On Computing the Expected Fisher Information Matrix for State-Space Model Parameters

by

Joseph E. Cavanaugh [1]
Department of Statistics, University of Missouri, Columbia

Robert H. Shumway
Division of Statistics, University of California, Davis

Abstract

A general, recursive algorithm is presented for computing the expected Fisher information matrix for state-space model parameters. Simulation results are featured where known Fisher information matrices corresponding to simple state-space models are estimated using both observed and expected information matrices. The accuracy of the two approaches is compared.

Keywords: EM algorithm, Kalman filter, recursive algorithm, time series.

[1] Direct correspondence to: Joseph E. Cavanaugh, Department of Statistics, 222 Math Sciences Bldg., University of Missouri, Columbia, MO, 65211.

1. Introduction

Suppose that $\hat{\theta}_n$ represents a vector of maximum likelihood estimates for a model parameterized by θ with associated likelihood $L(\theta\,|Y_n)$. Let θ_o represent the true value of θ, and let $I_n(\theta)$ denote the Fisher information matrix for θ based on Y_n; i.e.,

$$I_n(\theta) = E\left[-\frac{\partial^2 \ln L(\theta\,|Y_n)}{\partial\theta\partial\theta'}\right].$$

Let

$$\bar{I}(\theta_o) = \lim_{n\to\infty} \frac{1}{n} I_n(\theta_o),$$

where it is assumed that this limit exists, and that $\bar{I}(\theta_o)$ is positive definite.

It is well-known that in most settings, a set of non-restrictive regularity assumptions can be identified which will ensure that $\sqrt{n}(\hat{\theta}_n - \theta_o)$ is asymptotically multivariate normal with mean 0 and covariance matrix $\bar{I}(\theta_o)^{-1}$.

To estimate $I_n(\theta_o)$ for the approximation of $\bar{I}(\theta_o)^{-1}$, it is common to use either the *expected* or the *observed* Fisher information matrix. The expected information matrix is defined as

$$I_n(\hat{\theta}_n) = \left[E\left[-\frac{\partial^2 \ln L(\theta\,|Y_n)}{\partial\theta\partial\theta'}\right]\right]|_{\theta=\hat{\theta}_n}, \tag{1.1}$$

whereas the observed information matrix is defined as

$$\mathcal{I}_n(\hat{\theta}_n, Y_n) = \left[-\frac{\partial^2 \ln L(\theta\,|Y_n)}{\partial\theta\partial\theta'}\right]|_{\theta=\hat{\theta}_n}. \tag{1.2}$$

Perhaps the primary advantage of using $I_n(\hat{\theta}_n)$ over $\mathcal{I}_n(\hat{\theta}_n, Y_n)$ as an estimator of $I_n(\theta_o)$ is that $I_n(\hat{\theta}_n)$ is a maximum likelihood estimator (MLE) of $I_n(\theta_o)$. (This fact follows directly from the invariance property of MLE's.) Yet in many instances, evaluating the expectation in (1.1) is either unfeasible or impractical, making $\mathcal{I}_n(\hat{\theta}_n, Y_n)$ the estimator of choice. Such has generally been the case in the setting of the state-space model, where the structure of the Gaussian log-likelihood often makes $I_n(\hat{\theta}_n)$ inaccessible.

In this paper, we present a relatively simple, recursive algorithm for computing $I_n(\hat{\theta}_n)$ in the state-space setting. Our algorithm is general, and does not rely on restrictive simplifying assumptions. Moreover, it does not involve numerical differentiation, since all required derivatives are evaluated analytically. The algorithm is therefore both flexible and stable.

In Section 2, we provide a brief overview of the state-space model and maximum likelihood estimation in the state-space setting. In Section 3, we present our recursive algorithm for computing $I_n(\hat{\theta}_n)$. In Section 4, we discuss simulation results where known Fisher information matrices corresponding to simple state-space models are estimated using both $I_n(\hat{\theta}_n)$ and $\mathcal{I}_n(\hat{\theta}_n, Y_n)$. The accuracy of the two methods is compared. Finally, in Section 5, we present a short practical application.

2. The State-Space Model and ML Estimation

The state-space model is defined by the equations

$$y_t = A_t x_t + v_t, \tag{2.1}$$

$$x_t = \Phi x_{t-1} + w_t, \tag{2.2}$$

for $t = 1, \ldots, n$ time periods. In (2.1), y_t is an observed q-dimensional process, A_t is a known $q \times p$ design matrix, x_t is an unobserved p-dimensional state process, and the v_t are q-dimensional, zero-mean, independent disturbances with common covariance matrix R. In (2.2), Φ is a $p \times p$ transition matrix, and the w_t are p-dimensional, zero-mean, independent disturbances with common covariance matrix Q.

The mean and covariance matrix of x_o (the initial x_t) will be denoted by μ and Σ, respectively. It is often assumed that μ and Σ are known, since in many applications, μ and Σ are fixed at values predetermined by the investigator. Also, it is usually assumed that x_o, the w_t, and the v_t are mutually independent and multivariate normal.

The set of unknown parameters of the system generally consists of components of Φ, Q, and R. To represent these parameters, we will let θ denote a d-dimensional vector that uniquely determines the model coefficients and correlation structure: i.e., $\Phi = \Phi(\theta)$, $Q = Q(\theta)$, $R = R(\theta)$. We will let Y_t denote the observed data up until time t: i.e., $Y_t = [y_1, \ldots, y_t]$.

The likelihood $L(\theta \mid Y_n)$ is generally written in its *innovations form* (Schweppe, 1965). The Kalman (1960, 1961) filter innovation at time t is defined as

$$e_t(\theta) = y_t - A_t x_t^{t-1}(\theta) \quad \text{where} \quad x_t^{t-1}(\theta) = E(x_t \mid Y_{t-1}).$$

We will let $P_t^{t-1}(\theta)$ denote the conditional covariance matrix of $(x_t - x_t^{t-1}(\theta))$ given Y_{t-1}; i.e.,

$$P_t^{t-1}(\theta) = E((x_t - x_t^{t-1}(\theta))\ (x_t - x_t^{t-1}(\theta))^{'}\ |Y_{t-1}).$$

This will allow us to write the conditional covariance matrix of the innovation given Y_{t-1} as follows:

$$\Sigma_t(\theta) = E(e_t(\theta)\ e_t(\theta)^{'}\ |Y_{t-1}) = A_t P_t^{t-1}(\theta) A_t^{'} + R.$$

The complete Kalman filter equations provide us with a recursive algorithm for evaluating successive values of $x_t^{t-1}(\theta)$, $e_t(\theta)$, $P_t^{t-1}(\theta)$, and $\Sigma_t(\theta)$, as well as

$$x_t^t(\theta) = E(x_t\ |Y_t) \quad \text{and} \quad P_t^t(\theta) = E((x_t - x_t^t(\theta))\ (x_t - x_t^t(\theta))^{'}\ |Y_t).$$

The equations are as follows:

(i) $x_t^{t-1}(\theta) = \Phi x_{t-1}^{t-1}(\theta)$

(ii) $e_t(\theta) = y_t - A_t x_t^{t-1}(\theta)$

(iii) $P_t^{t-1}(\theta) = \Phi P_{t-1}^{t-1}(\theta)\Phi^{'} + Q$

(iv) $\Sigma_t(\theta) = A_t P_t^{t-1}(\theta) A_t^{'} + R$

(v) $K_t(\theta) = P_t^{t-1}(\theta) A_t^{'} \Sigma_t^{-1}(\theta)$

(vi) $x_t^t(\theta) = x_t^{t-1}(\theta) + K_t(\theta) e_t(\theta)$

(vii) $P_t^t(\theta) = P_t^{t-1}(\theta) - K_t(\theta) A_t P_t^{t-1}(\theta)$

The starting values $x_o^o = \mu$ and $P_o^o = \Sigma$ initialize the filter. The matrix $K_t(\theta)$ is called the *gain* or *weight* matrix.

Under the independence and normality assumptions on x_o, the w_t, and the v_t, the innovations are mutually independent and multivariate normal. Thus, for the log of the likelihood $L(\theta\ |Y_n)$, we can write

$$\ln L(\theta\ |Y_n) \propto -\frac{1}{2}\sum_{t=1}^{n} \ln |\Sigma_t(\theta)| - \frac{1}{2}\sum_{t=1}^{n} e_t(\theta)^{'} \Sigma_t^{-1}(\theta) e_t(\theta). \tag{2.3}$$

Since (2.3) is generally non-linear in the parameters, the maximum likelihood estimates $\hat{\theta}_n$ are usually found by using an iterative search algorithm. Maximum likelihood estimation can also be carried out via the EM algorithm. Details are provided in Shumway and Stoffer (1982).

As shown in Harvey (1989, pages 140-142), the $(i,j)^{th}$ element of the Fisher information matrix is given by

$$I_n(\theta;i,j) = E\left[-\frac{\partial^2 \ln L(\theta\ |Y_n)}{\partial\theta_i\partial\theta_j}\right] =$$
$$\frac{1}{2}\sum_{t=1}^{n} tr\left(\ \Sigma_t^{-1}(\theta)\frac{\partial\Sigma_t(\theta)}{\partial\theta_i}\Sigma_t^{-1}(\theta)\frac{\partial\Sigma_t(\theta)}{\partial\theta_j}\ \right) + \sum_{t=1}^{n} E\left(\ \left(\frac{\partial e_t(\theta)}{\partial\theta_i}\right)'\Sigma_t^{-1}(\theta)\left(\frac{\partial e_t(\theta)}{\partial\theta_j}\right)\ \right) \quad (2.4)$$

where $1 \le i,j \le d$. Recursive equations for analytically evaluating

$$\frac{\partial e_t(\theta)}{\partial\theta_i} \quad \text{and} \quad \frac{\partial\Sigma_t(\theta)}{\partial\theta_i}$$

for $i = 1,\ldots,d$ can be derived simply by differentiating each of the Kalman filter equations with respect to θ_i. (For details, see Harvey, 1989, pages 142-143.) These differentiated equations can be evaluated successively alongside the regular filter equations for $t = 1,\ldots,n$.

Sets of conditions which are sufficient to ensure the asymptotic normality of $\hat{\theta}_n$ are referenced and discussed in Harvey (1989, pages 128-130). Provided that a set of such conditions holds, we have the familiar result

$$\sqrt{n}(\hat{\theta}_n - \theta_o) \to N_d(0, \bar{I}(\theta_o)^{-1}).$$

3. Evaluating the Expected Fisher Information Matrix

Our goal is to estimate the Fisher information matrix at θ_o: $I_n(\theta_o)$.

A typical approach consists of evaluating

$$\Sigma_t^{-1}(\hat{\theta}_n), \quad \frac{\partial e_t(\theta)}{\partial\theta_i}|_{\theta=\hat{\theta}_n}, \quad \text{and} \quad \frac{\partial\Sigma_t(\theta)}{\partial\theta_i}|_{\theta=\hat{\theta}_n}$$

for $1 \le i \le d, 1 \le t \le n$, and using these quantities in (2.4) without retaining the expectation operator on the second term (Harvey, 1989, page 142). This essentially amounts to estimating $I_n(\theta_o)$ using $\mathcal{I}_n(\hat{\theta}_n, Y_n)$. Although

$$-\frac{\partial^2 \ln L(\theta\ |Y_n)}{\partial\theta_i\partial\theta_j} \ne \frac{1}{2}\sum_{t=1}^{n} tr\left(\ \Sigma_t^{-1}(\theta)\frac{\partial\Sigma_t(\theta)}{\partial\theta_i}\Sigma_t^{-1}(\theta)\frac{\partial\Sigma_t(\theta)}{\partial\theta_j}\ \right) + \sum_{t=1}^{n}\ \left(\frac{\partial e_t(\theta)}{\partial\theta_i}\right)'\Sigma_t^{-1}(\theta)\left(\frac{\partial e_t(\theta)}{\partial\theta_j}\right),$$

the left-hand and right-hand expressions differ by a factor having expectation zero. Ignoring this factor in computing $\mathcal{I}_n(\hat{\theta}_n, Y_n)$, therefore, amounts to discarding a quantity which estimates zero.

The primary motivation for estimating $I_n(\theta_o)$ with $\mathcal{I}_n(\hat{\theta}_n, Y_n)$ is that a straightforward evaluation of the expectation in (2.4) is not typically possible, unless the model is of a simple form (such

as a univariate autoregression). In what follows, we present a recursive algorithm for computing this expectation in a general setting. The algorithm provides us with an exact expression for (2.4) involving only θ. Evaluating this expression at $\hat{\theta}_n$ yields the expected information matrix $I_n(\hat{\theta}_n)$.

The development of our algorithm begins with representing each innovation $e_t(\theta)$ as an explicit linear combination of $y_t, y_{t-1}, \ldots, y_1$, and μ. With $y_o \equiv x_o^o = \mu$ and $K_o(\theta)$ defined as a $p \times p$ identity, we can write

$$e_t(\theta) = \sum_{k=0}^{t} F_k^t(\theta)\ y_k \quad (1 \le t \le n), \tag{3.1}$$

where the $F_k^t(\theta)$ are defined as follows:

$$\begin{aligned} F_t^t(\theta) &= I \quad (1 \le t \le n), \\ F_{t-1}^t(\theta) &= -A_t \Phi K_{t-1}(\theta) \quad (1 \le t \le n), \\ F_k^t(\theta) &= -A_t(\Phi(I - K_{t-1}(\theta)A_{t-1})) \cdots (\Phi(I - K_{k+1}(\theta)A_{k+1}))\Phi K_k(\theta) \\ &\quad (2 \le t \le n,\ 0 \le k \le t-2). \end{aligned}$$

(This result is easily verified by using the Kalman filter equations labeled (i), (ii), and (vi) to express $e_1(\theta)$ in terms of y_1 and $y_o \equiv x_o^o = \mu$; $e_2(\theta)$ in terms of y_2, y_1, and $y_o \equiv x_o^o = \mu$; $e_3(\theta)$ in terms of y_3, y_2, y_1, and $y_o \equiv x_o^o = \mu$; etc. See, for example, Watanabe, 1985.)

We can express the coefficients $F_k^t(\theta)$ using convenient recursions by writing

$$F_t^t(\theta) = R_t^t(\theta) \quad (1 \le t \le n), \tag{3.2}$$

$$F_k^t(\theta) = -A_t R_k^t(\theta) \quad (1 \le t \le n,\ 0 \le k \le t-1), \tag{3.3}$$

where

$$R_t^t(\theta) = I \quad (1 \le t \le n), \tag{3.4}$$

$$R_{t-1}^t(\theta) = \Phi K_{t-1}(\theta) \quad (1 \le t \le n), \tag{3.5}$$

$$\begin{aligned} R_k^t(\theta) &= \Phi(I - K_{t-1}(\theta)A_{t-1})R_k^{t-1}(\theta) \\ &\quad (2 \le t \le n,\ 0 \le k \le t-2). \end{aligned} \tag{3.6}$$

Now with representation (3.1) for $e_t(\theta)$, we can write

$$\frac{\partial e_t(\theta)}{\partial \theta_i} = \sum_{k=0}^{t} \left(\frac{\partial F_k^t(\theta)}{\partial \theta_i} \right)\ y_k \quad \text{for all } i \quad (1 \le t \le n), \tag{3.7}$$

where

$$\frac{\partial F_t^t(\theta)}{\partial \theta_i} = \frac{\partial R_t^t(\theta)}{\partial \theta_i} \quad \text{for all } i \quad (1 \le t \le n), \tag{3.8}$$

$$\frac{\partial F_k^t(\theta)}{\partial \theta_i} = -A_t \, \frac{\partial R_k^t(\theta)}{\partial \theta_i} \quad \text{for all } i \quad (1 \le t \le n,\ 0 \le k \le t-1), \tag{3.9}$$

and

$$\frac{\partial R_t^t(\theta)}{\partial \theta_i} = 0 \quad \text{for all } i \quad (1 \le t \le n), \tag{3.10}$$

$$\frac{\partial R_{t-1}^t(\theta)}{\partial \theta_i} = \frac{\partial \Phi}{\partial \theta_i} \, K_{t-1}(\theta) + \Phi \, \frac{\partial K_{t-1}(\theta)}{\partial \theta_i} \quad \text{for all } i \quad (1 \le t \le n), \tag{3.11}$$

$$\begin{aligned} \frac{\partial R_k^t(\theta)}{\partial \theta_i} = {} & \frac{\partial \Phi}{\partial \theta_i} \, (I - K_{t-1}(\theta) A_{t-1}) \, R_k^{t-1}(\theta) \\ & - \Phi \left(\frac{\partial K_{t-1}(\theta)}{\partial \theta_i} \, A_{t-1} \right) \, R_k^{t-1}(\theta) \\ & + \Phi \, (I - K_{t-1}(\theta) A_{t-1}) \, \frac{\partial R_k^{t-1}(\theta)}{\partial \theta_i} \quad \text{for all } i \\ & (2 \le t \le n,\ 0 \le k \le t-2). \end{aligned} \tag{3.12}$$

Note that in the preceding,

$$\frac{\partial K_o(\theta)}{\partial \theta_i} = 0 \quad \text{for all } i,$$

since $K_o(\theta)$ is an identity. For $1 \le t \le n$, the partials

$$\frac{\partial K_t(\theta)}{\partial \theta_i}$$

can be evaluated for all i using the recursions obtained by differentiating the Kalman filter equations with respect to θ_i (Harvey, 1989, page 143).

Representation (3.7), together with (3.8), (3.9), and (3.10), yields

$$\frac{\partial e_t(\theta)}{\partial \theta_i} = \sum_{k=0}^{t} -A_t \left(\frac{\partial R_k^t(\theta)}{\partial \theta_i} \right) \, y_k \quad \text{for all } i \quad (1 \le t \le n).$$

This expression allows us to write (2.4) as follows:

$$\begin{aligned} I_n(\theta; i, j) = {} & \frac{1}{2} \sum_{t=1}^{n} tr \left(\Sigma_t^{-1}(\theta) \frac{\partial \Sigma_t(\theta)}{\partial \theta_i} \Sigma_t^{-1}(\theta) \frac{\partial \Sigma_t(\theta)}{\partial \theta_j} \right) \\ & + \sum_{t=1}^{n} tr \left(\Sigma_t^{-1}(\theta) \left(\sum_{k=0}^{t} \sum_{l=0}^{t} A_t \, \left(\frac{\partial R_k^t(\theta)}{\partial \theta_i} \right) \, E\{ y_k y_l^{'} \} \, \left(\frac{\partial R_l^t(\theta)}{\partial \theta_j} \right)^{'} \, A_t^{'} \right) \right) \\ & (1 \le i, j \le d). \end{aligned} \tag{3.13}$$

Note that the only term in this representation for $I_n(\theta; i, j)$ which is not explicitly written in terms of θ is $E\{y_k y_l^{'}\}$ $(0 \leq k, l \leq n)$. Using the state-space equations (2.1) and (2.2) along with the convention $y_o \equiv \mu$, we can show that this expectation may be expressed as follows:

$$
\begin{aligned}
E\{y_k y_l^{'}\} &= A_k\{\sum_{s=0}^{k-1}\Phi^s Q(\Phi^{'})^{s+(l-k)}\}A_l^{'} + A_k\Phi^k\left(\Sigma+\mu\mu^{'}\right)(\Phi^{'})^l A_l^{'} \quad \text{when } 1\le k<l\le n,\\
&= A_k\{\sum_{s=0}^{l-1}\Phi^{s+(k-l)} Q(\Phi^{'})^s\}A_l^{'} + A_k\Phi^k\left(\Sigma+\mu\mu^{'}\right)(\Phi^{'})^l A_l^{'} \quad \text{when } 1\le l<k\le n,\\
&= A_k\{\sum_{s=0}^{k-1}\Phi^s Q(\Phi^{'})^s\}A_k^{'} + A_k\Phi^k\left(\Sigma+\mu\mu^{'}\right)(\Phi^{'})^k A_k^{'} + R \quad \text{when } 1\le k=l\le n,\\
E\{y_o y_l^{'}\} &= \mu\mu^{'}(\Phi^{'})^l A_l^{'} \quad \text{for } 1\le l\le n,\\
E\{y_k y_o^{'}\} &= A_k\Phi^k\mu\mu^{'} \quad \text{for } 1\le k\le n,\\
E\{y_o y_o^{'}\} &= \mu\mu^{'}. \qquad (3.14)
\end{aligned}
$$

Together, (3.13) and (3.14) provide us with an expression for $I_n(\theta;i,j)$ written entirely in terms of θ. Evaluating $I_n(\theta;i,j)$ at $\theta=\hat{\theta}_n$, therefore, will give us the elements of $I_n(\hat{\theta}_n)$.

The steps in the algorithm can be outlined as follows:

(a) Evaluate $E\{y_k y_l^{'}\}$ at $\theta=\hat{\theta}_n$ for $0\le k,l\le n$ using (3.14).

(b) Proceed through the Kalman filter equations for $t=1,\ldots,n$, holding θ fixed at $\hat{\theta}_n$. Expand the set of filter equations to evaluate the following at each t.

(i) Using the recursions in Harvey (1989, page 143), obtain

$$\frac{\partial\Sigma_t(\theta)}{\partial\theta_i}|_{\theta=\hat{\theta}_n} \quad \text{and} \quad \frac{\partial K_t(\theta)}{\partial\theta_i}|_{\theta=\hat{\theta}_n} \qquad (1\le i\le d).$$

(ii) Using (3.5), the recursions (3.6), (3.10), (3.11), and the recursions (3.12), obtain

$$R_{t-1}^t(\hat{\theta}_n),\ R_{t-2}^t(\hat{\theta}_n),\ \ldots,\ R_o^t(\hat{\theta}_n);$$

$$\frac{\partial R_t^t(\theta)}{\partial\theta_i}|_{\theta=\hat{\theta}_n}=0,\ \frac{\partial R_{t-1}^t(\theta)}{\partial\theta_i}|_{\theta=\hat{\theta}_n},\ \ldots,\ \frac{\partial R_o^t(\theta)}{\partial\theta_i}|_{\theta=\hat{\theta}_n} \qquad (1\le i\le d).$$

(iii) Obtain the contribution to the expected Fisher information matrix at time t:

$$
\begin{aligned}
I_n^t(\hat{\theta}_n;i,j) &= \frac{1}{2}\ tr\left(\Sigma_t^{-1}(\theta)\frac{\partial\Sigma_t(\theta)}{\partial\theta_i}\Sigma_t^{-1}(\theta)\frac{\partial\Sigma_t(\theta)}{\partial\theta_j}\right)|_{\theta=\hat{\theta}_n}\\
&\quad + tr\left(\Sigma_t^{-1}(\theta)\left(\sum_{k=0}^{t}\sum_{l=0}^{t}A_t\left(\frac{\partial R_k^t(\theta)}{\partial\theta_i}\right)E\{y_k y_l^{'}\}\left(\frac{\partial R_l^t(\theta)}{\partial\theta_j}\right)^{'}A_t^{'}\right)\right)|_{\theta=\hat{\theta}_n}\\
&\quad (1\le i,j\le d).
\end{aligned}
$$

(c) At the termination of the Kalman filter equations, obtain the elements of the expected Fisher information matrix $I_n(\hat{\theta}_n)$ by evaluating

$$I_n(\hat{\theta}_n; i, j) = \sum_{t=1}^{n} I_n^t(\hat{\theta}_n; i, j) \quad (1 \leq i, j \leq d). \tag{3.15}$$

It should be noted that in many applications, the number of computations required to evaluate $I_n(\hat{\theta}_n)$ can be considerably reduced by monitoring the powers of $\Phi(\hat{\theta}_n)$ in computing (3.14) (at $\theta = \hat{\theta}_n$), and by monitoring the values of $I_n^t(\hat{\theta}_n; i, j)$. Often, the terms in (3.14) involving "large" powers of $\Phi(\hat{\theta}_n)$ will be negligible, and the $I_n^t(\hat{\theta}_n; i, j)$ will converge to *steady-state* values as t increases (for each i, j). This should be the case when the eigenvalues of $\Phi(\hat{\theta}_n)$ are within the unit circle, meaning that the state process represented by the fitted model is *stable*. In such settings, it is efficient to ignore terms in (3.14) involving powers of $\Phi(\hat{\theta}_n)$ large enough to produce approximately the zero matrix, and to discontinue evaluating $I_n^t(\hat{\theta}_n; i, j)$ once t is large enough so that these terms are approximately constant. If the $I_n^t(\hat{\theta}_n; i, j)$ appear nearly constant by time period t_o and no further terms are computed, (3.15) can be replaced with

$$I_n(\hat{\theta}_n; i, j) = \sum_{t=1}^{t_o} I_n^t(\hat{\theta}_n; i, j) + (n - t_o)\ I_n^{t_o}(\hat{\theta}_n; i, j) \quad (1 \leq i, j \leq d).$$

4. Simulation Results

In this section, we present simulation results where known Fisher information matrices corresponding to some simple state-space models are estimated using both $I_n(\hat{\theta}_n)$ and $\mathcal{I}_n(\hat{\theta}_n, Y_n)$. The accuracy of the two methods is compared.

The type of model used in our simulations is the p^{th}-order univariate autoregression with observation noise. This model can be written as

$$y_t = z_t + v_t, \quad v_t \ \sim \ iid\ N(0, \sigma_R^2),$$

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + \epsilon_t, \quad \epsilon_t \ \sim \ iid\ N(0, \sigma_Q^2).$$

In state-space form, the model is expressed by writing the observation equation (2.1) as

$$y_t = (1, 0, \ldots, 0) \begin{pmatrix} z_t \\ z_{t-1} \\ \vdots \\ z_{t-p+1} \end{pmatrix} + v_t,$$

and the state equation (2.2) as

$$\begin{pmatrix} z_t \\ z_{t-1} \\ \vdots \\ z_{t-p+1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ z_{t-2} \\ \vdots \\ z_{t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Here, the covariance matrix Q of the state noise vector is a $p \times p$ matrix with all zero entries except for the entry in the upper left-hand corner, which is σ_Q^2. The observation noise is scalar, and has variance $R = \sigma_R^2$.

The parameter vector for the model is the $(p+2)$ x 1 vector

$$\theta = (\phi_1, \phi_2, \dots, \phi_p, \sigma_R^2, \sigma_Q^2)'.$$

The parameter estimates $\hat{\theta}_n$ are obtained using the EM algorithm (Shumway and Stoffer, 1982). In the algorithm, Σ is fixed at the identity, and μ is set equal to the Kalman smoothed estimate $x_o^n(\theta) = E(x_t \mid Y_n)$ evaluated at the current value of $\hat{\theta}_n$. (See Shumway, 1988, page 180.)

In each simulation set, 200 realizations of size n are generated from a model parameterized by a selected θ_o. For each realization, $I_n(\hat{\theta}_n)$ and $\mathcal{I}_n(\hat{\theta}_n, Y_n)$ are computed. The accuracy of $I_n(\hat{\theta}_n)$ and $\mathcal{I}_n(\hat{\theta}_n, Y_n)$ as estimators of $I_n(\theta_o)$ is ascertained by computing the mean square error (MSE) of the estimates of the ordered eigenvalues of $\frac{1}{n} I_n(\theta_o)$ provided by the ordered eigenvalues of $\frac{1}{n} I_n(\hat{\theta}_n)$ and $\frac{1}{n}\mathcal{I}_n(\hat{\theta}_n, Y_n)$. The MSE's are averaged over the 200 realizations, and the average MSE's are reported in Table 1. The ratio of the average MSE for $\frac{1}{n} I_n(\hat{\theta}_n)$ over the average MSE for $\frac{1}{n}\mathcal{I}_n(\hat{\theta}_n, Y_n)$ is also reported. The elements of $I_n(\theta_o)$ are determined by evaluating each of (3.13) and (3.14) at θ_o.

Results based on four different stable models (of orders $p = 1$ and $p = 2$) and three different sample sizes ($n = 50$, $n = 100$, and $n = 150$) are featured in Table 1. (The models were chosen so that the sets of eigenvalues of the four $\Phi(\theta_o)$ matrices have various magnitudes yet are all within the unit circle.) The results illustrate two phenomena of interest. First of all, the average MSE for $\frac{1}{n} I_n(\hat{\theta}_n)$ is consistently less than the average MSE for $\frac{1}{n}\mathcal{I}_n(\hat{\theta}_n, Y_n)$, which is not surprising since $\frac{1}{n} I_n(\hat{\theta}_n)$ serves as a maximum likelihood estimator of $\frac{1}{n} I_n(\theta_o)$ and $\frac{1}{n}\mathcal{I}_n(\hat{\theta}_n, Y_n)$ does not. Secondly, the ratio of the average MSE for $\frac{1}{n} I_n(\hat{\theta}_n)$ over the average MSE for $\frac{1}{n}\mathcal{I}_n(\hat{\theta}_n, Y_n)$ becomes close to one as the sample size increases, which should be expected since $\frac{1}{n} I_n(\hat{\theta}_n)$ and $\frac{1}{n}\mathcal{I}_n(\hat{\theta}_n, Y_n)$ are

asymptotically equivalent. Nevertheless, the results seem to indicate that in smaller samples, $I_n(\hat{\theta}_n)$ serves as a considerably more accurate estimator of $I_n(\theta_o)$ than $\mathcal{I}_n(\hat{\theta}_n, Y_n)$.

5. An Application

Shumway (1988, pages 180-181) considers a bivariate series consisting of 64 measurements on the temperature and the salt level of surface soil at equally spaced locations on a rectangular field. If the temperature measurements are modeled as a univariate series, a first-order autoregression with observation noise seems to provide an adequate fit to the data. For this model, the maximum likelihood estimates of the parameters are $\hat{\phi}_1 = 0.6779$, $\hat{\sigma}_R^2 = 0.1309$, and $\hat{\sigma}_Q^2 = 0.0881$.

Using the inverse of $I_n(\hat{\theta}_n)$ to obtain the estimated standard errors of the maximum likelihood estimates yields

$$\hat{SE}(\hat{\phi}_1) = 0.2075, \quad \hat{SE}(\hat{\sigma}_R^2) = 0.0677, \quad \hat{SE}(\hat{\sigma}_Q^2) = 0.0779.$$

Using the inverse of $\mathcal{I}_n(\hat{\theta}_n, Y_n)$ produces

$$\hat{SE}(\hat{\phi}_1) = 0.1985, \quad \hat{SE}(\hat{\sigma}_R^2) = 0.0671, \quad \hat{SE}(\hat{\sigma}_Q^2) = 0.0765.$$

In this application, note that the estimated standard errors are quite similar, and that the estimated standard errors based on $\mathcal{I}_n(\hat{\theta}_n, Y_n)$ are less that those based on $I_n(\hat{\theta}_n)$. In comparable sample-size applications, the latter is a common occurrence although the former is not. It is generally the case that the diagonal elements of $\mathcal{I}_n(\hat{\theta}_n, Y_n)$ exceed those of $I_n(\hat{\theta}_n)$, and that the diagonal elements of the inverse of $I_n(\hat{\theta}_n)$ exceed those of the inverse of $\mathcal{I}_n(\hat{\theta}_n, Y_n)$. Thus, estimated standard errors based on $\mathcal{I}_n(\hat{\theta}_n, Y_n)$ are generally less than those based on $I_n(\hat{\theta}_n)$, and may therefore have a tendency to be overly optimistic.

6. Conclusion

We have developed and summarized a general, recursive algorithm for computing the expected Fisher information matrix for state-space model parameters. We have also presented simulation results which indicate that in smaller sample settings, the expected information may more accurately estimate the true information matrix than the observed information. Our algorithm can be conveniently incorporated into any state-space model fitting routine based on maximum likelihood, making the expected information readily available for investigation and inference.

Table 1. Simulation results comparing $\frac{1}{n}I_n(\hat{\theta}_n)$ and $\frac{1}{n}\mathcal{I}_n(\hat{\theta}_n, Y_n)$ as estimators of $\frac{1}{n}I_n(\theta_o)$.

	n	p	θ_o	Ave. MSE for $\frac{1}{n}I_n(\hat{\theta}_n)$	Ave. MSE for $\frac{1}{n}\mathcal{I}_n(\hat{\theta}_n, Y_n)$	(Ave. MSE for $\frac{1}{n}I_n(\hat{\theta}_n)$) $\div$ (Ave. MSE for $\frac{1}{n}\mathcal{I}_n(\hat{\theta}_n, Y_n)$)
1	50	1	(0.90,0.50,1.00)	3.449	8.554	0.403
2	100	1	(0.90,0.50,1.00)	2.358	3.136	0.752
3	150	1	(0.90,0.50,1.00)	1.971	2.115	0.932
4	50	1	(-0.80,0.25,1.00)	0.786	1.284	0.612
5	100	1	(-0.80,0.25,1.00)	0.399	0.507	0.788
6	150	1	(-0.80,0.25,1.00)	0.262	0.287	0.914
7	50	2	(0.99,-0.80,0.50,1.00)	1.660	5.184	0.320
8	100	2	(0.99,-0.80,0.50,1.00)	1.092	1.973	0.553
9	150	2	(0.99,-0.80,0.50,1.00)	0.688	0.855	0.804
10	50	2	(1.40,-0.49,0.25,1.00)	53.179	95.961	0.554
11	100	2	(1.40,-0.49,0.25,1.00)	41.419	52.158	0.794
12	150	2	(1.40,-0.49,0.25,1.00)	22.594	23.762	0.951

Acknowledgements

The authors wish to express their gratitude to an anonymous referee for several insightful suggestions which helped to improve the original version of this manuscript.

Bibliography

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods (Second Edition)* (Springer-Verlag, New York.)

Cavanaugh, J. E. (1993), Small-sample model selection in the general state-space setting, Ph.D. dissertation, University of California, Davis, Division of Statistics. (Printed by UMI Dissertation Services, Ann Arbor, MI.)

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

Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *Trans. ASME Journal of Basic Engineering* **82**, 35–45.

Kalman, R. E. and Bucy, R. S. (1961), New results in linear filtering and prediction theory, *Trans. ASME Journal of Basic Engineering* **83**, 95–108.

Schweppe, F. (1965), Evaluation of likelihood functions for Gaussian signals, *IEEE Transactions on Information Theory* **11**, 61–70.

Shumway, R. H. (1988), *Applied Statistical Time Series Analysis* (Prentice-Hall, New Jersey).

Shumway, R. H. and Stoffer, D. S. (1982), An approach to time series smoothing and forecasting using the EM algorithm, *Journal of Time Series Analysis* **3**, 253–264.

Watanabe, N. (1985), Note on the Kalman filter with estimated parameters, *Journal of Time Series Analysis* **6**, 269–278.