.14) we obtain:

$$\boldsymbol{P}_{1}^{-1} = \Pi_{11} \tag{A.15}$$

$$\boldsymbol{P}_{1}^{-1}\boldsymbol{\mu} = \Pi_{11}\boldsymbol{P}_{12}(\boldsymbol{S}^{-1} + \boldsymbol{P}_{22})^{-1}\boldsymbol{S}^{-1}\boldsymbol{s} = -\Pi_{12}\boldsymbol{S}^{-1}\boldsymbol{s} \tag{A.16}$$

$$\boldsymbol{\mu}^{T}\boldsymbol{P}_{1}^{-1}\boldsymbol{\mu} = \boldsymbol{s}^{T}\boldsymbol{S}^{-1}\Pi_{21}\Pi_{11}^{+}\Pi_{12}\boldsymbol{S}^{-1}\boldsymbol{s} \tag{A.17}$$

where Π_{11}^{+} is the Moore-Penrose pseudo-inverse of Π_{11}, such that $\Pi_{11} = \Pi_{11}\Pi_{11}^{+}\Pi_{11}$ and $\boldsymbol{s}$ is defined in (19). By Theorem 3 of [12], the inverse of $\boldsymbol{P} = k\Omega^{N}(\Omega^{N})^{T} + \Omega^{E}\boldsymbol{P}^{E}(\Omega^{E})^{T}$ exists and is non-zero, see Eq. (20). Denoting $\boldsymbol{V} = \boldsymbol{P}^{-1}$, such that:

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}_{11} & \boldsymbol{P}_{12} \\ \boldsymbol{P}_{21} & \boldsymbol{P}_{22} \end{bmatrix}^{-1} = \boldsymbol{P}^{-1} \tag{A.18}$$

it is easy to prove that the blocks of Π in (A.10) and those of $\boldsymbol{V}$ in (A.18) are related by the following equalities:

$$\Pi_{11} = \boldsymbol{V}_{11} - \boldsymbol{V}_{12}(\boldsymbol{S} + \boldsymbol{V}_{22})^{-1}\boldsymbol{V}_{21} \tag{A.19}$$

$$\Pi_{12} = \boldsymbol{V}_{12}(\boldsymbol{S} + \boldsymbol{V}_{22})^{-1}\boldsymbol{S} \tag{A.20}$$

$$\Pi_{22} = \boldsymbol{V}_{22}(\boldsymbol{S} + \boldsymbol{V}_{22})^{-1}\boldsymbol{S} \tag{A.21}$$

and substituting (A.19)-(A.21) in (A.15)-(A.17), we obtain:

$$\boldsymbol{P}_1^{-1} = \boldsymbol{V}_{11} - \boldsymbol{V}_{12}(\boldsymbol{S} + \boldsymbol{V}_{22})^{-1}\boldsymbol{V}_{21} \tag{A.22}$$

$$\boldsymbol{P}_1^{-1}\boldsymbol{\mu} = -\boldsymbol{V}_{12}(\boldsymbol{S} + \boldsymbol{V}_{22})^{-1}\boldsymbol{s} \tag{A.23}$$

$$\boldsymbol{\mu}^T\boldsymbol{P}_1^{-1}\boldsymbol{\mu} = \boldsymbol{s}^T(\boldsymbol{S} + \boldsymbol{V}_{22})^{-1}\boldsymbol{V}_{21}\Pi_{11}^{+}\boldsymbol{V}_{12}(\boldsymbol{S} + \boldsymbol{V}_{22})^{-1}\boldsymbol{s} \tag{A.24}$$

where Π_{11}^{+} exists and is nonzero. Then (A.22)-(A.24) coincide with (21)-(23), respectively, and both quantities are always finite and non-zero. ■

References.

1. J. Terceiro, *Estimation of Dynamic Econometric Models with Errors in Variables*, Springer-Verlag, Heidelberg (1990).

2. A.C. Harvey, E. Ruiz and E. Sentana, Unobserved component time series models with ARCH disturbances. *Journal of Econometrics* **52** 129-157 (1992).

3. P.A.V.B. Swamy and G.S. Tavlas, Random coefficient models: theory and applications. *Journal of Economic Surveys* **9** 165-196 (1995).

4. P. De Jong, The likelihood for a state space model. *Biometrika* **75** 165-169 (1988).

5. F.C. Schweppe, Evaluation of likelihood functions for gaussian signals. *IEEE Transactions on Information Theory* **IT-11** 61-70 (1965).

6. A.C. Harvey and G.D.A. Phillips, Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika* **66** 49-58 (1979).

7. B.D.O. Anderson and J.B. Moore, *Optimal Filtering*, Prentice- Hall, New Jersey (1979).

8. G. Kitagawa, A nonstationary time series models and its fitting by a recursive filter. *Journal of Time Series Analysis* **81** 751-761 (1981).

9. W. Bell and S. Hillmer, Initializing the Kalman filter for nonstationary time series models. *Journal of Time Series Analysis* **12** 283-300 (1991).

10. C.F. Ansley and R. Kohn, Estimation, filtering and smoothing in state space models with incompletely specified initial conditions. *The Annals of Statistics* **13** 1286-1316 (1985).

11. S. J. Koopman, Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association* **92** 1630-1638 (1997).

12. P. De Jong and S. Chu-Chun-Lin, Stationary and non-stationary state space models. *Journal of Time Series Analysis* **15** 151-166 (1994).

13. P. Hr. Petkov, N.D. Christov and M.M. Konstantinov, *Computational Methods for Linear Control Systems*, Prentice-Hall, New Jersey (1991).

14. G.H. Golub and C.F. Van Loan, *Matrix Computations*, John Hopkins University Press (1996).

15. P. Van Overschee and B. De Moor, N4SID*: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica* **30** 75-93 (1994).

Table 1: Summary of simulation results of a state space model with stochastic input. Theoretical values: $\phi = 0.8$, $\omega_0 = 0.6$, $\omega_1 = 0.3$, $\delta_1 = 0.5$, $\sigma_\epsilon^2 = 1$ and $\sigma_a^2 = 0.1$. Replications: 1000. Model:

$$(1 - \phi B) u_t = \epsilon_t \quad ; \quad y_t = \frac{(\omega_0 + \omega_1 B)}{1 - \delta_1 B} u_t + a_t$$

Exact Initial Conditions [see equations (16)-(17)]

N	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\delta}_1$	$\hat{\sigma}_a^2$	$MSE(\hat{\omega}_0)$	$MSE(\hat{\omega}_1)$	$MSE(\hat{\delta}_1)$	$MSE(\hat{\sigma}_a^2)$
30	0.603 (0.063)	0.304 (0.090)	0.494 (0.052)	0.089 (0.024)	3.997	8.752	2.769	0.721
50	0.600 (0.048)	0.304 (0.071)	0.497 (0.041)	0.092 (0.018)	2.316	5.11	1.667	0.402
100	0.601 (0.032)	0.305 (0.048)	0.497 (0.026)	0.098 (0.014)	1.039	2.276	0.686	0.197
200	0.600 (0.023)	0.302 (0.034)	0.498 (0.019)	0.099 (0.010)	0.511	1.142	0.348	0.107

Standard Initial Conditions: $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{P_1} = \boldsymbol{A P_1 A}^T + \boldsymbol{E \Sigma_\epsilon E}^T$

N	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\delta}_1$	$\hat{\sigma}_a^2$	$MSE(\hat{\omega}_0)$	$MSE(\hat{\omega}_1)$	$MSE(\hat{\delta}_1)$	$MSE(\hat{\sigma}_a^2)$
30	0.650 (0.096)	0.215 (0.153)	0.533 (0.080)	0.156 (0.090)	11.763	30.579	7.521	11.19
50	0.631 (0.069)	0.250 (0.108)	0.519 (0.054)	0.134 (0.062)	5.721	14.225	3.24	4.965
100	0.617 (0.042)	0.278 (0.062)	0.507 (0.030)	0.119 (0.034)	2.067	4.251	0.943	1.501
200	0.609 (0.026)	0.288 (0.039)	0.504 (0.020)	0.110 (0.019)	0.749	1.665	0.917	0.461

Approximate Initial Conditions: $\boldsymbol{\mu} = \boldsymbol{W_N^{-1} w_N}$, $\boldsymbol{P_1} = \boldsymbol{A P_1 A}^T + \boldsymbol{E \Sigma_\epsilon E}^T$

N	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\delta}_1$	$\hat{\sigma}_a^2$	$MSE(\hat{\omega}_0)$	$MSE(\hat{\omega}_1)$	$MSE(\hat{\delta}_1)$	$MSE(\hat{\sigma}_a^2)$
30	0.603 (0.064)	0.306 (0.093)	0.492 (0.052)	0.086 (0.024)	4.051	8.704	2.792	0.761
50	0.600 (0.048)	0.305 (0.071)	0.496 (0.041)	0.090 (0.018)	2.318	5.109	1.668	0.421
100	0.601 (0.032)	0.306 (0.047)	0.496 (0.026)	0.097 (0.014)	1.038	2.28	0.686	0.199
200	0.600 (0.023)	0.302 (0.034)	0.498 (0.019)	0.098 (0.010)	0.512	1.421	0.349	0.108

N: Sample size
Standard errors in brackets
Mean Square Error (*MSE*) multiplied by 1000

Table 2: Summary of simulation results of a seasonal state space model with stochastic input. Theoretical values: $\phi = 0.8$, $\omega_0 = 0.6$, $\omega_1 = 0.3$, $\delta_1 = 0.5$, $\sigma^2_\epsilon = 1$ and $\sigma^2_a = 0.1$. Replications: 1000. Model:

$$(1 - \phi B^4) u_t = \epsilon_t \quad ; \quad y_t = \frac{(\omega_0 + \omega_1 B)}{1 - \delta_1 B^4} u_t + a_t$$

Exact Initial Conditions [see equations (16)-(17)]

N	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\delta}_1$	$\hat{\sigma}^2_a$	$MSE(\hat{\omega}_0)$	$MSE(\hat{\omega}_1)$	$MSE(\hat{\delta}_1)$	$MSE(\hat{\sigma}^2_a)$
30	0.608 (0.056)	0.303 (0.047)	0.492 (0.050)	0.088 (0.027)	3.146	2.23	2.56	0.848
50	0.604 (0.041)	0.300 (0.035)	0.496 (0.039)	0.093 (0.020)	1.707	1.217	1.541	0.453
100	0.560 (0.028)	0.300 (0.023)	0.500 (0.025)	0.098 (0.014)	0.78	0.506	0.62	0.203
200	0.600 (0.019)	0.300 (0.016)	0.499 (0.018)	0.098 (0.010)	0.372	0.251	0.31	0.102

Standard Initial Conditions: $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{P_1} = \boldsymbol{A P_1 A}^{\boldsymbol{T}} + \boldsymbol{E} \boldsymbol{\Sigma}_\epsilon \boldsymbol{E}^{\boldsymbol{T}}$

N	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\delta}_1$	$\hat{\sigma}^2_a$	$MSE(\hat{\omega}_0)$	$MSE(\hat{\omega}_1)$	$MSE(\hat{\delta}_1)$	$MSE(\hat{\sigma}^2_a)$
30	0.679 (0.098)	0.334 (0.069)	0.465 (0.080)	0.183 (0.084)	15.725	5.917	7.6	13.93
50	0.641 (0.061)	0.315 (0.045)	0.484 (0.053)	0.152 (0.055)	5.438	2.244	3.06	5.717
100	0.617 (0.035)	0.307 (0.026)	0.494 (0.029)	0.132 (0.035)	1.51	0.699	0.892	2.18
200	0.609 (0.022)	0.304 (0.017)	0.496 (0.019)	0.115 (0.018)	0.573	0.307	0.386	0.536

Approximate Initial Conditions: $\boldsymbol{\mu} = \boldsymbol{W_N}^{-1} \boldsymbol{w_N}$, $\boldsymbol{P_1} = \boldsymbol{A P_1 A}^{\boldsymbol{T}} + \boldsymbol{E} \boldsymbol{\Sigma}_\epsilon \boldsymbol{E}^{\boldsymbol{T}}$

N	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\delta}_1$	$\hat{\sigma}^2_a$	$MSE(\hat{\omega}_0)$	$MSE(\hat{\omega}_1)$	$MSE(\hat{\delta}_1)$	$MSE(\hat{\sigma}^2_a)$
30	0.610 (0.058)	0.304 (0.049)	0.488 (0.053)	0.076 (0.023)	3.474	2.423	2.985	1.119
50	0.606 (0.042)	0.301 (0.036)	0.493 (0.040)	0.085 (0.019)	1.811	1.305	1.65	0.565
100	0.601 (0.028)	0.301 (0.023)	0.498 (0.025)	0.094 (0.014)	0.8	0.521	0.627	0.222
200	0.601 (0.019)	0.301 (0.016)	0.499 (0.018)	0.096 (0.010)	0.375	0.256	0.315	0.11

N: Sample size
Standard errors in brackets
Mean Square Error (*MSE*) multiplied by 1000