Exact Initial Kalman Filtering and Smoothing for Nonstationary Time Series Models

Siem Jan KOOPMAN

This article presents a new exact solution for the initialization of the Kalman filter for state space models with diffuse initial conditions. For example, the regression model with stochastic trend, seasonal and other nonstationary autoregressive integrated moving average components requires a (partially) diffuse initial state vector. The proposed analytical solution is easy to implement and computationally efficient. The exact solution for smoothing is also given. Missing observations are handled in a straightforward manner. All proofs rely on elementary results.

KEY WORDS: Autoregressive integrated moving average component models; Diffuse initial conditions; Likelihood function and score vector; Missing observations; State space.

1. INTRODUCTION

Assume that a vector of observations $\mathbf{y}_t$ is generated by the Gaussian state-space model

$$\mathbf{y}_t = \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{G}_t\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathrm{N}(0, I), \quad t = 1, \ldots, n,$$

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}_t\boldsymbol{\alpha}_t + \mathbf{H}_t\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\alpha}_1 \sim \mathrm{N}(\mathbf{a}, \mathbf{P}), \qquad (1)$$

where $p \times 1$ vector $\boldsymbol{\alpha}_t$ is the state vector. The system matrices, $\mathbf{Z}_t, \mathbf{T}_t, \mathbf{G}_t$, and $\mathbf{H}_t$, for $t = 1, \ldots, n$, are assumed to be fixed and known. The former equation of (1) is referred to as the observation equation, whereas the latter equation is called the transition equation. A normally distributed random vector $\mathbf{x}$ with mean μ and variance matrix $\boldsymbol{\Lambda}$ is denoted by $\mathbf{x} \sim \mathrm{N}(\mu, \boldsymbol{\Lambda})$. The disturbance vector $\boldsymbol{\varepsilon}_t$ is serially uncorrelated. The appearance of $\boldsymbol{\varepsilon}_t$ in both equations is general rather than restrictive. The special case $\mathbf{H}_g\mathbf{G}_t' = 0$, for $t = 1, \ldots, n$, implies mutual independence between the two sets of disturbances. The mean vector $\mathbf{a}$ and variance matrix $\mathbf{P}$ of the initial state vector are assumed to be known.

The state-space model (1) is often used as a framework for representing linear time series models such as autoregressive integrated moving average (ARIMA) models, regression models with stochastic trend, seasonal or other ARIMA components, and models with time-varying regression parameters. The state space representation may contain some unknown elements in the system matrices. These hyperparameters are estimated by maximizing the likelihood function via some quasi-Newton optimization routine; see Section 4.

The Kalman filter produces the minimum mean squared linear estimator of the state vector $\boldsymbol{\alpha}_{t+1}$ using the set of observations $Y_t = \{\mathbf{y}_1, \ldots, \mathbf{y}_t\}$—that is, $\mathbf{a}_{t+1} = E(\boldsymbol{\alpha}_{t+1}|Y_t)$—and the corresponding variance matrix of the estimator $\mathbf{a}_{t+1}$—that is, $\mathbf{P}_{t+1} = \mathrm{var}(\boldsymbol{\alpha}_{t+1}|Y_t)$—for $t = 1, \ldots, n$. The Kalman filter is given by

$$\mathbf{v}_t = \mathbf{y}_t - \mathbf{Z}_t\mathbf{a}_t, \quad \mathbf{F}_t = \mathbf{Z}_t\mathbf{P}_t\mathbf{Z}_t' + \mathbf{G}_t\mathbf{G}_t',$$

$$\mathbf{M}_t = \mathbf{T}_t\mathbf{P}_t\mathbf{Z}_t' + \mathbf{H}_t\mathbf{G}_t, \qquad \mathbf{a}_{t+1} = \mathbf{T}_t\mathbf{a}_t + \mathbf{K}_t\mathbf{v}_t,$$

$$\mathbf{P}_{t+1} = \mathbf{T}_t\mathbf{P}_t\mathbf{T}_t' + \mathbf{H}_t\mathbf{H}_t' - \mathbf{C}_t, \qquad (2)$$

where $\mathbf{K}_t = \mathbf{M}_t\mathbf{F}_t^{-1}$ and $\mathbf{C}_t = \mathbf{M}_t\mathbf{F}_t^{-1}\mathbf{M}_t'$ for $t = 1, \ldots, n$. The Kalman filter is initialized by $\mathbf{a}_1 = \mathbf{a}$ and $\mathbf{P}_1 = \mathbf{P}$. (The proof of the Kalman filter can be found in Anderson and Moore 1979.) The one-step-ahead prediction error of the observation vector is $\mathbf{v}_t = \mathbf{y}_t -\mathrm{E}(\mathbf{y}_t|Y_{t-1})$ with variance matrix $\mathbf{F}_t = \mathrm{var}(\mathbf{y}_t|Y_{t-1}) = \mathrm{var}(\mathbf{v}_t)$. Furthermore, matrix $\mathbf{M}_t$ is defined as the covariance matrix $\mathbf{M}_t = \mathrm{cov}(\boldsymbol{\alpha}_{t+1}, \mathbf{y}_t|Y_t)$. The matrix $\mathbf{K}_t$ is called the Kalman gain matrix. The output of the Kalman filter can be used to compute the likelihood function of the state-space model (1) via the prediction error decomposition; see Section 4.

Minimum mean squared linear estimators using all observations Y_n are evaluated by a smoothing algorithm; that is, a set of backwards recursions that requires output generated by the Kalman filter. The basic smoothing algorithm of de Jong (1988) and Kohn and Ansley (1989) is given by

$$\mathbf{r}_{t-1} = \mathbf{Z}_t'\mathbf{F}_t'\mathbf{v}_t + \mathbf{L}_t'\mathbf{r}_t, \quad \mathbf{N}_{t-1} = \mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{Z}_t + \mathbf{L}_t'\mathbf{N}_t\mathbf{L}_t, \quad (3)$$

where $\mathbf{L}_t = \mathbf{T}_t - \mathbf{K}_t\mathbf{Z}_t, \mathbf{r_n} = \mathbf{0}$, and $\mathbf{N_n} = \mathbf{0}$, for $t = 1, \ldots, n$. The vector $\mathbf{r}_t$ and matrix $\mathbf{N}_t$ can be used for different purposes. For example, (a) de Jong (1988) and Kohn and Ansley (1989) applied (3) to obtain the smoothed estimator for $\boldsymbol{\alpha}_t$—that is, $E(\boldsymbol{\alpha}_t|Y_n) = \mathbf{a}_t + \mathbf{P}_t\mathbf{r}_{t-1}$, for $t = 1, \ldots, n$; (b) Koopman (1993) used it to obtain the smoothed estimator for $\boldsymbol{\varepsilon}_t$, that is $E(\boldsymbol{\varepsilon}_t|Y_n) = \mathbf{H}_t'\mathbf{r}_t + \mathbf{G}_t'\mathbf{e}_t$ where $\mathbf{e}_t = \mathbf{F}_t^{-1}\mathbf{v}_t - \mathbf{K}_t'\mathbf{r}_t$, for $t = 1, \ldots, n$; (c) Koopman and Shephard (1992) get the exact score function of the hyperparameter vector using (3); see Section 4.

The Kalman filter provides a general tool to handle missing observations in time series. When the vector of observations $\mathbf{y}_t$ is missing, the matrices $\mathbf{Z}_t$ and $\mathbf{G}_t$ are set equal to 0. When only some entries of $\mathbf{y}_t$ are missing, the matrices $\mathbf{Z}_t$ and $\mathbf{G}_t$ are adjusted appropriately by eliminating the rows corresponding to the missing entries in $\mathbf{y}_t$. The Kalman filter and the basic smoothing algorithm deal straightforwardly with different matrix dimensions of $\mathbf{Z}_t$ and $\mathbf{G}_t$.

Siem Jan Koopman is Assistant Professor, CentER, Tilburg University, 5000 LE Tilburg, The Netherlands. This work was carried out while the author was a lecturer at the Department of Statistics of the London School of Economics and Political Science and it was supported by ESRC grant R000235330. The author thanks J. Durbin, A. C. Harvey, J. R. Magnus, the associate editor, and a referee for their help and support in this research project.

Journal of the American Statistical Association
December 1997, Vol. 92, No. 440, Theory and Methods

The Kalman filter for a stationary time series model is initialized by the unconditional mean and variance matrix of $\boldsymbol{\alpha}_1$; that is, $\mathbf{a}_1 = \mathbf{a}$ and $\mathbf{P}_1 = \mathbf{P}$. Stationarity implies a time-invariant state space model [i.e., (1)], for which the system matrices are constant over time. We obtain $\text{var}(\boldsymbol{\alpha}_t) = \text{var}(\boldsymbol{\alpha}_1) = \mathbf{P} = \mathbf{TPT}' + \mathbf{HH}'$, which can be solved easily with respect to $\mathbf{P}$ (see Magnus and Neudecker 1988, chap. 2). Nonstationary time series models such as regression models with stochastic trend, seasonal, or other ARIMA components require noninformative prior conditions for the initial state vector $\boldsymbol{\alpha}_1$. In this article I consider the problem of initializing the Kalman filter for any nonstationary model in state-space form.

The initial state vector $\boldsymbol{\alpha}_1$ can generally be specified as

$$\boldsymbol{\alpha}_1 = \mathbf{a} + \mathbf{A}\eta + \mathbf{B}\boldsymbol{\delta}, \quad \boldsymbol{\eta} \sim \text{N}(0, \mathbf{I}), \quad \boldsymbol{\delta} \sim \text{N}(0, \kappa\mathbf{I}), \tag{4}$$

where matrices $\mathbf{A}$ and $\mathbf{B}$ are fixed and known. The $m \times 1$ vector $\boldsymbol{\delta}$ is referred to as the *diffuse* vector as $\kappa \to \infty$. This leads to

$$\boldsymbol{\alpha}_1 \sim \text{N}(\mathbf{a}, \mathbf{P}), \quad \mathbf{P} = \mathbf{P}_* + \kappa\mathbf{P}_\infty, \tag{5}$$

where $\mathbf{P}_* = \mathbf{AA}'$ and $\mathbf{P}_\infty = \mathbf{BB}'$. This specification implies that some elements of the initial state vector are not well defined. The Kalman filter (2) cannot be applied in cases where $\mathbf{P}_\infty$ is a nonzero matrix, because no real value can represent κ as $\kappa \to \infty$.

A numerical device for dealing with the diffuse initial state vector is to replace the scalar κ be some large value k; for example, 10^7 (see Burridge and Wallis 1985 and Harvey and Phillips 1979). This numerical solution is not exact and may generate inaccuracies due to numerical rounding errors. In this article I pursue an analytical approach where I express the Kalman filter quantities in terms of κ explicitly and then let $\kappa \to \infty$ to obtain the exact solution. Ansley and Kohn (1985, 1990; hereafter referred to as AK) first solved the initialization problem using an analytical approach. Their solution is general and deals with missing observations in the initial period. However, the general algorithm of AK is difficult to implement, standard software cannot be used, and the proof is long and complex. This article presents a new analytical solution that uses a trivial initialization. The proof is based on elementary results in calculus and matrix algebra.

An alternative exact approach was taken by de Jong (1991), which he labelled as the diffuse Kalman filter, hereafter referred to as DKF. The DKF augments the Kalman filter quantities $\mathbf{v}_t$ and $\mathbf{a}_{t+1}$ by m columns to get $\mathbf{V}_t$ and $\mathbf{A}_{t+1}$, with the initialization $\mathbf{A}_1 = (\mathbf{a}, \mathbf{B})$ and $\mathbf{P}_1 = \mathbf{P}_*$. Also, the DKF includes the matrix recursion $\mathbf{S}_{t+1} = \mathbf{S}_t + \mathbf{V}_t'\mathbf{F}_t^{-1}\mathbf{V}_t$ with $\mathbf{S}_1 = 0$. When the lower $m \times m$ block of $\mathbf{S}_d$ can be inverted, the appropriate corrections for $\mathbf{a}_{d+1}$ and $\mathbf{P}_{d+1}$ are made, for some $d < n$, and the augmented part disappears. The DKF is computationally inefficient due to the additional matrix operations and the inversion of a $m \times m$ matrix. Moreover, the corrections for the basic smoothing algorithm (3) are complicated and difficult to implement (see Chu-Chun-Lin and de Jong 1993).

Other solutions for the initialization problem of the Kalman filter are as follows:

- Anderson and Moore (1979) and Kitagawa (1981) discussed the information filter that evaluates the inverse of $\mathbf{P}_t$ recursively. This straightforward solution avoids the infinite case because $\kappa^{-1} \to 0$ as $\kappa \to \infty$. However, the information approach is not general because it cannot be applied to all cases.
- Harvey and Pierse (1984) treated the initial vector $\boldsymbol{\delta}$ as an unknown parameter and included it in the state vector. This approach is computationally inefficient due to the increased size of the state vector.
- Bell and Hillmer (1991) suggested applying the transformation approach of Ansley and Kohn (1985) directly to the initial dataset. This approach provides the appropriate initialization of the Kalman filter for the remaining dataset but needs modification when missing observations occur in the initial dataset. Gomez and Maravall (1994) rediscovered this approach for the specific application of ARIMA models.
- A specific square root algorithm based on fast Givens transformations can deal explicitly with $\mathbf{P}_t$ when it depends on $\kappa \to \infty$ (see Snyder and Saligari 1996). It is known that square root algorithms are numerically stable but have high computational costs. Moreover, it is not clear how to implement smoothing algorithms in this approach.

The article is organized as follows. Section 2 presents the exact initial Kalman filter, and Section 3 presents the exact initial smoothing algorithm. Section 4 applies the new results to compute the diffuse likelihood function and the score vector of the hyperparameter vector for nonstationary time series models. Section 5 gives some examples of how to implement the algorithm for univariate and multivariate cases, and Section 6 discusses some related issues and it gives an assessment of the computational efficiency of the new algorithms. Section 7 concludes.

2. THE EXACT INITIAL KALMAN FILTER

The formulation of the initial state variance matrix $\mathbf{P} = \mathbf{P}_* + \kappa\mathbf{P}_\infty$ implies a similar formulation for the Kalman filter quantities $\mathbf{F}_t = \mathbf{F}_{*,t} + \kappa\mathbf{F}_{\infty,t}$ and $\mathbf{M}_t = \mathbf{M}_{*,t} + \kappa\mathbf{M}_{\infty,t}$, where

$$\mathbf{F}_{*,t} = \mathbf{Z}_t\mathbf{P}_{*,t}\mathbf{Z}_t' + \mathbf{G}_t\mathbf{G}_t', \qquad \mathbf{F}_{\infty,t} = \mathbf{Z}_t\mathbf{P}_{\infty,t}\mathbf{Z}_t',$$

$$\mathbf{M}_{*,t} = \mathbf{T}_t\mathbf{P}_{*,t}\mathbf{Z}_t' + \mathbf{H}_t\mathbf{G}_t', \qquad \mathbf{M}_{\infty,t} = \mathbf{T}_t\mathbf{P}_{\infty,t}\mathbf{Z}_t'. \tag{6}$$

The Kalman update equations for $\mathbf{a}_{t+1}$ and $\mathbf{P}_{t+1} = \mathbf{P}_{*,t+1} + \kappa\mathbf{P}_{\infty,t+1}$ rely on the matrices $\mathbf{K}_t = \mathbf{M}_t\mathbf{F}_t^{-1}$ and $\mathbf{C}_t = \mathbf{M}_t\mathbf{F}_t^{-1}\mathbf{M}_t'$. In developing expressions for $\mathbf{K}_t$ and $\mathbf{C}_t$, we temporarily drop the time index t for notational convenience. For a properly defined model (1), the $N \times N$ matrix $\mathbf{F} = \mathbf{F}_* + \kappa\mathbf{F}_\infty$ is nonsingular, but matrices $\mathbf{F}_*$ and $\mathbf{F}_\infty = \mathbf{ZP}_\infty\mathbf{Z}'$ are not necessarily nonsingular and can be partially diagonalized as shown in Lemma 2 of the Appendix; that is,

$$(\mathbf{J}_1, \mathbf{J}_2)'\mathbf{F}_\infty(\mathbf{J}_1, \mathbf{J}_2) = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & 0 \end{bmatrix},$$

$$(\mathbf{J}_1, \mathbf{J}_2)'\mathbf{F}_*(\mathbf{J}_1, \mathbf{J}_2) = \begin{bmatrix} \mathbf{V}_* & 0 \\ 0 & \mathbf{I}_{N-r} \end{bmatrix}, \qquad (7)$$

where $\mathbf{J} = (\mathbf{J}_1, \mathbf{J}_2)$ is a nonsingular matrix, $r = r(\mathbf{F}_\infty)$, and $r(\mathbf{V}_*) \leq r$. Theorem 2 of the Appendix shows that the inverse of $\mathbf{F}$ can be expanded as

$$\mathbf{F}^{-1} = \mathbf{F}_*^- + \frac{1}{\kappa}\, \mathbf{F}_\infty^- - \frac{1}{\kappa^2}\, \mathbf{F}_\infty^-\mathbf{F}_*\mathbf{F}_\infty^- + O\left(\frac{1}{\kappa^3}\right), \qquad (8)$$

where $\mathbf{F}_\infty^- = \mathbf{J}_1\mathbf{J}_1'$ and $\mathbf{F}_*^- = \mathbf{J}_2\mathbf{J}_2'$. It follows from (7) that $\mathbf{J}_2'\mathbf{Z}\mathbf{P}_\infty = 0$, so that $\mathbf{M}_\infty\mathbf{J}_2 = 0$ and $\mathbf{M}_\infty\mathbf{F}_*^- = 0$. Then the Kalman gain matrix $\mathbf{K}$ is

$$\begin{aligned} \mathbf{K} &= (\mathbf{M}_* + \kappa\mathbf{M}_\infty)\mathbf{F}^{-1} \\ &= \mathbf{M}_*\mathbf{F}_*^- + \mathbf{M}_\infty\mathbf{F}_\infty^- + O\left(\frac{1}{\kappa}\right), \end{aligned} \qquad (9)$$

and matrix $\mathbf{C}$ is

$$\begin{aligned} \mathbf{C} &= (\mathbf{M}_* + \kappa\mathbf{M}_\infty)\mathbf{F}^{-1}(\mathbf{M}_* + \kappa\mathbf{M}_\infty)' \\ &= \mathbf{M}_*\mathbf{F}_*^-\mathbf{M}_*' + \mathbf{M}_\infty\mathbf{F}_\infty^-\mathbf{M}_*' + \mathbf{M}_*\mathbf{F}_\infty^-\mathbf{M}_\infty' \\ &\quad + \kappa\mathbf{M}_\infty\mathbf{F}_\infty^-\mathbf{M}_\infty' \\ &\quad - \mathbf{M}_\infty\mathbf{F}_\infty^-\mathbf{F}_*\mathbf{F}_\infty^-\mathbf{M}_\infty' + O\left(\frac{1}{\kappa}\right). \end{aligned} \qquad (10)$$

By reintroducing the time indices, the exact initial Kalman filter update equations of $\mathbf{a}_{t+1}$ and $\mathbf{P}_{t+1} = \mathbf{P}_{*,t+1} + \kappa\mathbf{P}_{\infty,t+1}$, when $\kappa \to \infty$, are given by

$$\mathbf{a}_{t+1} = \mathbf{T}_t\mathbf{a}_t + \mathbf{K}_{*,t}\mathbf{v}_t, \quad \mathbf{P}_{*,t+1} = \mathbf{T}_t\mathbf{P}_{*,t}\mathbf{T}_t' - \mathbf{C}_{*,t} + \mathbf{H}_t\mathbf{H}_t',$$

$$\mathbf{P}_{\infty,t+1} = \mathbf{T}_t\mathbf{P}_{\infty,t}\mathbf{T}_t' - \mathbf{C}_{\infty,t}, \qquad (11)$$

where $\mathbf{v}_t = \mathbf{y}_t - \mathbf{Z}_t\mathbf{a}_t$ and

$$\mathbf{K}_{*,t} = \mathbf{M}_{*,t}\mathbf{F}_{*,t}^- + \mathbf{M}_{\infty,t}\mathbf{F}_{\infty,t}^-,$$

$$\mathbf{C}_{*,t} = \mathbf{M}_{*,t}\mathbf{K}_{*,t}' + \mathbf{M}_{\infty,t}\mathbf{F}_{\infty,t}^-(\mathbf{M}_{*,t} - \mathbf{M}_{\infty,t}\mathbf{F}_{\infty,t}^-\mathbf{F}_{*,t})',$$

and

$$\mathbf{C}_{\infty,t} = \mathbf{M}_{\infty,t}\mathbf{F}_{\infty,t}^-\mathbf{M}_{\infty,t}'. \qquad (12)$$

The update of $\mathbf{a}_{t+1}$ in (11) follows from (2) and (9), because $\mathbf{v}_t$ does not depend on κ and the terms associated with $1/\kappa$ in (9) disappear when $\kappa \to \infty$. The update of $\mathbf{P}_{t+1}$ has two parts: $\mathbf{P}_{*,t+1}$ represents the part that does not rely on κ, whereas $\mathbf{P}_{\infty,t+1}$ is the part associated with κ. The specification for $\mathbf{C}_{*,t}$ in (12) follows from rearranging the terms not associated with κ in (10) and using the definition of $\mathbf{K}_{*,t}$. The remaining part of (10) is found in $\mathbf{C}_{\infty,t}$ as the terms associated with $1/\kappa$ disappear when $\kappa \to \infty$.

It is noted that the dimension N is generally small, and the usual Kalman filter requires the inversion of $\mathbf{F}_t$, for which some diagonalization routine is required. Thus diagonalization of matrices $\mathbf{F}_{\infty,t}$ and $\mathbf{F}_{*,t}$ in (7) does not impose an excessive additional computational burden compared to the usual Kalman filter. Moreover, the diagonalization (7) does not apply to univariate time series models, because $N = 1$. The remaining computations for (12) can be done very efficiently; see Section 6.

2.1 The Nonsingular and Univariate Case

The general solution can be simplified for cases where $\mathbf{F}_{\infty,t}$ is either a zero matrix or a nonsingular matrix for $t = 1, \ldots, n$. If $\mathbf{F}_{\infty,t}$ is a zero matrix, then $\mathbf{J}_{2,t} = \mathbf{J}_t$ and $\mathbf{J}_{1,t} = 0$, so $\mathbf{F}_{*,t}^- = \mathbf{F}_{*,t}^{-1}$ and $\mathbf{F}_{\infty,t}^- = 0$. If $\mathbf{F}_{\infty,t}$ is a nonsingular matrix, then $\mathbf{J}_{1,t} = \mathbf{J}_t$ and $\mathbf{J}_{2,t} = 0$ so $\mathbf{F}_{*,t}^- = 0$ and $\mathbf{F}_{\infty,t}^- = \mathbf{F}_{\infty,t}^{-1}$. For the cases (a) $\mathbf{F}_{\infty,t} = 0$ and (b) $r(\mathbf{F}_{\infty,t}) = N$, the matrices $\mathbf{K}_{*,t}, \mathbf{C}_{*,t}$ and $\mathbf{C}_{\infty,t}$ are given by

$$\begin{aligned} &\text{(a)} \quad \mathbf{K}_{*,t} = \mathbf{M}_{*,t}\mathbf{F}_{*,t}^{-1}, \\ &\qquad \mathbf{C}_{*,t} = \mathbf{M}_{*,t}\mathbf{K}_{*,t}', \\ &\qquad \mathbf{C}_{\infty,t} = 0, \\ &\text{(b)} \quad \mathbf{K}_{*,t} = \mathbf{M}_{\infty,t}\mathbf{F}_{\infty,t}^{-1}, \\ &\qquad \mathbf{C}_{*,t} = \mathbf{M}_{*,t}\mathbf{K}_{*,t}' + \mathbf{K}_{*,t}(\mathbf{M}_{*,t} - \mathbf{K}_{*,t}\mathbf{F}_{*,t})', \\ &\qquad \mathbf{C}_{\infty,t} = \mathbf{M}_{\infty,t}\mathbf{F}_{\infty,t}^{-1}\mathbf{M}_{\infty,t}'. \end{aligned} \qquad (13)$$

The expansion (8) for a nonsingular matrix $\mathbf{F}_{\infty,t}$ is given by Theorem 1 of the Appendix. Case (a) is the usual Kalman update (2). The modified Kalman filter of Ansley and Kohn (1990) for univariate models can be reformulated similar to (13).

These special cases are specifically relevant for univariate state-space models. When $N = 1, \mathbf{F}_{\infty,t}$ is a scalar that is 0 or positive. Thus the relevant equations for the univariate exact Kalman filter are (6), (11), and (13). The number of extra flops for case (b) of (13) compared to case (a) is limited to $p^2 + 2p$ in the univariate case.

2.2 Automatic Collapse to Kalman Filter

In the following it is shown that generally the rank of matrix $\mathbf{P}_{\infty,t+1}$ is equal to the rank of $\mathbf{P}_{\infty,t}$ minus the rank of $\mathbf{F}_{\infty,t}$. Specifically,

$$r(\mathbf{P}_{\infty,t+1}) \leq \min\{r(\mathbf{P}_{\infty,t}) - r(\mathbf{F}_{\infty,t}), r(\mathbf{T}_t)\}. \qquad (14)$$

The update for $\mathbf{P}_{\infty,t+1}$ in (11) can be rewritten as

$$\mathbf{P}_{\infty,t+1} = \mathbf{T}_t\mathbf{P}_{\infty,t}^\dagger\mathbf{T}_t', \qquad (15)$$

where $\mathbf{P}_{\infty,t}^\dagger = \mathbf{P}_{\infty,t} - \mathbf{P}_{\infty,t}\mathbf{Z}_t'\mathbf{F}_{\infty,t}^-\mathbf{Z}_t\mathbf{P}_{\infty,t} = \mathbf{P}_{\infty,t} - \mathbf{P}_{\infty,t}\mathbf{U}_{+,t}\mathbf{U}_{+,t}'\mathbf{P}_{\infty,t}$ with $\mathbf{F}_{\infty,t}^- = \mathbf{J}_{1,t}\mathbf{J}_{1,t}'$ and $\mathbf{U}_{+,t} = \mathbf{Z}_t'\mathbf{J}_{1,t}$; see also Equations (7) and (8). Matrix $\mathbf{P}_{\infty,t}$ is positive semidefinite and can be diagonalized so that matrix $\mathbf{U}_t'\mathbf{P}_{\infty,t}\mathbf{U}_t$ is diagonal with 0 and unity values for some nonsingular matrix $\mathbf{U}_t$. It is clear from (7) that $\mathbf{U}_{+,t}$ is a subset of the columns of $\mathbf{U}_t$. By noting that $r(\mathbf{F}_{\infty,t}) = r(\mathbf{U}_{+,t})$ and applying Lemma 4 of the Appendix to matrix $\mathbf{P}_{\infty,t}$, it follows that $r(\mathbf{P}_{\infty,t}^\dagger) = r(\mathbf{P}_{\infty,t}) - r(\mathbf{F}_{\infty,t})$. The proof of (14) is complete, because $r(\mathbf{AB}) \leq \min\{r(\mathbf{A}), r(\mathbf{B})\}$ for any matrix $\mathbf{A}$ and $\mathbf{B}$. Finally, it can be easily verified that $m = r(\mathbf{P}_\infty) = \sum_{t=1}^n r(\mathbf{F}_{\infty,t})$ for a properly defined state-space model (1) where $r(\mathbf{T}_t) \geq r(\mathbf{P}_{\infty,t}^\dagger)$, for $t = 1, \ldots, n$.

Result (14) ensures that matrix $\mathbf{P}_{\infty,d+1}$ for some time period $d < n$ is 0 and that $\mathbf{P}_t = \mathbf{P}_{*,t}$, for $t = d+1, \ldots, n$. Thus the Kalman update equations (2) are exact for $t = d+1, \ldots, n$, because the dependency of $\mathbf{P}_t$ on κ is eliminated after time $t = d$. The set of update equations (11) is termed the exact initial Kalman filter: a relatively simple but general algorithm for computing the exact initialization of the Kalman filter at time $t = d + 1$.

For univariate cases, the initialization period length d is equal to m when no missing observations are present in the initial dataset. When some data points or systematic sequences of data points are missing at the beginning of the dataset (e.g., all January observations are missing in the first 10 years of a monthly dataset), the initialization period d is larger than m. Under these circumstances, cases where $\mathbf{F}_{\infty,t} = \mathbf{Z}_t \mathbf{P}_{\infty,t} \mathbf{Z}_t' = 0$ but with $\mathbf{P}_{\infty,t} \neq 0$ occur, for $t = 1, \ldots, d$.

3. EXACT INITIAL SMOOTHING

The backward smoothing algorithm (3) is valid for $t = n, \ldots, d+1$, because the Kalman quantities do not depend on κ. Adjustments for dealing with κ in the initial period $t = d, \ldots, 1$, are derived as follows. The matrices $\mathbf{K}_t$ and $\mathbf{F}_t^{-1}$ in (3) are not multiplied by matrices depending on κ, so it follows from (8) and (12) that they are equal to $\mathbf{K}_{*,t}$ and $\mathbf{F}_{*,t}^-$, as $\kappa \to \infty$. This leads to the exact solution for smoothing; that is,

$$\mathbf{r}_{*,t-1} = \mathbf{Z}_t' \mathbf{F}_{*,t}^- \mathbf{v}_t + \mathbf{L}_{*,t}' \mathbf{r}_{*,t},$$

$$\mathbf{N}_{*,t-1} = \mathbf{Z}_t' \mathbf{F}_{*,t}^- \mathbf{Z}_t + \mathbf{L}_{*,t}' \mathbf{N}_{*,t} \mathbf{L}_{*,t}, \qquad (16)$$

with $\mathbf{L}_{*,t} = \mathbf{T}_t - \mathbf{K}_{*,t} \mathbf{Z}_t$, for $t = d, \ldots, 1$. The backward recursions (16) are initialized by $\mathbf{r}_{*,d} = \mathbf{r}_d$ and $\mathbf{N}_{*,d} = \mathbf{N}_d$. It is surprising that the exact initial smoother (16) is the same as the usual smoother (3), as it requires no extra storage or computing.

On the other hand, the smoother adjustments for the diffuse Kalman filter of de Jong (1991) are difficult to implement and are computationally inefficient (see Chu-Chun-Lin and de Jong 1993). This is a strong argument for using the exact initial Kalman filter of Section 2 when dealing with nonstationary time series models. The next section sets out the importance of Kalman filtering and smoothing for maximum likelihood estimation of state-space models.

4. LOG-LIKELIHOOD FUNCTION AND SCORE VECTOR

The diffuse likelihood function is the likelihood function of $\mathbf{y}$ invariant to the diffuse vector $\boldsymbol{\delta}$, where $\boldsymbol{\delta} \sim \mathrm{N}(0, \kappa \mathbf{I})$ and $\kappa \to \infty$. Specifically, it is the likelihood function of $\mathbf{My}$, where the rank of M is $r(\mathbf{M}) = Nn - m$, $\mathrm{cov}(\mathbf{My}, \boldsymbol{\delta}) = 0$, and $\log|\mathrm{var}(\mathbf{My})| = \sum_{t=1}^n \log|\mathbf{F}_t| - m \log|\kappa|$ (see Ansley and Kohn 1985, thm. 5.1, and de Jong 1991, thm. 4.2). The diffuse log-likelihood is formally defined as

$$\log L_\infty(\mathbf{y}) = \log L(\mathbf{y}) + \frac{m}{2} \log|\kappa|, \qquad (17)$$

where

$$\log L(\mathbf{y}) = \text{constant} - \frac{1}{2} \sum_{t=1}^n \log|\mathbf{F}_t| - \frac{1}{2} \sum_{t=1}^n \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t. \qquad (18)$$

The likelihood function for state-space models with $\mathbf{P}_\infty = 0$ is given by (18), which follows from the prediction error decomposition (see Schweppe 1965). The exact diffuse log-likelihood function (17) is relevant for nonstationary time series models and, as $\kappa \to \infty$, it can be expressed directly in terms of the initial Kalman filter quantities. First, it follows from (8) that $\sum_{t=1}^n \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t = \sum_{t=1}^n \mathbf{v}_t' \mathbf{F}_{*,t}^- \mathbf{v}_t$ as $\kappa \to \infty$. Also, note that $m = \sum_{t=1}^n r(\mathbf{F}_{\infty,t})$ for a properly defined state-space model (1). Applying Theorem 3 of the Appendix to $|\mathbf{F}_t| = |\mathbf{F}_{*,t} + \kappa \mathbf{F}_{\infty,t}|$ and taking logs yields

$$\log|\mathbf{F}_t| - r(\mathbf{F}_{\infty,t}) \log \kappa = -\log|\mathbf{F}_{*,t}^- + \mathbf{F}_{\infty,t}^-|,$$

so that

$$\log L_\infty(\mathbf{y}) = \text{constant} + \frac{1}{2} \sum_{t=1}^n \log|\mathbf{F}_{*,t}^- + \mathbf{F}_{\infty,t}^-| - \frac{1}{2} \sum_{t=1}^n \mathbf{v}_t' \mathbf{F}_{*,t}^- \mathbf{v}_t. \qquad (19)$$

The definitions of $\mathbf{F}_{*,t}^-$ and $\mathbf{F}_{\infty,t}^-$ are given below Equation (8), and for the special cases $\mathbf{F}_t^\infty = 0$ and $r(\mathbf{F}_t^\infty) = N$, they are given above Equation (13).

Unknown elements of the system matrices are placed in the hyperparameter vector ψ. Maximum likelihood estimation of ψ involves numerical maximization of the diffuse log-likelihood function (19). Quasi-Newton optimization methods require score information (see Gill, Murray and Wright 1981). Koopman and Shephard (1992) showed that the score vector with respect to ψ can be obtained by using an appropriate smoothing algorithm. For example, when the ith element of the hyperparameter vector ψ relates only to entries in $\mathbf{H}_t$ and $\mathbf{H}_t \mathbf{G}_t' = 0$, for $t = 1, \ldots, n$, its score value, evaluated at $\psi = \psi^*$, is given by

$$q_i(\psi^*) = \left. \frac{\partial \log L_\infty(\mathbf{y})}{\partial \psi_i} \right|_{\psi=\psi^*} = \mathrm{tr} \sum_{t=1}^n \frac{\partial \mathbf{H}_t}{\partial \psi_i} \mathbf{H}_t' (\mathbf{r}_t \mathbf{r}_t' - \mathbf{N}_t), \qquad (20)$$

where $\mathbf{r}_t$ and $\mathbf{N}_t$ are obtained from the smoother (3). System matrix $\mathbf{H}_t$ does not depend on κ, so $\mathbf{r}_t$ and $\mathbf{N}_t$ in (20) can be directly replaced by $\mathbf{r}_{*,t}$ and $\mathbf{N}_{*,t}$ of (16) for $t = d, \ldots, 1$, as $\kappa \to \infty$. It is convenient that nonstationary time series models do not impose a huge computational burden on exact likelihood and score evaluation.

5. SOME EXAMPLES

5.1 Local-Level Component Model

The univariate time series model with a local level component belongs to the class of unobserved components time series models (see Harrison and Stevens 1976 and Harvey 1989). The model is given by

$$y_t = \mu_t + [\sigma_y \quad 0] \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathrm{NID}(0, \mathbf{I}_2),$$

$$\mu_{t+1} = \mu_t + [0 \quad \sigma_\mu] \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, n, \qquad (21)$$

where μ_t is the stochastic time-varying level component and ε_t is the 2×1 vector of disturbances. It is straightforward to represent the local level model as a time-invariant state-space model (1): the (scalar) system matrices $\mathbf{Z}$ and $\mathbf{T}$ are set equal to unity, and $\mathbf{GG}' = \sigma_y^2$ and $\mathbf{HH}' = \sigma_\mu^2$. The initial state is $\alpha_1 \sim \mathrm{N}(0, \kappa)$, where $\kappa \to \infty$. The exact initial Kalman filter starts off with $a_1 = 0, P_{*,1} = 0$, and $P_{\infty,1} = 1$. Then the first update is given by

$$\begin{array}{lll} v_1 = y_1, & F_{*,1} = \sigma_y^2, & F_{\infty,1} = 1, \\ K_{*,1} = 1, & C_{*,1} = -\sigma_y^2, & C_{\infty,1} = 1, \\ a_2 = y_1, & P_{*,2} = \sigma_y^2 + \sigma_\mu^2, & P_{\infty,2} = 0. \end{array}$$

This example for model (21) shows that the initial Kalman filter allows the Kalman filter (2) to start off at $t = 2$ with $a_2 = y_1$ and $P_2 = \sigma_y^2 + \sigma_\mu^2$ (see also Harvey 1989, example 3.2.1).

5.2 Local Linear Trend Component Model

The time series model with a local linear trend component is given by

$$y_t = \mu_t + [\sigma_y \quad 0 \quad 0]\varepsilon_t, \quad \varepsilon_t \sim \mathrm{NID}(0, \mathbf{I}_3),$$

$$\mu_{t+1} = \mu_t + \beta_t + [0 \quad \sigma_\mu \quad 0]\varepsilon_t,$$

$$\beta_{t+1} = \beta_t + [0 \quad 0 \quad \sigma_\beta]\varepsilon_t, \quad t = 1, \ldots, n, \tag{22}$$

where the stochastic process μ_t represents a time-varying trend with a time-varying slope term β_t and ε_t is a 3×1 vector of disturbances (see Harvey 1989, sec. 2.3.2). The stochastic trend model (22) can be represented as a time-invariant state-space model, where the state vector is given by $\boldsymbol{\alpha}_t = (\mu_t, \beta_t)'$ and the system matrices are given by

$$\mathbf{Z} = [1 \quad 0], \quad \mathbf{T} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{GG}' = \sigma_y^2,$$

$$\mathbf{HH}' = \begin{bmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix},$$

with $\mathbf{HG}' = 0$. The initial Kalman filter starts off with $\mathbf{a}_1 = 0, \mathbf{P}_{*,1} = 0$, and $\mathbf{P}_{\infty,1} = \mathbf{I}_2$, and the first update is

$$\mathbf{K}_{*,1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{C}_{*,1} = -\sigma_y^2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

$$\mathbf{C}_{\infty,1} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} y_1 \\ 0 \end{bmatrix},$$

$$\mathbf{P}_{*,2} = \sigma_y^2 \begin{bmatrix} 1 + q_\mu & 0 \\ 0 & q_\beta \end{bmatrix}, \quad \mathbf{P}_{\infty,2} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

where $q_\mu = \sigma_\mu^2/\sigma_y^2$ and $q_\beta = \sigma_\beta^2/\sigma_y^2$. Note that the rank of $\mathbf{P}_{\infty,2}$ equals 1. The next update is

$$\mathbf{K}_{*,2} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \mathbf{C}_{*,2} = -\sigma_y^2 \begin{bmatrix} 4 & 3 + q_\mu \\ 3 + q_\mu & 2 + q_\mu \end{bmatrix},$$

$$\mathbf{C}_{\infty,2} = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 2y_2 - y_1 \\ y_2 - y_1 \end{bmatrix},$$

$$\mathbf{P}_{*,3} = \sigma_y^2 \begin{bmatrix} 5 + 2q_\mu + q_\beta & 3 + q_\mu + q_\beta \\ 3 + q_\mu + q_\beta & 2 + q_\mu + 2q_\beta \end{bmatrix},$$

$$\mathbf{P}_{\infty,3} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

such that $\mathbf{P}_3 = \mathbf{P}_{*,3}$. It is shown that the initial Kalman filter lets the usual Kalman filter (2) start at $t = 3$ with the initializations $\mathbf{a}_3$ and $\mathbf{P}_3 = \mathbf{P}_{*,3}$ as given.

5.3 Common-Level Component Model

Consider the bivariate common-level component model

$$\mathbf{y}_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} 0 \\ \bar{\mu} \end{pmatrix} + \begin{pmatrix} 1 \\ \theta \end{pmatrix} \mu_t + \begin{bmatrix} \sigma_1 & 0 & 0 \\ \sigma_{12} & \sigma_2 & 0 \end{bmatrix} \varepsilon_t,$$

$$\varepsilon_t \sim \mathrm{NID}(0, \mathbf{I}_3),$$

$$\mu_{t+1} = \mu_t + [0 \quad 0 \quad \sigma_\mu]\varepsilon_t, \quad t = 1, \ldots, n, \tag{23}$$

where ε_t is the 3×1 disturbance vector and $\bar{\mu}$ is a fixed constant. This model lets both series in $\mathbf{y}_t$ depend on the same underlying trend. The state-space form of the bivariate model is based on the state vector $\boldsymbol{\alpha}_t = (\mu_t, \bar{\mu})'$ with the initial state $\boldsymbol{\alpha}_1 \sim \mathrm{N}(0, \kappa \mathbf{I}_2)$, where $\kappa \to \infty$. Note that the constant $\bar{\mu}$ is treated as a diffuse random vector. The exact Kalman filter initialization is $\mathbf{a}_1 = 0, \mathbf{P}_{*,1} = 0$, and $\mathbf{P}_{\infty,1} = \mathbf{I}_2$. For simplicity, we set $\sigma_1 = \sigma_2 = 1$ and $\sigma_{12} = 0$ such that $\mathbf{GG}' = \mathbf{I}_2$. The initial Kalman step is

$$\mathbf{v}_1 = \begin{bmatrix} y_{1,1} \\ y_{2,1} \end{bmatrix}, \quad \mathbf{F}_{*,1} = \mathbf{I}_2,$$

$$\mathbf{F}_{\infty,1} = \begin{bmatrix} 1 & \theta \\ \theta & 1 + \theta^2 \end{bmatrix}, \quad \mathbf{K}_{*,1} = \begin{bmatrix} 1 & 0 \\ -\theta & 1 \end{bmatrix},$$

$$\mathbf{C}_{*,1} = -\begin{bmatrix} 1 & -\theta \\ -\theta & 1 + \theta^2 \end{bmatrix}, \quad \mathbf{C}_{\infty,1} = \mathbf{I}_2,$$

$$\mathbf{a}_2 = \begin{bmatrix} y_{1,1} \\ y_{2,1} - \theta y_{1,1} \end{bmatrix},$$

$$\mathbf{P}_{*,2} = \begin{bmatrix} 1 + \sigma_\mu^2 & -\theta \\ -\theta & 1 + \theta^2 \end{bmatrix}, \quad \mathbf{P}_{\infty,2} = 0.$$

This shows that the Kalman filter (2) can start from $t = 2$ onward.

The restriction $\bar{\mu} = 0$ enforces the ratio of both levels to be constant for all time periods $t = 1, \ldots, n$. The state-space form of the restricted model is given by

$$\mathbf{y}_t = \begin{bmatrix} 1 \\ \theta \end{bmatrix} \alpha_t + \begin{bmatrix} \alpha_1 & 0 & 0 \\ \sigma_{12} & \sigma_2 & 0 \end{bmatrix} \varepsilon_t,$$

$$\alpha_{t+1} = \alpha_t + [0 \quad 0 \quad \sigma_\mu]\varepsilon_t, \tag{24}$$

with the scalar state $\alpha_t = \mu_t$ and the initial state $\alpha_1 \sim \mathrm{N}(0, \kappa)$, where $\kappa \to \infty$. The intialization is $a_1 = 0, P_{*,1} = 0$, and $P_{\infty,1} = 1$. For this case, matrix $\mathbf{F}_{\infty,1}$ is singular; that is, $r(\mathbf{F}_{\infty,1}) = 1$. The initial Kalman step, when $\mathbf{GG}' = \mathbf{I}_2$, is given by

$$\mathbf{v}_1 = \begin{bmatrix} y_{1,1} \\ y_{2,1} \end{bmatrix}, \quad \mathbf{F}_{*,1} = \mathbf{I}_2, \quad \mathbf{F}_{\infty,1} = \begin{bmatrix} 1 & \theta \\ \theta & \theta^2 \end{bmatrix},$$

$$\mathbf{J} = \begin{bmatrix} \varphi & -\theta\sqrt{\varphi} \\ \theta\varphi & \sqrt{\varphi} \end{bmatrix}, \quad \mathbf{F}_{*,1}^{-} = \varphi \begin{bmatrix} \theta^2 & -\theta \\ -\theta & 1 \end{bmatrix},$$

$$\mathbf{F}_{\infty,1}^{-} = \varphi^2 \begin{bmatrix} 1 & \theta \\ \theta & \theta^2 \end{bmatrix},$$

$$\mathbf{K}_{*,1} = \varphi[1 \quad \theta], \qquad C_{*,1} = -\varphi, \qquad C_{\infty,1} = 1,$$

$$a_2 = \varphi(y_{1,1} + \theta y_{2,1}), \qquad P_{*,2} = \varphi + \sigma_\mu, \qquad P_{\infty,2} = 0,$$

where $\varphi = (1 + \theta^2)^{-1}$. Note that $\mathbf{F}_{\infty,1}\mathbf{F}_{*,1}^{-} = 0$. Again, the Kalman filter (2) is applied from $t = 2$ onward. (More details on the common trend components model can be found in Harvey and Koopman 1997.)

6. MISCELLANEOUS ISSUES

This section discusses some practical issues: computational costs, numerical performance of the usual Kalman filter and missing values. These matters are illustrated by using the following unobserved components time series models:

1. Local level model (see Sec. 5 or Harvey 1989, p. 102)
2. Local linear trend model (see Sec. 5 or Harvey 1989, p. 170)
3. Trend model with quarterly seasonals (see Harvey 1989, p. 172)
4. Trend model with monthly seasonals (see Harvey 1989, p. 172)
5. Trend model with stochastic cycle (see Harvey 1989, p. 171)
6. Bivariate local level model (see Harvey and Koopman 1997)
7. Bivariate local linear trend model (see Harvey and Koopman 1997).

6.1 Computational Costs

The exact initial Kalman filter requires more computations than the usual Kalman filter. However, after a limited number of updates, the initial Kalman filter reduces to the usual Kalman filter, and no extra computations are required. The additional computations for the initial Kalman filter are caused by the updating of matrix $\mathbf{P}_{\infty,t}$ and the computing of matrices $\mathbf{K}_{*,t}$ and $\mathbf{C}_{*,t}$ when $r(\mathbf{F}_t^{\infty}) \neq 0$, for $t = 1, \ldots, d$. For many practical state-space models, the system matrices $\mathbf{Z}_t$ and $\mathbf{T}_t$ are sparse selection matrices containing many 0s and ones such that the updating of matrix $\mathbf{P}_{\infty,t}$ is computationally cheap. Table 1 reports the number of extra flops

Table 1. Number of Extra Flops

Model	Initial	Diffuse
Likelihood evaluation		
1. Local level (N = 1)	3	7
2. Local linear trend (N = 1)	18	46
3. Trend plus cycle (N = 1)	60	100
4. Basic seasonal (s = 4, N = 1)	225	600
5. Basic seasonal (s = 12, N = 1)	3,549	9,464
7. Local level (N = 2)	44	76
8. Local linear trend (N = 2)	272	488
Score vector calculation		
1. Local level (N = 1)	0	11
2. Local linear trend (N = 1)	0	105
4. Basic seasonal (s = 4, N = 1)	0	1,800

(compared to the usual Kalman filter) required for likelihood evaluation using the approach of Section 2 and using the DKF approach of de Jong (1991) as described in Section 1.

The results in Table 1 show that on average, the additional number of computations for the initial Kalman filter is 50% less compared to the additional computational effort for the DKF when the set of models 1–7 is considered. Regarding computation of the score vector, the results in Table 1 for the approach of Section 3 are even more favorable. The initial smoothing algorithm requires no additional computing, whereas the DKF smoothing algorithm needs additional computing for the initial period (see Chu-Chun-Lin and de Jong 1993). For example, DKF smoothing requires an additional 1,800 flops when it is used to calculate the score vector for a trend model with quarterly seasonals; see Table 1. This will slow down the maximum likelihood estimation of such models considerably.

6.2 Missing Values

The initial Kalman filter can deal with any pattern of missing data in the set of observations. For example, the strategy of initializing the Kalman filter for the local linear trend model in Section 5 is still valid when observation 2 is missing. In this case, the Kalman filter (2) starts at $t = 4$ with the initialization

$$\mathbf{a}_4 = \begin{bmatrix} 1.5y_3 - .5y_1 \\ .5y_3 - .5y_1 \end{bmatrix},$$

$$\mathbf{P}_4 = \sigma_y^2 \begin{bmatrix} 2.5 + 1.5q_\mu + 1.25q_\beta & 1 + .5q_\mu + 1.25q_\beta \\ 1 + .5q_\mu + 1.25q_\beta & .5 + .5q_\mu + 2.25q_\beta \end{bmatrix},$$

where $q_\mu = \sigma_\mu^2/\sigma_y^2$ and $q_\beta = \sigma_\beta^2/\sigma_y^2$. The solution can be verified by applying the initial Kalman filter (11). Note that when the full observation vector $\mathbf{y}_t$ is missing, the initial Kalman quantities $\mathbf{K}_t, \mathbf{C}_{*,t}$, and $\mathbf{C}_{\infty,t}$ are set to 0.

To illustrate how the initialization deals with missing data, Table 2 considers a univariate time series and a bivariate time series. Missing entries in the time series are indicated by "y." The rank of the matrices $\mathbf{F}_{\infty,t}$ and $\mathbf{P}_{\infty,t}$ for $t = 1, \ldots, n$, are reported for a selection of time series models. Note that $\mathbf{F}_{\infty,t}$ is a scalar in the univariate case.

Table 2. Rank Reduction When Missing Observations are Present

	Univariate Kalman filter														
t	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
Missing	n	y	n	y	n	y	n	n	n	y	n	n	n	n	n
Model 2, 5															
$r(\mathbf{F}_{\infty,t})$	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
$r(\mathbf{P}_{\infty t})$	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Model 3															
$r(\mathbf{F}_{\infty t})$	1	0	1	0	1	0	0	1	0	0	0	0	0	1	0
$r(\mathbf{P}_{\infty,t})$	5	4	4	3	3	2	2	2	1	1	1	1	1	1	0

	Bivariate Kalman filter				
t	*1*	*2*	*3*	*4*	*5*
Missing Y_1	n	y	n	n	n
Missing Y_2	y	n	y	n	n
Model 6					
$r(\mathbf{F}_{\infty,t})$	1	1	0	0	0
$r(\mathbf{P}_{\infty,t})$	2	1	0	0	0
Model 7					
$r(\mathbf{F}_{\infty,t})$	1	1	1	1	0
$r(\mathbf{P}_{\infty,t})$	4	3	2	1	0

Table 2 confirms that the initial Kalman filter must be applied as long as the rank of covariance matrix $\mathbf{P}_{\infty,t}$ is larger than 0. The rank of matrix $\mathbf{F}_{\infty,t}$ determines the rank reduction of $\mathbf{P}_{\infty,t+1}$; see Section 2. The conclusion is that the presence of missing values in the initial period does not alter our strategy of initializing the Kalman filter.

6.3 Numerical Performance

The initial Kalman filter requires an extra computational effort. It would be computationally more efficient to apply the usual Kalman filter and replace κ by some large value k, though this strategy may give numerical inaccuracies and is not an exact solution. Bell and Hillmer (1991) assessed the performance of the large k initialization by comparing the innovation variance of the numerical approach and the exact approach. The performance of the numerical approach can be measured by the statistic

$$\log k \|\mathbf{F}_t^{(k)} - \mathbf{F}_{\infty,t}\|, \quad t = d+1, \ldots n,$$

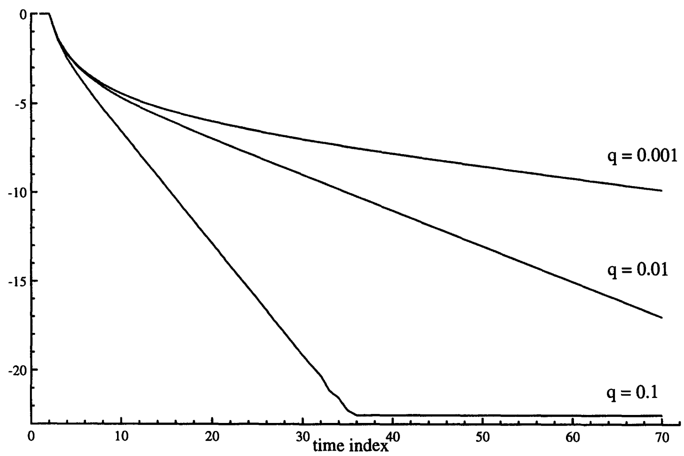

Figure 1. Numerical Accuracy of the Kalman Filter for the Local Level Model.

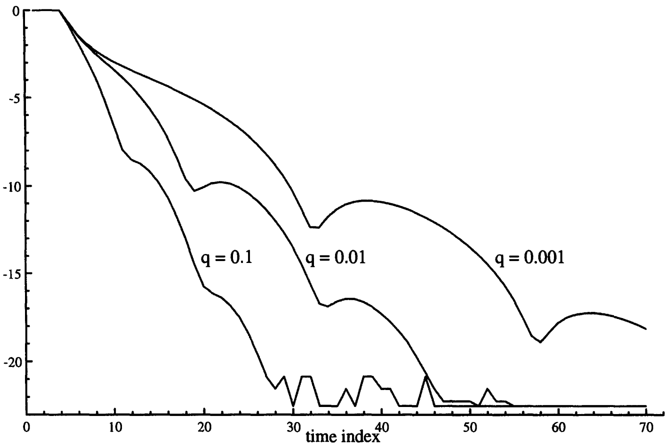

Figure 2. Numerical Accuracy of the Kalman Filter for the Local Trend Model.

where $\mathbf{F}_{\infty,t}$ is the variance matrix $\mathbf{F}_t$ of the Kalman filter with the exact initialization and $\mathbf{F}_t^{(k)}$ is the corresponding $\mathbf{F}_t$ matrix of the Kalman filter with the large k initialization for $t = d+1, \ldots, n$. Note that $\|\mathbf{A}\|$ denotes the absolute value of the determinant of square matrix $\mathbf{A}$. The statistic should be as negative as possible to achieve numerical precision for the usual Kalman filter. The statistic depends on the numerical accuracy of the computer; the most negative value obtained on my Pentium computer was -22.5.

The statistic is calculated for models 1, 2, and 3 with variance matrix $\mathbf{HH}'$ equal to $q\mathbf{I}$, where q is some predetermined value. Figure 1 plots the statistic for model 1 and three different values of q. The choice of k is 10^7, but the statistic is almost invariant to k when $k > 10^4$. Similarly, Figures 2 and 3 are constructed for models 2 and 3. They show that the numerical performance of the usual Kalman filter is unsatisfactory, because most values are too far away from the computer accuracy level of -22.5. Also, the statistic depends very much on the choice of the value q.

7. CONCLUSIONS

This article has presented an analytical solution for the initialization of the Kalman filter for state-space models

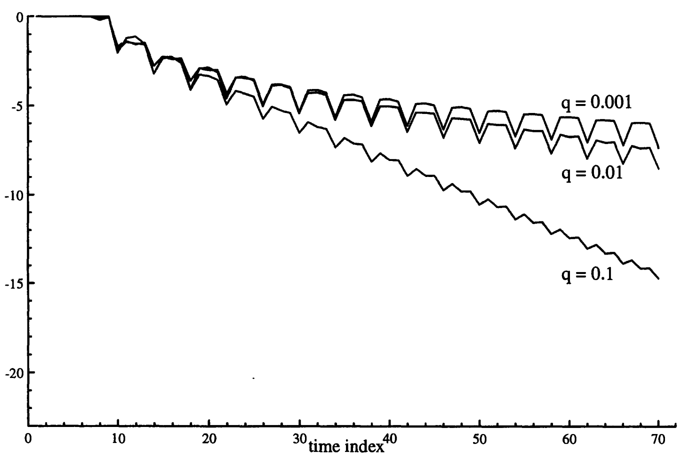

Figure 3. Numerical Accuracy of the Kalman Filter for the Local Level Model.

with diffuse initial conditions. The univariate initial Kalman filter is similar to the modified Kalman filter of Ansley and Kohn (1990). However, more efficient and straightforward methods are developed for univariate and multivariate state-space models. The proofs are simple and transparent. Also, it is shown that likelihood and score evaluation for nonstationary time series models requires only a small number of extra computations. The new results have important implications for the computational efficiency of statistical algorithms dealing with univariate and multivariate time series models.

Other findings of this article are as follows:

- The alternative solution of de Jong (1991) is computationally less efficient when applied to a wide class of time series models with ARIMA components.
- The performance of the Kalman filter with the diffuse prior replaced by some large value is numerically unsatisfactory.
- The exact initial filtering and smoothing algorithms deal with any pattern of missing data entries in the time series.

APPENDIX: TECHNICAL PROOFS

Notation. The dimensionality of vectors or matrices is not given when it is obvious or when it is not relevant. Most of the notation in this article is standard, but a selection is summarized here for convenience:

- The identity matrix is $\mathbf{I}$ and the $m \times m$ identity matrix is $\mathbf{I}_m$.
- The rank of matrix $\mathbf{A}$ is $r(\mathbf{A})$, its determinant is $|\mathbf{A}|$, and its inverse is $\mathbf{A}^{-1}$.
- A block diagonal matrix is denoted by $\text{diag}\{\mathbf{A}, \mathbf{B}\}$ with $\mathbf{A}$ and $\mathbf{B}$ both square; psd refers to a symmetric positive (semi) definite matrix.

Lemma 1. Simultaneous Diagonalization of a psd matrix $\mathbf{A}$ and a pd matrix $\mathbf{M}$. Let $\mathbf{A}$ be a $p \times p$ psd matrix and let $\mathbf{M}$ be a $p \times p$ pd matrix. A nonsingular matrix $\mathbf{P}$ and a diagonal positive semidefinite matrix $\mathbf{\Lambda}$ exist so that

$$\mathbf{A} = \mathbf{P}'\mathbf{\Lambda}\mathbf{P}, \qquad \mathbf{M} = \mathbf{P}'\mathbf{P}.$$

It follows that

$$\mathbf{Q}'\mathbf{A}\mathbf{Q} = \mathbf{\Lambda}, \qquad \mathbf{Q}'\mathbf{M}\mathbf{Q} = \mathbf{I}, \qquad \mathbf{M}^{-1} = \mathbf{Q}\mathbf{Q}',$$

where $\mathbf{Q} = \mathbf{P}^{-1}$ (see Magnus and Neudecker 1988, sec. 1.18, them. 23).

Lemma 2. Partial Diagonalization of two psd matrices $\mathbf{A}$ and $\mathbf{B}$. Let $\mathbf{A}$ and $\mathbf{B}$ be two $p \times p$ psd matrices for which $\mathbf{A} + \mathbf{B}$ is pd and $r(\mathbf{B}) = q \leq p$. Then

$$\mathbf{Q}'\mathbf{B}\mathbf{Q} = \begin{bmatrix} \mathbf{I}_q & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{Q}'\mathbf{A}\mathbf{Q} = \begin{bmatrix} \mathbf{C}_1 & 0 \\ 0 & \mathbf{I}_{p-q} \end{bmatrix},$$

where $\mathbf{Q}$ is nonsingular and $r(\mathbf{C}_1) \leq q$.

Proof. Matrix $\mathbf{B}$ is diagonalized using nonsingular matrix $\mathbf{K}$, so

$$\mathbf{K}'\mathbf{B}\mathbf{K} = \begin{bmatrix} \mathbf{I}_q & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{K}'\mathbf{A}\mathbf{K} = \begin{bmatrix} \mathbf{E}_1 & \mathbf{E}_{12} \\ \mathbf{E}_{12}' & \mathbf{E}_2 \end{bmatrix},$$

where $r(\mathbf{E}_2) = p - q$ because $r(\mathbf{A} + \mathbf{B}) = p$. Also,

$$\mathbf{L}'\mathbf{K}'\mathbf{A}\mathbf{K}\mathbf{L} = \begin{bmatrix} \mathbf{C}_1 & 0 \\ 0 & \mathbf{I}_{p-q} \end{bmatrix},$$

where $\mathbf{C}_1$ is a $q \times q$ matrix. The operations with respect to the nonsingular matrix $\mathbf{L}$ apply only to the last $p - q$ columns and rows of matrix $\mathbf{K}'\mathbf{A}\mathbf{K}$, so that matrix $\mathbf{L}$ has no effect on $\mathbf{K}'\mathbf{B}\mathbf{K}$, because the last $p - q$ columns and rows of $\mathbf{K}'\mathbf{B}\mathbf{K}$ are 0. Setting $\mathbf{Q} = \mathbf{K}\mathbf{L}$ completes the proof (see also Searle (1982), appendix to chap. 11, sec. 2.c, thm. 3).

Lemma 3. Expansion of ratio $1/(1 + y)$. The infinite geometric progression formula is

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n, \quad |x| < 1.$$

An expansion for

$$\frac{1}{1+y}, \quad y > 1,$$

is obtained through the identities

$$\frac{1}{1+y} = 1 + \left(\frac{1}{1+y} - 1\right) = 1 - \left(\frac{y}{1+y}\right) = 1 - \left(\frac{1}{1+\frac{1}{y}}\right).$$

Setting $x = -1/y$, it follows that

$$\frac{1}{1+y} = \frac{1}{y} - \frac{1}{y^2} + \frac{1}{y^3} - \cdots, \qquad y > 1.$$

Lemma 4. Rank reduction of a psd matrix $\mathbf{A}$. Let $\mathbf{A}$ be a $p \times p$ psd matrix that can be diagonalized such that $\mathbf{Q}'\mathbf{A}\mathbf{Q} = \mathbf{D}$, where $\mathbf{Q}$ is nonsingular and $\mathbf{D}$ is diagonal with unity and zero values. Then

$$r(\mathbf{A} - \mathbf{A}\mathbf{Q}_+\mathbf{Q}_+'\mathbf{A}) = r(\mathbf{A}) - q,$$

where matrix $\mathbf{Q}_+$ is a subset of columns of $\mathbf{Q}$ such that $\mathbf{Q}_+'\mathbf{A}\mathbf{Q}_+ = \mathbf{I}_q$ and $q \leq r(\mathbf{A})$.

Proof. Matrix $\mathbf{A}$ is diagonable, and its rank equals the number of nonzero eigenvalues (see Searle 1982, Appendix to chap. 11, sec. 2b). It follows that $r(\mathbf{A}) = r(\mathbf{Q}'\mathbf{A}\mathbf{Q}) = r(\mathbf{D})$, because $\mathbf{Q}$ is nonsingular. Matrix $\mathbf{Q}$ can be scaled such that matrix $\mathbf{D}$ contains only unity and zero values. The rank of $\mathbf{A}$ is equal to the number of unity values of $\mathbf{D}$. The proof is complete, because the diagonal matrix $\mathbf{Q}'\mathbf{A}\mathbf{Q} - \mathbf{Q}'\mathbf{A}\mathbf{Q}_+\mathbf{Q}_+'\mathbf{A}\mathbf{Q}$ has q unity values less compared to matrix $\mathbf{D} = \mathbf{Q}'\mathbf{A}\mathbf{Q}$.

Theorem 1. Expansion of $(\mathbf{A}+\kappa\mathbf{M})^{-1}$ where $\mathbf{A}$ is a psd matrix and $\mathbf{M}$ is a pd matrix. Let $\mathbf{A}$ be a $p \times p$ psd matrix, and let $\mathbf{M}$ be a $p \times p$ pd matrix. Then

$$(\mathbf{A} + \kappa\mathbf{M})^{-1} = \frac{1}{\kappa}\mathbf{M}^{-1}\frac{1}{\kappa^2}\mathbf{M}\mathbf{A}\mathbf{M}^{-1} + O\left(\frac{1}{\kappa^3}\right),$$

with κ large enough.

Proof. It follows from Lemma 1 that the inverse of $(\mathbf{A}+\kappa\mathbf{M})$ can be written as

$$\begin{aligned} (\mathbf{A} + \kappa\mathbf{M})^{-1} &= (\mathbf{P}'\mathbf{\Lambda}\mathbf{P} + \kappa\mathbf{P}'\mathbf{P})^{-1} \\ &= \{\mathbf{P}'(\mathbf{\Lambda} + \kappa\mathbf{I})\mathbf{P}\}^{-1} \\ &= \mathbf{Q}(\mathbf{\Lambda} + \kappa\mathbf{I})^{-1}\mathbf{Q}', \end{aligned}$$

where $\mathbf{P}$ is nonsingular and $\mathbf{Q} = \mathbf{P}^{-1}$. The diagonal matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_p)$ may contain 0s, as it is a psd matrix. Any element $1/(\lambda + \kappa)$ of matrix $(\mathbf{\Lambda} + \kappa\mathbf{I})^{-1}$ can be expanded by using Lemma 3 as

$$\frac{1}{\lambda + \kappa} = \frac{1}{\lambda}\left(\frac{1}{1+y}\right) = \frac{1}{\lambda}\left(\frac{1}{y} - \frac{1}{y^2} + \frac{1}{y^3} - \cdots\right),$$

where $y = \kappa/\lambda$ and $\kappa > \lambda$. Resubstituting y back leads to

$$\frac{1}{\lambda+\kappa} = \frac{1}{\kappa} - \frac{\lambda}{\kappa^2} + \frac{\lambda^2}{\kappa^3} - \cdots,$$

so that

$$(\mathbf{\Lambda} + \kappa\mathbf{I})^{-1} = \frac{1}{\kappa}\,\mathbf{I} - \frac{1}{\kappa^2}\,\mathbf{\Lambda} + \frac{1}{\kappa^3}\,\mathbf{\Lambda}^2 - \cdots.$$

By writing

$$(\mathbf{A} + \kappa\mathbf{M})^{-1} = \mathbf{Q}\left[\frac{1}{\kappa}\,\mathbf{I} - \frac{1}{\kappa^2}\,\mathbf{\Lambda} + O\left(\frac{1}{\kappa^3}\right)\right]\mathbf{Q}'$$

and by applying Lemma 1, which states that $\mathbf{QQ}' = \mathbf{M}^{-1}$ and $\mathbf{\Lambda} = \mathbf{Q}'\mathbf{AQ}$, the proof is completed. Note that scalar κ must be subject to $\kappa > \max(\lambda_1, \ldots, \lambda_p)$.

Theorem 2. Expansion of $(\mathbf{A} + \kappa\mathbf{B})^{-1}$ where $\mathbf{A}$ and $\mathbf{B}$ are psd matrices. Let $\mathbf{A}$ and $\mathbf{B}$ be two $p \times p$ psd matrices such that $\mathbf{A} + \kappa\mathbf{B}$ is pd as $\kappa > 0$ and $r(\mathbf{B}) = q \leq p$. Then

$$(\mathbf{A} + \kappa\mathbf{B})^{-1} = \mathbf{A}^- + \frac{1}{\kappa}\,\mathbf{B}^{-1} - \frac{1}{\kappa^2}\,\mathbf{B}^-\mathbf{AB}^- + O\left(\frac{1}{\kappa^3}\right),$$

where $\mathbf{A}^- = \mathbf{Q}_2\mathbf{Q}_2', \mathbf{B}^- = \mathbf{Q}_1\mathbf{Q}_1'$ and $\mathbf{Q} = [\mathbf{Q}_1 \quad \mathbf{Q}_2]$ is the partial diagonalization matrix of Lemma 2 so that $r(\mathbf{Q}_1'\mathbf{AQ}_1) = q$.

Proof. Lemma 2 shows that matrix $\mathbf{A} + \mathbf{B}$ can be partially diagonalized by the nonsingular matrix $\mathbf{Q} = [\mathbf{Q}_1 \quad \mathbf{Q}_2]$, where $\mathbf{Q}_1$ is a $p \times q$ matrix and $\mathbf{Q}_2$ is a $p \times p - q$ matrix, so that $r(\mathbf{Q}_1'\mathbf{AQ}_1) = q, \mathbf{Q}_2'\mathbf{AQ}_2 = \mathbf{I}_{p-q}, \mathbf{Q}_1'\mathbf{BQ}_1 = \mathbf{I}_q$, and

$$\mathbf{QQ}' = [\mathbf{Q}_1 \quad \mathbf{Q}_2]\begin{bmatrix}\mathbf{Q}_1' \\ \mathbf{Q}_2'\end{bmatrix} = \mathbf{B}^- + \mathbf{A}^-,$$

with $\mathbf{A}^-$ and $\mathbf{B}^-$ as defined earlier. Define the partial diagonal matrix $\mathbf{C} = \mathbf{Q}'(\mathbf{A} + \kappa\mathbf{B})\mathbf{Q}$, and define the nonsingular matrix $\mathbf{C}_1 = \mathbf{Q}_1'\mathbf{AQ}_1$. It follows from Lemma 2 that $\mathbf{C} = \text{diag}\{\mathbf{C}_1 + \kappa\mathbf{I}_q, \mathbf{I}_{p-q}\}$, and by applying Theorem 1 to matrix $\mathbf{C}_1 + \kappa\mathbf{I}_q$, we obtain that $\mathbf{C}^{-1} = \text{diag}\{[(1/\kappa)\mathbf{I}_q] - [(1/\kappa^2)\mathbf{C}_1] + O(1/\kappa^3), \mathbf{I}_{p-q}\}$ or

$$\mathbf{C}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{I}_{p-q}\end{bmatrix} + \frac{1}{\kappa}\begin{bmatrix}\mathbf{I}_q & 0 \\ 0 & 0\end{bmatrix} - \frac{1}{\kappa^2}\begin{bmatrix}\mathbf{C}_1 & 0 \\ 0 & 0\end{bmatrix} + O\left(\frac{1}{\kappa^3}\right),$$

with κ large enough. The definition of $\mathbf{C}$ implies that $(\mathbf{A} + \kappa\mathbf{B})^{-1} = \mathbf{QC}^{-1}\mathbf{Q}'$, which completes the proof. Note that $\mathbf{Q}_1\mathbf{C}_1\mathbf{Q}_1' = \mathbf{Q}_1\mathbf{Q}_1'\mathbf{AQ}_1\mathbf{Q}_1' = \mathbf{B}^-\mathbf{AB}^-$.

Theorem 3. Determinant of $\mathbf{A} + \kappa\mathbf{B}$ where $\mathbf{A}$ and $\mathbf{B}$ are psd matrices. Let $\mathbf{A}$ and $\mathbf{B}$ be two $p \times p$ psd matrices such that $\mathbf{A} + \kappa\mathbf{B}$ is pd as $\kappa > 0$ and $r(\mathbf{B}) = q \leq p$. Then

$$\kappa^{-q}|\mathbf{A} + \kappa\mathbf{B}| \to |\mathbf{A}^- + \mathbf{B}^-|^{-1}, \quad \text{as} \quad \kappa \to \infty,$$

where $\mathbf{A}^- = \mathbf{Q}_2\mathbf{Q}_2'$ and $\mathbf{B}^- = \mathbf{Q}_1\mathbf{Q}_1'$ with matrix $\mathbf{Q} = [\mathbf{Q}_1 \quad \mathbf{Q}_2]$ from Lemma 2 so that $\mathbf{Q}_1'\mathbf{BQ}_1 = \mathbf{I}_q$.

Proof. This result follows from the proof of Theorem 2, because $\mathbf{C} = \mathbf{Q}'(\mathbf{A}+\kappa\mathbf{B})\mathbf{Q} = \text{diag}\{\mathbf{C}_1 + \kappa\mathbf{I}_q, \mathbf{I}_{p-q}\}$ and $\mathbf{QQ}' = \mathbf{A}^- + \mathbf{B}^-$, so that

$$\begin{aligned}\kappa^{-q}|\mathbf{A} + \kappa\mathbf{B}| &= \kappa^{-q}|\mathbf{C}||\mathbf{QQ}'|^{-1} \\ &= \kappa^{-q}|\mathbf{C}_1 + \kappa\mathbf{I}_q||\mathbf{A}^- + \mathbf{B}^-|^{-1} \\ &= |\kappa^{-1}\mathbf{C}_1 + \mathbf{I}_q||\mathbf{A}^- + \mathbf{B}^-|^{-1}.\end{aligned}$$

Letting $\kappa \to \infty$ completes the proof.

[Received November 1995. Revised May 1997.]

REFERENCES

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

Ansley, C. F., and Kohn, R. (1985), "Estimation, Filtering and Smoothing in State Space Models With Incompletely Specified Initial Conditions," *The Annals of Statistics*, 13, 1286–1316.

——— (1990), "Filtering and Smoothing in State Space Models With Partially Diffuse Initial Conditions," *Journal of Time Series Analysis*, 11, 275–293.

Bell, W., and Hillmer, S. (1991), "Initializing the Kalman Filter for Nonstationary Time Series Models," *Journal of Time Series Analysis*, 12, 283–300.

Burridge, P., and Wallis, K. F. (1985), "Calculating the Variance of Seasonally Adjusted Series," *Journal of the American Statistical Association*, 80, 541–552.

Chu-Chun-Lin, S., and de Jong, P. (1993), "A Note on Fast Smoothing," unpublished paper, University of British Columbia, Vancouver.

de Jong, P. (1988), "A Cross-Validation Filter for Time Series Models," *Biometrika*, 75, 594–600.

——— (1991), "The Diffuse Kalman Filter," *The Annals of Statistics*, 19, 1073–1083.

Gill, P. E., Murray, W., and Wright, M. H. (1981), *Practical Optimization*, New York: Academic Press.

Gomez, V., and Maravall, A. (1994), "Estimation, Prediction and Interpolation for Nonstationary Series With the Kalman Filter," *Journal of the American Statistical Association*, 89, 611–624.

Harrison, P. J., and Stevens, C. F. (1976), "Bayesian Forecasting," *Journal of the Royal Statistical Society*, Ser. B, 38, 205–247.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge, U.K.: Cambridge University Press.

Harvey, A. C., and Koopman, S. J. (1997), "Multivariate Structural Time Series Models," in *Systematic Dynamics in Economic and Financial Models*, C. Heij, H. Schumacher, B. Hanzon, and C. Praagman, eds., Chichester, U.K.: Wiley.

Harvey, A. C., and Phillips, G. D. A. (1979), "Maximum Likelihood Estimation of Regression Models With Autoregressive-Moving Average Disturbances," *Biometrika*, 66, 49–58.

Harvey, A. C., and Pierse, R. G. (1984), "Estimating Missing Observations in Economic Time Series," *Journal of the American Statistical Association*, 79, 125–131.

Kitagawa, G. (1981), "A Nonstationary Time Series Model and Its Fitting by a Recursive Filter," *Journal of Time Series Analysis*, 2, 103–116.

Kohn, R., and Ansley, C. F. (1989), "A Fast Algorithm for Signal Extraction, Influence and Cross-Validation in State Space Models," *Biometrika*, 76, 65–79.

Koopman, S. J. (1993), "Disturbance Smoother for State Space Models," *Biometrika*, 80, 117–126.

Koopman, S. J., and Shephard, N. (1992), "Exact Score for Time Series Model in State Space Form," *Biometrika*, 79, 823–826.

Magnus, J. R., and Neudecker, H. (1988), *Matrix Differential Calculus With Applications in Statistics and Econometrics*, Chichester, U.K.: Wiley.

Searle, S. R. (1982), *Matrix Algebra Useful for Statistics*, New York: Wiley.

Schweppe, C. F. (1965), "Evaluation of Likelihood Functions for Gaussian Signals," *IEEE Transactions on Information Theory*, 11, 61–70.

Snyder, R. D., and Saligari, G. R. (1996), "Initialization of the Kalman Filter With Partially Diffuse Initial Conditions," *Journal of Time Series Analysis*, 17, 409–424.