Documento de trabajo

A Fast and Stable Method to Compute the Likelihood of State-Space Models with Unit Roots

José Casal
Sonia Sotoca
Miguel Jerez

No.9901 Junio 1999

ICAE
Instituto Complutense de Análisis Económico
UNIVERSIDAD COMPLUTENSE
FACULTAD DE ECONOMICAS
Campus de Somosaguas
28223 MADRID

Teléfono 91 394 26 11 - FAX 91 294 26 13
Internet: http://www.ucm.es/info/icae/
E-mail: icaesec@ccee.ucm.es

ICAE
Instituto Complutense de Análisis Económico
UNIVERSIDAD COMPLUTENSE

A FAST AND STABLE METHOD TO COMPUTE THE LIKELIHOOD OF STATE-SPACE MODELS WITH UNIT ROOTS

José Casals
Sonia Sotoca
Miguel Jerez

Universidad Complutense de Madrid
Campus de Somosaguas
28223 Madrid

ABSTRACT

We propose two fast and stable methods to compute the likelihood of econometric models in state-space form, allowing for unit roots. The first one exploits the properties of the Kalman filter when applied to models in steady-state innovations form. Afterwards we derive a procedure with similar properties that can be applied to any state-space model satisfying weak assumptions.

RESUMEN

En este trabajo se proponen dos métodos rápidos y eficientes para evaluar la función de verosimilitud de modelos econométricos en forma de espacio de los estados, permitiendo raíces unitarias. El primero de ellos aprovecha las propiedades del filtro de Kalman cuando se aplica a modelos en forma *steady-state innovations*. Posteriormente se deriva un procedimiento con propiedades similares que puede aplicarse a cualquier modelo en espacio de los estados que satisfaga algunos supuestos poco restrictivos.

Key words: State-Space models, exact likelihood, Kalman filter, unit roots

JEL classification: C32; C40.

Mailing address: Departamento de Fundamentos del Análisis Económico II, Universidad Complutense, Campus de Somosaguas, 28223 Madrid, Spain, E-mail: mjerez@ccee.ucm.es.

1. Introduction.

Methods to compute the likelihood of the parameters in State-Space (SS) models are important for three main reasons. First, the SS formulation includes as particular cases most time series models (Aoki, 1990). Second, the SS framework is adequate for nonstandard models like those with rational expectations (Watson, 1989) or errors-in-variables (Terceiro, 1990). Third, literature provides the main results required for model estimation and inference (Terceiro, 1990). On the other hand, when compared with standard algorithms, SS methods often increase the computational cost and the risk of numerical degradation.

Increased computational overhead is due to the sparse structure of the matrices in SS representations, which causes most computer code to expend large amounts of time in calculating the result of trivial floating-point operations (*e.g.*, zero-by-zero).

The numerical risk is a consequence of the intrinsic instability of the Kalman Filter (KF) Riccati equation, which is prone to generate conditional covariances with negative eigenvalues (see, *e.g.*, Bierman, 1977). Most approaches to this problem factorize the unstable matrices, paying again a cost in computer runtime.

In this article we address both issues, first, by exploiting the properties of the KF when applied to models with a special structure and, later, by extending these results to the general case. We begin in Section 2 by defining common notation and describing a likelihood computation procedure due to De Jong (1988). This method - which will be referred hereafter as 'conventional method' - has the important property of allowing for unit roots in the state equation. It will be used, first, as a starting point to derive our methods and, second, as a benchmark to measure computational gains.

Section 3 centers in the many models that can be expressed in a SS form known as 'steady-state innovations'. When a SS model has this structure, the KF second-order moments are known. Therefore, direct substitution of these variables by their analytical solutions simplifies drastically the KF propagation providing both, numerical stability and computational efficiency. Section 4 derives a similar method that can be applied to a general SS model satisfying weak assumptions. Section 5 compares the algorithms described in sections 2, 3 and 4 from the computational point of view. The Appendix contains the proof of the Theorem in Section 4.

2. A standard approach to likelihood evaluation.

Consider the model:

$$x_{t+1} = \Phi x_t + \Gamma u_t + E w_t \tag{1}$$

$$z_t = H x_t + D u_t + C v_t \tag{2}$$

where z_t is a $(m\times 1)$ vector of observable variables, u_t is a $(r\times 1)$ vector of inputs, x_t is a $(n\times 1)$ state vector, the errors w_t and v_t are independent of the initial state x_1 and such that $w_t \sim \text{iid}N(0,Q)$, $v_t \sim \text{iid}N(0,R)$, $\text{cov}(w_t,v_t) = S$ for all $t = 1,2,...,N$ and $x_1 \mid u_1,...,u_N \sim N(\bar{x}_1, P_1)$. The values in Φ, Γ, E, H, D, Q, R, S, $\bar{x}_1$ and P_1 are unknown.

There are many ways to compute the likelihood of (1)-(2). De Jong (1988) derives the expression:

$$l(Z \mid U,\theta) = \log|P_1| + \sum_{t=1}^{N} \log|B_t| + \sum_{t=1}^{N} \tilde{z}_t^T B_t^{-1} \tilde{z}_t + \log|P_1^{-1} + W_N| - w_N^T (P_1^{-1} + W_N)^{-1} w_N \tag{3}$$

where $Z = [z_1, z_2, ..., z_N]$, $U = [u_1, u_2, ..., u_N]$ and θ is a vector that contains the unknown values in Φ, Γ, E, H, D, Q, R and S. The innovations $\tilde{z}_t$ and their conditional covariance matrices B_t in (3) result from a KF initialized with $P_1 = 0$, from now on $\text{KF}(\bar{x}_1, 0)$:

$$\hat{x}_{t+1|t} = \Phi \hat{x}_{t|t-1} + \Gamma u_t + K_t \tilde{z}_t \tag{4}$$

$$\tilde{z}_t = z_t - H \hat{x}_{t|t-1} - D u_t \tag{5}$$

$$K_t = \left(\Phi P_{t|t-1} H^T + E S C^T\right) B_t^{-1} \tag{6}$$

$$P_{t+1|t} = \Phi P_{t|t-1} \Phi^T + E Q E^T - K_t B_t K_t^T \tag{7}$$

$$B_t = H P_{t|t-1} H^T + C R C^T \tag{8}$$

where $\hat{x}_{t+1|t}$ and $\hat{x}_{t|t-1}$ are the one-step-ahead conditional expectations of x_{t+1} and x_t, $P_{t+1|t}$ and $P_{t|t-1}$ are the corresponding conditional covariances and K_t is the KF gain. The variables w_N and W_N in (3) result from the recursions:

$$w_t = w_{t-1} + \bar{\Phi}_{t-1}^T H^T B_t^{-1} \tilde{z}_t \; ; \; w_0 = 0 \tag{9}$$

$$W_t = W_{t-1} + \bar{\Phi}_{t-1}^T H^T B_t^{-1} H \bar{\Phi}_{t-1} \; ; \; W_0 = 0 \tag{10}$$

$$\bar{\Phi}_t = (\Phi - K_t H) \bar{\Phi}_{t-1} \; ; \; \bar{\Phi}_0 = I \tag{11}$$

Note that:

1) The terms $\log|P_1|$, $\log|P_1^{-1} + W_N|$ and $w_N^T (P_1^{-1} + W_N)^{-1} w_N$ in (3) depend of the initial state covariance. When the system is stationary, all of them can be computed. If there are some unit roots, the first term can be approximated taking into account only its stationary part, see Casals and Sotoca (1997). The other expressions are not problematic because P_1^{-1} is finite, see De Jong and Chu-Chun-Lin (1994).

2) The values $\bar{x}_1$ and P_1 should be computed taking into account not only the stationarity of the series, but also the stochastic structure of u_t (Casals and Sotoca, 1997).

3. Likelihood computation for models in steady-state innovations form.

Many time series models, including transfer functions, VAR and VARMAX, can be represented in the steady-state innovations form (Aoki, 1990):

$$x_{t+1} = \Phi x_t + \Gamma u_t + E\varepsilon_t \tag{12}$$

$$z_t = Hx_t + Du_t + \varepsilon_t \tag{13}$$

where ε_t is a $(m\times 1)$ vector of errors independent of the initial state and such that $\varepsilon_t \sim$ iid$N(0,Q)$. Note that (12)-(13) is a special case of (1)-(2), where $w_t = v_t = \varepsilon_t$ and $C = I$. When a model is in this form, the KF computations can be done in a simplified manner, see Anderson and Moore (1979, pp. 232-235). In the rest of this section we combine these ideas with the likelihood function (3) to obtain a simplified version of the algorithm given by Eqs. (3)-(11).

When the model is (12)-(13), the likelihood function (3) becomes:

$$l(Z|U,\theta) = \log|P_1| + N\log|Q| + \sum_{t=1}^{N} \tilde{z}_t^T Q^{-1}\tilde{z}_t + \log|P_1^{-1} + W_N| - w_N^T(P_1^{-1} + W_N)^{-1}w_N \tag{14}$$

and the KF$(\bar{x}_1, 0)$ propagation simplifies to:

$$\hat{x}_{t+1|t} = \Phi\hat{x}_{t|t-1} + \Gamma u_t + E\tilde{z}_t \tag{15}$$

$$\tilde{z}_t = z_t - H\hat{x}_{t|t-1} - Du_t \tag{16}$$

because the Riccati equation (7) has an exact and strong solution $\bar{P} = 0$, which implies $P_{t+1|t} = 0$, $B_t = Q$ and $K_t = E$ for all $t = 1,2,...,N$. A proof of these results is immediate by induction. Starting in $t=2$ and substituting $P_1 = 0$ in the right-hand-side of (6), (8) and (7), yields $K_2 = EQQ^{-1} = E$, $B_2 = Q$ and $P_{2|1} = EQE^T - EQE^T = 0$, respectively, and these solutions remain the same for all $t>2$.

Last, expressions (9)-(11) become:

$$w_t = w_{t-1} + \bar{\Phi}_{t-1}^T H^T Q^{-1}\tilde{z}_t \qquad ; w_0 = 0 \tag{17}$$

$$W_t = W_{t-1} + \bar{\Phi}_{t-1}^T H^T Q^{-1} H\bar{\Phi}_{t-1} \quad ; W_0 = 0 \tag{18}$$

$$\bar{\Phi}_t = (\Phi - EH)\bar{\Phi}_{t-1} \qquad ; \bar{\Phi}_0 = I \tag{19}$$

The efficiency and stability of the algorithm (14)-(19) versus (3)-(11) is due to the fact that it avoids the recursive calculation of $P_{t+1|t}$, B_t and K_t replacing these variables by exact values. There are also minor computational gains in (14) and (17)-(19) because of the time-invariant nature of B_t and K_t.

4. Simplified computation of the likelihood of general state-space models.

The main limitation of the procedure described in Section 3 is that the direct SS representation of relevant models does not conform with Eqs. (12)-(13). Some of these are structural time series models (Harvey, 1989) or VARMAX models with observation errors (Terceiro, 1990). The following theorem defines weak sufficient conditions for a general model to be represented in an equivalent steady-state innovations form. This result allows one to simplify the KF recursions in a similar way to that of previous section.

Theorem: Consider the SS model (1)-(2) and assume that:

i) *The pair* $(H,\bar{\Phi})$, *where* $\bar{\Phi} = \Phi - ESC^T(CRC^T)^{-1}H$, *is detectable.*
ii) $P_1 \geq \bar{P}$

In these conditions, z_t *is also the output of the steady-state innovations model:*

$$x_{t+1}^* = \Phi x_t^* + \Gamma u_t + \bar{K}\varepsilon_t \tag{20}$$

$$z_t = Hx_t^* + Du_t + \varepsilon_t \tag{21}$$

where $\varepsilon_t \sim$ iid$N(0,\bar{B})$ *and* $x_1^* \sim$ iid$N(\bar{x}_1, P_1 - \bar{P})$, *and:*

$$\bar{P} = \Phi\bar{P}\Phi^T + EQE^T - \bar{K}\bar{B}\bar{K}^T \tag{22}$$

$$\bar{K} = (\Phi\bar{P}H^T + ESC^T)\bar{B}^{-1} \tag{23}$$

$$\bar{B} = H\bar{P}H^T + CRC^T \tag{24}$$

The proof of this result is in the Appendix. As an immediate corollary, it follows that the likelihood of (1)-(2) and (20)-(21) is the same i.i.f. i) and ii) hold.

Assumption ii) is not very restrictive, as it always holds under the time immemorial hypothesis (De Jong and Chu-Chun-Lin, 1994). The hypothesis i) implies that the strong solution of the Riccati equation (7) can be expressed as (22)-(24), see Chan *et al.* (1984).

The Theorem implies that, if hypotheses i) and ii) hold, the likelihood of (20)-(21) can be written as:

$$l(Z \mid U,\theta) = \log|P_1 - \bar{P}| + N\log|\bar{B}| + \sum_{t=1}^{N} \tilde{z}_t^T \bar{B}^{-1} \tilde{z}_t + \\ + \log|(P_1 - \bar{P})^{-1} + W_N| - w_N^T[(P_1 - \bar{P})^{-1} + W_N]^{-1} w_N \quad (25)$$

and in the partial nonstationary case (*i.e.*, when the model contains stationary and nonstationary roots) the term $(P_1 - \bar{P})^{-1}$ can be computed by a direct application of Theorem 3 in De Jong and Chu-Chun-Lin (1994), which implies that this matrix is finite and nonzero.

The innovations $\tilde{z}_t$ in (25) result from a KF$(\bar{x}_1, \bar{P})$ which, when applied to (20)-(21), simplifies to:

$$\hat{x}^*_{t+1|t} = \Phi \hat{x}^*_{t|t-1} + \Gamma u_t + \bar{K}\tilde{z}_t \quad (26)$$

$$\tilde{z}_t = z_t - H\hat{x}^*_{t|t-1} - D u_t \quad (27)$$

Finally, the sequences w_N and W_N are obtained from the recursions:

$$w_t = w_{t-1} + \bar{\Phi}^T_{t-1} H^T \bar{B}^{-1} \tilde{z}_t \qquad ; w_0 = 0 \quad (28)$$

$$W_t = W_{t-1} + \bar{\Phi}^T_{t-1} H^T \bar{B}^{-1} H \bar{\Phi}_{t-1} \qquad ; W_0 = 0 \quad (29)$$

$$\bar{\Phi}_t = (\Phi - \bar{K}H)\bar{\Phi}_{t-1} \qquad ; \bar{\Phi}_0 = I \quad (30)$$

Therefore, the KF equations required to compute K_t and B_t are replaced by the exact values (23)-(24) and the propagation of $P_{t+1|t}$ is substituted by the solution of (22). The literature proposes efficient and stable algorithms to solve this equation (Ionescu *et al.*, 1997).

5. Computational advantages.

The computational cost of the algorithms described in Sections 2, 3 and 4 - denoted respectively by A0, A1 and A2 - can be measured by the number of elementary floating point operations (flops) required to propagate the corresponding filters. Table 1 compares the flops required by Eqs. (4)-(11) of A0 and Eqs. (15)-(19) of A1 for common econometric models. It is immediate to see that the relative efficiency of A1 is a direct function of the dimension of the state vector (n) and the number of observable variables (m). The computational gains range from 43%-136% in the case of univariate models, to 130-220% in the case of bivariate models and to 1000-2000% in the case of models with a monthly seasonal part. The sample size does not affect the flops ratio in this case, as the cost of both filters is proportional to N.

Table 2 compares the load of A0 with the number of flops required to compute Eqs. (22)-(24) and (26)-

(30) of A2, for two models which does not conform directly with the steady-state innovations structure. In this case the relative efficiency of A2 increases with the sample size and the gains range from 26% (for N=100) to 250% (for N=1000).

[Insert Tables 1 and 2]

The advantages of A1 and A2 in terms of numerical stability arise from the fact that they avoid the propagation of the Riccati equations (7)-(8), which the literature (see, *e.g.*, Bierman, 1977) unanimously recognizes as the main source of numerical degradation of the KF. Additional stability is obtained by avoiding the explicit inversion of P_1 through its Cholesky factors.

Acknowledgements.

Sonia Sotoca acknowledges financial support from Fundación Caja Madrid and the CICYT project PB95-0912/95 (nº 5290806).

References

Anderson, B.D.O. and J.B. Moore, 1979. Optimal Filtering, (Prentice-Hall, Englewood Cliffs, N.J.).

Aoki, M., 1990. State Space Modeling of Time Series, (Springer-Verlag, Heidelberg).

Bierman, G.J., 1977. Factorization Methods for Discrete Sequential Estimation, (Academic Press, Orlando).

Casals, J. and S. Sotoca, 1997. Exact Initial Conditions for Maximum Likelihood Estimation of State Space Models with Stochastic Inputs. Economics Letters 57, 3, 261-267.

Chan, S.W., G.C. Goodwin and K.S. Sin, 1984. Convergence Properties of the Riccati Difference Equation in Optimal Filtering of Nonstabilizable Systems. IEEE Transactions on Automatic Control, 29, 110-118.

De Jong, P., 1988. The likelihood for a state space model. Biometrika 75, 1, 165-169.

De Jong, P. and S. Chu-Chun-Lin, 1994. Stationary and Non-Stationary State Space Models. Journal of Time Series Analysis 15, 2, 151-166.

Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter, (Cambridge

University Press, Cambridge).

Ionescu, V., Oara, C. and M. Weiss, 1997. General Matrix Pencil Techniques for the Solution of Algebraic Riccati Equations: A Unified Approach. IEEE Transactions on Automatic Control 42, 8, 1085-1097.

Terceiro, J., 1990. Estimation of dynamic econometric models with errors in variables, (Springer-Verlag, Heidelberg).

Watson, M.W., 1989. Recursive solution methods for dynamic linear rational expectations models., Journal of Econometrics 41, 65-89.

the first and second-order conditional moments of z_1 are:

$$E(z_1 \mid x_1^*) = HE(x_1 \mid x_1^*) = H(x_1^* + \bar{x}_1) \tag{A.10}$$

$$\mathrm{cov}(z_1 \mid x_1^*) = H\mathrm{cov}(x_1 \mid x_1^*)H^T + CRC^T = H\bar{P}H^T + CRC^T \tag{A.11}$$

3) Finally, the *last term* of (A.3) depends of $E(x_1^* \mid z)$ and $\mathrm{var}(x_1^* \mid z)$. To compute these moments note that:

$$z = \bar{H}x_1 + z^* \tag{A.12}$$

where z is the observable output corresponding to the initial conditions x_1 and P_1; z^* is the output corresponding to a zero initial state with no uncertainty, and: $\bar{H} = \left[H\Phi \;\; H\Phi^2 \;\; \ldots \;\; H\Phi^{N-1}\right]^T$.

Eq. (A.12) can be written as:

$$z = \bar{H}x_1^* + z^* + \bar{H}(x_1 - x_1^*) = \bar{H}x_1^* + z^{**} \tag{A.13}$$

where z^{**} is the output of:

$$z_t^{**} = Hx_t^{**} + Cv_t \tag{A.14}$$

$$x_{t+1}^{**} = \Phi x_t^{**} + Ew_t \tag{A.15}$$

with an initial state $x_1^{**} = x_1 - x_1^*$ such that $x_1^{**} \sim N(\bar{x}_1, \bar{P})$. Then, the application of a KF$(\bar{x}_1, \bar{P})$ to (A.14)-(A.15) yields a sequence of uncorrelated innovations $\tilde{z}_t^{**}$, which covariance is $\bar{B}$, and (A.14) can be expressed as:

$$\tilde{z} = Xx_1^* + \tilde{z}^{**} \tag{A.16}$$

where X is a matrix which block-row is $H\bar{\Phi}_{t-1}$ and the sequence $\bar{\Phi}_t = (\Phi - \bar{K}H)\bar{\Phi}_{t-1}$ is initialized with $\bar{\Phi}_1 = I$. Building on these results, it is easy to prove (De Jong, 1988) that:

$$E(x_1^* \mid \tilde{z}) = [(P_1 - \bar{P})^{-1} + W]^{-1}w \tag{A.17}$$

$$\mathrm{cov}(x_1^* \mid \tilde{z}) = [(P_1 - \bar{P})^{-1} + W]^{-1} \tag{A.18}$$

where $W = X^T\bar{B}^{-1}X$ and $w = X^T\bar{B}^{-1}\tilde{z}$.

From the results in steps 1), 2) and 3), it follows that the log-likelihood of (1)-(2) can be written as:

Appendix.

Consider model (1)-(2) and assume, without loss of generality, that there are no exogenous variables (*i.e.*, $\Gamma = \mathbf{0}$ and $\boldsymbol{D} = \mathbf{0}$). Let be the state x_1^* which is related with the true initial state of (1)-(2), x_1, by:

$$x_1^* = (\boldsymbol{P}_1 - \bar{\boldsymbol{P}})\boldsymbol{P}_1^{-1}(x_1 - \bar{x}_1) + \xi \tag{A.1}$$

where ξ is a random vector such that $E(\xi) = \mathbf{0}$, $E(\xi x_1^T) = \mathbf{0}$, $\bar{x}_1 = E(x_1)$; $\bar{\boldsymbol{P}}$ is given by Eq. (22) and:

$$\mathrm{cov}(\xi) = (\boldsymbol{P}_1 - \bar{\boldsymbol{P}}) - (\boldsymbol{P}_1 - \bar{\boldsymbol{P}})\boldsymbol{P}_1^{-1}(\boldsymbol{P}_1 - \bar{\boldsymbol{P}}) \tag{A.2}$$

Note that assumptions i) and ii) of the Theorem assure that $\mathrm{cov}(\xi)$ is positive-semidefinite. Then, the log-likelihood function of the sample is:

$$l(z \mid \theta) = l(x_1^*) + l(z \mid x_1^*) - l(x_1^* \mid z) \tag{A.3}$$

Under the gaussian assumption and discarding constants, the terms in (A.3) can be expressed as shown in the following steps 1), 2) and 3):

1) The *first term* is:

$$l(x_1^*) = \log|\boldsymbol{P}_1 - \bar{\boldsymbol{P}}| + (x_1^*)^T(\boldsymbol{P}_1 - \bar{\boldsymbol{P}})^{-1}x_1^* \tag{A.4}$$

because $E(x_1^*) = \mathbf{0}$ and $\mathrm{var}(x_1^*) = (\boldsymbol{P}_1 - \bar{\boldsymbol{P}})\boldsymbol{P}_1^{-1}(\boldsymbol{P}_1 - \bar{\boldsymbol{P}}) + \mathrm{var}(\xi) = \boldsymbol{P}_1 - \bar{\boldsymbol{P}}$, see (A.1)- (A.2).

2) The *second term* is:

$$l(z \mid x_1^*) = N\log|\bar{\boldsymbol{B}}| + \sum_{t=1}^{N} \tilde{z}_t \bar{\boldsymbol{B}}^{-1}\tilde{z}_t \tag{A.5}$$

where $\tilde{z}_t$ and $\bar{\boldsymbol{B}}$ result from the KF$(\bar{x}_1, \bar{\boldsymbol{P}})$:

$$\hat{x}_{t+1|t} = \Phi\hat{x}_{t|t-1} + \bar{\boldsymbol{K}}\tilde{z}_t \tag{A.6}$$

$$\tilde{z}_t = z_t - \boldsymbol{H}\hat{x}_{t|t-1} \tag{A.7}$$

$$\boldsymbol{P}_t = \bar{\boldsymbol{P}} \tag{A.8}$$

$$\boldsymbol{B}_t = \bar{\boldsymbol{B}} \tag{A.9}$$

where $\bar{\boldsymbol{K}}$ and $\bar{\boldsymbol{B}}$ are given by (23) and (24), respectively. The initialization $(\bar{x}_1, \bar{\boldsymbol{P}})$ is adequate because

$$l(Z \mid \theta) = \log|\boldsymbol{P}_1 - \bar{\boldsymbol{P}}| + N\log|\bar{\boldsymbol{B}}| + \sum_{t=1}^{N} \tilde{z}_t\boldsymbol{B}_t^{-1}\tilde{z}_t + \log|(\boldsymbol{P}_1 - \bar{\boldsymbol{P}})^{-1} + \boldsymbol{W}| - w^T[(\boldsymbol{P}_1 - \bar{\boldsymbol{P}})^{-1} + \boldsymbol{W}]^{-1}w \tag{A.19}$$

which coincides with the log-likelihood of (20)-(21), see Eq. (25). ■

Table 1: Computational efficiency of A1 versus A0.

Dimensions	Models	Flops (A0) / Flops (A1)
Univariate:		
$n=1\,;\,m=1$	$AR(1)$, $MA(1)$, $ARMA(1,1)$	1.44
$n=2\,;\,m=1$	$AR(2)$, $MA(2)$, $ARMA(2,1)$	2.36
Univariate (with seasonal part):		
$n=4\,;\,m=1$	$AR(1)_4$, $MA(1)_4$, $AR(1)\times MA(1)_4$	3.95
$n=5\,;\,m=1$	$MA(1)\times MA(1)_4$, $ARMA(1,1)\times MA(1)_4$	4.77
$n=8\,;\,m=1$	$AR(2)_4$, $MA(2)_4$, $ARMA(2,1)_4$	7.29
$n=12;\,m=1$	$AR(1)_{12}$, $MA(1)_{12}$, $AR(1)\times MA(1)_{12}$	10.69
$n=13;\,m=1$	$MA(1)\times MA(1)_{12}$, $ARMA(1,1)\times MA(1)_{12}$	11.54
$n=24;\,m=1$	$AR(2)_{12}$, $MA(1)_{12}$, $MA(1)\times AR(2)_{12}$	20.94
Bivariate:		
$n=2\,;\,m=2$	$VAR(1)$, $VMA(1)$, $VARMA(1,1)$	2.30
$n=4\,;\,m=2$	$VAR(2)$, $VMA(2)$, $VARMA(1,2)$, $VARMA(2,1)$	3.22

Table 2: Computational efficiency of A2 versus A0.

Dimensions	Model	Flops (A0) / Flops (A2)		
		N=100	N=500	N=1000
$n=2\,;\,m=1$	$ARMA(2,1)$ + observation error	1.26	2.77	3.46
	Structural time series model†	1.28	3.30	3.61

† The model is:

$$T_t = T_{t-1} + D_t$$
$$D_t = D_{t-1} + \eta_t$$
$$z_t = T_t + \varepsilon_t$$

where T_t is the trend, D_t is the change of trend and the noise terms are such that $\eta_t \sim \text{iid}(0,\sigma_\eta^2)$, $\varepsilon_t \sim \text{iid}(0,\sigma_\varepsilon^2)$ and $\text{E}(\eta_t\,\varepsilon_t) = 0$.