2002

A state space approach to bootstrapping conditional forecasts in ARMA models

Wall, Kent D.

Blackwell Publishers

Journal of Time Series Analysis, v. 23, no. 6, 2002, pp. 733-751
http://hdl.handle.net/10945/47794

A STATE SPACE APPROACH TO BOOTSTRAPPING CONDITIONAL FORECASTS IN ARMA MODELS

By Kent D. Wall and David S. Stoffer

Naval Postgraduate School, University of Pittsburgh

First Version received May 2000

Abstract. A bootstrap approach to evaluating conditional forecast errors in ARMA models is presented. The key to this method is the derivation of a reverse-time state space model for generating conditional data sets that capture the salient stochastic properties of the observed data series. We demonstrate the utility of the method using several simulation experiments for the MA(q) and ARMA(p, q) models. Using the state space form, we are able to investigate conditional forecast errors in these models quite easily whereas the existing literature has only addressed conditional forecast error assessment in the pure AR(p) form. Our experiments use short data sets and non-Gaussian, as well as Gaussian, disturbances. The bootstrap is found to provide useful information on error distributions in all cases and serves as a broadly applicable alternative to the asymptotic Gaussian theory.

Keywords. Bootstrap; state space; forecasting; prediction errors; simulation.

1. INTRODUCTION

This paper is concerned with assessing the conditional forecast accuracy of ARMA models using a state approach and the Monte Carlo bootstrap. Our work is motivated by four considerations. First, the state space model provides a convenient unifying representation for the AR(p), MA(q) and ARMA(p, q) models. Second, the actual practice of forecasting involves the prediction of a future point on an observed sample path, thus conditional forecast error assessment is of most interest. Third, real-life applications involving time series data are often characterized by short data sets and lack of distributional information. Asymptotic theory provides little help here and often there are no compelling reasons to assume Gaussian distributions apply. Finally, the utility and applicability already demonstrated by the bootstrap for prediction of AR processes suggests that it has much to offer in the prediction of ARMA processes.

Early application of the bootstrap to assess conditional forecast errors is found in Findley (1986), Stine (1987), Thombs and Schuchany (1990), Kabaila (1993) and McCullough (1994, 1996). Interest in the evaluation of confidence intervals for conditional forecast errors has led to methodological problems because a backward, or reverse-time, set of residuals must be generated. Findley (1986) first

0143-9782/02/06 733–751 JOURNAL OF TIME SERIES ANALYSIS Vol. 23, No. 6

discussed this problem and Breidt *et al.* (1992) offer a solution that is implemented in the work of McCullough (1994, 1996). To date, there is a well grounded methodology for AR models and this work has established the utility of the bootstrap.

A similar state of affairs appears not to exist with moving-average (MA) or ARMA models, although these models are just as important and useful as their AR relatives. We suspect that this is due to the difficulty with which one can identify mechanisms required to generate bootstrap data sets, whether forwards or backwards in time. For AR models, this is easily accomplished because the required initial, or terminal (in the case of conditional forecasts), conditions are given in terms of the observed series. With MA or ARMA models this is not the case because the models require solutions of difference equations involving unobserved disturbances.

The state space model and its related innovations filter offer a way around this difficulty. It is worthwhile, therefore, to investigate how well this can be done in practice. Stoffer and Wall (1991) found such a combination to be of use in assessing parameter estimation error, and this naturally leads to the same question being asked in relation to conditional prediction errors. We find that the bootstrap is as useful in evaluating conditional forecast errors as it has proven to be in assessing parameter estimation errors, particularly in a non-Gaussian environment.

Our paper is organized into six sections. Section 2 defines the state space representation used for ARMA models, sets the notation for what follows, and outlines the parameter estimation problem that is an integral part of the method we present. Section 3 outlines our solution to the problem of obtaining backward, or reverse-time, representations of state space models so that conditional bootstrap data can be generated. Section 4 reviews the bootstrap procedure for conditional forecast evaluation and Section 5 is devoted to simulation examples. Section 6 contains a summary and conclusions. Details of the state space model derivation are given in an appendix.

2. ARMA MODELS IN STATE SPACE FORM

The state space model used in this paper is defined by the equations

$$\mathbf{s}(t+1) = \mathbf{F}\mathbf{s}(t) + \mathbf{G}\mathbf{x}(t) + \mathbf{w}(t) \tag{1}$$

and

$$\mathbf{y}(t) = \mathbf{H}\mathbf{s}(t) + \mathbf{D}\mathbf{x}(t) + \mathbf{v}(t) \tag{2}$$

where $\mathbf{s}(t)$ is an $n \times 1$ vector of unobserved state variables, $\mathbf{y}(t)$ is a $m \times 1$ vector of observed outputs or endogenous variables, and $\mathbf{x}(t)$ is an $r \times 1$ vector of observed inputs or exogenous variables. The constant matrices $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$ and $\mathbf{D}$ represent the model coefficients and have dimensions compatible with the matrix operations required in (1) and (2). The two terms $\mathbf{w}(t)$ and $\mathbf{v}(t)$ represent zero-mean random processes that are each independent and identically distributed (i.i.d.) with

$$E\{\mathbf{w}(t)\mathbf{w}(t)'\} = \mathbf{Q} \qquad E\{\mathbf{v}(t)\mathbf{v}(t)'\} = \mathbf{R} \qquad E\{\mathbf{w}(t)\mathbf{v}(t)'\} = \mathbf{S} \tag{3}$$

where $\mathbf{Q}$ is an $n \times n$ nonnegative definite matrix and $\mathbf{R}$ is a $m \times m$ nonnegative definite matrix.

The ARMA(p, q) process that we represent in state space form is defined by

$$y(t) + a_1 y(t-1) + \cdots + a_p y(t-p) = \varepsilon(t) + b_1 \varepsilon(t-1) + \cdots + b_q \varepsilon(t-q)$$

where $\varepsilon(t)$ is an i.i.d. process with finite variance σ_ε^2. Its realization in state space form is achieved by defining $n = \max\{p, q\}, m = 1$ and employing an observable canonical form in the definition of the state space coefficient matrices:

$$\mathbf{F} = \begin{bmatrix} 0 & \dots & 0 & 0 & 0 & -a_n \\ 1 & \dots & 0 & 0 & 0 & -a_{n-1} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & 0 & 0 & -a_3 \\ 0 & \dots & 0 & 1 & 0 & -a_2 \\ 0 & \dots & 0 & 0 & 1 & -a_1 \end{bmatrix}$$

and

$$\mathbf{H} = [\,0 \quad \dots \quad 0 \quad 0 \quad 0 \quad 1\,]$$

$\mathbf{G} = 0$ and $\mathbf{D} = 0$. If $n > p$ then $a_\ell = 0$ for $\ell > p$, and if $n > q$ then $b_\ell = 0$ for $\ell > q$. The random processes are defined by $\mathbf{v}(t) = \varepsilon(t)$ and $\mathbf{w}(t) = \mathbf{g_0}\varepsilon(t)$ where

$$\mathbf{g_0} = [\,b_n - a_n \quad b_{n-1} - a_{n-1} \quad \dots \quad b_3 - a_3 \quad b_2 - a_2 \quad b_1 - a_1\,]'$$

The variance-covariance matrices are given by

$$\mathbf{Q} = \sigma_\varepsilon^2 \mathbf{g_0}\mathbf{g_0}' \qquad \mathbf{R} = [\sigma_\varepsilon^2] \qquad \mathbf{S} = \sigma_\varepsilon^2 \mathbf{g_0}$$

Note that $\mathbf{w}(t)$ and $\mathbf{v}(t)$ are always correlated in state representations of ARMA models.

The model coefficients and the correlation structure are assumed to be uniquely parameterized by a $k \times 1$ vector $\boldsymbol{\theta}$; that is, $\mathbf{F} = \mathbf{F}(\boldsymbol{\theta}), \mathbf{G} = \mathbf{G}(\boldsymbol{\theta}), \mathbf{H} = \mathbf{H}(\boldsymbol{\theta})$, etc. The vector $\boldsymbol{\theta}$ is assumed to be an element of some compact space, $\mathcal{P}$, usually a subset of $\mathbb{R}^k$. The use of the observable canonical form ensures the model is completely identified – see, for example Wall (1987) – and the parameterization is unique once we impose the usual invertability and stability conditions on the autoregressive and moving-average operators.

Let $\mathbf{s}(t+1|t)$ denote the best linear predictor of $\mathbf{s}(t+1)$ based on the data $\mathcal{Y}^t = \{\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(t)\}$ and $\mathcal{X}^t = \{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(t)\}$, obtained via the Kalman filter. Also obtained from the Kalman filter are the innovations, the innovations covariance matrix, and the Kalman filter gain matrix,

$$\epsilon(t) = \mathbf{y}(t) - \mathbf{Hs}(t|t-1) - \mathbf{Dx}(t) \tag{4}$$

$$\Sigma(t) = \mathbf{HP}(t|t-1)\mathbf{H}' + \mathbf{R} \tag{5}$$

$$\mathbf{K}(t) = [\mathbf{FP}(t|t-1)\mathbf{H}' + \mathbf{S}]\Sigma(t)^{-1} \tag{6}$$

respectively, where $\mathbf{P}(t|t-1)$ is the covariance matrix of the state estimation error, $\mathbf{s}(t) - \mathbf{s}(t|t-1)$. This matrix is generated recursively according to

$$\mathbf{P}(t+1|t) = [\mathbf{F} - \mathbf{K}(t)\mathbf{H}]\mathbf{P}(t|t-1)[\mathbf{F} - \mathbf{K}(t)\mathbf{H}]' + \mathbf{Q} + \mathbf{K}(t)\mathbf{RK}(t)' - \mathbf{SK}(\mathbf{t})' - \mathbf{K}(\mathbf{t})\mathbf{S}'$$

The model innovations give rise to the innovations form representation of the observations:

$$\mathbf{s}(t+1|t) = \mathbf{Fs}(t|t-1) + \mathbf{Gx}(t) + \mathbf{K}(t)\epsilon(t) \tag{7}$$

$$\mathbf{y}(t) = \mathbf{Hs}(t|t-1) + \mathbf{Dx}(t) + \epsilon(t) \tag{8}$$

Parameter estimation is accomplished via Gaussian maximum likelihood (GML). The essential part of the logarithm of the Gaussian likelihood function is

$$L(\boldsymbol{\theta}|\mathcal{Y}^T, \mathcal{X}^T) = -\sum_{t=1}^{T}\{\ln(\det \Sigma(t, \boldsymbol{\theta})) + \epsilon(t, \boldsymbol{\theta})'\Sigma(t, \boldsymbol{\theta})^{-1}\epsilon(t, \boldsymbol{\theta})\} \tag{9}$$

Parameter estimation is achieved by minimizing this function with respect to $\boldsymbol{\theta}$. We employ a BFGS variable metric algorithm to accomplish this, proceeding iteratively from some initial guess, $\boldsymbol{\theta}^0$, to convergence. The value of $\boldsymbol{\theta}$ at convergence constitutes the GML estimate and is denoted by $\hat{\boldsymbol{\theta}}$. The numerical approximation to the inverse Hessian of $L(\boldsymbol{\theta}|\mathcal{Y}^T, \mathcal{X}^T)$ at convergence (an automatic by-product of the BFGS algorithm) gives the variance-covariance matrix of the parameter estimates.

3. GENERATING REVERSE-TIME DATASETS

The generation of bootstrap data sets in forward time is easy. Given an initial condition or prior, $\mathbf{s}(t_0|t_0-1)$ and $\mathbf{P}(t_0|t_0-1)$, and a bootstrap sample, $\epsilon(t)^*$ with $\epsilon(t_0)^* = \epsilon(t_0)$, the innovations form, (7) and (8), is solved recursively from $t = 1$ through $t = T$ to produce realizations passing through the given initial observation. Such computations are all that is required in obtaining bootstrap estimates of parameter estimation error statistics or unconditional forecast error statistics (Stoffer and Wall, 1991). The generation of bootstrap data sets for assessing *conditional* forecast errors is not so straightforward because they must be generated backward and this requires a *backward-time state space model*.

An early discussion of the problems related to backward time models in assessing conditional forecast errors is found in Findley (1986). Further consideration of the problem is found in Breidt *et al.* (1992). This literature stresses the need to properly construct a set of 'backward' residuals and Breidt *et al.* (1992) provide an algorithm for this that solves the problem for AR(p) models. A similar result is needed for state space models, but development of backward-time representations has not received much attention in the literature. Notable exceptions are the elegant presentation found in Caines (1988, ch. 4) and a derivation in Aoki (1989, ch. 5). Our work requires an extension of their results to the time-varying case.

The key system in generating bootstrap data sets is the innovations, filter form, (7) and (8), rewritten here using the standardized residuals, $\mathbf{e}(t) = \Sigma(t)^{-\frac{1}{2}}\boldsymbol{\epsilon}(t)$, and the stacked vector $\boldsymbol{\xi}(t) = [\mathbf{s}'(t+1|t) : \mathbf{y}'(t)]'$:

$$\boldsymbol{\xi}(t) = \begin{bmatrix} \mathbf{F} & \mathbf{0} \\ \mathbf{H} & \mathbf{0} \end{bmatrix} \boldsymbol{\xi}(t-1) + \begin{bmatrix} \mathbf{G} \\ \mathbf{D} \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} \mathbf{G}_1(t) \\ \mathbf{D}_1(t) \end{bmatrix} \mathbf{e}(t) \tag{10}$$

where $\mathbf{G}_1(t) = \mathbf{K}(t)\Sigma(t)^{\frac{1}{2}}$ and $\mathbf{D}_1(t) = \Sigma(t)^{\frac{1}{2}}$. We require a backward-time representation of this system. All the problems highlighted by Findley (1986) and Breidt *et al.* (1992) appear here. For example, the first block row of (10) cannot be solved backwards in time by simply expressing $\mathbf{s}(t|t-1)$ in terms of $\mathbf{s}(t+1|t)$. First, $\mathbf{F}$ is not always invertible; e.g., MA(q) models. Second, even when $\mathbf{F}$ is invertible, $\mathbf{F}^{-1}$ has characteristic roots outside the unit circle whenever $\mathbf{F}$ has its characteristic roots inside the unit circle. This situation is intolerable in generating reverse time trajectories because of the explosive nature of the solutions for $\mathbf{s}(t|t-1)$. In addition, we now have a time-varying system; i.e., the last terms of each of the equations immediately above depend on t.

These difficulties are overcome by building on the method found in Caines (1988, pp. 236–7). Special attention must be given to the way in which the time-varying matrices propagate through the derivations and proper account must be taken of the effects of the known, or observed input sequence $\mathbf{x}(t)$. First, assume $\mathbf{x}(t) \equiv \mathbf{0}$ (the case when $\mathbf{x}(t) \neq \mathbf{0}$ is addressed below). Application of the symmetry of minimal splitting subspaces yields the following reverse-time state space representation for $t = T-1, T-2, \ldots, 1$:

$$\begin{aligned} \mathbf{r}(t) &= \mathbf{F}'\mathbf{r}(t+1) + \mathbf{A}(t)\mathbf{s}(t|t-1) - \mathbf{B}(t)\mathbf{e}(t) \\ \mathbf{y}(t) &= \mathbf{N}(t)\mathbf{r}(t+1) - \mathbf{L}(t)\mathbf{s}(t|t-1) + \mathbf{M}(t)\mathbf{e}(t) \end{aligned}$$

where

$$\begin{aligned} \mathbf{A}(t) &= \mathbf{V}(t)^{-1} - \mathbf{F}'\mathbf{V}(t+1)^{-1}\mathbf{F} \\ \mathbf{B}(t) &= \mathbf{F}'\mathbf{V}(t+1)^{-1}\mathbf{G}_1(t) \\ \mathbf{C}(t) &= \mathbf{I} - \mathbf{G}_1(t)'\mathbf{V}(t+1)^{-1}\mathbf{G}_1(t) \\ \mathbf{L}(t) &= \mathbf{D}_1(t)\mathbf{B}'(t) - \mathbf{H}\mathbf{V}(t)\mathbf{A}(t) \\ \mathbf{M}(t) &= \mathbf{D}_1(t)\mathbf{C}(t) - \mathbf{H}\mathbf{V}(t)\mathbf{B}(t) \\ \mathbf{N}(t) &= \mathbf{H}\mathbf{V}(t)\mathbf{F}' + \mathbf{D}_1(t)\mathbf{G}_1'(t) \end{aligned}$$

and

$$\mathbf{V}(t+1) = \mathbf{F}\mathbf{V}(t)\mathbf{F}' + \mathbf{G}_1(t)\mathbf{G}_1'(t)$$

The reverse-time state vector is $\mathbf{r}(t)$. The backward recursion begins at $t = T-1$ using $\mathbf{r}(T) = \mathbf{V}(T)^{-1}\mathbf{s}(T|T-1)$. Details of the derivation are given in Appendix A.

The algorithm specified above is equivalent to that given in Breidt *et al.* (1992) when the model is restricted to the stationary AR(p) case. For example, consider the AR(1) model. Its state space representation is defined by $\mathbf{F} = \mathbf{F}' = -a_1, \mathbf{H} = 1, \mathbf{G} = \mathbf{D} = 0, \mathbf{Q} = a_1^2\sigma_\varepsilon^2, \mathbf{R} = \sigma_\varepsilon^2$ and $\mathbf{S} = -a_1\sigma_\varepsilon^2$. The solutions to (4)–(8) in the stationary case are obtained when $t_0 \to -\infty$ in all difference equations. In steady-state $\mathbf{P}(t|t-1) = \mathbf{0}$, the filter innovations are identical to the $\varepsilon(t)$ process, $\Sigma(t) = \sigma_\varepsilon^2$ and $\mathbf{K}(t) = -a_1$. The coefficient matrices in our reverse time state space model reduce to constants:

$$\begin{aligned}
\mathbf{V}(t) &= \mathbf{V} = a_1^2\left[1 - a_1^2\right]^{-1}\sigma_\varepsilon^{-2} \\
\mathbf{A}(t) &= \mathbf{A} = \left[1 - a_1^2\right]^2 a_1^{-2}\sigma_\varepsilon^{-2} \\
\mathbf{B}(t) &= \mathbf{B} = \left[1 - a_1^2\right]\sigma_\varepsilon^{-1} \\
\mathbf{C}(t) &= \mathbf{C} = a_1^2 \\
\mathbf{L}(t) &= \mathbf{L} = 0 \\
\mathbf{M}(t) &= \mathbf{M} = 0 \\
\mathbf{N}(t) &= \mathbf{N} = -a_1\left[1 - a_1^2\right]^{-1}\sigma_\varepsilon^2
\end{aligned}$$

The associated reverse time difference equation for $y(t)$ is obtained by manipulating the equations for $\mathbf{r}(t)$ and $\mathbf{y}(t)$ given above. First, advance the time index in the equation for $\mathbf{r}(t)$ and substitute the result into the equation for $\mathbf{y}(t)$. Second, substitute

$$\mathbf{r}(t+2) = \mathbf{N}^{-1}[y(t+2) + \mathbf{L}s(t+1)|t) - \mathbf{M}\mathbf{e}(t+1)]$$

from the equation defining $\mathbf{y}(t)$ when its time index has been advanced by one period. The result is the first-order backwards difference equation:

$$y(t) = -a_1 y(t+1) - \frac{1 - a_1^2}{a_1} s(t+1|t) + a_1\sigma_\varepsilon e(t+1)$$

In the stationary case

$$\sigma_\varepsilon e(t+1) = \varepsilon(t+1)$$

and

$$s(t+1|t) = -a_1 \sum_{j=-\infty}^{t} (-a_1)^{t-j} \varepsilon(j)$$

from solution of (7). Making these substitutions gives

$$\begin{aligned} y(t) &= -a_1 y(t+1) + a_1 \varepsilon(t+1) + \sum_{j=-\infty}^{t} (-a_1)^{t-j} \varepsilon(j) - \sum_{j=-\infty}^{t} (-a_1)^{t+2-j} \varepsilon(j) \\ &= -a_1 y(t+1) + a_1 \varepsilon(t+1) + \varepsilon(t) + \sum_{j=-\infty}^{t} (-a_1)^{t+1-j} [\varepsilon(j-1) + a_1 \varepsilon(j)] \end{aligned}$$

The method of Breidt *et al.* solves the backward difference equation

$$a(L^{-1}) y(t) = w(t)$$

where

$$w(t) = \frac{a(L^{-1})}{a(L)} \varepsilon(t)$$

For the AR(1) model, we have $a(L^{-1}) = 1 + a_1 L^{-1}$ so the backward difference equation of Breidt *et al.* is

$$y(t) = -a_1 y(t+1) + a_1 \varepsilon(t+1) + \varepsilon(t) - a_1 w(t-1)$$

The last term in this equation can be represented in terms of $\varepsilon(j)$ using the rational lag definition for $w(t)$:

$$\begin{aligned} w(t-1) &= a(L^{-1}) \sum_{j=0}^{\infty} (-a_1 L)^j \varepsilon(t-1) \\ &= \sum_{j=0}^{\infty} (-a_1 L)^j a(L^{-1}) \varepsilon(t-1) \\ &= \sum_{j=0}^{\infty} (-a_1)^j [\varepsilon(t-1-j) + a_1 \varepsilon(t-j)] \\ &= \sum_{j=-\infty}^{t} (-a_1)^{t-j} [\varepsilon(j-1) + a_1 \varepsilon(j)]. \end{aligned}$$

Thus

$$-a_1 w(t-1) = \sum_{j=-\infty}^{t} (-a_1)^{t+1-j} [\varepsilon(j-1) + a_1 \varepsilon(j)]$$

and the Breidt *et al.* backward difference equation is identical to that obtained with our algorithm.

Non-zero observed inputs, $\mathbf{x}(t)$, are incorporated easily by applying the above derivation to the model written in terms of variations, $\tilde{\mathbf{y}}(t) = \mathbf{y}(t) - \mathbf{y}_f(t)$ and $\tilde{\mathbf{s}}(t) = \mathbf{s}(t) - \mathbf{s}_f(t)$, taken about the 'forced' response. First, one computes the solution to

$$\begin{aligned} \mathbf{s}_f(t+1) &= \mathbf{F}\mathbf{s}_f(t) + \mathbf{G}\mathbf{x}(t) \\ \mathbf{y}_f(t) &= \mathbf{H}\mathbf{s}_f(t) + \mathbf{D}\mathbf{x}(t) \end{aligned}$$

with $\mathbf{s}_f(0) = \mathbf{0}$. Next, one sets $\mathbf{x}(t) \equiv \mathbf{0}$ and replaces $\mathbf{y}(t)$ and $\mathbf{s}(t|t-1)$ by $\tilde{\mathbf{y}}(t)$ and $\tilde{\mathbf{s}}(t)$. After applying the above derivation to the variational model, the complete backward trajectory is obtained by 'adding back in' the forced response, i.e., by using $\tilde{\mathbf{y}}(t) + \mathbf{y}_f(t)$ as the backward data set.

The above specifies a three-step procedure for the generation of backward time data sets (written here for $\mathbf{x}(t) \equiv \mathbf{0}$):

1. Generate $\mathbf{V}(t), \mathbf{A}(t), \mathbf{B}(t), \mathbf{C}(t), \mathbf{L}(t), \mathbf{M}(t)$ and $\mathbf{N}(t)$ *forwards* in time from $t = 1$ through $t = T$ given

$$\mathbf{V}(1) = E\{\mathbf{s}(1|0)\mathbf{s}(1|0)'\} \tag{11}$$

2. For given $\{\mathbf{e}^*(t); 1 \leqslant t \leqslant T-1\}$, set $\mathbf{s}^*(1) = \mathbf{0}$ and generate $\{\mathbf{s}^*(t); 1 \leqslant t \leqslant T\}$ *forwards* in time from $t = 1$ through $t = T$ via

$$\mathbf{s}^*(t+1) = \mathbf{F}\mathbf{s}^*(t) + \mathbf{G}_1(t)\mathbf{e}^*(t) \tag{12}$$

3. Set $\mathbf{r}(T) = \mathbf{V}(T)^{-1}\mathbf{s}(T|T-1)$ and generate $\{\mathbf{y}^*(t); 1 \leqslant t \leqslant T\}$ *backwards* in time from $t = T-1$ through $t = 1$ via the reverse time state space model

$$\mathbf{r}(t) = \mathbf{F}'\mathbf{r}(t+1) + \mathbf{A}(t)\mathbf{s}^*(t) - \mathbf{B}(t)\mathbf{e}^*(t) \tag{13}$$

$$\mathbf{y}^*(t) = \mathbf{N}(t)\mathbf{r}(t+1) - \mathbf{L}(t)\mathbf{s}^*(t) + \mathbf{M}(t)\mathbf{e}^*(t) \tag{14}$$

This procedure assumes one already has drawn randomly, with replacement, from the model estimated standardized residuals to obtain a set of $T-1$ residuals denoted $\{\mathbf{e}^*(t); 1 \leqslant t \leqslant T-1\}$. The last residual is kept set at $\mathbf{e}^*(T) = \mathbf{e}(T)$ in order to ensure the conditioning requirement is met on $\boldsymbol{\xi}^*(T)$, i.e., that $\boldsymbol{\xi}^*(T) = \boldsymbol{\xi}(T)$. This requirement follows from the autoregressive structure of (10). The creation of an arbitrary number of bootstrap data sets is accomplished by repeating the above for each set of bootstrap residuals $\{\mathbf{e}^*(t); 1 \leqslant t \leqslant T-1; \mathbf{e}^*(T) = \mathbf{e}(T)\}$. Figure 1 presents a sample of 100 reverse-time trajectories for the ARMA(2,1) model

$$y(t) = 1.4y(t-1) - 0.85y(t-2) + \varepsilon(t) + 0.6\varepsilon(t-1)$$

with $\sigma_\varepsilon = 0.2$ and $T = 49$. The original, observed sample is plotted with the bold line.

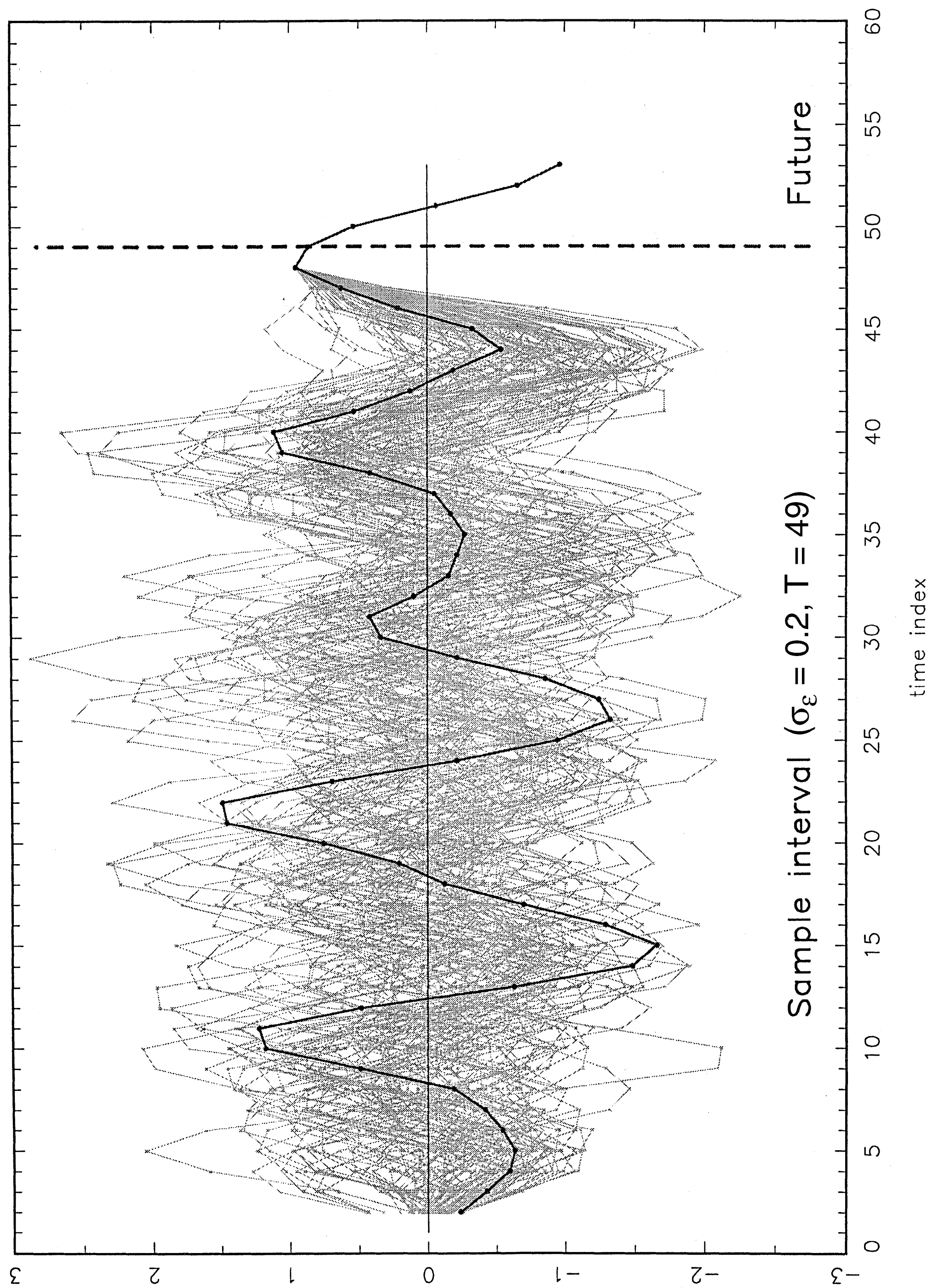

FIGURE 1. Reverse-time realizations for an ARMA(2,1) model.

4. COMPUTING FORECAST ERRORS VIA THE BOOTSTRAP

We first obtain $\hat{\boldsymbol{\theta}}$ via GML estimation using the original data. The associated residuals are denoted $\{\hat{\boldsymbol{\epsilon}}(t); 1 \leqslant t \leqslant T\}$ and their standardized values are denoted by $\{\hat{\mathbf{e}}(t); 1 \leqslant t \leqslant T\}$. For $b = 1, 2, 3, \ldots, B$ (where B is the number of bootstrap replications) we execute the following six steps:

1. Construct a sequence of $T + L$ standardized residuals $\{\mathbf{e}^b(t); 1 \leqslant t \leqslant T + L\}$ via random draws, with replacement, from the standardized residuals $\{\hat{\mathbf{e}}(t); 1 \leqslant t \leqslant T\}$. This sequence is formed as follows:
   (i) Use $T-1$ vectors to form $\{\mathbf{e}^b(t); 1 \leqslant t \leqslant T - 1\}$.
   (ii) Fix $\mathbf{e}^b(T) = \hat{\mathbf{e}}(T)$.
   (iii) Use the remaining L vectors to form $\{\mathbf{e}^b(t); T + 1 \leqslant t \leqslant T + L\}$.
2. Generate data $\{\mathbf{y}^b(t); 1 \leqslant t \leqslant T - 1\}$ via the backward state space model (13) and (14) with $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ using the residuals $\{\mathbf{e}^b(t); 1 \leqslant t \leqslant T - 1\}$. Set $\mathbf{y}^b(T) = \mathbf{y}(T)$.
3. Generate data $\{\mathbf{y}^b(t); T + 1 \leqslant t \leqslant T + L\}$ via the forward state space model (10) with $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\mathbf{s}^b(T|T-1) = \mathbf{s}(T|T-1)$ using the residuals $\{\mathbf{e}^b(t); T + 1 \leqslant t \leqslant T + L\}$.
4. Compute model parameter estimates $\boldsymbol{\theta}^b$ via GML using the data $\{\mathbf{y}^b(t); 1 \leqslant t \leqslant T\}$ and $\{\mathbf{x}(t); 1 \leqslant t \leqslant T\}$.
5. Compute the bootstrap conditional forecasts $\{\hat{\mathbf{y}}^b(T + \ell, \boldsymbol{\theta}^b); 1 \leqslant \ell \leqslant L\}$ via the forward time state space model (10) with $\boldsymbol{\theta} = \boldsymbol{\theta}^b$, $\mathbf{s}^b(T|T-1) = \mathbf{s}(T|T-1)$ and $\mathbf{e}^b(t) = \mathbf{0}$ for $t \geq T + 1$.
6. Compute the bootstrap conditional forecast errors via

$$\mathbf{d}^b(\ell) = \mathbf{y}^b(T + \ell, \hat{\boldsymbol{\theta}}) - \hat{\mathbf{y}}^b(T + \ell, \boldsymbol{\theta}^b) \qquad 1 \leqslant \ell \leqslant L$$

The extent to which the bootstrap captures the behaviour of the actual forecast errors derives from the extent to which these errors mimic the stochastic process

$$\mathbf{d}(\ell) = \mathbf{y}(T + \ell, \boldsymbol{\theta}) - \hat{\mathbf{y}}(T + \ell, \hat{\boldsymbol{\theta}}) \qquad 1 \leqslant \ell \leqslant L$$

5. SIMULATION EXPERIMENTS

We now present some evidence of the value of the bootstrap in conditional forecast error estimation via simulation experiments. These begin with a known model and complete information concerning the distributions for $\mathbf{w}(t)$ and $\mathbf{v}(t)$. Thus, we approximate the true conditional forecast error distribution in an informative way by direct sampling. This requires careful attention be given to the generation of backward sample paths because we seek to evaluate conditional forecast errors. Since an understanding of this approach is essential to the interpretation of our results, we give the details of how this is done before proceeding.

We first re-write the underlying model, (1)–(2), in a form similar to the innovations representation so that the backward time generation algorithm can be applied to it in a straightforward manner:

$$\mathbf{s}(t+1) = \mathbf{F}\mathbf{s}(t) + \mathbf{G}\mathbf{x}(t) + \mathbf{G}_2\mathbf{e}(t) \tag{15}$$

$$\mathbf{y}(t) = \mathbf{H}\mathbf{s}(t) + \mathbf{D}\mathbf{x}(t) + \mathbf{D}_2\mathbf{e}(t) \tag{16}$$

where

$$\mathbf{e}(t) = \mathbf{\Psi}^{-1/2}[\mathbf{w}(t)'\mathbf{v}(t)']'$$

$$\mathbf{G}_2 = \begin{bmatrix}\mathbf{I}_p & 0\end{bmatrix}\mathbf{\Psi}^{1/2}$$

$$\mathbf{D}_2 = \begin{bmatrix}0 & \mathbf{I}_q\end{bmatrix}\mathbf{\Psi}^{1/2}$$

and $\mathbf{\Psi}$ is the variance-covariance matrix of the joint process $[\mathbf{w}(t)'\mathbf{v}(t)']'$. Using this notation for the data generating process, the algorithm used to simulate 'true' conditional forecast errors is as follows:

Given a sample path $\{\mathbf{y}^0(t); 1 \leqslant t \leqslant T\}$ from which conditional forecasts (initialized at $\mathbf{y}^0(T)$) are projected, let m denote the simulation index. For $m = 1, 2, \ldots, M$ execute the following five steps.

1. Make $T + L$ draws from the known distributions for $\mathbf{w}(t)$ and $\mathbf{v}(t)$.
2. Generate a conditional sample path passing through $\mathbf{y}^0(T)$:
   (a) Use the forward time model, (15) and (16), with $\boldsymbol{\theta} = \boldsymbol{\theta}^0, \mathbf{y}^m(T) = \mathbf{y}^0(T)$ and $\mathbf{s}(T) = \mathbf{s}^0(T)$, to generate $\{\mathbf{y}^m(T+\ell); 1 \leqslant \ell \leqslant L\}$. This utilizes $\{\mathbf{w}^m(T+\ell)$ and $\mathbf{v}^m(T+\ell); 1 \leqslant \ell \leqslant L\}$ to produce the future data to be forecasted.
   (b) Use the backward time algorithm applied to (15)–(16). This produces the data $\{\mathbf{y}^m(t); T \geqslant t \geqslant 1\}$ with $\mathbf{y}^m(T) = \mathbf{y}^0(T)$ to use in estimating the parameter vector.
3. Estimate the model parameter vector via GML to obtain $\hat{\boldsymbol{\theta}}^m$.
4. Compute the conditional forecast of the future data using the estimated model: Set $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^m, \mathbf{s}(T+1|T) = \mathbf{s}(T+1|T; \hat{\boldsymbol{\theta}}^m)$ and $\mathbf{e}(t) = \mathbf{0}$ in the forward time model, (10), and solve for $\{\hat{\mathbf{y}}^m(T+\ell, \hat{\boldsymbol{\theta}}^m); 1 \leqslant \ell \leqslant L\}$.
5. Compute and store the conditional forecast error:

$$\mathbf{d}^m(\ell) = \mathbf{y}^m(T+\ell, \boldsymbol{\theta}) - \hat{\mathbf{y}}^m(T+\ell, \hat{\boldsymbol{\theta}}^m) \qquad 1 \leqslant \ell \leqslant L$$

Note that the original sample path is only used to fix the point from which all backward data sets originate and all forecasts propagate. In the simulation experiments below, we use M = 2,000 and L = 4. The approximate 'true' distribution is then given by the relative frequency histogram of the observed conditional forecast errors.

Each of the experiments presented below is summarized by two sets of four histograms. One set presents the approximate 'true' relative frequency histograms for each forecast lead time, while the other set presents the relative frequency histograms obtained from application of the bootstrap. Superimposed on each is the Gaussian density that follow from application of the asymptotic Gaussian theory. All experiments use short data sets with $T = 49$ to emphasize the efficacy of the bootstrap when the use of asymptotics is questionable and where bias is a factor in the forecasts. Prediction intervals follow immediately from the data sumarized in the histograms. Although we choose to present only the histograms, the percentile, the bias-corrected (BC), and the accelerated bias-corrected (BC_a) method all are applicable for generating confidence intervals using the generated data (Efron, 1987).

5.1. *Moving average models*

Our first experiment uses the MA(2) model:

$$y(t) = v(t) - 0.40v(t-1) - 0.45v(t-2)$$

where $v(t) = 0.2z(t)$ and $z(t)$ is a centered exponential with unit variance. The model is second-order with

$$\mathbf{F} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad \mathbf{G} = \begin{bmatrix} -0.40 \\ -0.45 \end{bmatrix} \qquad \mathbf{H} = [1 \quad 0] \qquad \mathbf{D} = 0$$

The issue here is not the accuracy of the bootstrap relative to the Gaussian theory (the Gaussian theory is clearly inapplicable) but the fidelity of the bootstrap relative to the 'true' distribution. We want to know how informative is the bootstrap when we know the asymptotic Gaussian theory is inapplicable. Figure 2 depicts the results for the 'true' distribution of conditional forecast errors and Figure 3 gives the results of the bootstrap. The bootstrap represents clearly the salient features of the true situation; it indicates the small sample bias, the asymmetry and peakedness of the true forecast errors.

5.2. *Autoregressive-moving average models*

Our next simulation experiment involves the ARMA(1,1):

$$y(t) = 0.7y(t-1) + v(t) + 0.10v(t-1)$$

where $v(t) = 0.2z(t)$ and $z(t)$ is a mixture of 90% N(–1/9, .15) and 10% N(1, .15). The model is first-order with

$$\mathbf{F} = [0.70] \qquad \mathbf{G} = [0.80] \qquad \mathbf{H} = [1] \qquad \mathbf{D} = 0$$

Figures 4 and 5 reveal the value of the bootstrap. Indication of the mixture distribution is striking in both the 'true' and the bootstrap; the bimodality and asymmetry are clearly evident.

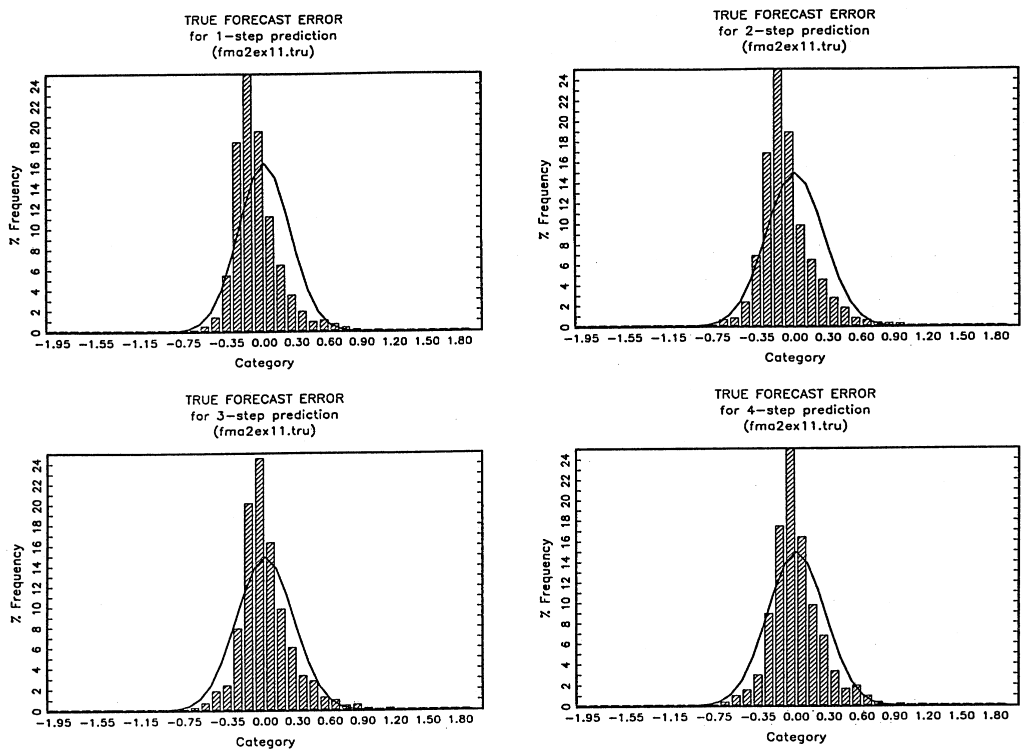

FIGURE 2. 'True' histograms for the MA(2) example.

6 CONCLUSIONS

Our presentation demonstrates the utility of the bootstrap in assessing the conditional forecast errors of state space models. Since these model representations include as special cases the MA(q), ARMA(p,q), dynamic factor, and time-varying parameter models, we also provide means to evaluate conditional forecast error in models that have not received much attention in the literature. Simulation experiments show the bootstrap yields informative estimates of the conditional forecast error distribution for MA(q), ARMA(p,q), and state space models. With as few as 1000 bootstrap replication we are able to detect non-Gaussian situations and bias. The ability of the bootstrap to detect non-Gaussian situations is startling. Even in cases where all the disturbances are Gaussian, the bootstrap is able to detect departures from the asymptotic situation. Our empirical application suggests that invoking the asymptotic Gaussian theory is not justified, even with sample series as long as 50 observations when only two parameters are estimated. We note, in passing, that our results are more suggestive than complete; we must await the results of large scale simulation studies with M on the order of 10^4 with coverage probabilities computed on the basis of ensembles composed of 1000 or more time series.

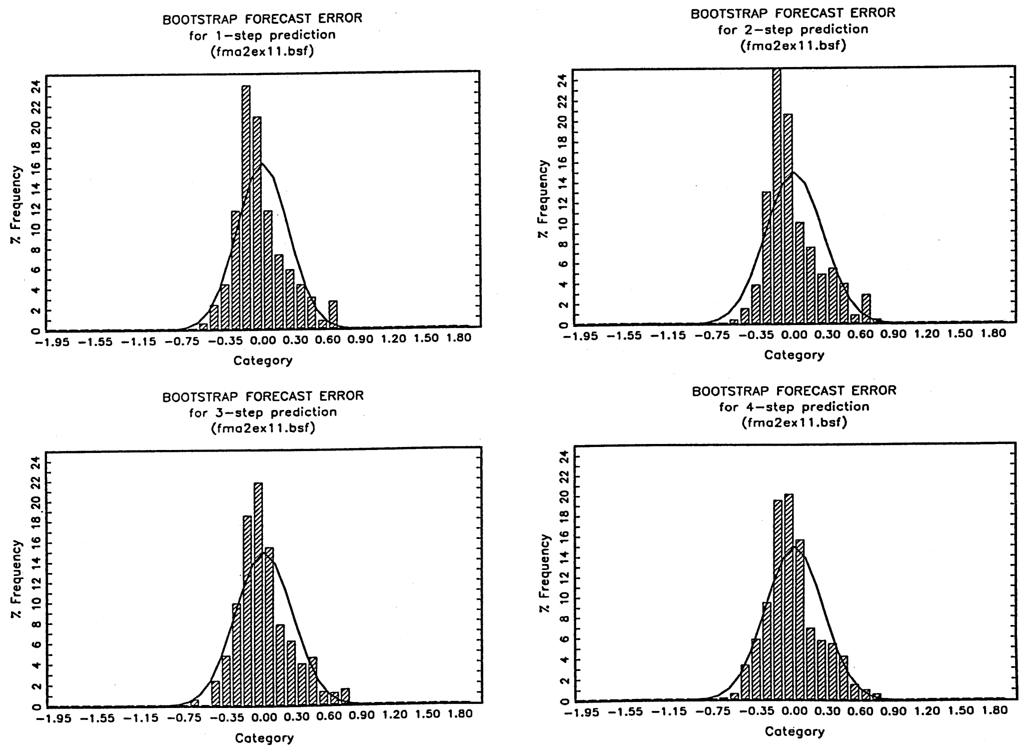

FIGURE 3. Bootstrap histograms for the MA(2) example.

The computational demands of bootstrapping are not burdensome. An implementation in GAUSS on a 400MHz Pentium II desktop PC requires approximately 27 minutes for a model with three or four estimated parameters when $B = 2000$. One resample of the model residuals, the generation of a conditional (backward) dataset and the estimation of the model parameters is completed in less than one second in our empirical example, and this involves additional time to set up the time-varying matrices within the Kalman filter.

APPENDIX

A: DERIVATION OF THE REVERSE-TIME INNOVATIONS FILTER

The derivation assumes $\mathbf{x}(t) = 0$. If $\mathbf{x}(t) \neq 0$, the derivation is applied to the variational model, as outlined in the text. With this in mind, the reverse-time state space model corresponding to (10) is

$$\mathbf{r}(t) = \mathbf{F}_B(t)\mathbf{r}(t+1) + \mathbf{G}_{1,B}(t)\mathbf{e}_B(t)$$
$$\mathbf{y}(t) = \mathbf{H}_B(t)\mathbf{r}(t+1) + \mathbf{D}_{1,B}(t)\mathbf{e}_B(t)$$

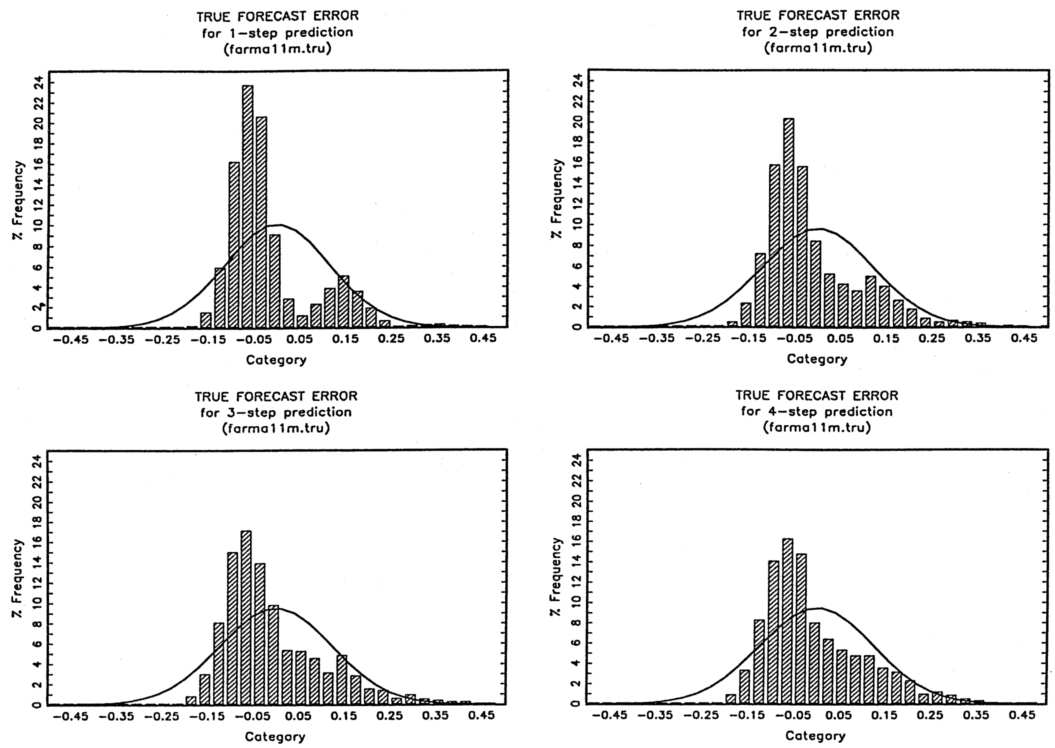

FIGURE 4. 'True' histograms for the ARMA(1,1) example.

Allowance is made for time variation in all matrices. The objective is to express $\mathbf{F}_B(t), \mathbf{G}_{1,B}(t), \mathbf{H}_B(t), \mathbf{D}_{1,B}(t)$ and $\mathbf{e}_B(t)$ in terms of the coefficient matrices and variables of the forward-time model (10). We do this first for the state equation, deriving $\mathbf{F}_B(t)$ and the joint term $\mathbf{G}_{1,B}(t)\mathbf{e}_B(t)$. We next address the output equation, deriving $\mathbf{H}_B(t)$ and the joint term $\mathbf{D}_{1,B}(t)\mathbf{e}_B(t)$. The terms involving the backward residual, $\mathbf{e}_B(t)$, are found to be a function of both the forward residual and the forward state. To simplify notation, we use $\mathbf{s}(t)$ to denote $\mathbf{s}(t|t-1)$ and assume the reader has no difficulty identifying $\mathbf{N}(t)$ with $\mathbf{H}_B(t)$.

Let $\overline{\mathcal{S}}_t = \mathcal{H}_t^+ + \mathcal{X}_t$ where $\mathcal{H}_t^+$ is the linear span of $\{\mathbf{y}(k); k \geq t\}$ and $\mathcal{X}_t$ is the minimal state space; see Caines (1988, pp. 216–35) for details. Let $(\mathbf{v}|\mathcal{W})$ denote orthogonal projection of $\mathbf{v}$ onto the space $\mathcal{W}$ and define $\mathbf{r}(t) = \mathbf{v}(t)^{-1}\mathbf{s}(t)$. By definition $\mathbf{r}(t)$ is orthogonal to $\mathcal{H}_t^+$ but not $\mathcal{X}_t$, in fact

$$
\begin{aligned}
(\mathbf{r}(t)|\overline{\mathcal{S}}_{t+1}) &= (\mathbf{r}(t)|\mathcal{X}_{t+1}) \\
&= \mathbf{V}(t)^{-1}E\{\mathbf{s}(t)\mathbf{s}'(t+1)\}[E\{\mathbf{s}(t+1)\mathbf{s}'(t+1)\}]^{-1}\mathbf{s}(t+1) \\
&= \mathbf{V}(t)^{-1}E\{\mathbf{s}(t)[\mathbf{s}'(t)\mathbf{F}' + \mathbf{e}'(t)\mathbf{G}_1'(t)]\}\mathbf{V}(t+1)^{-1}\mathbf{s}(t+1) \\
&= \mathbf{V}(t)^{-1}[\mathbf{V}(t)\mathbf{F}' + \mathbf{0}]\mathbf{V}(t+1)^{-1}\mathbf{s}(t+1) \\
&= \mathbf{F}'\mathbf{V}(t+1)^{-1}\mathbf{s}(t+1) \\
&= \mathbf{F}'\mathbf{r}(t+1)
\end{aligned}
$$

Hence $\mathbf{F}_B(t) = \mathbf{F}'$. Now

$$\mathbf{r}(t) = (\mathbf{r}(t)|\overline{\mathcal{S}}_{t+1}) + \mathbf{z}(t)$$

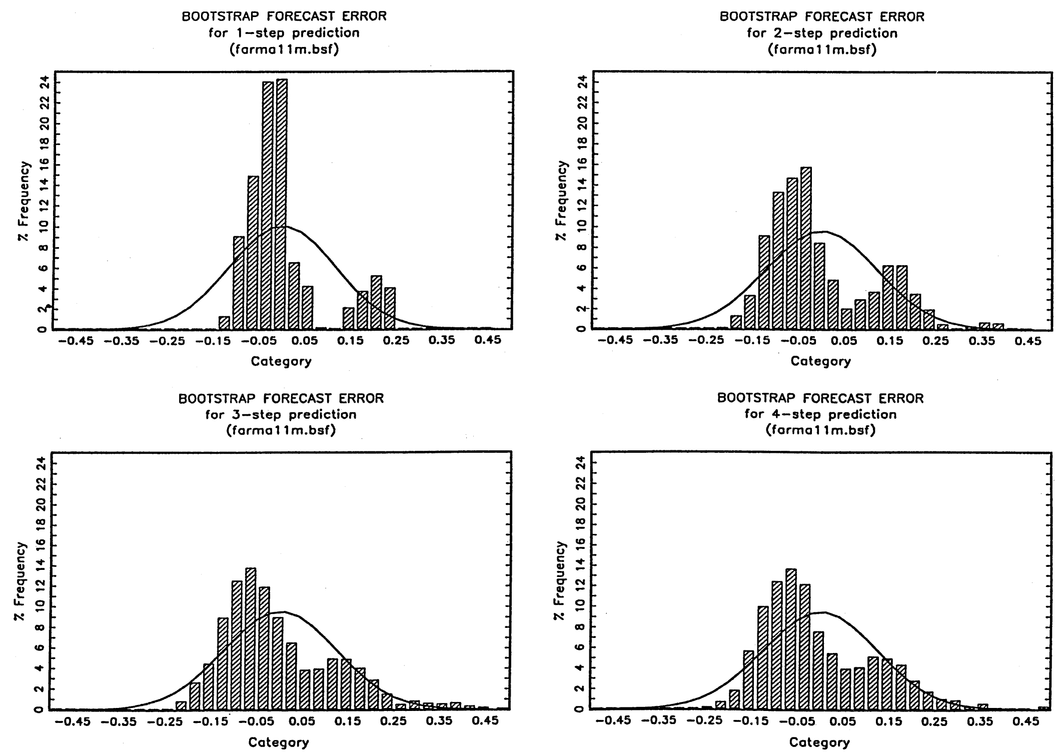

FIGURE 5. Bootstrap histograms for the ARMA(1,1) example.

where $\mathbf{z}(t) = [\mathbf{r}(t) - (\mathbf{r}(t)|\overline{\mathcal{S}}_{t+1})$ is the projection error; i.e., that component of $\mathbf{r}(t)$ orthogonal to $\overline{\mathcal{S}}_{t+1}$. Thus we write

$$\mathbf{r}(t) = \mathbf{F}'\mathbf{r}(t+1) + [\mathbf{r}(t) - (\mathbf{r}(t)|\overline{\mathcal{S}}_{t+1})]$$

where

$$\begin{aligned}\mathbf{r}(t) - (\mathbf{r}(t)|\overline{\mathcal{S}}_{t+1}) &= \mathbf{V}(t)^{-1}\mathbf{s}(t) - \mathbf{F}'\mathbf{V}(t+1)^{-1}\mathbf{s}(t+1)\\ &= \mathbf{V}(t)^{-1}\mathbf{s}(t) - \mathbf{F}'\mathbf{V}(t+1)^{-1}[\mathbf{F}\mathbf{s}(t) + \mathbf{G}_1(t)\mathbf{e}(t)]\\ &= [\mathbf{V}(t)^{-1} - \mathbf{F}'\mathbf{V}(t+1)^{-1}\mathbf{F}]\mathbf{s}(t) - \mathbf{F}'\mathbf{V}(t+1)^{-1}\mathbf{G}_1(t)\mathbf{e}(t)\end{aligned}$$

Note that, by definition, this is equal to $\mathbf{G}_{1,B}(t)\mathbf{e}_B(t)$. Explicit derivation of $\mathbf{G}_{1,B}(t)$ and $\mathbf{e}_B(t)$ is not required for implementation and is not pursued. The interested reader may refer to Caines(1998) and Aoki(1989) for further details.

The output equation derivation follows a similar route by making use of the orthogonality between $\mathbf{r}(t+1)$ and $\mathbf{y}(t)$:

$$\begin{aligned}E\{\mathbf{y}(t)\mathbf{r}'(t+1)\} &= E\{[\mathbf{H}_B(t)\mathbf{r}(t+1) + \mathbf{D}_{1,B}(t)\mathbf{e}_B(t)]\mathbf{r}'(t+1)\}\\ &= \mathbf{H}_B(t)E\{\mathbf{V}(t+1)^{-1}\mathbf{s}(t+1)\mathbf{s}'(t+1)\mathbf{V}(t+1)^{-1}\} + \mathbf{D}_{1,B}(t)\cdot\mathbf{0}\\ &= \mathbf{H}_B(t)\mathbf{V}(t+1)^{-1}\end{aligned}$$

but we also have

$$\begin{aligned}E\{\mathbf{y}(t)\mathbf{r}'(t+1)\} &= E\{[\mathbf{Hs}(t)+\mathbf{D}_1(t)\mathbf{e}(t)]\mathbf{s}'(t+1)\}\mathbf{V}(t+1)^{-1}\\ &= \mathbf{H}E\{\mathbf{s}(t)[\mathbf{s}'(t)\mathbf{F}'+\mathbf{e}'(t)\mathbf{G}_1'(t)]\}\mathbf{V}(t+1)^{-1}\\ &\quad +\mathbf{D}_1(t)E\{\mathbf{e}(t)[\mathbf{s}'(t)\mathbf{F}'+\mathbf{e}'(t)\mathbf{G}_1'(t)]\}\mathbf{V}(t+1)^{-1}\\ &= \mathbf{HV}(t)\mathbf{F}'\mathbf{V}(t+1)^{-1}+\mathbf{D}_1(t)\cdot\mathbf{I}\cdot\mathbf{G}_1'(t)\mathbf{V}(t+1)^{-1}\end{aligned}$$

The implied identity gives

$$\mathbf{H}_B(t)=\mathbf{HV}(t)\mathbf{F}'+\mathbf{D}_1(t)\mathbf{G}_1'(t)$$

Our implementation is complete by obtaining an expression for $\mathbf{D}_{1,B}(t)\mathbf{e}_B(t)=\mathbf{y}(t)-E\{\mathbf{y}(t)|\overline{\mathcal{S}}_{t+1}\}$. Since $E\{\mathbf{y}(t)|\overline{\mathcal{S}}_{t+1}\}=\mathbf{H}_B(t)\mathbf{r}(t+1)$ this is yields

$$\begin{aligned}\mathbf{D}_{1,B}(t)\mathbf{e}_B(t) &= \mathbf{Hs}(t)+\mathbf{D}_1(t)\mathbf{e}(t)-\mathbf{H}_B(t)\mathbf{V}(t+1)^{-1}\mathbf{s}(t+1)\\ &= \mathbf{Hs}(t)+\mathbf{D}_1(t)\mathbf{e}(t)-[\mathbf{HV}(t)\mathbf{F}'+\mathbf{D}_1(t)\mathbf{G}_1'(t)][\mathbf{Fs}(t)+\mathbf{G}_1(t)\mathbf{e}(t)]\\ &= [\mathbf{HV}(t)\mathbf{A}(t)-\mathbf{D}_1(t)\mathbf{B}'(t)]\mathbf{s}(t)+[\mathbf{D}_1(t)\mathbf{C}(t)-\mathbf{HV}(t)\mathbf{B}(t)]\mathbf{e}(t)\end{aligned}$$

Once again, we do not derive explicit expressions for $\mathbf{D}_{1,B}(t)$ and $\mathbf{e}_B(t)$ since they are not required for implementation.

B: STEADY-STATE JUSTIFICATION OF THE PROCEDURE

We assume that we have n observations, $\{\mathbf{y}(T-n+1),\ldots,\mathbf{y}(T)\}$, and that n is large. Throughout this appendix, we let $\hat{\boldsymbol{\theta}}_n$ denote the (assumed consistent as $n\to\infty$) Gaussian MLE of $\boldsymbol{\theta}$, and let $\boldsymbol{\theta}_n^*$ denote a bootstrap parameter estimate. We assume that the eigenvalues of $\mathbf{F}(\boldsymbol{\theta})$ are within the unit circle and the system is controllable and observable; these assumptions are enough to ensure the asymptotic stability of the filter. For one-step-ahead forecasting, the process $\boldsymbol{\xi}(t)$ is given by

$$\boldsymbol{\xi}(T+1)=\mathbf{A}(\boldsymbol{\theta})\boldsymbol{\xi}(T)+\mathbf{B}(\boldsymbol{\theta})\mathbf{x}(T+1)+\mathbf{C}(\boldsymbol{\theta})\mathbf{e}(T+1) \tag{B.1}$$

where

$$\boldsymbol{\xi}(t)=\begin{bmatrix}\mathbf{s}(t+1|t)\\ \mathbf{y}(t)\end{bmatrix} \tag{B.2}$$

and the matrices $\mathbf{A}(\boldsymbol{\theta}),\mathbf{B}(\boldsymbol{\theta})$ and $\mathbf{C}(\boldsymbol{\theta})$ are defined by the matrices appearing in (10) of the text (in steady-state $\mathbf{C}(\boldsymbol{\theta})$ is independent of t). Recall that $\{\mathbf{x}(t)\}$ is a fixed and known input process. For convenience, we have dropped the parameter from the notation when representing a filtered value that depends upon $\boldsymbol{\theta}$. For example, in (B.1), we write $\boldsymbol{\xi}(T)=\boldsymbol{\xi}(T,\boldsymbol{\theta})$ and $\mathbf{e}(T)=\mathbf{e}(T,\boldsymbol{\theta})$. The process $\{\mathbf{e}(t)\}$ is the standardized, steady-state innovations sequence so that $E\{\mathbf{e}(t)\}=\mathbf{0}$ and $E\{\mathbf{e}(t)\mathbf{e}'(t)\}=\mathbf{I}_q$.

The one-step-ahead conditional forecast estimate is given by

$$\tilde{\boldsymbol{\xi}}(T+1)=\mathbf{A}(\hat{\boldsymbol{\theta}}_n)\hat{\boldsymbol{\xi}}(T)+\mathbf{B}(\hat{\boldsymbol{\theta}}_n)\mathbf{x}(T+1) \tag{B.3}$$

where, in keeping consistent with the notation, we have written $\hat{\boldsymbol{\xi}}(t)=\boldsymbol{\xi}(t,\hat{\boldsymbol{\theta}}_n)$. The conditional forecast estimate is labelled with a tilde. Watanabe (1985) showed that, under the assumed conditions and notation,

$$\hat{\mathbf{s}}(T+1/T)=\mathbf{s}(T+1/T)+\mathbf{o}_p(1)(n\to\infty),$$

and, consequently we write

$$\hat{\boldsymbol{\xi}}(T) = \boldsymbol{\xi}(T) + \mathbf{o}_p(1)$$

noting that the final q elements of $\hat{\boldsymbol{\xi}}(T)$ and $\boldsymbol{\xi}(T)$ are identical. Hence, the conditional prediction error can be written as

$$\begin{aligned}\Delta_n &\equiv \boldsymbol{\xi}(T+1) - \tilde{\boldsymbol{\xi}}(T+1) \\ &= [\mathbf{A}(\boldsymbol{\theta}) - \mathbf{A}(\hat{\boldsymbol{\theta}}_n)]\hat{\boldsymbol{\xi}}(T) + [\mathbf{B}(\boldsymbol{\theta}) - \mathbf{B}(\hat{\boldsymbol{\theta}}_n)]\mathbf{x}(T+1) \\ &\quad + \mathbf{A}(\boldsymbol{\theta})\mathbf{o}_p(1) + \mathbf{C}(\boldsymbol{\theta})\mathbf{e}(T+1) \end{aligned} \tag{B.4}$$

From (B.4), we see the two sources of variation, namely the variation due to estimating the parameter $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}_n$ and the variation due to estimating the innovation value $\mathbf{e}(T+1)$ by zero.

In the conditional bootstrap procedure, we mimic (B.1) and obtain a pseudo observation

$$\boldsymbol{\xi}^*(T+1) = \mathbf{A}(\hat{\boldsymbol{\theta}}_n)\hat{\boldsymbol{\xi}}(T) + \mathbf{B}(\hat{\boldsymbol{\theta}}_n)\mathbf{x}(T+1) + \mathbf{C}(\hat{\boldsymbol{\theta}}_n)\mathbf{e}^*(T+1) \tag{B.5}$$

where we hold $\hat{\boldsymbol{\xi}}(T)$ fixed throughout the resampling procedure. Note that, because the filter is in steady-state, the data, $\{\mathbf{y}(T-n+1), \ldots, \mathbf{y}(T)\}$, completely determine $\hat{\boldsymbol{\theta}}_n$ and consequently $\hat{\boldsymbol{\xi}}(T)$. For finite sample lengths, the data and the initial conditions determine $\hat{\boldsymbol{\theta}}_n$. As a practical matter, if precise initial conditions are unknown, one can drop the first few data points from the estimation of $\boldsymbol{\theta}$ so that changing the initial state conditions does not change $\hat{\boldsymbol{\theta}}_n$ nor $\hat{\boldsymbol{\xi}}(T)$. We remark that while the data $\{\mathbf{y}(T-n+1), \ldots, \mathbf{y}(T)\}$ completely determine $\hat{\boldsymbol{\xi}}(T)$, the reverse is not true; that is, fixing $\hat{\boldsymbol{\xi}}(T)$ in no way fixes the entire data sequence $\{\mathbf{y}(T-n+1), \ldots, \mathbf{y}(T)\}$. For example, in the AR(1) model, fixing $\hat{\boldsymbol{\xi}}(T)$ is equivalent to fixing $\mathbf{y}(T)$ only. In addition, $\mathbf{e}^*(T+1)$ is a random draw from the empirical distribution of the standardized steady-state innovations $\{\hat{\mathbf{e}}(T-n+1), \ldots, \hat{\mathbf{e}}(T)\}$ where, as above, we have written $\hat{\mathbf{e}}(t) \equiv \mathbf{e}(t, \hat{\boldsymbol{\theta}}_n)$. Under the mixing conditions of Gastwirth and Rubin (1975), the empirical distribution of the standardized steady-state innovations $\{\hat{\mathbf{e}}(T-n+1), \ldots, \hat{\mathbf{e}}(T)\}$ converges weakly $(n \to \infty)$ to the standardized steady-state innovations distribution.

To mimic the conditional forecast in (B.3), the bootstrap estimated conditional forecast is given by

$$\tilde{\boldsymbol{\xi}}^*(T+1) \equiv \mathbf{A}(\boldsymbol{\theta}_n^*)\hat{\boldsymbol{\xi}}(T) + \mathbf{B}(\boldsymbol{\theta}_n^*)\mathbf{x}(T+1) \tag{B.6}$$

which yields the bootstrapped conditional forecast error

$$\begin{aligned}\Delta_n^* &\equiv \boldsymbol{\xi}^*(T+1) - \tilde{\boldsymbol{\xi}}^*(T+1) \\ &= [\mathbf{A}(\hat{\boldsymbol{\theta}}_n) - \mathbf{A}(\boldsymbol{\theta}_n^*)]\hat{\boldsymbol{\xi}}(T) + [\mathbf{B}(\hat{\boldsymbol{\theta}}_n) - \mathbf{B}(\boldsymbol{\theta}_n^*)]\mathbf{x}(T+1) \\ &\quad + \mathbf{C}(\hat{\boldsymbol{\theta}}_n)\mathbf{e}^*(T+1) \end{aligned} \tag{B.7}$$

Comparison of (B.4) and (B.7) shows why, in finite samples, the bootstrap works; that is, (B.7) is a sample-based imitation of (B.4). Letting $n \to \infty$ in (B.4), while holding $\hat{\boldsymbol{\xi}}(T)$ fixed, we see that if $\hat{\boldsymbol{\theta}}_n \to_p \boldsymbol{\theta}$ then $\Delta_n \Rightarrow \mathbf{C}(\boldsymbol{\theta})\mathbf{u}$ where $\mathbf{u}$ is a random vector that is distributed according to the steady-state standardized innovations distribution ($\Rightarrow$ denotes weak convergence). In addition, if the innovations are mixing and if conditional on the data, $\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n \to_p \mathbf{0}$, then $\Delta_n^* \Rightarrow \mathbf{C}(\boldsymbol{\theta})\mathbf{u}$ as $n \to \infty$. Extending these results to k-step-ahead forecasts follows easily by induction. Stoffer and Wall (1991) established conditions under which $\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n \to_p \mathbf{0}$ as $n \to \infty$ when the forward innovations are resampled. It remains to determine the conditions under which this result holds when the backward innovations are resampled.

REFERENCES

AOKI, M. (1989) *Optimization of Stochastic Systems: Topics in Discrete-time Dynamics* (2nd ed). New York: Academic Press.

BREIDT, J., DAVIS, R. and DUNSMUIR, W. (1992) On backcasting in linear time series. *In New Directions in Time Series Analysis*, Vol. 1, (eds D. Brillinger et al.). New York: Springer-Verlag, 25–40.

CAINES, P. E. (1988) *Linear Stochastic Systems*. New York: John Wiley & Sons.

EFRON, B. (1987) Better bootstrap confidence intervals. *Journal of the American Statistical Association* 82, 171–85.

FINDLEY, D. (1986) On bootstrap estimates of forecast mean square errors for autoregressive processes. *In Computer Science and Statistics*. The interface, (ed. D. Allen), Amsterdam, North-Holland, 11–17.

GASTWIRTH, J. L. and RUBIN, H. (1975) The asymptotic distribution theory of the empiric c.d.f. for mixing processes. *Annals of Statistics* 3, 809–24.

KABAILA, P. (1993) On bootstrap predictive inference for autoregressive processes. *Journal of Time Series Analysis* 14 (5), 473–84.

MCCULLOUGH, B. D. (1994) Bootstrapping forecasting intervals: An application to AR(p) Models. *Journal of Forecasting* 13 (1), 51–66.

———(1996) Consistent forecast intervals when the forecast-period exogenous variables are stochastic. *Journal of Forecasting* 15 (4), 293–304.

STINE, R. A. (1987) Estimating properties of autoregressive forecasts. *Journal of the American Statistical Association* 82, 1072–8.

STOFFER, D. S. and WALL, K. D. (1991) Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter. *Journal of the American Statistical Association* 86 (416), 1024–33.

THOMBS, L. and SCHUCHANY, W. (1990) Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association* 85, 486–92.

WALL, K. D. (1987) Identification theory for varying coefficient regression models. *Journal of Time Series Analysis* 8 (3), 359–71.

WATANABE, N. (1985) Note on the Kalman filter with estimated parameters. *Journal of Time Series Analysis* 6 (3), 269–78.