Resampling in State Space Models*

David S. Stoffer
Department of Statistics
University of Pittsburgh
Pittsburgh, PA 15260 USA

Kent D. Wall
Defense Resources Management Institute
Naval Postgraduate School
Monterey, CA 93943 USA

Abstract. Resampling the innovations sequence of state space models has proved to be a useful tool in many respects. For example, while under general conditions, the Gaussian MLEs of the parameters of a state space model are asymptotically normal, several researchers have found that samples must be fairly large before asymptotic results are applicable. Moreover, problems occur if the any of parameters are near the boundary of the parameter space. In such situations, the bootstrap applied to the innovation sequence can provide an accurate assessment of the sampling distributions of the parameter estimates. We have also found that a resampling procedure can provide insight into the validity of the model. In addition, the bootstrap can be used to evaluate conditional forecast errors of state space models. The key to this method is the derivation of a reverse-time innovations form of the state space model for generating conditional data sets. We will provide some theoretical insight into our procedures that show why resampling works in these situations, and we provide simulations and data examples that demonstrate our claims.

Key words. ARMAX models, Bootstrap, Finite sample distributions, Forecasting, Innovations filter, Kalman filter, Reverse-time state space model, Stochastic regression, Stochastic volatility.

1 Introduction

A very general model that seems to subsume a whole class of special cases of interest is the state space model or the dynamic linear model, which was introduced in Kalman (1960) and Kalman and Bucy (1961). Although the model was originally developed as a method primarily for use in aerospace-related research, it has been applied to modeling data from such diverse fields as economics (e.g. Harrison and Stevens, 1976, Harvey and Pierse, 1984, Harvey and Todd, 1983, Kitagawa and Gersch 1984, Shumway and Stoffer, 1982), medicine (e.g. Jones, 1984) and molecular biology (e.g. Stultz, et al, 1993). An excellent modern treatment of time series analysis based on the state space model is the text by Durbin and Koopman (2001). We note, in particular, that ARMAX models can be written in state space form (see e.g. Shumway and Stoffer, 2000, §4.6), so anything we say and do here regarding state space models applies equally to ARMAX models.

Here, we write the state space model as

$$\boldsymbol{x}_{t+1} = \Phi \boldsymbol{x}_t + \Upsilon \boldsymbol{u}_t + \boldsymbol{w}_t \qquad t = 0, 1, ..., n \tag{1}$$

$$\boldsymbol{y}_t = A_t \boldsymbol{x}_t + \Gamma \boldsymbol{u}_t + \boldsymbol{v}_t \qquad t = 1, ..., n \tag{2}$$

where $\boldsymbol{x}_t$ represents the p-dimensional state vector, and $\boldsymbol{y}_t$ represents the q-dimensional observation vector. In the state equation (1), the initial state $\boldsymbol{x}_0$ has mean $\boldsymbol{\mu}_0$ and variance-covariance matrix

*Chapter 9 of *State Space and Unobserved Component Models: Theory and Applications,* A. Harvey, S.J. Koopman and N. Shephard (Eds). Cambridge University Press, 2004

Σ_0; Φ is $p \times p$, Υ is $p \times r$, and $\boldsymbol{u}_t$ is an $r \times 1$ vector of fixed inputs. In the observation equation (2), A_t is $q \times p$ and Γ is $q \times r$. Here, $\boldsymbol{w}_t$ and $\boldsymbol{v}_t$ are white noise series (both independent of $\boldsymbol{x}_0$), with $\text{var}(\boldsymbol{w}_t) = Q$, $\text{var}(\boldsymbol{v}_t) = R$, but we also allow the state noise and observation noise to be correlated at time t; that is, $\text{cov}(\boldsymbol{w}_t, \boldsymbol{v}_t) = S$, and zero otherwise. Note, S is a $p \times q$ matrix. Throughout, we assume the model coefficients and the correlation structure of the model are uniquely parameterized by a $k \times 1$ parameter vector Θ; thus, $\Phi = \Phi(\Theta)$, $\Upsilon = \Upsilon(\Theta)$, $Q = Q(\Theta)$, $A_t = A_t(\Theta)$, $\Gamma = \Gamma(\Theta)$, $R = R(\Theta)$, and $S = S(\Theta)$.

We denote the best linear predictor of $\boldsymbol{x}_{t+1}$ given the data $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t\}$ as $\boldsymbol{x}_{t+1}^t$, and denote the covariance matrix of the prediction error, $(\boldsymbol{x}_{t+1} - \boldsymbol{x}_{t+1}^t)$, as P_{t+1}^t. The Kalman filter (e.g. Anderson and Moore, 1979) can be used to obtain the predictors and their covariance matrices successively as new observations become available. The innovation sequence, $\{\boldsymbol{\epsilon}_t;\ t = 1, \ldots, n\}$, is defined to be the sequence of errors in the best linear prediction of $\boldsymbol{y}_t$ given the data $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}\}$. The innovations are

$$\boldsymbol{\epsilon}_t = \boldsymbol{y}_t - A_t \boldsymbol{x}_t^{t-1} - \Gamma \boldsymbol{u}_t, \qquad t = 1, \ldots, n, \tag{3}$$

where the innovation variance-covariance matrix is given by

$$\Sigma_t = A_t P_t^{t-1} A_t' + R, \qquad t = 1, \ldots, n. \tag{4}$$

The innovations form of the Kalman filter, for $t = 1, \ldots, n$, is given by the following equations with initial conditions $\boldsymbol{x}_1^0 = \Phi \boldsymbol{\mu}_0 + \Upsilon \boldsymbol{u}_0$ and $P_1^0 = \Phi \Sigma_0 \Phi' + Q$:

$$\begin{aligned} \boldsymbol{x}_{t+1}^t &= \Phi \boldsymbol{x}_t^{t-1} + \Upsilon \boldsymbol{u}_t + K_t \boldsymbol{\epsilon}_t, & (5)\\ P_{t+1}^t &= \Phi P_t^{t-1} \Phi' + Q - K_t \Sigma_t K_t', & (6)\\ K_t &= (\Phi P_t^{t-1} A_t' + S) \Sigma_t^{-1}. & (7) \end{aligned}$$

In this article, we will work with the standardized innovations

$$\boldsymbol{e}_t = \Sigma_t^{-1/2} \boldsymbol{\epsilon}_t, \tag{8}$$

so we are guaranteed these innovations have, at least, the same first two moments. In (8), $\Sigma_t^{1/2}$ denotes the unique square root matrix of Σ_t defined by $\Sigma_t^{1/2} \Sigma_t^{1/2} = \Sigma_t$. We now define the $(p+q) \times 1$ vector

$$\boldsymbol{\xi}_t = \begin{bmatrix} \boldsymbol{x}_{t+1}^t \\ \boldsymbol{y}_t \end{bmatrix}.$$

Combining (3) and (5) results in a vector first-order equation for $\boldsymbol{\xi}_t$ given by

$$\boldsymbol{\xi}_t = F_t \boldsymbol{\xi}_{t-1} + G \boldsymbol{u}_t + H_t \boldsymbol{e}_t, \tag{9}$$

where

$$F_t = \begin{bmatrix} \Phi & 0 \\ A_t & 0 \end{bmatrix}, \quad G = \begin{bmatrix} \Upsilon \\ \Gamma \end{bmatrix}, \quad H_t = \begin{bmatrix} K_t \Sigma_t^{1/2} \\ \Sigma_t^{1/2} \end{bmatrix}.$$

Estimation of the model parameters Θ is accomplished by Gaussian quasi-maximum likelihood. The innovations form of the Gaussian likelihood (ignoring a constant) is

$$\begin{aligned} -\ln L_Y(\Theta) &= \frac{1}{2} \sum_{t=1}^{n} \Big(\ln |\Sigma_t(\Theta)| + \boldsymbol{\epsilon}_t(\Theta)' \Sigma_t(\Theta)^{-1} \boldsymbol{\epsilon}_t(\Theta) \Big) \\ &= \frac{1}{2} \sum_{t=1}^{n} \Big(\ln |\Sigma_t(\Theta)| + \boldsymbol{e}_t(\Theta)' \boldsymbol{e}_t(\Theta) \Big), \end{aligned} \tag{10}$$

where $L_Y(\Theta)$ denotes the likelihood of Θ given the data $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ assuming normality; note that we have emphasized the dependence of the innovations on the parameters Θ. We stress the fact that it is not necessary for the data to be Gaussian to consider (10) as the criterion function to be used for parameter estimation. Furthermore, under certain rare conditions, the Gaussian quasi-MLE of Θ when the process is non-Gaussian is asymptotically optimal; details can be found in Caines (1988, Chapter 8).

2 Assessing the Finite Sample Distribution of Parameter Estimates

Although, under general conditions (which we assume to hold in this section), the MLEs of the parameters of the model, Θ, are consistent and asymptotically normal, time series data are often of short or moderate length. Several researchers have found evidence that samples must be fairly large before asymptotic results are applicable (Dent and Min, 1978; Ansley and Newbold, 1980). Moreover, it is well known that problems occur if the parameters are near the boundary of the parameter space. In this section, we discuss an algorithm for bootstrapping state space models to assess the finite sample distribution of the model parameters. This algorithm and its justification, including the non-Gaussian case, along with examples and simulations, can be found in Stoffer and Wall (1991).

Let $\widehat{\Theta}$ denote the Gaussian quasi-MLE of Θ, that is, $\widehat{\Theta} = \text{argmax}_{\Theta} L_Y(\Theta)$, where $L_Y(\Theta)$ is given in (10); of course, if the process is Gaussian, $\widehat{\Theta}$ is the MLE. Let $\boldsymbol{\epsilon}_t(\widehat{\Theta})$ and $\Sigma_t(\widehat{\Theta})$ be the innovation values obtained by running the filter under $\widehat{\Theta}$. Once this has been done, the bootstrap procedure is accomplished by the following steps.

1. Construct the standardized innovations

$$\boldsymbol{e}_t(\widehat{\Theta}) = \Sigma_t^{-1/2}(\widehat{\Theta})\boldsymbol{\epsilon}_t(\widehat{\Theta}).$$

2. Sample, with replacement, n times from the set $\{\boldsymbol{e}_1(\widehat{\Theta}), ..., \boldsymbol{e}_n(\widehat{\Theta})\}$ to obtain $\{\boldsymbol{e}_1^*, ..., \boldsymbol{e}_n^*\}$, a bootstrap sample of standardized innovations.

3. To construct a bootstrap data set $\{\boldsymbol{y}_1^*, ..., \boldsymbol{y}_n^*\}$, solve (9) using $\boldsymbol{e}_t^*$ in place of $\boldsymbol{e}_t$; that is, solve

$$\boldsymbol{\xi}_t^* = F_t(\widehat{\Theta})\boldsymbol{\xi}_{t-1}^* + G(\widehat{\Theta})\boldsymbol{u}_t + H_t(\widehat{\Theta})\boldsymbol{e}_t^*, \tag{11}$$

   for $t = 1, \ldots, n$. The exogenous variables $\boldsymbol{u}_t$ and the initial conditions of the Kalman filter remain fixed at their given values, and the parameter vector is held fixed at $\widehat{\Theta}$. Note that a bootstrapped observation $\boldsymbol{y}_t^*$ is obtained from the final q rows of the $(p+q) \times 1$ vector $\boldsymbol{\xi}_t^*$. Because of startup irregularities, it is sometimes a good idea to set $\boldsymbol{y}_t^* \equiv \boldsymbol{y}_t$ for the first few values of t, say $t = 1, 2, \ldots, t_0$, where t_0 is small, and to sample from $\{\boldsymbol{e}_{t_0+1}(\widehat{\Theta}), ..., \boldsymbol{e}_n(\widehat{\Theta})\}$. That is, do not bootstrap the first few data points; typically setting t_0 to 4 or 5 will suffice.

4. Using the bootstrap data set $\{\boldsymbol{y}_t^*;\ t = 1, ..., n\}$, construct a likelihood, $L_{Y^*}(\Theta)$, and obtain the MLE of Θ, say, $\widehat{\Theta}^*$.

5. Repeat steps 2 through 4, a large number, B, of times, obtaining a bootstrapped set of parameter estimates $\{\widehat{\Theta}_b^*;\ b = 1, ..., B\}$. The finite sample distribution of $(\widehat{\Theta} - \Theta)$ may be approximated by the distribution of $(\widehat{\Theta}_b^* - \widehat{\Theta})$, for $b = 1, ..., B$.

2.1 Stochastic Regression

An interesting application of the state-space model was given in Newbold and Bos (1985, pp. 61-73). Of the several alternative models they investigate, we focus on the one specified by their equations (4.7a) and (4.7b). Their model has one output variable, the nominal interest rate recorded for three-month treasury bills, y_t. The output equation is specified by

$$y_t = \alpha + \beta_t z_t + v_t,$$

where z_t is the quarterly inflation rate in the Consumer Price Index, α is a fixed constant, β_t is a stochastic regression coefficient, and v_t is white noise with variance σ_v^2. The stochastic regression term, which comprises the state variable, is specified by a first-order autoregression,

$$(\beta_{t+1} - b) = \phi(\beta_t - b) + w_t,$$

where b is a constant, and w_t is white noise with variance σ_w^2. The noise processes, v_t and w_t, are assumed to be uncorrelated.

Using the notation of the state space model (1) and (2), we have in the state equation, $\boldsymbol{x}_t = \beta_t$, $\Phi = \phi$, $\boldsymbol{u}_t \equiv 1$, $\Upsilon = (1 - \phi)b$, $Q = \sigma_w^2$, and in the observation equation, $A_t = z_t$, $\Gamma = \alpha$, $R = \sigma_v^2$, and $S = 0$. The parameter vector is $\Theta = (\phi, \alpha, b, \sigma_w, \sigma_v)'$.

We consider the first estimation exercise reported in Table 4.3 of Newbold and Bos. This exercise covers the period from the first quarter of 1953 through the second quarter of 1965, $n = 50$ observations. We repeat their analysis so our results can be compared to their results. In addition, we focus on this analysis because it demonstrates that the bootstrap applied to the innovation sequence can provide an accurate assessment of the sampling distributions of the parameter estimates when analyzing short time series. Moreover, this analysis demonstrates that a resampling procedure can provide insight into the validity of the model.

The results of the Newton–Raphson estimation procedure are listed in Table 1. The MLEs obtained in Newbold and Bos are in agreement with our values, and differ only in the fourth decimal place; the differences are attributed to the fact that we use a different numerical optimization routine. Included in Table 1 are the asymptotic standard errors reported in Newbold and Bos. Also shown in the Table 1 are the corresponding standard errors obtained from $B = 500$ runs of the bootstrap. These standard errors are simply the square root of $\sum_{b=1}^{B}(\widehat{\Theta}_{ib}^* - \widehat{\Theta}_i)^2/(B-1)$, where Θ_i, represents the ith parameter, $i = 1, ..., 5$, and $\widehat{\Theta}_i$ is the MLE of Θ_i.

The asymptotic standard errors listed in Table 1 are typically smaller than those obtained from the bootstrap. This result is the most pronounced in the estimates of ϕ, σ_w, and σ_v, where the bootstrapped standard errors are about 50% larger than the corresponding asymptotic value. Also, asymptotic theory prescribes the use of normal theory when dealing with the parameter estimates. The bootstrap, however, allows us to investigate the small sample distribution of the estimators and, hence, provides more insight into the data analysis.

For example, Figure 1 shows the bootstrap distribution of the estimator of ϕ. This distribution is highly skewed with values concentrated around 0.8, but with a long tail to the left. Some quantiles

Table 1: Comparison of Asymptotic Standard Errors (SE) and Bootstrapped Standard Errors ($B = 500$).

Parameter	MLE	Asymptotic SE	Newbold & Bos SE	Bootstrap SE
ϕ	0.841	0.200	0.212	0.304
α	-0.771	0.645	0.603	0.645
b	0.858	0.278	0.259	0.277
σ_w	0.127	0.092	NA	0.182
σ_v	1.131	0.142	NA	0.217

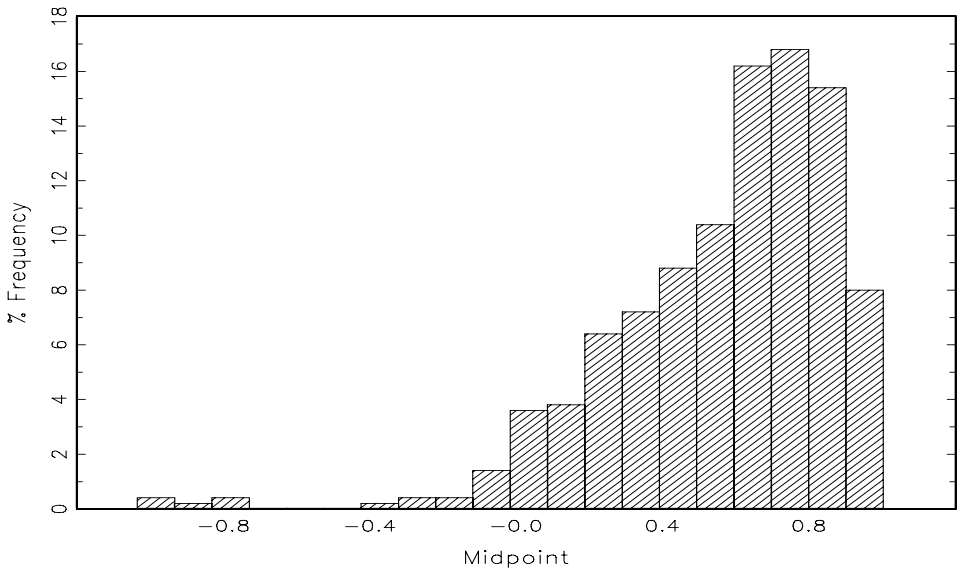

Figure 1: Bootstrap distribution, $B = 500$, of the estimator of ϕ.

of the bootstrapped distribution of ϕ are -0.09 (2.5%), 0.03 (5%), 0.16 (10%), 0.87 (90%), 0.92 (95%), 0.94 (97.5%), and they can be used to obtain confidence intervals. For example, a 90% confidence interval for ϕ would be approximated by (0.03, 0.92). This interval is rather wide, and we will interpret this after we discuss the results of the estimation of σ_w.

Figure 2 shows the bootstrap distribution of the estimator of σ_w. The distribution is concentrated at two locations, one at approximately $\widehat{\sigma}_w^* = 0.15$ and the other at $\widehat{\sigma}_w^* = 0$. The cases in which $\widehat{\sigma}_w^* \approx 0$ correspond to deterministic state dynamics. When $\sigma_w = 0$ and $|\phi| < 1$, then $\beta_t \approx b$ for large t, so the approximately 25% of the cases in which $\widehat{\sigma}_w^* \approx 0$ suggest a fixed state, or constant coefficient model. The cases in which $\widehat{\sigma}_w^*$ is away from zero would suggest a truly stochastic regression parameter. To investigate this matter further, Figure 3 shows the joint bootstrapped estimates, $(\widehat{\phi}^*, \widehat{\sigma}_w^*)$, for non-negative values of $\widehat{\phi}^*$. The joint distribution suggests $\widehat{\sigma}_w^* > 0$ corresponds to $\widehat{\phi}^* \approx 0$. When $\phi = 0$, the state dynamics are given by $\beta_t = b + w_t$. If, in addition, σ_w is small

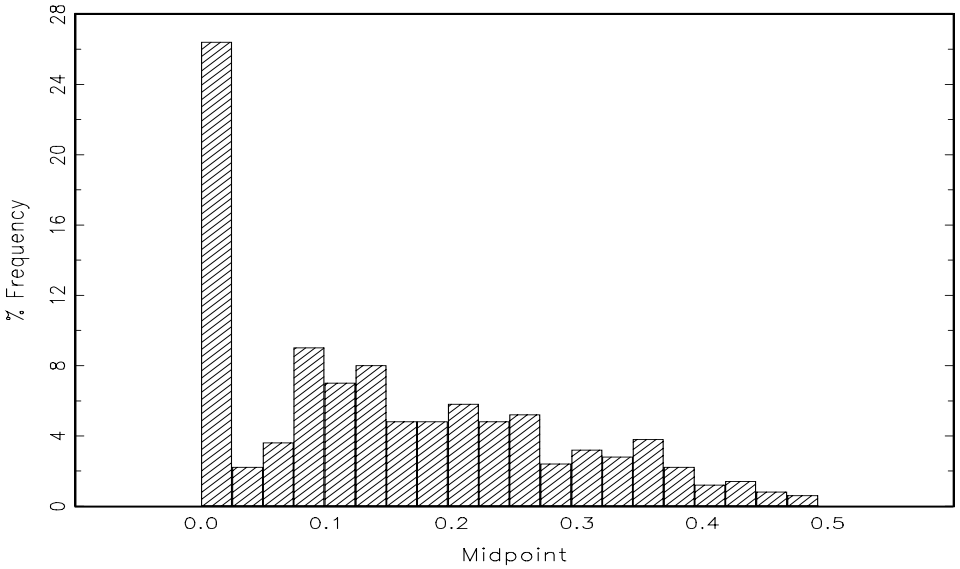

Figure 2: Bootstrap distribution, $B = 500$, of the estimator of σ_w.

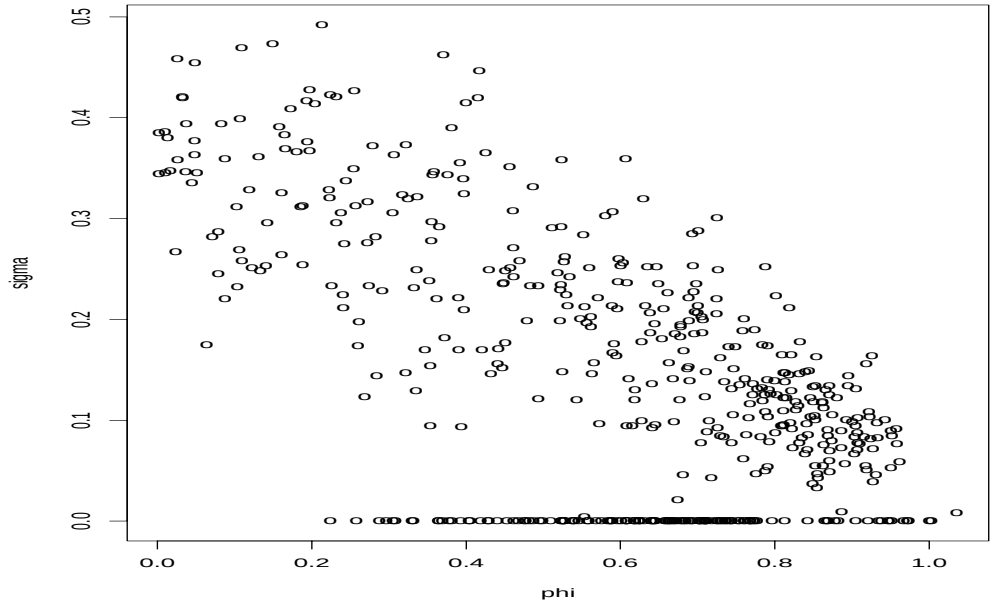

Figure 3: Joint bootstrap distribution, $B = 500$, of the estimators of ϕ and σ_w. Only the values corresponding to $\widehat{\phi}^* \geq 0$ are shown.

relative to b (as it appears to be in this case), the system is nearly deterministic; that is, $\beta_t \approx b$. Considering these results, the bootstrap analysis leads us to conclude the dynamics of the data are best described in terms of a *fixed*, rather than stochastic, regression effect.

If, however, we use the same model for the entire data set presented in Newbold and Bos (that is, 110 quarters of three-month treasury bills and inflation rate, covering 1953:I to 1980:II), stochastic regression appears to be appropriate. In this case the estimates using Newton-Raphson with estimated standard errors ("asymptotic" | "bootstrap") are:

$$\widehat{\phi} = 0.896 \ (0.067 \mid 0.274), \qquad \widehat{\alpha} = -0.970 \ (0.475 \mid 0.538), \qquad \widehat{b} = 1.090 \ (0.158 \mid 0.221),$$

$$\widehat{\sigma}_w = 0.117 \ (0.037 \mid 0.122), \qquad \widehat{\sigma}_v = 1.191 \ (0.108 \mid 0.171).$$

We note that the asymptotic standard error estimates are still too small, and the bootstrapped distribution of $\widehat{\phi}$ is still markedly skewed. In particular, a 90% bootstrap confidence interval for ϕ is (.46, .92).

2.2 Stochastic Volatility

This problem is somewhat different than the previous section in that it is not a straight-forward application of the algorithm. In this example, we consider the stochastic volatility model due to Harvey, Ruiz and Shephard (1994). Let r_t denote the return or growth rate of a process of interest. For example, if s_t is the value of a stock at time t, the return or relative gain of the stock is $r_t = \ln(s_t/s_{t-1})$. Typically, it is $\text{var}(r_t) = \sigma_t^2$ that is of interest. In the stochastic volatility model, we model $h_t = \ln \sigma_t^2$ as an AR(1), that is,

$$h_{t+1} = \phi_0 + \phi_1 h_t + w_t, \tag{12}$$

where w_t is white Gaussian noise with variance σ_w^2; this comprises the state equation. The observations are taken to be $y_t = \ln r_t^2$, and y_t is related to the state via

$$y_t = \alpha + h_t + v_t. \tag{13}$$

Together, (12) and (13) make up the stochastic volatility model, where h_t represents the unobserved volatility of the process y_t. If v_t was Gaussian white noise, (12)–(13) would form a Gaussian state space model, and we could then use standard results to fit the model to data. Unfortunately, $y_t = \ln r_t^2$ is rarely normal, so one typically assumes that $v_t = \ln z_t^2$ where z_t is standard Gaussian white noise. In this case, $\ln z_t^2$ is distributed as the log of a chi-squared random variable with one degree of freedom. Kim, Shephard and Chib (1998) proposed modeling the log of a chi-squared random variable by a mixture of normals.

Various approaches to the fitting of stochastic volatility models have been examined; these methods include a wide range of assumptions on the observational noise process. A good summary of the proposed techniques, both Bayesian (via MCMC) and non-Bayesian approaches (such as quasi-maximum likelihood estimation and the EM algorithm), can be found in Jacquier et al (1994), and Shephard (1996). Simulation methods for classical inference applied to stochastic volatility models are discussed in Danielson (1994) and Sandmann and Koopman (1998).

In an effort to keep matters simple, our method (see Shumway and Stoffer, 2000, §4.10) of fitting stochastic volatility models is to retain the Gaussian state equation (12), but in the observation equation (13), we consider v_t to be white noise, and distributed as a mixture of two normals, one centered at zero. In particular, we write

$$v_t = (1 - \eta_t) z_{t0} + \eta_t z_{t1}, \tag{14}$$

where η_t is an iid Bernoulli process, $\Pr\{\eta_t = 0\} = \pi_0$, $\Pr\{\eta_t = 1\} = \pi_1$, with $\pi_0 + \pi_1 = 1$, and where $z_{t0} \sim$ iid $\mathrm{N}(0, \sigma_0^2)$, and $z_{t1} \sim$ iid $\mathrm{N}(\mu_1, \sigma_1^2)$.

The advantage of this model is that it is fairly easy to fit because it uses normality. The model specified by equations (12)–(14), and the corresponding filter, are similar to those presented in Peña and Guttman (1988), who used the idea to obtain a robust Kalman filter, and, as previously mentioned, Kim, Shephard and Chib (1998). In addition, this technique is similar to technique discussed in Shumway and Stoffer (2000, §4.8). In particular, the filtering equations for this model are:

$$
\begin{aligned}
h_{t+1}^t &= \phi_0 + \phi_1 h_t^{t-1} + \sum_{j=0}^{1} \pi_{tj} K_{tj} \epsilon_{tj}, && (15)\\
P_{t+1}^t &= \phi_1^2 P_t^{t-1} + \sigma_w^2 - \sum_{j=0}^{1} \pi_{tj} K_{tj}^2 \Sigma_{tj}, && (16)\\
\epsilon_{t0} &= y_t - \alpha - h_t^{t-1}, && (17)\\
\epsilon_{t1} &= y_t - \alpha - h_t^{t-1} - \mu_1, && (18)\\
\Sigma_{t0} &= P_t^{t-1} + \sigma_0^2, && (19)\\
\Sigma_{t1} &= P_t^{t-1} + \sigma_1^2, && (20)\\
K_{t0} &= \phi_1 P_t^{t-1} \,/\, \Sigma_{t0}, && (21)\\
K_{t1} &= \phi_1 P_t^{t-1} \,/\, \Sigma_{t1}. && (22)
\end{aligned}
$$

To complete the filtering, we must be able to assess the probabilities $\pi_{t1} = \Pr(\eta_t = 1 \mid y_1, \ldots, y_t)$, for $t = 1, \ldots, n$; of course, $\pi_{t0} = 1 - \pi_{t1}$. Let $f_j(t \mid t-1)$ denote the conditional density of y_t given the past $y_1, ..., y_{t-1}$, and $\eta_t = j$ $(j = 0, 1)$. Then,

$$
\pi_{t1} = \frac{\pi_1 f_1(t \mid t-1)}{\pi_0 f_0(t \mid t-1) + \pi_1 f_1(t \mid t-1)}, \qquad (23)
$$

where we assume the distribution π_j, for $j = 0, 1$ has been specified *a priori*. If the investigator has no reason to prefer one state over another the choice of uniform priors, $\pi_1 = 1/2$, will suffice. Unfortunately, it is computationally difficult to obtain the exact values of $f_j(t \mid t-1)$; although we can give an explicit expression of $f_j(t \mid t-1)$, the actual computation of the conditional density is prohibitive. A viable approximation, however, is to choose $f_j(t \mid t-1)$ to be the normal density, $\mathrm{N}(h_t^{t-1} + \mu_j,\, \Sigma_{tj})$, for $j = 0, 1$ and $\mu_0 = 0$; see Shumway and Stoffer (2000, §4.8) for details.

The innovations filter given in (15)–(23) can be derived from the Kalman filter by a simple conditioning argument. For example, to derive (15), we write

$$
\begin{aligned}
\mathrm{E}\left(h_{t+1} \mid y_1, \ldots, y_t\right) &= \sum_{j=0}^{1} \mathrm{E}\left(h_{t+1} \mid y_1, \ldots, y_t, \eta_t = j\right) \Pr(\eta_t = j \mid y_1, \ldots, y_t)\\
&= \sum_{j=0}^{1} \left(\phi_0 + \phi_1 h_t^{t-1} + K_{tj}\epsilon_{tj}\right) \pi_{tj}\\
&= \phi_0 + \phi_1 h_t^{t-1} + \sum_{j=0}^{1} \pi_{tj} K_{tj} \epsilon_{tj}.
\end{aligned}
$$

Estimation of the parameters, $\Theta = (\phi_0, \phi_1, \sigma_0^2, \mu_1, \sigma_1^2, \sigma_w^2)'$, is accomplished via MLE based on the likelihood given by

$$\ln L_Y(\Theta) = \sum_{t=1}^{n} \ln \left(\sum_{j=0}^{1} \pi_j \ f_j(t \mid t-1) \right), \tag{24}$$

where the densities for $f_j(t \mid t-1)$ are approximated by the normal densities previously mentioned.

To perform the bootstrap, we develop a vector first-order equation, as was done in (9). First, using (17)–(18), and noting that $y_t = \pi_{t0} y_t + \pi_{t1} y_t$, we may write

$$y_t = \alpha + h_t^{t-1} + \pi_{t0}\epsilon_{t0} + \pi_{t1}(\epsilon_{t1} + \mu_1). \tag{25}$$

Consider the standardized innovations

$$e_{tj} = \Sigma_{tj}^{-1/2} \epsilon_{tj}, \qquad j = 0, 1, \tag{26}$$

and define the 2×1 vector

$$\boldsymbol{e}_t = \begin{bmatrix} e_{t0} \\ e_{t1} \end{bmatrix}.$$

Also, define the 2×1 vector

$$\boldsymbol{\xi}_t = \begin{bmatrix} h_{t+1}^t \\ y_t \end{bmatrix}.$$

Combining (15) and (25) results in a vector first-order equation for $\boldsymbol{\xi}_t$ given by

$$\boldsymbol{\xi}_t = F\boldsymbol{\xi}_{t-1} + G_t + H_t \boldsymbol{e}_t, \tag{27}$$

where

$$F = \begin{bmatrix} \phi_1 & 0 \\ 1 & 0 \end{bmatrix}, \quad G_t = \begin{bmatrix} \phi_0 \\ \alpha + \pi_{t1}\mu_1 \end{bmatrix}, \quad H_t = \begin{bmatrix} \pi_{t0} K_{t0} \Sigma_{t0}^{1/2} & \pi_{t1} K_{t1} \Sigma_{t1}^{1/2} \\ \pi_{t0} \Sigma_{t0}^{1/2} & \pi_{t1} \Sigma_{t1}^{1/2} \end{bmatrix}.$$

Hence, the steps in bootstrapping for this case are the same as steps 1 through 5 previously described, but with (11) replaced by the following first-order equation:

$$\boldsymbol{\xi}_t^* = F(\widehat{\Theta})\boldsymbol{\xi}_{t-1}^* + G_t(\widehat{\Theta}; \widehat{\pi}_{t1}) + H_t(\widehat{\Theta}; \widehat{\pi}_{t1})\boldsymbol{e}_t^*, \tag{28}$$

where $\widehat{\Theta} = (\widehat{\phi}_0, \widehat{\phi}_1, \widehat{\sigma}_0^2, \widehat{\alpha}, \widehat{\mu}_1, \widehat{\sigma}_1^2, \widehat{\sigma}_w^2)'$ is the MLE of Θ, and $\widehat{\pi}_{t1}$ is estimated via (23), replacing $f_1(t \mid t-1)$ and $f_0(t \mid t-1)$ by their respective estimated normal densities ($\widehat{\pi}_{t0} = 1 - \widehat{\pi}_{t1}$).

To examine the efficacy of the bootstrap for the stochastic volatility model, we generated $n = 200$ observations from the following stochastic volatility model:

$$h_t = .95 h_{t-1} + w_t, \tag{29}$$

where w_t is white Gaussian noise with variance $\sigma_w^2 = 1$. The observations were then generated as

$$y_t = h_t + v_t, \tag{30}$$

where the observational white noise process, v_t, is distributed as the log of a chi-squared random variable with one degree of freedom. The density of v_t is given by

$$f_v(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(e^x - x\right) \right\} \quad -\infty < x < \infty, \tag{31}$$

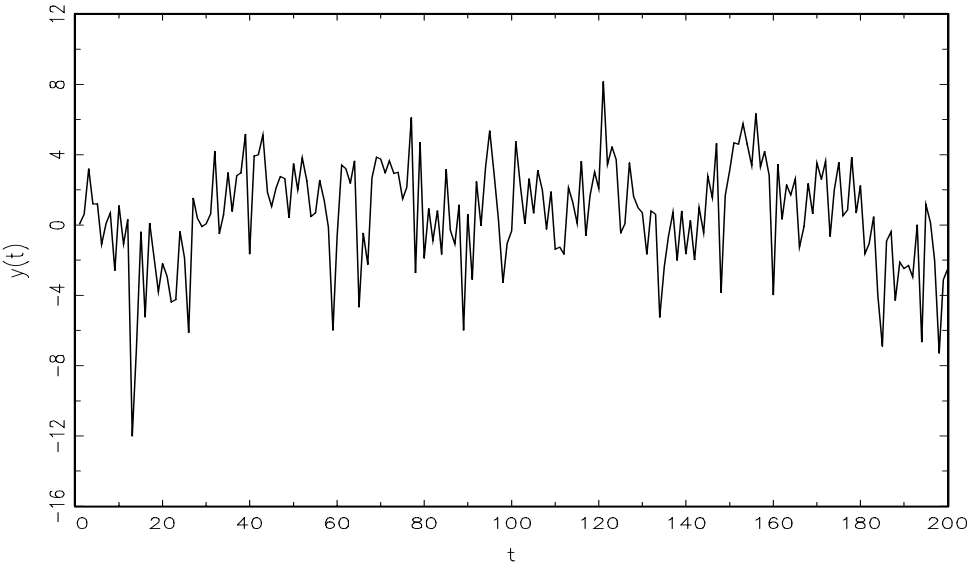

Figure 4: Simulated data, $n = 200$, from the stochastic volatility model (29)–(30).

and its mean and variance are -1.27 and $\pi^2/2$, respectively; the density (31) is highly skewed with a long tail on the left. The data are shown in Figure 4. Then, we assumed the true error distribution was unknown to us, and we fit the model (12)–(14) using the Gauss BFGS variable metric algorithm to maximize the likelihood. The results for the state parameters are given in Table 2 in the columns marked *MLE* and *Asymptotic SE*. Next, we bootstrapped the data, $B = 500$ times, using the incorrect model (12)–(14) to assess the finite sample standard errors (SE). The results are listed in Table 2 in the column marked *Bootstrap SE*. Finally, using the correct model, (29)–(30), we simulated 500 processes, estimated the parameters based on the model (12)–(14) also via a BFGS variable metric algorithm, and assessed the SEs of the estimates of the actual state parameters. These values are listed in Table 2 in the column labeled *"True" SE*.

Table 2: Stochastic Volatility Simulation Results.

State Parameter	Actual Value	MLE	Asymptotic SE	Bootstrap SE†	"True" SE‡
ϕ	0.95	0.963	0.032	0.032	0.036
σ_w	1	1.042	0.279	0.215	0.252

† Based on 500 bootstrapped samples. ‡ Based on 500 replications.

In Table 2 we notice that the bootstap SE and the asymptotic SE of ϕ are about the same; also, both estimates are slightly smaller than the "true" value. The interest here, however, is not so much in the SEs, but in the actual sampling distribution of the estimates. To explore the finite sample distribution of the estimate of ϕ, Figure 5 shows the centered bootstrap histogram: $(\widehat{\phi}^*_b - \widehat{\phi})$, for $b = 1, \ldots, 500$ bootstrapped replications [the bars are filled with lines of positive slope], the centered

"true" histogram: $(\widehat{\phi}_j - \phi)$, where $\widehat{\phi}_j$ is the MLE obtained on the j-th iteration, for $j = 1, \ldots, 500$ Monte Carlo replications [the bars are filled with flat lines], and the centered asymptotic normal distribution of $(\widehat{\phi} - \phi)$ [appropriately scaled for comparison with the histograms], superimposed on eachother. Clearly, the bootstrap distribution is closer to the "true" distribution than the estimated asymptotic normal distribution; the bootstrap distribution captures the positive kurtosis (peakedness) and asymmetry of the "true" distribution.

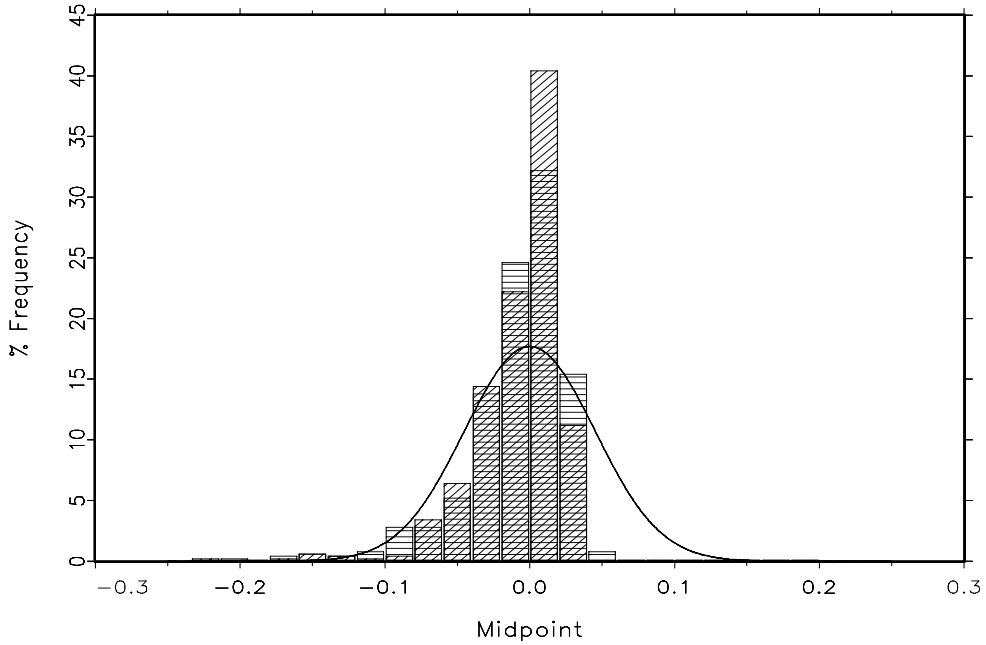

Figure 5: Sampling distributions of the estimate of ϕ; simulated data example: The centered bootstrap histogram (lines with positive slope), the centered "true" histogram (flat lines), and the centered asymptotic normal distribution.

In an example using actual data, we consider the analysis of quarterly U.S. GNP from 1947(1) to 2002(3), $n = 223$. The data are seasonally adjusted and were obtained from the Federal Reserve Bank of St. Louis (`http://research.stlouisfed.org/fred/data/gdp/gnpc96`). The growth rate is plotted in Figure 6 and appears to be a stable process. Analysis of the data indicates the growth rate is an MA(2) [for more details of this part of the analysis, see Shumway and Stoffer, 2000, §2.8], however, the residuals of that fit, which appear to be white, suggest that there is volatility.

Figure 7 shows the log of the squared residuals, say y_t, from the MA(2) fit on the U.S. GNP series. The stochastic volatility model (12)–(14) was then fit to y_t. Table 3 shows the MLEs of the model parameters along with their asymptotic SEs assuming the model is correct. Also displayed in Table 3 are the means and SEs of $B = 500$ bootstrapped samples. As in the simulation, there is some amount of agreement between the asymptotic values and the bootstrapped values. Based on the previous simulation, we would be more prone to focus on the actual sampling distributions, rather than assume normality. For example, Figure 8 compares the bootstrap histogram and asymptotic

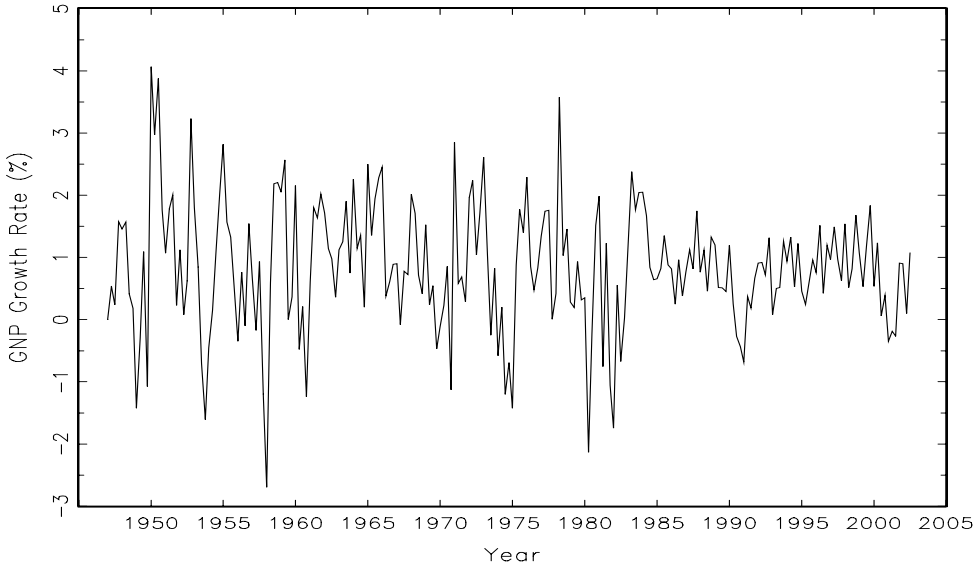

Figure 6: U.S. GNP quarterly growth rate.

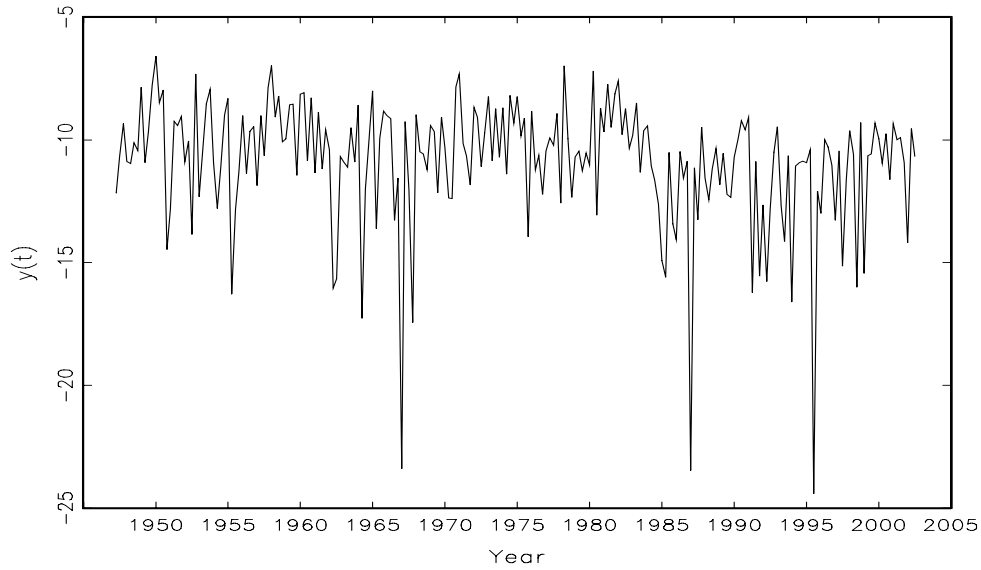

Figure 7: Log of the squared residuals from an MA(2) fit on GNP growth rate.

normal distribution of $\widehat{\phi}_1$. In this case, as in the simulation, the bootstrap distribution exhibits positive kurtosis and skewness which is missed by the assumption of asymptotic normality. Based on the simulation, we would be prone to believe the results of the bootstrap are fairly accurate.

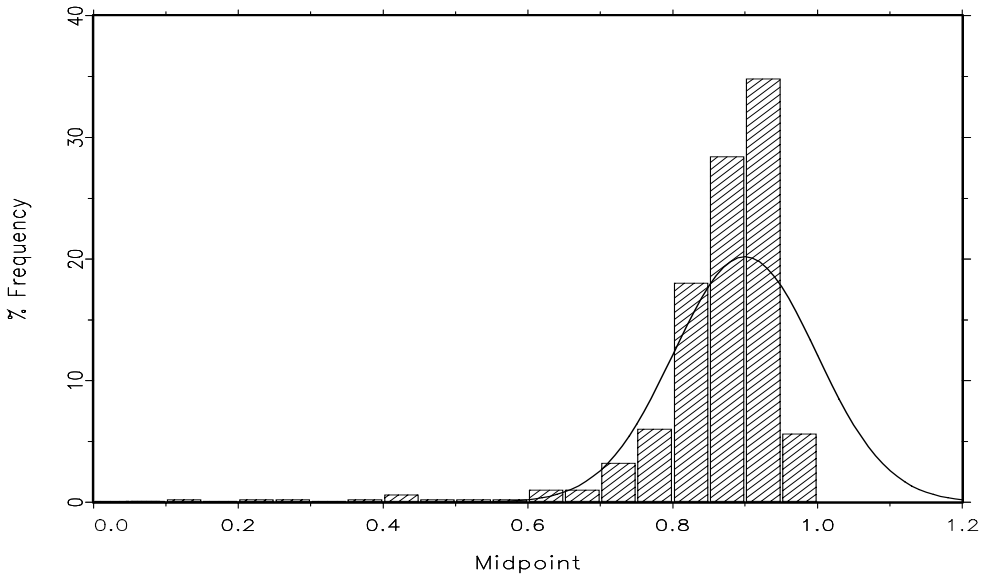

Figure 8: Bootstrap histogram and asymptotic distribution of $\widehat{\phi}_1$ for the US GNP example.

Table 3: Estimates and Their Asymptotic and Bootstrap Standard Errors for US GNP Example.

Parameter	MLE	Asymptotic SE	Bootstrap Mean†	Bootstrap SE†
ϕ_0	0.068	0.274	−0.010	0.353
ϕ_1	0.900	0.099	0.864	0.102
σ_w	0.378	0.208	0.696	0.375
α	−10.524	2.321	−10.792	0.748
μ_1	−2.164	0.567	−1.941	0.416
σ_1	3.007	0.377	2.891	0.422
σ_0	0.935	0.198	0.692	0.362

† Based on 500 bootstrapped samples.

3 Assessing the Finite Sample Distribution of Conditional Forecasts

In this section we focus on assessing the conditional forecast accuracy of time series models using a state space approach and resampling methods. Our work is motivated by the following considerations. First, the state space model provides a convenient unifying representation for various models, including ARMA(p, q) models. Second, the actual practice of forecasting involves the prediction of a future point based 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 other processes.

Early application of the bootstrap to assess conditional forecast errors can be 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 discussed this problem and Breidt, Davis and Dunsmuir (1992, 1995) offered 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 for other time series models. We suspect 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 other time series models this may not be 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. In §2, such a combination was 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 presentation is based on the work of Wall and Stoffer (2002).

3.1 Generating Reverse Time Datasets

As seen in §2, the generation of bootstrap data sets in forward time is easy. Given an initial condition or prior, (11) is solved recursively for $t = 1, \ldots, n$ to produce realizations passing through the given initial condition. Such computations are all that is required in obtaining bootstrap estimates of parameter estimation error statistics or unconditional forecast error statistics. The generation of bootstrap data sets for assessing *conditional* forecast errors is not so straight forward 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, Davis and Dunsmuir (1992, 1995). This literature stresses the need to properly construct a set of "backward" residuals and Breidt, Davis and Dunsmuir (1992, 1995) 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, (9); recall

$$\boldsymbol{\xi}_t = F_t \boldsymbol{\xi}_{t-1} + G\boldsymbol{u}_t + H_t \boldsymbol{e}_t, \tag{9}$$

where

$$\boldsymbol{\xi}_t = \begin{bmatrix} \boldsymbol{x}_{t+1}^t \\ \boldsymbol{y}_t \end{bmatrix}, \quad F_t = \begin{bmatrix} \Phi & 0 \\ A_t & 0 \end{bmatrix}, \quad G = \begin{bmatrix} \Upsilon \\ \Gamma \end{bmatrix}, \quad H_t = \begin{bmatrix} K_t \Sigma_t^{1/2} \\ \Sigma_t^{1/2} \end{bmatrix}.$$

We require a backward-time representation of this system. All the problems highlighted by Findley (1986) and Breidt, Davis and Dunsmuir (1992, 1995) appear here. For example, the first p rows of (9) cannot be solved backwards in time by simply expressing $\boldsymbol{x}_t^{t-1}$ in terms of $\boldsymbol{x}_{t+1}^t$. First, Φ is not always invertible; e.g., MA(q) models. Second, even when Φ is invertible, Φ^{-1} has characteristic roots outside the unit circle whenever Φ 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 $\boldsymbol{\xi}_t$. In addition, we now have a time-varying system.

These difficulties are overcome by building on the method found in Caines (1988, pp. 236-237). 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 $\boldsymbol{u}_t$. For ease, we will assume here that $\boldsymbol{u}_t \equiv \mathbf{0}$; the general case is presented in Wall and Stoffer (2002). Application of the symmetry of minimal splitting subspaces yields the following reverse-time state space representation for $t = n-1, n-2, \ldots, 1$:

$$\boldsymbol{r}_t = \Phi' \boldsymbol{r}_{t+1} + B_t \boldsymbol{x}_t^{t-1} - C_t \boldsymbol{e}_t$$

$$\boldsymbol{y}_t = N_t \boldsymbol{r}_{t+1} - L_t \boldsymbol{x}_t^{t-1} + M_t \boldsymbol{e}_t$$

where

$$\begin{aligned} B_t &= V_t^{-1} - \Phi' V_{t+1}^{-1} \Phi \\ C_t &= \Phi' V_{t+1}^{-1} K_t \Sigma_t^{-1/2}, \\ D_t &= I - \Sigma_t^{-1/2} K_t' V_{t+1}^{-1} K_t \Sigma_t^{-1/2}, \\ L_t &= \Sigma_t^{-1/2} C_t' - A_t V_t B_t, \\ M_t &= \Sigma_t^{-1/2} D_t - A_t V_t C_t, \\ N_t &= A_t V_t \Phi' + \Sigma_t^{-1} K_t', \end{aligned}$$

and

$$V_{t+1} = \Phi V_t \Phi' + K_t \Sigma_t^{-1} K_t'.$$

The reverse-time state vector is $\boldsymbol{r}_t$. The backward recursion is initialized by $\boldsymbol{r}_n = V_n^{-1} \boldsymbol{x}_n^{n-1}$. Details of the derivation are given in Wall and Stoffer (2002).

The above recursion specifies a three step procedure for the generation of backward time data sets (written here for $\boldsymbol{u}_t \equiv \mathbf{0}$):

1. Generate $V_t, B_t, C_t, D_t, L_t, M_t$ and N_t *forwards* in time, $t = 1, \ldots, n$, with initial condition

$$V_1 = P_1^0 \tag{32}$$

2. For given $\{\boldsymbol{e}_t^*;\ 1 \le t \le n-1\}$, set $\boldsymbol{x}_1^* = \mathbf{0}$ and generate $\{\boldsymbol{x}_t^*;\ 1 \le t \le n\}$ *forwards* in time, $t = 1, \ldots, n$, via

$$\boldsymbol{x}_{t+1}^* = \Phi \boldsymbol{x}_t^* + K_t \Sigma_t^{1/2} \boldsymbol{e}_t^* \tag{33}$$

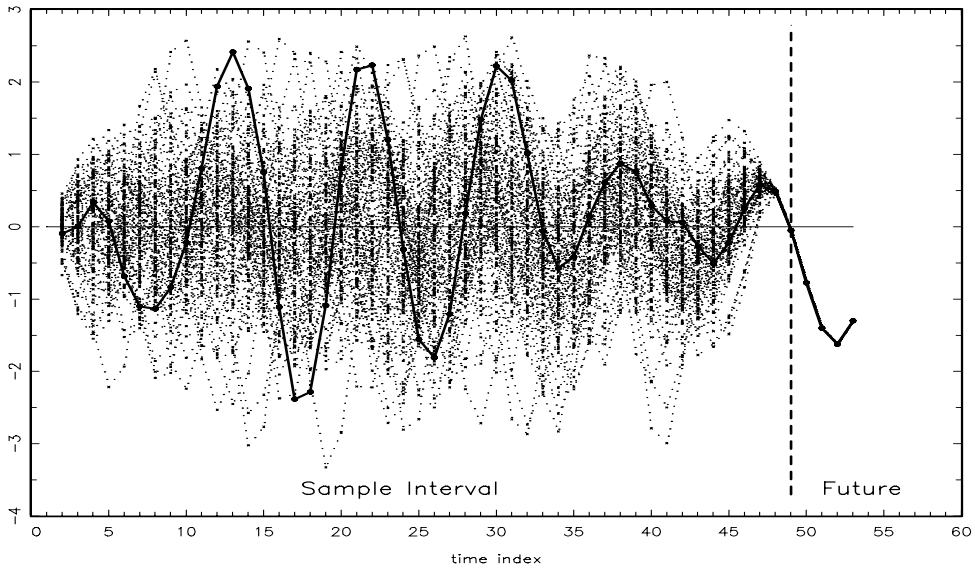

Figure 9: Reverse time realizations of the ARMA$(2,1)$ process given in (36).

3. Set $\boldsymbol{r}_n^* = \boldsymbol{r}_n = V_n^{-1}\boldsymbol{x}_n^{n-1}$ and generate $\{\boldsymbol{y}_t^*;\ 1 \le t \le n\}$ *backwards* in time, $t = n-1, n-2 \ldots, 1$, via the reverse time state space model

$$\boldsymbol{r}_t^* = \Phi' \boldsymbol{r}_{t+1}^* + B_t \boldsymbol{x}_t^* - C_t \boldsymbol{e}_t^* \tag{34}$$

$$\boldsymbol{y}_t^* = N_t \boldsymbol{r}_{t+1} - L_t \boldsymbol{x}_t^* + M_t \boldsymbol{e}_t^* \tag{35}$$

This procedure assumes one already has drawn randomly, with replacement, from the model estimated standardized residuals to obtain a set of $n-1$ residuals denoted $\{\boldsymbol{e}_t^*;\ 1 \le t \le n-1\}$. The last residual is kept set at $\boldsymbol{e}_n^* = \boldsymbol{e}_n$ in order to ensure the conditioning requirement is met on $\boldsymbol{\xi}_n^*$; that is, $\boldsymbol{\xi}_n^* = \boldsymbol{\xi}_n$. This requirement follows from the autoregressive structure of (9); for details, see §4.2. The creation of an arbitrary number of bootstrap data sets is accomplished by repeating the above for each set of bootstrap residuals $\{\boldsymbol{e}_t^*;\ 1 \le t \le n-1;\ \boldsymbol{e}_n^* = \boldsymbol{e}_n\}$.

As an example, consider the univariate ARMA(p,q) process given by

$$y_t + a_1 y_{t-1} + \cdots + a_p y_{t-p} = v_t + b_1 v_{t-1} + \cdots + b_q v_{t-q}$$

where v_t is an i.i.d. process with variance σ_v^2. This process can be represented in state space form, (1)–(2), in various ways. For example, let $m = \max\{p,q\}$, let $\boldsymbol{x}_t$ be an m-dimensional state vector, and write the state space coefficient matrices as

$$\Phi = \begin{bmatrix} 0 & \cdots & 0 & 0 & 0 & -a_m \\ 1 & \cdots & 0 & 0 & 0 & -a_{m-1} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 & -a_3 \\ 0 & \cdots & 0 & 1 & 0 & -a_2 \\ 0 & \cdots & 0 & 0 & 1 & -a_1 \end{bmatrix},$$

$$A \;=\; \left[\begin{array}{ccccccc} 0 & \cdots & 0 & 0 & 0 & & 1 \end{array}\right],$$

$\Upsilon = 0$ and $\Gamma = 0$. The state noise process is defined by the $\boldsymbol{w}_t = \boldsymbol{g}\, v_t$ where

$$\boldsymbol{g} = \left[\begin{array}{ccccccc} b_m - a_m, & b_{m-1} - a_{m-1}, & \cdots & b_3 - a_3, & b_2 - a_2, & b_1 - a_1 \end{array}\right]'.$$

If $m > p$ then $a_\ell = 0$ for $\ell > p$, and if $m > q$ then $b_\ell = 0$ for $\ell > q$. The variance-covariance matrices are given by

$$Q = \sigma_v^2\, \boldsymbol{g}\, \boldsymbol{g}', \quad R = \sigma_v^2, \quad S = \sigma_v^2\, \boldsymbol{g}.$$

Figure 9 presents a sample of 100 reverse-time trajectories for the Gaussian ARMA$(2,1)$ model

$$y_t = 1.4 y_{t-1} - 0.85 y_{t-2} + v_t + 0.6 v_{t-1}, \tag{36}$$

with $\sigma_v = 0.2$ and $n = 49$. The original, observed sample is plotted with the bold line.

3.2 Computing Forecast Errors via the Bootstrap

At this point we assume we have n observations, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$, and we wish to forecast m time points into the future. In addition, we have the MLEs of the model parameters Θ, say $\widehat{\Theta}$, based on the data. The associated standardized innovation values are denoted by $\{\boldsymbol{e}_t(\widehat{\Theta});\ 1 \le t \le n\}$; note, to avoid any possible confusion, we emphasize the dependence of the values on the parameters. For $b = 1, 2, \ldots, B$ (where B is the number of bootstrap replications) we execute the following six steps:

1. Construct a sequence of $n + m$ standardized residuals

   $$\{\boldsymbol{e}_t^b(\widehat{\Theta});\ 1 \le t \le n + m\}$$

   via $n + m - 1$ random draws, with replacement, from the standardized residuals $\{\boldsymbol{e}_t(\widehat{\Theta});\ 1 \le t \le n\}$. This sequence is formed as follows: (i) use $n-1$ vectors to form $\{\boldsymbol{e}_t^b(\widehat{\Theta});\ 1 \le t \le n-1\}$; (ii) fix $\boldsymbol{e}_n^b(\widehat{\Theta}) = \boldsymbol{e}_n(\widehat{\Theta})$; and (iii) use the remaining m vectors to form $\{\boldsymbol{e}_t^b(\widehat{\Theta});\ n+1 \le t \le n+m\}$.

2. Generate data

   $$\{\boldsymbol{y}_t^b(\widehat{\Theta});\ 1 \le t \le n-1\}$$

   via the backward state space model (34) and (35) with $\Theta = \widehat{\Theta}$ using the residuals $\{\boldsymbol{e}_t^b(\widehat{\Theta});\ 1 \le t \le n-1\}$. Set $\boldsymbol{y}_n^b(\widehat{\Theta}) = \boldsymbol{y}_n$.

3. Generate data

   $$\{\boldsymbol{y}_t^b(\widehat{\Theta});\ n+1 \le t \le m+n\}$$

   via the forward state space model (9) with $\Theta = \widehat{\Theta}$ and with $\boldsymbol{x}_t^{t-1;b} = \boldsymbol{x}_t^{t-1}(\widehat{\Theta})$ and using the residuals $\boldsymbol{e}_t^b(\widehat{\Theta})$, for $n + 1 \le t \le n + m$.

4. Compute model parameter estimates Θ^b via MLE using the data $\{\boldsymbol{y}_t^b(\widehat{\Theta});\ 1 \le t \le n\}$.

5. Compute the bootstrap conditional forecasts

$$\{\boldsymbol{y}_t^b(\Theta^b);\ n+1 \le t \le m+n\}$$

via the forward time state space model (9) with $\Theta = \Theta^b$, and with $\boldsymbol{x}_t^{t-1;b} = \boldsymbol{x}_t^{t-1}(\Theta^b)$ and $\boldsymbol{e}_t^b = \mathbf{0}$ for $n+1 \le t \le n+m$.

6. Compute the bootstrap conditional forecast errors via:

$$\boldsymbol{\delta}_\ell^b = \boldsymbol{y}_{n+\ell}^b(\widehat{\Theta}) - \boldsymbol{y}_{n+\ell}^b(\Theta^b);\ 1 \le \ell \le m.$$

The extent to which the bootstrap captures the behavior of the actual forecast errors derives from the extent to which these errors mimic the stochastic process $\boldsymbol{\delta}_\ell = \boldsymbol{y}_{n+\ell}(\Theta) - \boldsymbol{y}_{n+\ell}(\widehat{\Theta});\ 1 \le \ell \le m$.

As an example, consider the univariate ARMA$(1,1)$ process given by

$$y_t = 0.7y_{t-1} + v_t + 0.10v_{t-1} \tag{37}$$

where $v_t = 0.2z_t$ and z_t is a mixture of 90% N$(\mu = -1/9, \sigma = .15)$ and 10% N$(\mu = 1, \sigma = .15)$. To demonstrate the benefits of resampling, we will assume that we do not know the true distribution of v_t and will act as if it was normal. The model is first-order with

$$\Phi = [0.70] \quad A = [1] \quad \text{and} \quad \boldsymbol{g} = [0.80]\,,$$

in the notation of previous example.

In this simulation we use $B = 2,000$ and $m = 4$. The approximate "true" distribution is then given by the relative frequency histogram of the observed conditional forecast errors. The results of the simulation is summarized by two sets of four histograms. One set (Figure 10) presents the approximate "true" relative frequency histograms for each forecast lead time, while the other set (Figure 11) presents the relative frequency histograms obtained from application of the bootstrap. Superimposed on each is the Gaussian density that follows from application of the asymptotic Gaussian theory. The simulation uses a short data set with $n = 49$ to emphasize the efficacy of the bootstrap when the use asymptotics is questionable and where bias is a factor in the forecasts. Prediction intervals follow immediately from the data summarized 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 (see Efron, 1987).

Figures 10 and 11 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.

3.3 Stochastic Regression

We now illustrate the use of the bootstrap in assessing forecast errors in the data set analyzed in §2.1. Recall, the treasury bill interest rate is modeled as being linearly related to quarterly inflation as

$$y_t = \alpha + \beta_t z_t + v_t,$$

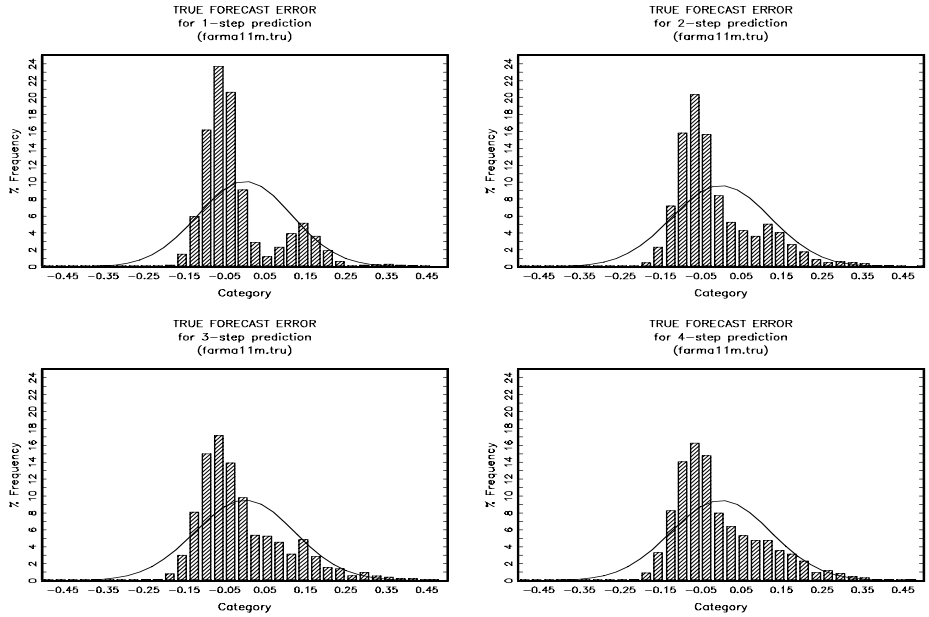

Figure 10: "True" forecast histograms for the ARMA$(1,1)$ process given in (37).

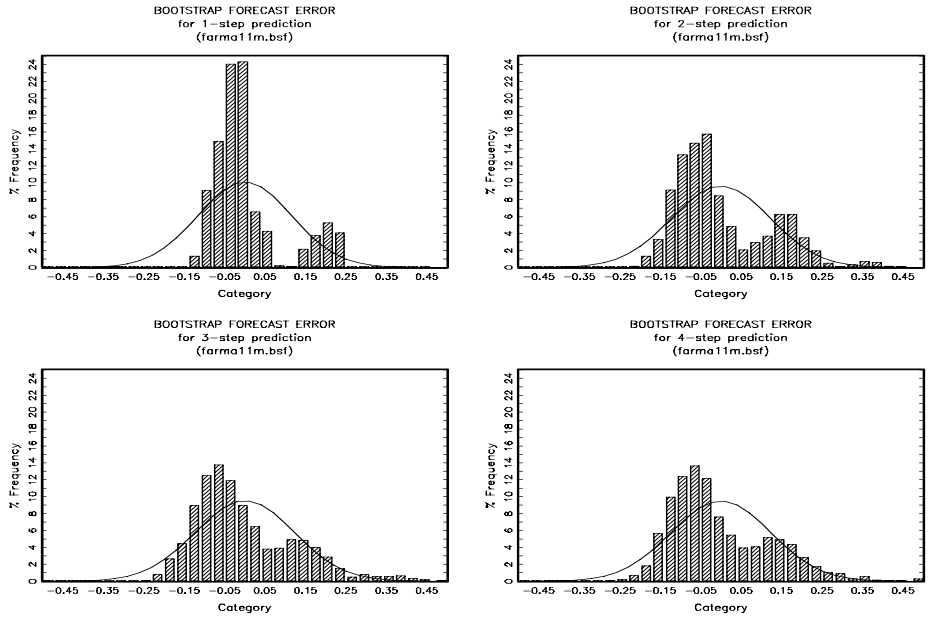

Figure 11: Bootstrap forecast histograms for the ARMA$(1,1)$ process given in (37).

where α is a fixed constant, β_t is a stochastic regression coefficient, and v_t is white noise with variance σ_v^2. The stochastic regression term, which comprises the state variable, is specified by a first-order autoregression,

$$(\beta_t - b) = \phi(\beta_{t-1} - b) + w_t,$$

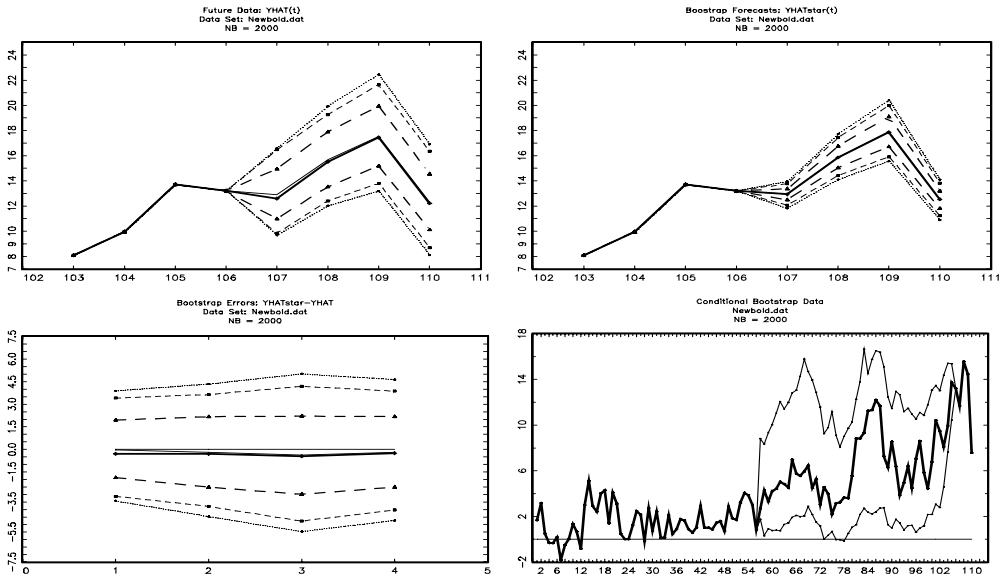

Figure 12: Dymanic behavior of the quantiles of $y^b_{n+\ell}(\widehat{\Theta})$ [upper left panel], the quantiles of $y^b_{n+\ell}(\Theta^b)$ [upper right panel], and the quantiles of the bootstrap conditional forecast errors $y^b_{n+\ell}(\widehat{\Theta}) - y^b_{n+\ell}(\Theta^b)$, for $\ell = 1, 2, 3, 4$ [lower left panel]. The y_t series [lower right panel] as a bold line and the envelope of the backward data series as fine lines above and below the observed sample in the stochastic regression example.

where b is a constant, and w_t is white noise with variance σ_w^2. The noise processes, v_t and w_t, are assumed to be uncorrelated.

The model parameter vector contains five elements, $\Theta = (\phi, \alpha, b, \sigma_w, \sigma_v)'$ and is estimated via Gaussian quasi-maximum likelihood using data from the first quarter of 1967 through the second quarter of 1979 (49 observations). The MLEs and their estimated standard errors (in parentheses) were:

$$\widehat{\phi} = 0.898_{\,(0.101)} \quad \widehat{\alpha} = -0.615_{\,(1.457)} \quad \widehat{b} = 1.195_{\,(0.278)}$$

$$\widehat{\sigma}_w = 0.092_{\,(0.049)} \quad \widehat{\sigma}_v = 1.287_{\,(0.197)}$$

Among the many forecast error assessment questions that can be asked concerning this model are ones concerning the properties of the conditional forecast error distribution *assuming that we know the future values of the inflation rate*. In particular, is a Gaussian assumption warranted when assume we know the actual future values of z_t? Such questions may arise within the context of a "rational expectations" framework wherein economic agents are assumed so well informed that they "know" the inflation rate. The bootstrap, coupled with our methodology here, can shed some light on just such a question as this.

Figure 12 depicts the bootstrap results with $B = 2000$. The upper left panel presents the dynamic behavior of the quantiles (specifically, 2.5%, 5%, 16%, 50%, 84%, 95%, 97.5%) of $y^b_{n+\ell}(\widehat{\Theta})$ and the upper right panel presents the quantiles of $y^b_{n+\ell}(\Theta^b)$. Given the significant variability in the upper right panel, it is clear that the variability due to the additive disturbances (upper left

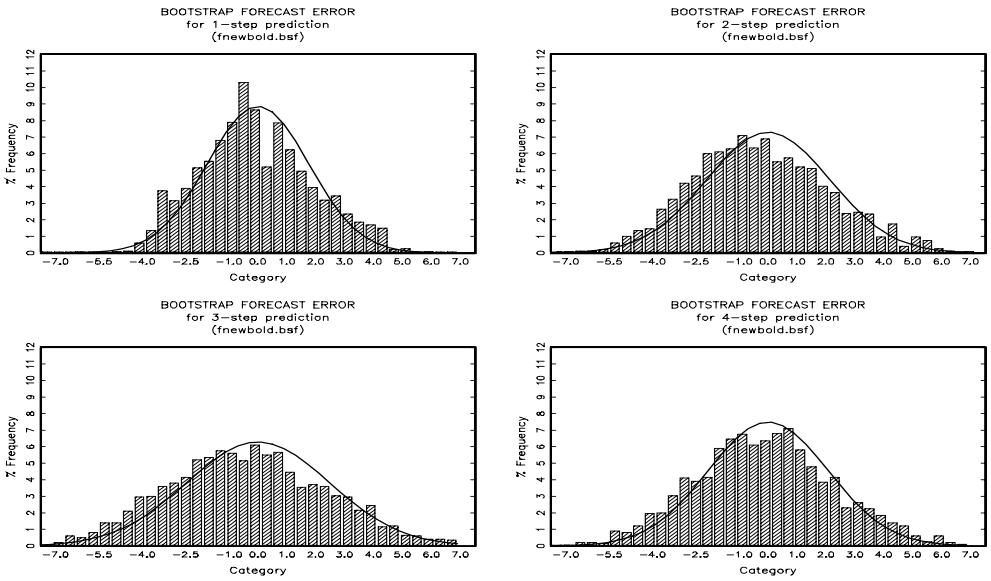

Figure 13: Histograms of four conditional forecast errors, $B = 2000$, in the stochastic regression example.

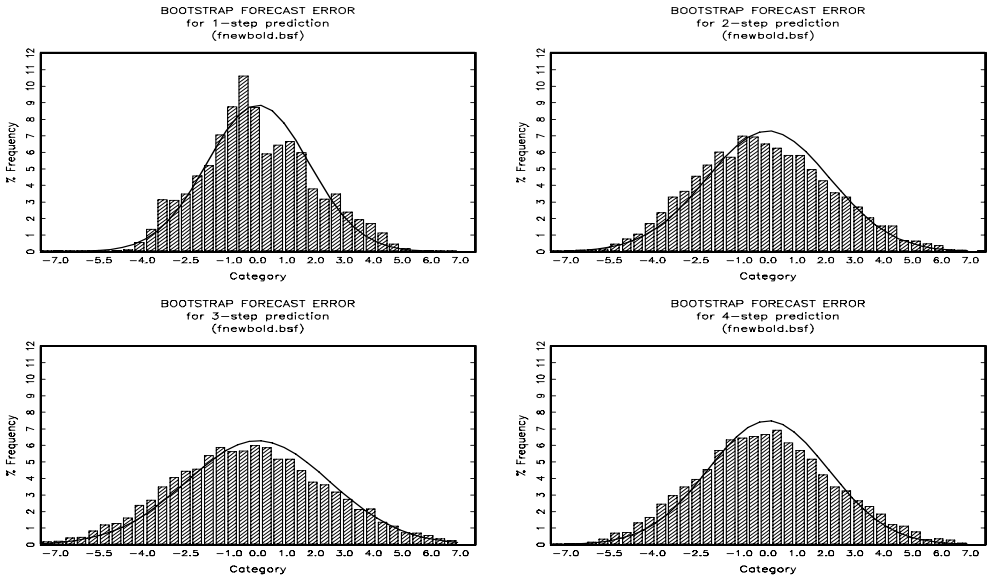

Figure 14: Histograms of four conditional forecast errors, $B = 10,000$, in the stochastic regression example.

panel) is not the dominant factor in the forecast uncertainty that it is so often assumed to be. The lower left panel the depicts dynamic behavior of the quantiles of the bootstrap conditional forecast errors $y^b_{n+\ell}(\widehat{\Theta}) - y^b_{n+\ell}(\Theta^b)$, for $\ell = 1, 2, 3, 4$. The lower right panel plots the y_t series as a bold line

and the envelope of the backward data series as fine lines above and below the observed sample. We find the backward generated series to be highly representative of the stochastic properties of the observed series.

Figure 13 presents histograms of the conditional forecast errors, $y^b_{n+\ell}(\widehat{\Theta}) - y^b_{n+\ell}(\Theta^b)$, for $\ell = 1, 2, 3, 4$, when $B = 2000$ and Figure 14 presents the histograms when $B = 10,000$. Each picture gives indication of the problems in assuming that the asymptotic theory applies. Negative bias is indicated and t-tests reject zero means for $\ell = 2, 3, 4$, in both bootstrap experiments. A Kolmogorov-Smirnov test rejects the asymptotic Gaussian distribution (which are also displayed in the figures) for all forecast lead times for both values of B. It appears little is gained in extending the bootstrap replications beyond $B = 2000$, other than the more "smooth" appearance of the histograms.

4 Discussion

The state space model provides a convenient unifying representation for various time domain models. This article demonstrates the utility of resampling the innovations of time domain models via state space models and the Kalman (innovations) filter. We have based our presentation primarily on the material in two articles, Stoffer and Wall (1991) and Wall and Stoffer (2002).

In Stoffer and Wall (1991) we developed a resampling scheme to assess the finite sample distribution of parameter estimates for general time domain models. This algorithm uses the elegance of the state space model in innovations form to construct a simple resampling scheme. The key point is that while under general conditions, the MLEs of the model parameters are consistent and asymptotically normal, time series data are often of short or moderate length so that the use of asymptotics may lead to wrong conclusions. Moreover, it is well known that problems occur if the parameters are near the boundary of the parameter space. We have provided additional examples here that emphasize the usefulness of the algorithm. We have also explained, heuristically, why the resampling scheme is asymptotically correct under appropriate conditions.

We have also discussed conditional forecast accuracy of time domain models using a state space approach and resampling methods that was first presented in Wall and Stoffer (2002). Applications involving time series data are often characterized by short data sets and lack of distributional information; asymptotic theory provides little help here and frequently there are no compelling reasons to assume Gaussian distributions apply. Interest in the evaluation of confidence intervals for conditional forecast errors in AR models led to methodological problems because a backward, or reverse-time, set of residuals must be generated. This problem was eventually solved and there is now a well grounded methodology for AR models. Researchers were confined to AR models because the required initial, or terminal (in the case of conditional forecasts), conditions are given in terms of the observed series. With other time series models this may not be the case because the models require solutions of difference equations involving unobserved disturbances. The state space model and its related innovations filter offered a way around this difficulty. We have exhibited a reverse-time state space in innovations form. We have presented additional examples here that demonstrate resampling as useful in evaluating conditional forecast errors as it has proven to be in assessing parameter estimation errors, particularly in a non-Gaussian environment. In the Appendix, we explain, heuristically, why resampling works in large samples.

Appendix: Large Sample Heuristics

In §2, resampling techniques were used to determine the finite sample distributions of the parameter estimates when the use of asymptotics was questionable. In §3, we used resampling to assess the finite sample distributions of the forecast errors. The extent to which resampling the innovations does what it is supposed to do can be measured in various ways. In the finite sample case, we can perform simulations—where the true distributions are known—and compare the bootstrap results to the known results. If the bootstrap works well in simulations, we may feel confident that the bootstrap will work well in similar situations, but, of course, we have no guarantee that it works in general. In this way, the examples in §2 and §3 help demonstrate the validity of the resampling procedures discussed in those sections.

Another approach is to ask if the bootstrap will give the correct asymptotic answer. That is, if we have an infinite amount of data and can resample an infinite amount of times, do we get the correct asymptotic distribution (typically, we require asymptotic normality). If the answer is no, we can assume that resampling will not work with small samples. If the answer is yes, we can only hope that resampling will work with small samples, but again, we have no guarantee. For state space models, how well the resampling techniques perform in finite samples hinges on at least three things. First, the techniques are conditional on the data, so the success of the resampling depends on how typical the data set is for the particular model. Second, we assume the model is correct (at least approximately); if the proposed model is far from the truth, the results of the resampling will also be incorrect. Finally, assuming the data set is typical and the model is correct, the success of the resampling depends on how close the empirical distribution of the innovations is to the actual distribution of the innovations. We are guaranteed such closeness in large samples if the innovations are stable and mixing in the sense of Gastwirth and Rubin (1975).

Section 2 Heuristics

Stoffer and Wall (1991) established the asymptotic justification of the procedure presented in §2 under general conditions (including the case where the process is non-Gaussian). To keep matters simple, we assume here that the state space model, (1)–(2) with $A_t \equiv A$, is Gaussian, observable and controllable, and the eigenvalues of Φ are within the unit circle. We denote the true parameters by Θ_0, and we assume the dimension of Θ_0 is the dimension of the parameter space. Let $\widehat{\Theta}_n$ be the consistent estimator of Θ_0 obtained by maximizing the Gaussian innovations likelihood, $L_Y(\Theta)$, given in (10). Then, under general conditions ($n \to \infty$),

$$\sqrt{n}\left(\widehat{\Theta}_n - \Theta_0\right) \sim \text{AN}\left[0,\ \mathcal{I}_n(\Theta_0)^{-1}\right],$$

where $\mathcal{I}_n(\Theta)$ is the information matrix given by

$$\mathcal{I}_n(\Theta) = n^{-1} E\left[-\partial^2 \ln L_Y(\Theta) \,/\, \partial\Theta\ \partial\Theta'\right].$$

Precise details and the proof of this result are given in Caines (1988, Chapter 7) and in Hannan and Deistler (1988, Chapter 4).

Let $\widehat{\Theta}_n^*$ denote the parameter estimates obtained from the resampling procedure of §2. Let B_n be the number of bootstrap replications and, for ease, we take $B_n = n$. Then, Stoffer and Wall (1991) established that, under certain regularity conditions ($n \to \infty$),

$$\sqrt{n}\left(\widehat{\Theta}_n^* - \widehat{\Theta}_n\right) \sim \text{AN}\left[0,\ \mathcal{I}_n^*(\widehat{\Theta}_n)^{-1}\right],$$

where $\mathcal{I}_n^*(\Theta)$ is the information matrix given by

$$\mathcal{I}_n^*(\Theta) = n^{-1} E_*\left[-\partial^2 \ln L_Y(\Theta) \,/\, \partial\Theta\ \partial\Theta'\right],$$

and E_* denotes expectation with respect to the empirical distribution of the innovations. It was then shown that

$$\mathcal{I}_n(\Theta_0) - \mathcal{I}_n^*(\widehat{\Theta}_n) \to 0 \tag{38}$$

almost surely, as $n \to \infty$; hence, the resampling procedure is asymptotically correct.

It is informative to examine, at least partially, why (38) holds. Let $Z_{ta} \equiv Z_{ta}(\Theta) = \partial(\boldsymbol{e}_t' \boldsymbol{e}_t)/\partial\theta_a$ where $\boldsymbol{e}_t \equiv \boldsymbol{e}_t(\Theta)$ is the standardized innovation, (8), and θ_a is the a-th component of Θ. Similarly, let $Z^*_{ta} \equiv Z^*_{ta}(\Theta) = \partial(\boldsymbol{e}_t^{*'} \boldsymbol{e}_t^*)/\partial\theta_a$, where $\boldsymbol{e}_t^* \equiv \boldsymbol{e}_t^*(\Theta)$ is the resampled standardized innovation. The (a,b)-th element of $\mathcal{I}_n(\Theta_0)$ is

$$n^{-1}\sum_{t=1}^{n}\{E(Z_{ta}Z_{tb}) - E(Z_{ta})E(Z_{tb})\}\,\big|_{\,\Theta=\Theta_0} \tag{39}$$

whereas the (a,b)-th element of $\mathcal{I}^*_n(\widehat{\Theta}_n)$ is

$$n^{-1}\sum_{t=1}^{n}\{E_*(Z^*_{ta}Z^*_{tb}) - E_*(Z^*_{ta})E_*(Z^*_{tb})\}\,\big|_{\,\Theta=\widehat{\Theta}_n}\,. \tag{40}$$

The terms in (40) are

$$E_*(Z^*_{ta}) = n^{-1}\sum_{j=1}^{n} Z_{ja} \quad \text{and} \quad E_*(Z^*_{ta}Z^*_{tb}) = n^{-1}\sum_{j=1}^{n} Z_{ja}Z_{jb}. \tag{41}$$

Hence, (39) contains population moments, whereas (40) contains the corresponding sample moments. It should be clear that under appropriate conditions, (39) and (40) are asymptotically ($n \to \infty$) equivalent. Details of these results can be found in Stoffer and Wall (1991, Appendix).

Section 3 Heuristics

As in the previous part, to keep matters simple, we assume the state space model (1)–(2), with $A_t \equiv A$, is observable and controllable, and the eigenvalues of Φ are within the unit circle; these assumptions ensure the asymptotic stability of the filter. We assume that we have N observations, $\{\boldsymbol{y}_{n-N+1}, \ldots, \boldsymbol{y}_n\}$ available, and that N is large. We let $\widehat{\Theta}_N$ denote the (assumed consistent as $N \to \infty$) Gaussian MLE of Θ, and let $\widehat{\Theta}^*_N$ denote a bootstrap parameter estimate.

For one-step-ahead forecasting the model specifies that the process $\boldsymbol{\xi}_t$, which we assume is in steady-state at time $n+1$, is given by

$$\boldsymbol{\xi}_{n+1} = F(\Theta)\boldsymbol{\xi}_n + G(\Theta)\boldsymbol{u}_n + H(\Theta)\boldsymbol{e}_{n+1} \tag{42}$$

where

$$\boldsymbol{\xi}_t = \begin{bmatrix} \boldsymbol{x}_{t+1}^t \\ \boldsymbol{y}_t \end{bmatrix}, \tag{43}$$

and

$$F = \begin{bmatrix} \Phi & 0 \\ A & 0 \end{bmatrix}, \quad G = \begin{bmatrix} \Upsilon \\ \Gamma \end{bmatrix}, \quad H = \begin{bmatrix} K\Sigma^{1/2} \\ \Sigma^{1/2} \end{bmatrix};$$

K and Σ represent the steady-state gain and innovation variance-covariance matrices, respectively. Recall that $\{\boldsymbol{u}_t\}$ is a fixed and known input process.

For convenience, we have dropped the parameter from the notation when representing a filtered value that depends on Θ. For example, in (42) we wrote $\boldsymbol{\xi}_t \equiv \boldsymbol{\xi}_t(\Theta)$ and $\boldsymbol{e}_t \equiv \boldsymbol{e}_t(\Theta)$. The process $\boldsymbol{e}_t$ is the standardized, steady-state innovation sequence so that $E\{\boldsymbol{e}_t\} = \boldsymbol{0}$ and $E\{\boldsymbol{e}_t\boldsymbol{e}_t'\} = I_q$.

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

$$\widetilde{\boldsymbol{\xi}}_{n+1} = F(\widehat{\Theta}_N)\widehat{\boldsymbol{\xi}}_n + G(\widehat{\Theta}_N)\boldsymbol{u}_n, \tag{44}$$

where, in keeping consistent with the notation, we have written $\widehat{\boldsymbol{\xi}}_n \equiv \boldsymbol{\xi}_n(\widehat{\Theta}_N)$. The conditional forecast estimate is labeled with a tilde. Watanabe (1985) showed that, under the assumed conditions and notation,

$\boldsymbol{x}_{n+1}^n(\widehat{\Theta}_N) = \boldsymbol{x}_{n+1}^n(\Theta) + \boldsymbol{o}_p(1)$ $(N \to \infty)$, and consequently, we write $\widehat{\boldsymbol{\xi}}_n = \boldsymbol{\xi}_n + \boldsymbol{o}_p(1)$, noting that the final q elements of $\widehat{\boldsymbol{\xi}}_n$ and $\boldsymbol{\xi}_n$ are identical. Hence, the conditional prediction error can be written as

$$\begin{aligned} \Delta_N &\equiv \boldsymbol{\xi}_{n+1} - \widetilde{\boldsymbol{\xi}}_{n+1} \\ &= [F(\Theta) - F(\widehat{\Theta}_N)]\widehat{\boldsymbol{\xi}}_n + + F(\Theta)\boldsymbol{o}_p(1) + [G(\Theta) - G(\widehat{\Theta}_N)]\boldsymbol{u}_n + H(\Theta)\boldsymbol{e}_{n+1} \end{aligned} \tag{45}$$

From (45) we see the two sources of variation, namely the variation due to estimating the parameter Θ by $\widehat{\Theta}_N$, and the variation due to the predicting the innovation value $\boldsymbol{e}_{n+1}$ by zero.

In the conditional bootstrap procedure, we mimic (42) and obtain a pseudo observation

$$\boldsymbol{\xi}_{n+1}^* = F(\widehat{\Theta}_N)\widehat{\boldsymbol{\xi}}_n + G(\widehat{\Theta}_N)\boldsymbol{u}_n + H(\widehat{\Theta}_N)\boldsymbol{e}_{n+1}^*, \tag{46}$$

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

To mimic the forecast in (44), the bootstrap estimated conditional forecast is given by

$$\widetilde{\boldsymbol{\xi}}_{n+1}^* = F(\widehat{\Theta}_N^*)\widehat{\boldsymbol{\xi}}_n + G(\widehat{\Theta}_N^*)\boldsymbol{u}_n, \tag{47}$$

which yields the bootstrapped conditional forecast error

$$\begin{aligned} \Delta_N^* &\equiv \boldsymbol{\xi}_{n+1}^* - \widetilde{\boldsymbol{\xi}}_{n+1}^* \\ &= [F(\widehat{\Theta}_N) - F(\widehat{\Theta}_N^*)]\widehat{\boldsymbol{\xi}}_n + [G(\widehat{\Theta}_N) - G(\widehat{\Theta}_N^*)]\boldsymbol{u}_n + H(\widehat{\Theta}_N)\boldsymbol{e}_{n+1}^*. \end{aligned} \tag{48}$$

Comparison of (45) and (48) shows why, in finite samples, the bootstrap works; that is, (48) is a sample-based imitation of (45). Letting $N \to \infty$ in (45), while holding $\widehat{\boldsymbol{\xi}}_n$ fixed, we have that, if $\widehat{\Theta}_N \to_p \Theta$, then $\Delta_N \Rightarrow H(\Theta)\boldsymbol{e}$ where $\boldsymbol{e}$ is a random vector that is distributed according to the steady-state standardized innovation distribution ($\Rightarrow$ denotes weak convergence). In addition, if conditional on the data, $\widehat{\Theta}_N^* - \widehat{\Theta}_N \to_p \boldsymbol{0}$, then $\Delta_N^* \Rightarrow H(\Theta)\boldsymbol{e}$ as $N \to \infty$. Extending these results to m-step-ahead forecasts follows easily by induction. Stoffer and Wall (1991) established conditions under which $\widehat{\Theta}_N^* - \widehat{\Theta}_N \to_p \boldsymbol{0}$ as $N \to \infty$ when the forward bootstrapped samples are used. It remains to determine the conditions under which this result holds when the backward bootstrap data are used.

ACKNOWLEDGMENTS: The work of D.S. Stoffer was supported, in part, by the National Science Foundation under Grant No. DMS-0102511.

References

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

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

Ansley, C.F. and P. Newbold (1980). Finite sample properties of estimators for autoregressive moving average processes. *J. Econ.*, 13, 159-183.

Breidt, J., Davis, R., and Dunsmuir, W. (1992). On Backcasting in Linear Time Series, in *New Directions in Time Series Analysis*, Vol. 1, Brillinger, D. et al. (eds.), New York: Springer-Verlag, pp.25-40.

Breidt, J., Davis, R., and Dunsmuir, W. (1995). Time-reversibility, identifiability, and independence of innovations in stationary time series. *J. Time Series Anal.*, 13, 377-390.

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

Danielson, J. (1994). Stochastic volatility in asset prices: Estimation with simulated maximum likelihood. *J. Econometrics*, 61, 375-400.

Dent, W., and A.-S. Min. (1978). A Monte Carlo study of autoregressive-integrated-moving average processes. *J. Econ.*, 7, 23-55.

Durbin, J. and Koopman, S.J. (2001). *Time Series Analysis by State Space Models*. Oxford: Oxford University Press.

Efron, B. (1987). Better bootstrap confidence intervals. *J. Amer. Stat. Assoc.*, 82, 171-185.

Findley, D. (1986). On Bootstrap Estimates of Forecast Mean Square Errors for Autoregressive Processes. In *Computer Science and Statistics*, D. Allen (ed). Amsterdam: The Interface, North-Holland, pp. 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, 809-824.

Hannan, E.J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*, New York: John Wiley & Sons.

Harrison, P.J. and C.F. Stevens (1976). Bayesian forecasting (with discussion). *J. R. Stat. Soc. B*, 38, 205-247.

Harvey, A.C. and P.H.J. Todd (1983). Forecasting economic time series with structural and Box-Jenkins models: A case study. *J. Bus. Econ. Stat.*, 1, 299-307.

Harvey, A. C. and R. G. Pierse (1984). Estimating missing observations in economic time series. *J. Am. Stat. Assoc.*, 79, 125-131.

Harvey, A.C. (1991). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey A.C., E. Ruiz and N. Shephard (1994). Multivariate stochastic volatility models", *Rev. Economic Studies*, 61, 247-264.

Jacquier, E., N.G. Polson, and P.E. Rossi (1994). Bayesian analysis of stochastic volatility models. *J. Bus. Econ. Stat.*, 12, 371-417.

Jones, R.H. (1984). Fitting multivariate models to unequally spaced data. In *Time Series Analysis of Irregularly Observed Data*, pp 158-188. E. Parzen ed. Lecture Notes in Statistics, 25, New York: Springer-Verlag.

Kabaila, P. (1993). On bootstrap predictive inference for autoregressive processes. *J. Time Series Anal.*, 14, 473-484.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Trans ASME J. Basic Eng.*, 82, 35-45.

Kalman, R.E. and R.S. Bucy (1961). New results in filtering and prediction theory. *Trans. ASME J. Basic Eng.,* 83, 95-108.

Kim S., N. Shephard and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Rev. Economic Studies*, 65, p.361-393.

Kitagawa, G. and W. Gersch (1984). A smoothness priors modeling of time series with trend and seasonality. *J. Am. Stat. Assoc.*, 79, 378-389.

McCullough, B. D. (1994). Bootstrapping forecasting intervals: An application to AR(p) models. *J. Forecasting*, 13, 51-66.

McCullough, B. D. (1996). Consistent forecast intervals when the forecast-period exogenous variables are stochastic. *J. Forecasting*, 15, 293-304.

Newbold, P., and Bos, T. (1985). *Stochastic Parameter Regression Models*, Beverly Hills: Sage Publications.

Peña, D. and I. Guttman (1988). A Bayesian approach to robustifying the Kalman filter. In *Bayesian Analysis of Time Series and Dynamic Linear Models*, pp. 227-254. J.C. Spall ed. New York: Marcel Dekker.

Sandmann, G. and S.J. Koopman (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *J. Econometrics*, 87 (no. 2), 271-301

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In *Time Series Models in Econometrics, Finance and Other Fields* , pp 1-67. D.R. Cox, D.V. Hinkley, and O.E. Barndorff-Nielson eds. London: Chapman and Hall.

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

Shumway, R.H. and D.S. Stoffer (2000). *Time Series Analysis and Its Applications*. New York: Springer.

Stine, R.A. (1987). Estimating properties of autoregressive forecasts. *J. Amer. Stat. Assoc.*, 82, 1072-1078.

Stoffer, D.S. and Wall, K.D. (1991). Bootstrapping state space models: Gaussian maximum likelihood estimation and the Kalman filter. *J. Amer. Stat. Assoc.*, 86, 1024-1033.

Stultz, C.M., White, J.V. and Smith, T.F. (1993). Structural analysis based on state space modeling. *Protein Sci.*, 2, 305-314.

Thombs, L. and Schuchany, W. (1990). Bootstrap prediction intervals for autoregression. *J. Amer. Stat. Assoc.*, 85, 486-492.

Wall, K.D. and Stoffer, D.S. (2002). A state space approach to bootstrapping conditional forecasts in ARMA models. *J. Time Series Anal.*, 23, 733-751.

Watanabe, N. (1985). Note on the Kalman filter with estimated parameters. *J. Time Series Anal.*, 6, 269-278.