Bootstrapping State-Space Models: Gaussian Maximum Likelihood Estimation and the Kalman Filter

DAVID S. STOFFER and KENT D. WALL*

The bootstrap is proposed as a method for assessing the precision of Gaussian maximum likelihood estimates of the parameters of linear state-space models. Our results also apply to autoregressive moving average models, since they are a special case of state-space models. It is shown that for a time-invariant, stable system, the bootstrap applied to the innovations yields asymptotically consistent standard errors. To investigate the performance of the bootstrap for finite sample lengths, simulation results are presented for a two-state model with 50 and 100 observations; two cases are investigated, one with real characteristic roots and one with complex characteristic roots. The bootstrap is then applied to two real data sets, one used in a test for efficient capital markets and one used to develop an autoregressive integrated moving average model for quarterly earnings data. We find the bootstrap to be of definite value over the conventional asymptotics.

KEY WORDS: ARMA models; Empirical sampling distributions; Monte Carlo simulation; Stochastic regression.

1. INTRODUCTION

State-space models and Kalman filtering have become important and powerful tools for the statistician and the econometrician. Together they provide the researcher with a modeling framework and a computationally efficient way to compute parameter estimates over a wide range of situations. Problems involving stationary and nonstationary stochastic processes (Goodrich and Caines 1979), systematically or stochastically varying parameters (Pagan 1980), and unobserved or latent variables (as in signal extraction problems) all have been fruitfully approached with these tools (Burmeister, Wall, and Hamilton 1986). In addition, smoothing problems and time series with missing observations have been studied with methodologies based on this combination (Shumway and Stoffer 1982). Many authors have exploited the state-space model and Kalman filter recursions for estimation and prediction of autoregressive moving average (ARMA) processes (Harvey and Phillips 1979; Gardner, Harvey, and Phillips 1980; Jones 1980; Harvey and Pierse 1984) and of structural models (Harvey and Todd 1983; Kitagawa and Gersch 1984; Harvey and Durbin 1986). In each of these instances the state-space formulation and the Kalman filter has yielded a modeling and estimation methodology that is much less cumbersome than the more traditional regression-based approach.

Problems of inference in state-space models estimated using the Kalman filter are made tractable by the existence of an asymptotic theory. Under appropriate conditions, both the parameter estimates obtained by maximum likelihood techniques and the state estimates from the Kalman filter have been shown to be consistent and asymptotically normal (Ljung and Caines 1979; Spall and Wall 1984). Time series data, however, is often of short to moderate length, and the use of asymptotic methods is suspect. For ARMA models, several researchers have found evidence that samples must be fairly large before asymptotic results are applicable (Dent and Min 1978; Ansley and Newbold 1980) and one should expect a similar situation in the case of state-space models.

Our approach employs the nonparametric Monte Carlo bootstrap suggested by Efron (1979) and focuses on the Gaussian maximum likelihood estimator. We propose the nonparametric bootstrap because we feel it is more useful in practice; it does not rely on distributional assumptions that cannot be adequately checked in small- to moderate-size samples. The value of the nonparametric bootstrap has been demonstrated in a regression framework by Freedman (1981) and Freedman and Peters (1984a,b). Other work using the nonparametric bootstrap to study forecast errors is reported in Findley (1985) and Stine (1985). Nevertheless, if the family of distributions for the model can be specified, we would suggest a parametric bootstrap. We use Gaussian likelihood estimation because it is the method of choice found in the literature and has desirable asymptotic properties.

In Section 2 we begin with a description of the state-space model and an outline of the parameter estimation problem for these models, then Section 3 gives the bootstrap procedure for such models. Section 4 presents some empirical studies that illustrate the utility of the bootstrap in small- to moderate-size samples, underscoring its value in empirical research, and Section 5 briefly describes some computational aspects of the procedure. Of fundamental concern is that the bootstrap is asymptotically correct, that is, the bootstrap is at least as good as the conventional asymptotic theory; the asymptotic justification of the procedure is given in the Appendix.

2. THE STATE–SPACE MODEL AND ESTIMATION

The *state-space model* is defined by the equations

$$\mathbf{s}(t+1) = \mathbf{F}\mathbf{s}(t) + \mathbf{G}\mathbf{x}(t) + \mathbf{w}(t) \tag{2.1}$$

and

$$\mathbf{y}(t) = \mathbf{H}\mathbf{s}(t) + \mathbf{D}\mathbf{x}(t) + \mathbf{v}(t), \tag{2.2}$$

where $\mathbf{s}(t)$ is a $p \times 1$ vector of unobserved state variables, $\mathbf{y}(t)$ is a $q \times 1$ vector of observed outputs or endogenous variables, and $\mathbf{x}(t)$ is an $r \times 1$ vector of observed inputs or

* David S. Stoffer is Associate Professor, Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260. Kent D. Wall is Professor, Defense Resources Management Education Center, Naval Postgraduate School, Monterey CA 93943. The work of David S. Stoffer was supported in part by National Science Foundation Grant DMS-9000522. The authors thank the referees and the associate editor for helpful comments and additional references.

Journal of the American Statistical Association
December 1991, Vol. 86, No. 416, Theory and Methods

exogenous variables. The constant matrices **F**, **G**, **H**, and **D** represent the model coefficients of dimensions compatible with the matrix operations required in (2.1) and (2.2). The two terms $\mathbf{w}(t)$ and $\mathbf{v}(t)$ represent zero-mean random processes that are each independent and identically distributed with

$$E\{\mathbf{w}(t)\mathbf{w}'(t)\} = \mathbf{Q}, \quad E\{\mathbf{v}(t)\mathbf{v}'(t)\} = \mathbf{R}, \quad E\{\mathbf{w}(t)\mathbf{v}'(t)\} = \mathbf{0}, \tag{2.3}$$

where **Q** is a $p \times p$ nonnegative definite matrix and **R** is a $q \times q$ nonnegative definite matrix. While **R** is usually required to be positive definite, this is not imposed here so that we may include some popular representations of ARMA processes (Harvey and Pierse 1984; Jones 1980). It should be noted, however, that restricting **R** to be positive definite in no way precludes the use of (2.1)–(2.3) in representing ARMA models (Anderson and Moore 1979, p. 236; Caines 1988, p. 115). The assumptions concerning the definiteness of **R** are only a matter of personal choice in representing ARMA processes via (2.1)–(2.3).

The model coefficients and the correlation structure are assumed to be uniquely parameterized by a $k \times 1$ vector $\boldsymbol{\theta}$; that is, $\mathbf{F} = \mathbf{F}(\boldsymbol{\theta})$, $\mathbf{G} = \mathbf{G}(\boldsymbol{\theta})$, $\mathbf{H} = \mathbf{H}(\boldsymbol{\theta})$, $\mathbf{D} = \mathbf{D}(\boldsymbol{\theta})$, $\mathbf{Q} = \mathbf{Q}(\boldsymbol{\theta})$, and $\mathbf{R} = \mathbf{R}(\boldsymbol{\theta})$. The vector $\boldsymbol{\theta}$ is assumed to be an element of some compact space, $\mathscr{P}$, usually a subset of $\mathscr{R}^k$. Furthermore, it is assumed that the parameterization is such that the model is completely identified (Wall 1987; Pagan 1980).

Let $\mathbf{s}(t+1 \mid t)$ denote the best linear predictor of $\mathbf{s}(t+1)$ based on the data $\mathscr{Y}^t = \{\mathbf{y}(1), \cdots \mathbf{y}(t)\}$ and $\mathscr{X}^t = \{\mathbf{x}(1), \ldots, \mathbf{x}(t)\}$, obtained via the Kalman filter (Anderson and Moore 1979, p. 44). Also obtained from the Kalman filter are the *innovations*, the *innovations covariance matrix*, and the *Kalman gain matrix*,

$$\boldsymbol{\epsilon}(t) = \mathbf{y}(t) - \mathbf{Hs}(t \mid t-1) - \mathbf{Dx}(t), \tag{2.4a}$$

$$\boldsymbol{\Sigma}(t) = \mathbf{HP}(t \mid t-1)\mathbf{H}' + \mathbf{R}, \tag{2.4b}$$

and

$$\mathbf{K}(t) = \mathbf{P}(t \mid t-1)\mathbf{H}'\boldsymbol{\Sigma}(t)^{-1}, \tag{2.4c}$$

respectively, where $\mathbf{P}(t \mid t-1)$ is the covariance matrix of $\mathbf{s}(t) - \mathbf{s}(t \mid t-1)$. The model innovations from the Kalman filter give rise to the *innovations form representation* (Anderson and Moore 1979, p. 231) of the observations:

$$\mathbf{s}(t+1 \mid t) = \mathbf{Fs}(t \mid t-1) + \mathbf{Gx}(t) + \mathbf{FK}(t)\boldsymbol{\epsilon}(t) \tag{2.5}$$

and

$$\mathbf{y}(t) = \mathbf{Hs}(t \mid t-1) + \mathbf{Dx}(t) + \boldsymbol{\epsilon}(t). \tag{2.6}$$

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

$$\mathbf{L}(\boldsymbol{\theta} \mid \mathscr{Y}^T, \mathscr{X}^T) = -\sum_{t=1}^{T} \{\log|\boldsymbol{\Sigma}(t, \boldsymbol{\theta})| + \boldsymbol{\epsilon}(t, \boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(t, \boldsymbol{\theta})\boldsymbol{\epsilon}(t, \boldsymbol{\theta})\}, \tag{2.7}$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{\epsilon}$ are generated from (2.4) and $|\cdot|$ denotes the determinant. The influence of $\boldsymbol{\theta}$ is made explicit here to emphasize how the quantities defined in (2.4) depend on the parameterization. Maximizing (2.7) with respect to $\boldsymbol{\theta}$ yields the Gaussian maximum likelihood estimate $\hat{\boldsymbol{\theta}} = \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta} \mid \mathscr{Y}^T, \mathscr{X}^T)$. Iterative procedures for maximizing this function exist in many forms and are easily implemented (Burmeister and Wall 1982).

3. THE MONTE CARLO BOOTSTRAP FOR STATE–SPACE MODELS

The Monte Carlo bootstrap procedure for state-space models is defined by a four-step algorithm. We assume that the model estimation has been completed and that the Kalman filter has been run with $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ so that the estimated innovations, $\boldsymbol{\epsilon}(t, \hat{\boldsymbol{\theta}})$, are available.

1. Construct the standardized innovations by setting

$$\mathbf{e}(t, \hat{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}^{-1/2}(t, \hat{\boldsymbol{\theta}})\boldsymbol{\epsilon}(t, \hat{\boldsymbol{\theta}}), \tag{3.1}$$

where $\boldsymbol{\Sigma}^{-1/2}(t, \boldsymbol{\theta})$ is the inverse of the unique square-root matrix of $\boldsymbol{\Sigma}(t, \boldsymbol{\theta})$. By using (3.1) we are guaranteed that all model residuals have, at least, the same two first moments.

2. Sample, with replacement, T times from $\{\mathbf{e}(t, \hat{\boldsymbol{\theta}});\ 1 \le t \le T\}$ to obtain $\{\mathbf{e}^*(t, \hat{\boldsymbol{\theta}});\ 1 \le t \le T\}$, a bootstrap sample of standardized innovations.

3. Using the innovations form representation, (2.5) and (2.6), construct a bootstrap data set $\{\mathbf{y}^*(t);\ 1 \le t \le T\}$ as follows: Define the $(p + q) \times 1$ vector $\boldsymbol{\xi}(t) = [\mathbf{s}'(t+1 \mid t) \mid \mathbf{y}'(t)]'$. Stacking (2.5) and (2.6) results in a vector first-order equation in $\boldsymbol{\xi}(t)$,

$$\boldsymbol{\xi}(t) = \mathbf{A}\boldsymbol{\xi}(t-1) + \mathbf{Bx}(t) + \mathbf{C}(t)\mathbf{e}(t, \boldsymbol{\theta}), \tag{3.2}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{F} & \mathbf{0} \\ \mathbf{H} & \mathbf{0} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{G} \\ \mathbf{D} \end{bmatrix}, \quad \mathbf{C}(t) = \begin{bmatrix} \mathbf{FK}(t)\boldsymbol{\Sigma}^{1/2}(t, \boldsymbol{\theta}) \\ \boldsymbol{\Sigma}^{1/2}(t, \boldsymbol{\theta}) \end{bmatrix}.$$

Thus, to construct a bootstrap data set, $\{\mathbf{y}^*(t);\ 1 \le t \le T\}$, simply solve (3.2) using $\{\mathbf{e}^*(t, \hat{\boldsymbol{\theta}});\ 1 \le t \le T\}$ in place of $\{\mathbf{e}(t, \boldsymbol{\theta});\ 1 \le t \le T\}$. The exogenous variables, $\{\mathbf{x}(t);\ 1 \le t \le T\}$, and the initial conditions of the Kalman filter remain fixed at their given values while the parameter vector $\boldsymbol{\theta}$ is held fixed at $\hat{\boldsymbol{\theta}}$.

4. Repeat steps (2) and (3) a large number, N, of times, obtaining a set of replications, $\{\hat{\boldsymbol{\theta}}^{*i};\ 1 \le i \le N\}$. Estimate the distribution of $\hat{\boldsymbol{\theta}}$ from the distribution of the $\hat{\boldsymbol{\theta}}^{*i}$.

4. EMPIRICAL STUDIES

While the results of the Appendix establish the fundamental fact that the bootstrap is asymptotically correct, it does not recommend the bootstrap over the standard asymptotic theory. To investigate the properties of the procedure for small- to moderate-samples, we have conducted several simulation experiments. We find the bootstrap to be superior to the standard asymptotics—better standard errors for the parameter estimates can be obtained with the bootstrap. In addition, the bootstrap provides valuable information for establishing interval estimates—something of vital interest if the small-sample estimator yields skewed distributions. In our investigation we consider the estimation of two state-space models using simulated data and then proceed to two real data sets.

4.1 Simulation Experiments

Our simulation experiments concern a two-state model in observable canonical form with one random disturbance in the state equation, one random error in the output equation and Gaussian noise. The coefficient matrices of (2.1)–(2.3) take the form

$$\mathbf{F} = \begin{bmatrix} 0 & f_{12} \\ 1 & f_{22} \end{bmatrix}, \qquad \mathbf{G} = \begin{bmatrix} 0 \\ .3 \end{bmatrix}, \qquad \mathbf{Q} = \begin{bmatrix} 0 & 0 \\ 0 & .0025 \end{bmatrix},$$
$$\mathbf{H} = [0 \quad 1], \qquad \mathbf{D} = [0], \qquad \mathbf{R} = [.01].$$

All simulated data use a zero initial state [i.e., $\mathbf{s}(0) = (0, 0']$ and an input series $\{\mathrm{x}(t);\ 1 \le t \le T\}$ drawn from a uniform distribution on the interval $[-.5, .5]$.

To avoid imposing nonnegativity conditions on the parameters in $\mathbf{Q}$ and $\mathbf{R}$, we write the stochastic specification for the disturbances in terms of their unique (lower triangular) Cholesky factors and estimate the square roots of the respective variances. Thus we estimate q_{22} and r_{11} in the Cholesky factors

$$\mathbf{Q} = \mathbf{Q}_L\mathbf{Q}_L' = \begin{bmatrix} q_{11} & 0 \\ q_{21} & q_{22} \end{bmatrix} \begin{bmatrix} q_{11} & q_{21} \\ 0 & q_{22} \end{bmatrix}$$

and

$$\mathbf{R} = \mathbf{R}_L\mathbf{R}_L' = [r_{11}][r_{11}],$$

where $q_{11} = q_{21} = 0$. The induced parameterization is $\boldsymbol{\theta} = [f_{12}, f_{22}, g_{21}, q_{22}, r_{11}]'$.

Two different situations are covered using this model, each distinguished from the other by the latent roots of the characteristic equation associated with $\mathbf{F}$. The first case has complex roots specified by $\lambda_{1,2} = .7 \pm .6i$ ($i = \sqrt{-1}$); the corresponding elements of $\mathbf{F}$ are $f_{12} = -.85$ and $f_{22} = 1.40$. The second case has real roots $\lambda_1 = .4$ and $\lambda_2 = .8$, so that $f_{12} = -.32$ and $f_{22} = 1.20$. In each case, $\{y(t)\}$ is nearly a parameter redundant ARMA (2, 2) process; that is, $(1 - \lambda_1 B)(1 - \lambda_2 B)y(t) \approx (1 - \lambda_1 B)(1 - \lambda_2 B)v(t)$, where B is the usual backward shift operator, and we should expect instability of the estimates for small sample sizes (Box and Jenkins 1976, pp. 248–250).

For each model we carry out two simulation experiments, the first experiment using a short sample of length $T = 50$ and the second experiment using a moderate sample of length $T = 100$. In each experiment $N = 1{,}000$. Thus there are four experiments that reveal the influence of both sample length and dynamic behavior of the underlying model.

Each experiment employed the bootstrap procedure of Section 3, modified to exclude random sampling of the $\mathbf{e}(t, \hat{\boldsymbol{\theta}})$ for $1 \le t \le 3$. Thus $\mathbf{e}^*(t, \hat{\boldsymbol{\theta}}) = \mathbf{e}(t, \hat{\boldsymbol{\theta}})$ for $1 \le t \le 3$ with random sampling applied only over $4 \le t \le T$. This modification was necessary to avoid start-up problems caused by the transient behavior of the Kalman filter over the interval $1 \le t \le 3$; in these examples, $\boldsymbol{\Sigma}(t)$ and $\mathbf{K}(t)$ were an order of magnitude larger for $1 \le t \le 3$ than they were for $t > 3$. We recommend this modification, as needed, to avoid obtaining unrepresentative bootstrap samples $\{\mathbf{y}^*(t)\}$ due to unusual values of $\boldsymbol{\Sigma}(t)$ and $\mathbf{K}(t)$ being present in the generation of $\mathbf{y}^*(t)$ via (3.2).

Small-Sample Case, T = 50, Complex Roots. Table 1 presents the results for the small-sample case, $T = 50$. The first two columns in this and subsequent tables give the parameter estimates and nominal standard errors that result from GML estimation; the third and fourth columns present the bootstrap sample means and standard deviations; the last two columns display the means and standard deviations of an approximation to the true sampling distribution. These latter quantities were obtained by a parametric Monte Carlo experiment in which GML estimation was carried out on an additional 1,000 data sets, each generated by simulating the model (2.1)–(2.3) with Gaussian noise, $\boldsymbol{\theta}$ fixed at its true value, $\mathbf{s}(0) = \mathbf{0}$, and $\{\mathbf{x}(t)\text{: } 1 \le t \le 50\}$ fixed at the values used in the original data generation. The result is an attempt to approximate the true small-sample distribution of $\hat{\boldsymbol{\theta}}$. We label the last two columns as "true" confident that 1,000 replications is adequate for our purposes.

For the complex root case, we find the nominal asymptotic standard errors to seriously understate the actual variability of the estimator for the coefficients in $\mathbf{F}$, $\mathbf{Q}$, and $\mathbf{R}$. The nominal standard errors for f_{12} and f_{22} are approximately 60% of the true values, while the bootstrap estimates are almost equivalent to the true values. The nominal standard errors for the remaining parameters fare better in terms of the g_{21} parameter, but fall significantly short of the true values for q_{22} and r_{11}. In all parameters we find the bootstrap yields more accurate assessments of estimator error.

Small-Sample Case, T = 50, Real Roots. A similar result holds for the real-root case presented in Table 2. The asymptotic standard errors produced by the GML estimator perform poorly in all comparisons. The bootstrap appears to overstate the true standard errors, however, we do not consider this to be a significant tendency.

Figures 1 and 2 depict the sample histograms for the estimated parameters in $\mathbf{F}$ derived from the bootstrap and the approximation to the true sampling distribution. In our small sample, the true sampling distribution is far from normal, as the asymptotic theory would predict. Figure 1 indicates

Table 1. Small Sample (T = 50) With Complex Roots

	Gaussian ML		*Bootstrap*		*True*	
	Estimate	*Nominal SE*	*Mean*	*SD*	*Mean*	*Nominal SE*
f_{12}	−.9008	.0388	−.8387	.0661	−.8381	.0642
f_{22}	1.4148	.0355	1.3953	.0605	1.3896	.0606
g_{21}	.3434	.0475	.3072	.0485	.3075	.0577
q_{22}	.0513	.0124	.0394	.0182	.0393	.0202
r_{11}	.0879	.0125	.1003	.0149	.0999	.0160

Table 2. Small Sample (T = 50) With Real Roots

	Gaussian ML		Bootstrap		True	
	Estimate	Nominal SE	Mean	SD	Mean	Nominal SE
f_{12}	−.3222	.1548	−.2210	.1969	−.2258	.1953
f_{22}	1.2186	.1595	1.1166	.1988	1.1025	.1963
g_{21}	.2954	.0573	.3236	.0666	.3225	.0638
q_{22}	.0518	.0183	.0489	.0264	.0450	.0242
r_{11}	.1018	.0153	.0976	.0192	.0973	.0168

what appears to be a mixture distribution with concentrations at 0 and $-.3$ for f_{12}, and masses concentrated at .9 and 1.2 for f_{22}. Examination of the estimates reveals that when f_{22} is close to .9, f_{12} is close to 0. Thus the true sampling distribution indicates that there is a significant probability of obtaining a sample in which the smaller root $\lambda =$

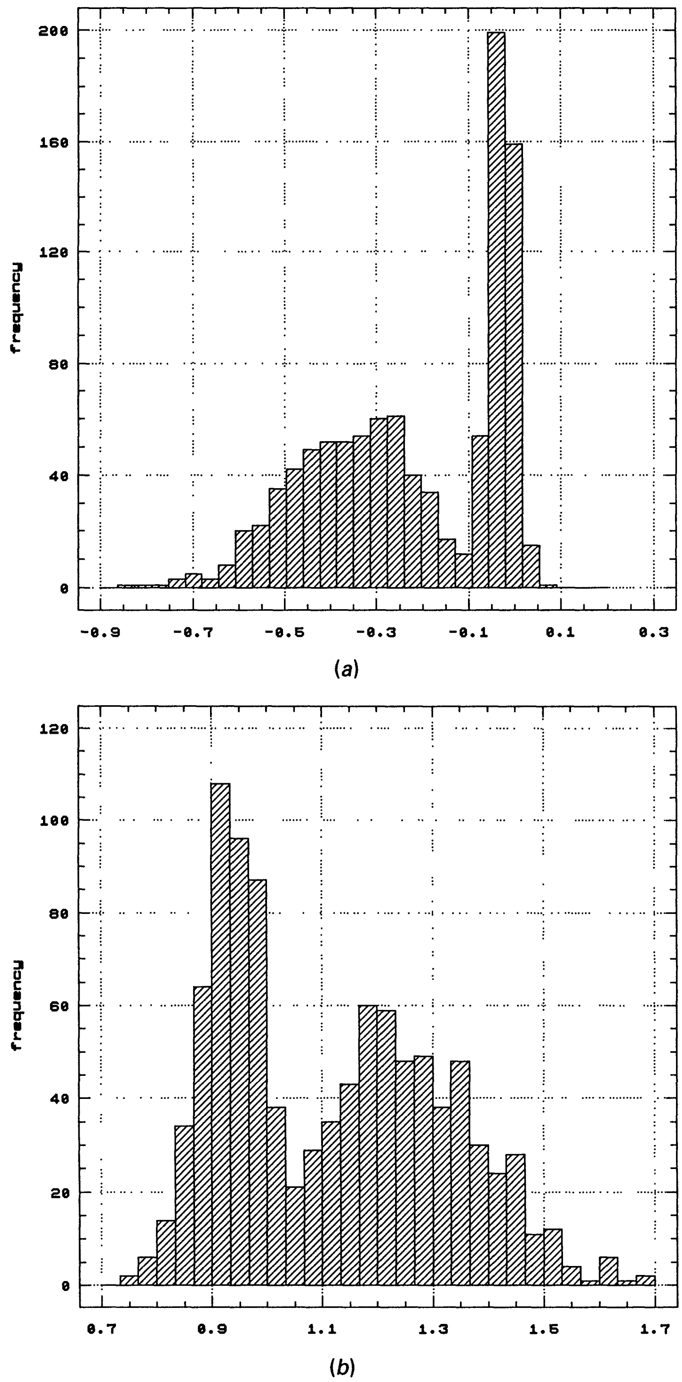

Figure 1. Bootstrap Histograms for Elements of F With T = 50, Real Roots: (a) element f_{12}; (b) element f_{22}.

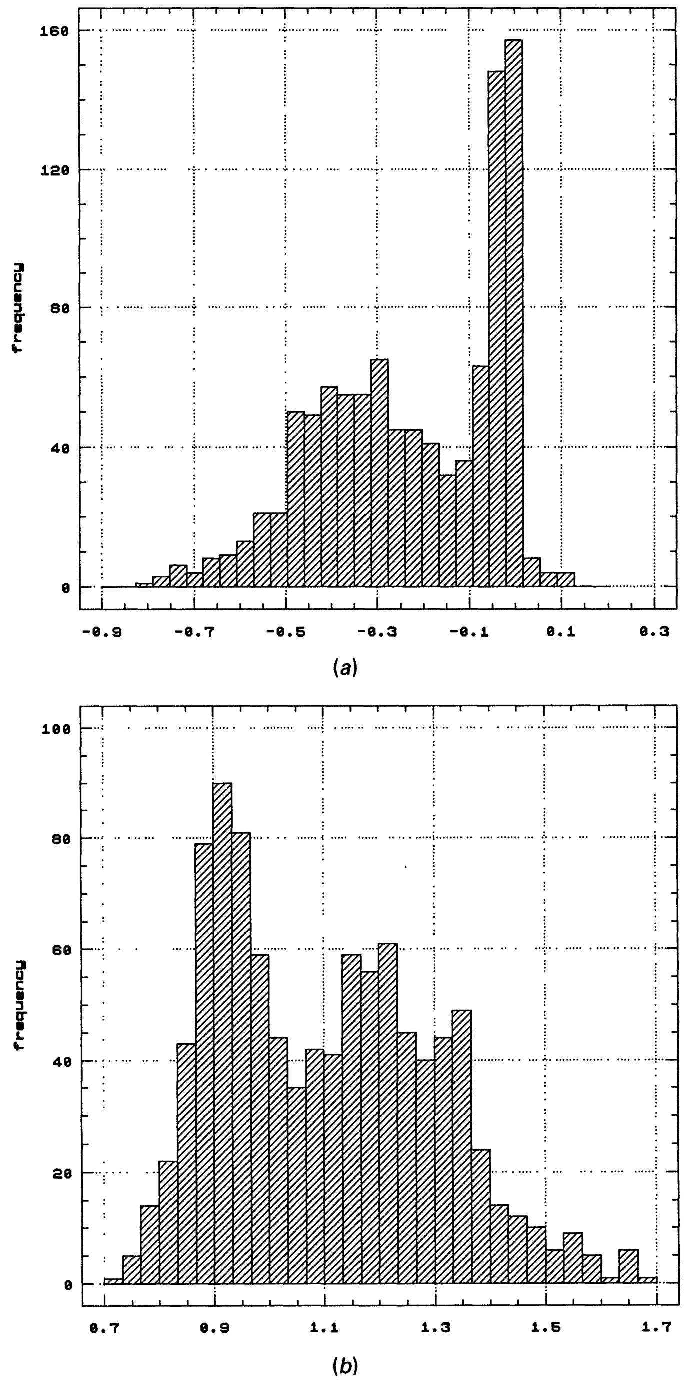

Figure 2. True Histograms for Elements of F With T = 50, Real Roots: (a) element f_{12}; (b) element f_{22}.

Table 3. Medium Sample (T = 100) With Complex Roots

	Gaussian ML		Bootstrap		True	
	Estimate	Nominal SE	Mean	SD	Mean	Nominal SE
f_{12}	−.7927	.0367	−.7885	.0371	−.8434	.0388
f_{22}	1.3642	.0370	1.3594	.0366	1.3927	.0397
g_{21}	.3699	.0437	.3697	.0406	.3045	.0389
q_{22}	.0484	.0089	.0440	.0116	.0464	.0099
r_{11}	.1119	.0106	.1103	.0119	.1000	.0100

.4 cannot be identified. By employing the bootstrap, we obtain vital information concerning the problems with model specification due to near parameter redundancy when sample sizes are small.

Medium-Sample Case, $T = 100$, Complex Roots. Table 3 summarizes the results for the complex-root case using a medium-sized sample of 100 observations. With complex roots and samples of moderate size, it appears that the GML estimator does a better job in assessing the standard errors of the estimates. There is no definite pattern indicating bias in these statistics, with the GML and the bootstrap both agreeing closely with the "true" variability of the estimator. It is reasonable to conclude that with 100 observations and complex roots we have a situation in which convergence to the asymptotic result is close.

Medium-Sample Case $T = 100$, Real Roots. Table 4 summarizes the results for the real-root case under a moderate-sized sample. Here we find continuing problems with the GML nominal standard errors. Understatement of the standard errors for both **F** parameters is indicated. This is also true for the **Q** and **R** parameters, while for g_{21} the GML standard errors appear satisfactory. The problems in estimating the elements of **F** with two real roots persisted in this case; while these problems were less severe with $T = 100$ than with $T = 50$, it appeared that one would require samples of at least 200 observations before all modes of dynamic response make themselves felt in the data. From the results of the Appendix we know that as $T \to \infty$, both sets of histograms would become unimodal (and normal) about $f_{12} = -.32$ and $f_{22} = 1.2$.

4.2 Stochastic Regression

An interesting application of GML state-space model estimation is given by Newbold and Bos (1985, pp. 61–73). Of the several alternative models they investigate, we focus on that specified by their equations (4.7a) and (4.7b). This model has one output variable, the (nominal) interest rate recorded for three-month treasury bills denoted by $i(t)$. The output equation in this model is specified as $i(t) = \alpha + \beta(t)r(t) + v(t)$, where α is a constant intercept, $r(t)$ denotes the observed quarterly inflation rate in the Consumer Price Index, $\beta(t)$ is a stochastically varying regression coefficient, and $v(t)$ is an additive zero-mean iid process with finite variance σ_v^2. The slope coefficient comprises the state variable and is specified to follow a first-order autoregressive process $\beta(t+1) - b = \varphi[\beta(t) - b] + w(t)$. The state noise, $w(t)$, is assumed to follow a zero-mean iid process with finite variance σ_w^2. In our notation, we have $\mathbf{F} = \varphi$, $\mathbf{G} = (1 - \varphi)b$, $\mathbf{Q} = \sigma_w^2$, $\mathbf{H} = \mathbf{H}(t) = r(t)$, $\mathbf{D} = \alpha$, $\mathbf{R} = \sigma_v^2$, $\mathbf{x}(t) \equiv 1$, and $\boldsymbol{\theta} = [\varphi, b, \alpha, \sigma_w, \sigma_v]'$.

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, $T = 50$ observations. The results of our estimation are presented in Table 5. The differences between our GML estimates and those reported in Newbold and Bos are basically attributable to the fact that we use a different numerical optimization routine. We obtain agreement to at least three decimal places in the parameter estimates, and our log-likelihood value is −81.9425, while theirs is −81.95. The bootstrap results reported in Table 5 were obtained by application of the procedure of Section 3 with $N = 1{,}000$ and without modification for start-up problems since there was no chance of the procedure being confounded by initial transient behavior of the Kalman filter.

Review of the bootstrap standard errors indicates that the asymptotic nominal standard errors are biased downward in three of the five parameters. The nominal asymptotic standard errors for the constants α and b perform quite well with respect to the bootstrap results, but the same cannot be said for φ, σ_w, and σ_v. The nominal standard errors for these three parameters are between 60% and 70% of their bootstrap counterparts. The GML estimator performs quite well with respect to the two intercept terms (α in the output equation and b in the state equation) but underestimates variability in the transition matrix parameter and the two variance parameters.

Table 4. Medium Sample (T = 100) With Real Roots

	Gaussian ML		Bootstrap		True	
	Estimate	Nominal SE	Mean	SD	Mean	Nominal SE
f_{12}	−.3379	.1212	−.2540	.1400	−.2520	.1486
f_{22}	1.1705	.1209	1.0923	.1458	1.1259	.1536
g_{21}	.2929	.0516	.3139	.0505	.3139	.0460
q_{22}	.0541	.0107	.0554	.0163	.0508	.0155
r_{11}	.1032	.0099	.0993	.0122	.0981	.0112

Table 5. Stochastic Regression Example (T = 50)

	Gaussian ML		Bootstrap		Newbold and Bos	
	Estimate	Nominal SE	Mean	SD	Estimate	Nominal SE
φ	.8414	.1997	.5897	.2775	.8414	.2122
b	.8584	.2776	.8416	.2737	.8584	.2591
α	−.7714	.6449	−.7652	.6315	−.7715	.6033
σ_W	.1269	.0924	.1562	.1272	.1268[a]	NA
σ_V	1.1306	.1419	1.0121	.2421	1.1306[b]	NA

[a]Square root of σ_a^2 estimate reported in Newbold and Bos.
[b]Square root of σ_e^2 estimate reported in Newbold and Bos.

Apart from the standard errors, the bootstrap highlights other important aspects of the estimation problem. Table 5 shows the mean of the bootstrap estimates for φ to be markedly different from the GML estimator; Figure 3 presents the bootstrap histogram for the estimate of φ. This figure reveals some significant departures from the normality of the parameter estimates that is maintained in the asymptotic theory. Figure 3 shows the estimator of φ to have a skewed distribution, with values concentrated around .8 but possessing a long tail to the left, which explains the discrepancy between the bootstrap mean and the GML estimate. The right-hand tail is greatly attenuated because φ cannot go above unity without the onset of instability in the state equation. This behavior is similar to that observed in studies of the estimates of parameters in ARMA models near the boundaries of the stability or invertibility regions (Box and Jenkins 1976, p. 224). This situation suggests the use of standard errors for inference and hypothesis testing has less meaning than the use of interval estimates obtained through quantiles.

The bootstrap histograms for the estimates of b and α depicted symmetric distributions suggestive of the normal case maintained in the asymptotic theory. In Figure 4, the histogram pertaining to σ_w is concentrated about two distinct locations, one relating to a structure logically consistent with the original model specification and the other to a model structure quite different from that originally intended. The concentration about $\sigma_w = .13$, with a long tail toward the right, is to be expected given our desire to fit a model with stochastic coefficient $\beta(t)$. The concentration at $\sigma_w = 0$ is consistent with *deterministic* state dynamics. When $\sigma_w = 0$ and $|\varphi| < 1$, $\beta(t) \approx b$ for large t so that the approximately 225 cases where $\sigma_w = 0$ suggest a large number of samples correspond to a fixed state, or a constant coefficient configuration. The histogram for σ_v indicated a slight asymmetry with a long tail toward the left.

4.3 ARIMA Modeling

Shumway (1988, pp. 186–190) used state-space methodology to develop and analyze a structural model for the quarterly data, $z(t)$, on earnings per share for the U.S. company Johnson and Johnson, fourth quarter 1970 to first quarter 1980, $T = 38$ observations. We consider a special case of Shumway's model that was shown by Harvey (1981, p. 180) to be an ARIMA $(0, 1, 1) \times (0, 1, 1)_4$ model, that is, $y(t) = \beta_0 + (1 - \beta_1 B)(1 - \beta_4 B^4)v(t)$, where $y(t) = (1$

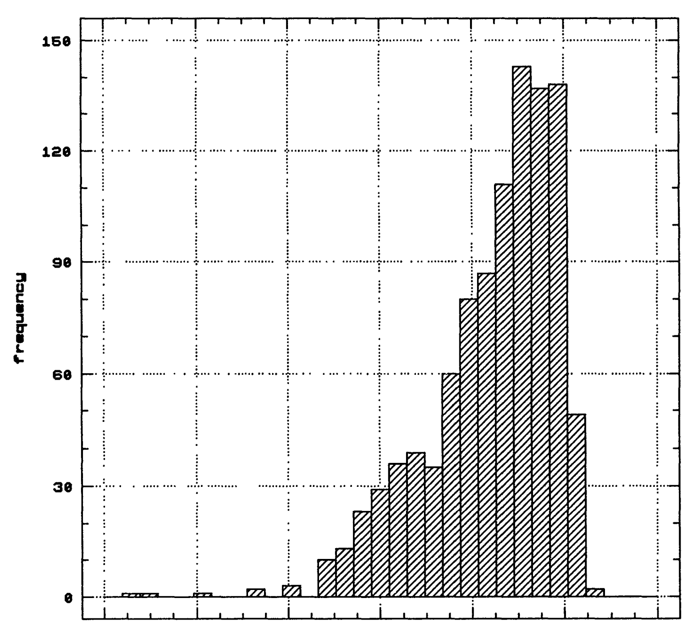

Figure 3. Bootstrap Histogram for φ—Stochastic Regression Example.

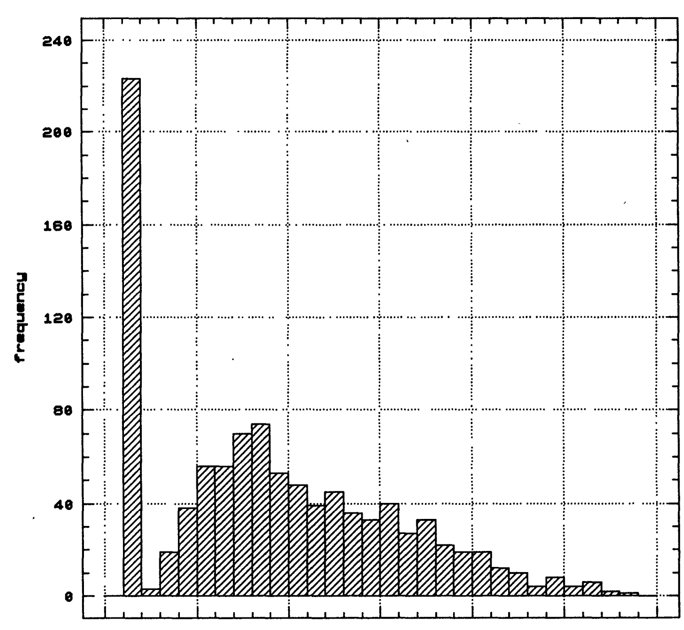

Figure 4. Bootstrap Histogram for σ_w—Stochastic Regression Example.

Table 6. ARIMA Modeling Example (T = 38)

	Gaussian ML		Bootstrap		Unconditional LS	
	Estimate	Nominal SE	Mean	SD	Estimate	Nominal SE
β_0	.0312	.0115	.0240	.0106	.0271	.0108
β_1	.9851	.1321	.9869	.0943	.9285	.0931
β_4	−.3136	.1627	−.2850	.2174	−.3293	.1733
σ_V	.5449	.0741	.4841	.1802	.5741	NA

$- B)(1 - B^4)z(t)$, $\text{var}\{v(t)\} = \sigma_v^2$, and B is the usual backward shift operator. For this problem, we used the state-space representation for ARIMA models described in Anderson and Moore (1979, pp. 113–114); the induced parameterization is $\boldsymbol{\theta} = [\beta_0, \beta_1, \beta_4, \sigma_v]$.

Table 6 compares the estimation results from GML, unconditional least squares (ULS), and the bootstrap. For small samples, ULS is inferior to GML as is evident, for example, by the difference between the ULS and the GML estimates of β_1. Review of the standard errors reveals that, except for the constant term β_0, the asymptotic nominal standard errors obtained via GML are considerably different from the bootstrap standard errors.

Based on GML, the t ratio for the estimate of β_4 is -1.93, which is borderline significant, whereas the t ratio for the GML estimate of β_4 based on the bootstrap standard error is -1.44, which is not even significant at the 15% level. Further investigation revealed that the removal of this parameter did not seriously affect the results of the analysis, although there was some indication (based on the autocorrelation of the residuals) that this parameter would be needed for a longer series length if all else remained fairly constant. Moreover, the GML estimate of β_1 is very close to the noninvertibility region and, as discussed in Section 4.2, estimation near the boundary of invertibility will be unstable. This situation suggests that inspection of the bootstrap distribution will be more meaningful than inference based on the standard error associated with the estimate of β_1.

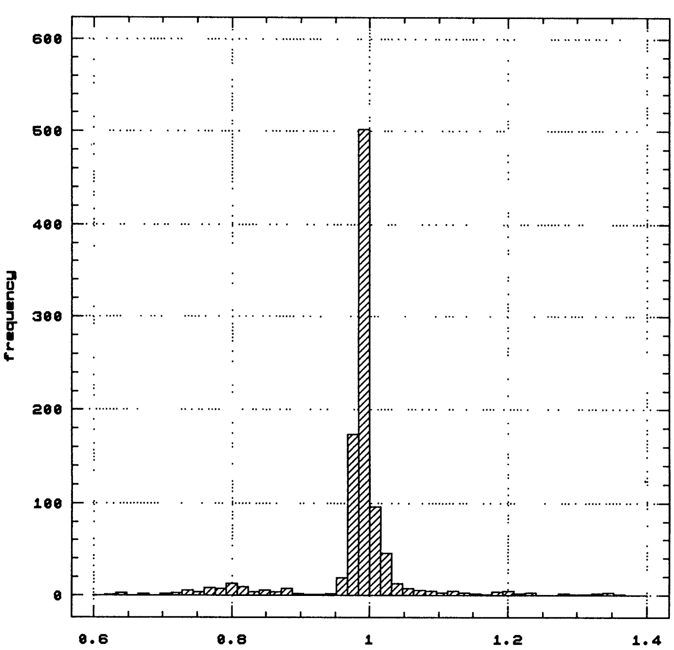

Figure 5. Bootstrap Histogram for β_1—ARIMA Modeling Example.

Figure 5 presents the bootstrap histogram corresponding to the estimate of β_1, which, as expected, demonstrates the departure from normality of the GML estimator due to the boundary problem. It is interesting to note that the lower and upper quartiles of this highly peaked distribution are .983 and .997, respectively, and approximately 22% of the estimates lead to a noninvertible model. The distributions corresponding to the other model parameters showed fairly symmetric distributions, and estimation and inference based on the bootstrap standard errors is comparable with estimation and inference based on the quantiles of these distributions.

5. COMPUTATIONAL CONSIDERATIONS

All our work was carried out on a Compaq Deskpro 386/20 using the Gauss programming language (version 2.0). The solution to the maximization of (2.7) was obtained via the Broyden–Fletcher–Goldfarb–Shanno update algorithm using a back step one-dimensional search (Dennis and Schnabel 1983). All gradients were calculated numerically using finite first differencing. On average, one iteration with this algorithm took between 2.5 and 3.0 seconds when $T = 50$ and approximately 5.0 seconds when $T = 100$ in the simulated data experiments. The Newbold and Bos data required approximately 2.8 seconds per iteration. The Johnson and Johnson data required approximately 4.0 seconds per iteration (ULS was performed using Minitab). In most samples, convergence to a parameter estimate was obtained in less than 15 iterations so that total time for one parameter estimation was on the order of 45 seconds for samples of length $T = 50$ and on the order of 75 seconds for $T = 100$. Thus, to generate all bootstrap estimates when $N = 1{,}000$, approximately 12.5 hours of computation were required when $T = 50$, while approximately 21 hours were required when $T = 100$. In the small-sample case this means that a complete bootstrap study can be completed overnight. One merely initializes the bootstrap program before leaving the office for the evening and returns in the morning to a complete set of estimates; in effect, this type of computational burden is not costly.

APPENDIX: ASYMPTOTIC JUSTIFICATION OF THE PROCEDURE

We show that under appropriate conditions (namely, the conditions needed to obtain an asymptotic theory for the GML estimator), the bootstrap procedure applied to the innovations yields asymptotically consistent standard errors. In particular, we show that, if the number of bootstrap replications N is such that $N \rightarrow$

∞ while the sample length $T \to \infty$, the asymptotic distributions of the GML estimator and the bootstrap estimator are equivalent. For simplicity and ease of notation we put $N = T$ throughout the Appendix. Furthermore, we assume that $\mathbf{F}(\boldsymbol{\theta})$ has all its eigenvalues within the unit circle and $\mathbf{H}(\boldsymbol{\theta})\mathbf{Q}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})' + \mathbf{R}(\boldsymbol{\theta})$ is positive definite for $\boldsymbol{\theta} \in \mathscr{P}$—these assumptions assure the global asymptotic stability of the filter (Watanabe 1985, proposition 1).

Since our results for the bootstrap depend on the existence of an asymptotic theory for the GML estimator in the state-space model, we briefly outline the results on the consistency and asymptotic normality of the GML estimator. For details, refer to Ljung and Caines (1979). Let $\boldsymbol{\theta}_0$ be the finite-dimensional unknown, $k \times 1$, true parameter vector that uniquely determines $\mathbf{F} = \mathbf{F}(\boldsymbol{\theta}_0)$, $\mathbf{G} = \mathbf{G}(\boldsymbol{\theta}_0)$, $\mathbf{H} = \mathbf{H}(\boldsymbol{\theta}_0)$, $\mathbf{D} = \mathbf{D}(\boldsymbol{\theta}_0)$, $\mathbf{Q} = \mathbf{Q}(\boldsymbol{\theta}_0)$ and $\mathbf{R} = \mathbf{R}(\boldsymbol{\theta}_0)$. Let $\hat{\boldsymbol{\theta}}_T$ be the consistent, asymptotically normal estimate of $\boldsymbol{\theta}_0$ obtained by minimizing

$$V_T(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^{T} \{\log|\boldsymbol{\Sigma}(t, \boldsymbol{\theta})| + \mathbf{e}'(t, \boldsymbol{\theta})\mathbf{e}(t, \boldsymbol{\theta})\}$$
$$= -T^{-1}L(\boldsymbol{\theta} \mid \mathscr{Y}^T, \mathscr{X}^T)$$

[cf. (2.7)]. Note that $\mathbf{V}_T^{(1)}(\hat{\boldsymbol{\theta}}_T) = \partial V_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_T} = \mathbf{0}$, where superscript (1) denotes differentiation. Let $W_T(\boldsymbol{\theta}) = E[V_T(\boldsymbol{\theta})]$ and assume that $W_T(\boldsymbol{\theta})$ has a unique global minimum at $\bar{\boldsymbol{\theta}}_T$. Then, under appropriate conditions (Ljung and Caines 1979, theorem 1), as $T \to \infty$, $\hat{\boldsymbol{\theta}}_T - \bar{\boldsymbol{\theta}}_T \to 0$ almost surely and

$$T^{1/2}\mathbf{B}_T^{-1/2}(\hat{\boldsymbol{\theta}}_T - \bar{\boldsymbol{\theta}}_T) \xrightarrow{\mathscr{D}} N(\mathbf{0}, \mathbf{I}), \tag{A.1}$$

where

$$\mathbf{B}_T = [\mathbf{W}_T^{(2)}(\bar{\boldsymbol{\theta}}_T)]^{-1}\mathbf{U}_T(\bar{\boldsymbol{\theta}}_T)[\mathbf{W}_T^{(2)}(\bar{\boldsymbol{\theta}}_T)]^{-1} \tag{A.2}$$

and

$$\mathbf{U}_T(\bar{\boldsymbol{\theta}}_T) = TE\{\mathbf{V}_T^{(1)}(\bar{\boldsymbol{\theta}}_T)\mathbf{V}_T^{(1)}(\bar{\boldsymbol{\theta}}_T)'\}; \tag{A.3}$$

in (A.2) the superscript (2) refers to the $k \times k$ matrix of second-order partials of $W_T(\boldsymbol{\theta})$ with respect to the $k \times 1$ vector $\boldsymbol{\theta}$.

If, in addition, $W_T(\boldsymbol{\theta}) \to W(\boldsymbol{\theta})$ uniformly in $\boldsymbol{\theta}$ as $T \to \infty$ and $W(\boldsymbol{\theta})$ has a global minimum at $\boldsymbol{\theta}_0$, then under appropriate conditions (Ljung and Caines 1979, corollary 2),

$$T^{1/2}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{\mathscr{D}} N(\mathbf{0}, \mathbf{B}), \tag{A.4}$$

where

$$\mathbf{B} = [\mathbf{W}^{(2)}(\boldsymbol{\theta}_0)]^{-1}\mathbf{U}(\boldsymbol{\theta}_0)[\mathbf{W}^{(2)}(\boldsymbol{\theta}_0)]^{-1} \tag{A.5}$$

and $\mathbf{U}(\boldsymbol{\theta}_0) = \lim_{T\to\infty} \mathbf{U}_T(\bar{\boldsymbol{\theta}}_T)$.

We shall also make use of the following results given in Ljung and Caines (1979, lemma 1). Under the conditions for which (A.1) is true,

$$V_T(\boldsymbol{\theta}) - W_T(\boldsymbol{\theta}) \to 0 \quad \text{almost surely} \tag{A.6a}$$

and

$$\mathbf{V}_T^{(l)}(\boldsymbol{\theta}) - \mathbf{W}_T^{(l)}(\boldsymbol{\theta}) \to \mathbf{0}, \quad l = 1, 2, \tag{A.6b}$$

almost surely, as $T \to \infty$ uniformly in $\boldsymbol{\theta} \in \mathscr{P}$.

We now establish the asymptotic justification of the bootstrap procedure discussed in Section 3. Henceforth we assume all the conditions necessary to establish (A.1)–(A.6) for the GML estimator $\hat{\boldsymbol{\theta}}_T$ (as listed in Ljung and Caines 1979). Let $\{\mathbf{e}^*(1, \boldsymbol{\theta}), \ldots, \mathbf{e}^*(T, \boldsymbol{\theta})\}$ be a bootstrap sample of standardized innovations, put $V_T^*(\boldsymbol{\theta}) = T^{-1}\Sigma_{t=1}^T \{\log|\boldsymbol{\Sigma}(t, \boldsymbol{\theta})| + \mathbf{e}^{*\prime}(t, \boldsymbol{\theta})\mathbf{e}^*(t, \boldsymbol{\theta})\}$, and let $W_T^*(\boldsymbol{\theta}) = E^*\{V_T^*(\boldsymbol{\theta})\}$, where E^* denotes expectation with respect to the bootstrap distribution. We now establish the following lemma.

Lemma 1. $W_T^*(\boldsymbol{\theta}) = V_T(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \mathscr{P}$, and hence $\hat{\boldsymbol{\theta}}_T$ minimizes $W_T^*(\boldsymbol{\theta})$.

Proof. First, note that

$$E^*\{\mathbf{e}^{*\prime}(t, \boldsymbol{\theta})\mathbf{e}^*(t, \boldsymbol{\theta})\} = T^{-1}\sum_{j=1}^{T}\mathbf{e}'(j, \boldsymbol{\theta})\mathbf{e}(j, \boldsymbol{\theta}),$$

from which it follows that

$$V_T(\boldsymbol{\theta}) - W_T^*(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T}[\mathbf{e}'(t, \boldsymbol{\theta})\mathbf{e}(t, \boldsymbol{\theta})$$
$$- E^*\{\mathbf{e}^{*\prime}(t, \boldsymbol{\theta})\mathbf{e}^*(t, \boldsymbol{\theta})\}] = 0.$$

Let $\hat{\boldsymbol{\theta}}_T^*$ minimize $V_T^*(\boldsymbol{\theta})$. Then $\mathbf{0} = [\partial V_T^*(\boldsymbol{\theta})/\partial\theta_1, \ldots, \partial V_T^*(\boldsymbol{\theta})/\partial\theta_k]'|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_T^*} = \mathbf{V}_T^{*(1)}(\hat{\boldsymbol{\theta}}_T^*)$. Expand $\mathbf{V}_T^{*(1)}(\boldsymbol{\theta})$ at $\hat{\boldsymbol{\theta}}_T^*$ to obtain

$$\mathbf{0} = \mathbf{V}_T^{*(1)}(\hat{\boldsymbol{\theta}}_T^*) = \mathbf{V}_T^{*(1)}(\hat{\boldsymbol{\theta}}_T) + \mathbf{V}_T^{*(2)}(\boldsymbol{\eta}_T)(\hat{\boldsymbol{\theta}}_T^* - \hat{\boldsymbol{\theta}}_T), \tag{A.7}$$

where $\mathbf{V}_T^{*(2)}(\boldsymbol{\eta}_T)$ denotes the $k \times k$ matrix of second-order partials $\partial\mathbf{V}_T^{(1)}(\boldsymbol{\eta}_T)/\partial\boldsymbol{\theta}$ such that $\boldsymbol{\eta}_T$ is on the segment connecting the points $\hat{\boldsymbol{\theta}}_T^*$ and $\hat{\boldsymbol{\theta}}_T$ in $\mathscr{P}$. Then, whenever $\mathbf{V}_T^{*(2)}(\boldsymbol{\eta}_T)$ is invertible we have $(\hat{\boldsymbol{\theta}}_T^* - \hat{\boldsymbol{\theta}}_T) = [\mathbf{V}_T^{*(2)}(\boldsymbol{\eta}_T)]^{-1}\mathbf{V}_T^{*(1)}(\hat{\boldsymbol{\theta}}_T)$. Now, by Ljung and Caines (1979, theorem 1), we have

$$T^{1/2}\mathbf{B}_T^{*-1/2}(\hat{\boldsymbol{\theta}}_T^* - \hat{\boldsymbol{\theta}}_T) \xrightarrow{\mathscr{D}} N(\mathbf{0}, \mathbf{I})$$

as $T \to \infty$, where

$$\mathbf{B}_T^* = [\mathbf{W}_T^{*(2)}(\hat{\boldsymbol{\theta}}_T)]^{-1}\mathbf{U}_T^*(\hat{\boldsymbol{\theta}}_T)[\mathbf{W}_T^{*(2)}(\hat{\boldsymbol{\theta}}_T)]^{-1}, \tag{A.8}$$

$$\mathbf{U}_T^*(\hat{\boldsymbol{\theta}}_T) = TE^*\{[\mathbf{V}_T^{*(1)}(\hat{\boldsymbol{\theta}}_T)][\mathbf{V}_T^{*(1)}(\hat{\boldsymbol{\theta}}_T)]'\}, \tag{A.9}$$

and $\mathbf{W}_T^{*(2)}(\boldsymbol{\theta}) = E^*\{\mathbf{V}_T^{*(2)}(\boldsymbol{\theta})\}$.

To establish the bootstrap principle, we show that $\mathbf{B}_T^* - \mathbf{B}_T \to \mathbf{0}$ almost surely as $T \to \infty$ [cf. (A.2) and (A.8)]. Now, by Lemma 1, $\mathbf{W}_T^{*(2)}(\hat{\boldsymbol{\theta}}_T) = \mathbf{V}_T^{(2)}(\hat{\boldsymbol{\theta}}_T)$, and hence using (A.6), $\mathbf{W}_T^{*(2)}(\hat{\boldsymbol{\theta}}_T) - \mathbf{W}_T^{(2)}(\bar{\boldsymbol{\theta}}_T) \to 0$ almost surely as $T \to \infty$. Therefore, it remains to show that $\mathbf{U}_T^*(\hat{\boldsymbol{\theta}}_T)$ and $\mathbf{U}_T(\bar{\boldsymbol{\theta}}_T)$ differ by a small amount, component-wise, for sufficiently large T [cf. (A.3) and (A.9)]. We now state and prove this final result.

Lemma 2. $\mathbf{U}_T^*(\hat{\boldsymbol{\theta}}_T) - \mathbf{U}_T(\bar{\boldsymbol{\theta}}_T) \to \mathbf{0}$ almost surely as $T \to \infty$.

Proof. We shall suppress the argument $\boldsymbol{\theta}$ as needed to provide notational convenience and where its exclusion is obvious. Let $U_T^*(a, b) = TE^*\{(\partial V^*/\partial\theta_a)(\partial V^*/\partial\theta_b)\}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_T}$ be the (a, b)th element of the matrix $\mathbf{U}_T^*(\hat{\boldsymbol{\theta}}_T)$, for a, b, = 1, ..., k. To further ease the notation, let $C_{ta} = \partial[\log|\boldsymbol{\Sigma}(t)|]/\partial\theta_a$ and $Z_{ta}^* = \partial[\mathbf{e}^{*\prime}(t)\mathbf{e}^*(t)]/\partial\theta_a$ so that $TU_T^*(a, b) = E^*\{\Sigma_{t=1}^T\Sigma_{s=1}^T (C_{ta} + Z_{ta}^*)(C_{sb} + Z_{sb}^*)\}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_T}$. Now write

$$TU_T^*(a, b) = \sum_{t=1}^{T}\sum_{s=1}^{T} \{(C_{ta} + E^*Z_{ta}^*)(C_{sb} + E^*Z_{sb}^*)$$
$$- E^*(Z_{ta}^*)E^*(Z_{sb}^*) + E^*(Z_{ta}^*Z_{sb}^*)\}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_T}$$

so that

$$U_T^*(a, b) = TE^*(\partial V^*/\partial\theta_a)E^*(\partial V^*/\partial\theta_b)|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_T}$$
$$+ T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} \{E^*(Z_{ta}^*Z_{sb}^*) - E^*(Z_{ta}^*)E^*(Z_{sb}^*)\}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_T}. \tag{A.10}$$

The first term in (A.10) is the (a, b)th element of $T[\mathbf{W}_T^{*(1)}(\hat{\boldsymbol{\theta}}_T)][\mathbf{W}_T^{*(1)}(\hat{\boldsymbol{\theta}}_T)]'$ and hence is zero by Lemma 1. To evaluate the second term in (A.10) note that

$$E^*(Z_{ta}^*) = T^{-1}\sum_{j=1}^{T} Z_{ja}, \tag{A.11a}$$

$$E^*(Z_{ta}^*Z_{tb}^*) = T^{-1}\sum_{j=1}^{T} Z_{ja}Z_{jb}, \tag{A.11b}$$

and

$$E^*(Z_{ta}^*Z_{sb}^*) = T^{-2}\sum_{i=1}^{T}\sum_{j=1}^{T} Z_{ia}Z_{jb}, \quad s \neq t, \tag{A.11c}$$

where $Z_{ta} = \partial[\mathbf{e}'(t)\mathbf{e}(t)]/\partial\theta_a$. Substituting (A.11) into (A.10) we have

$$U_T^*(a, b) = \left\{\left[T^{-1}\sum_{t=1}^{T} Z_{ta}Z_{tb}\right] - \left[T^{-1}\sum_{t=1}^{T} Z_{ta}\right]\left[T^{-1}\sum_{s=1}^{T} Z_{sb}\right]\right\}\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_T}. \quad (A.12)$$

Next we evaluate

$$U_T(a, b) = TE\{(\partial V/\partial\theta_a)(\partial V/\partial\theta_b)\}|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}_T}.$$

Arguing as in (A.10), we may show that

$$U_T(a, b) = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\{E(Z_{ta}Z_{sb}) - E(Z_{ta})E(Z_{sb})\}|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}_T} \quad (A.13)$$

where we have used the fact that $\mathbf{W}_T^{(1)}(\bar{\boldsymbol{\theta}}_T) = \mathbf{0}$. Since $\mathbf{U}_T(\bar{\boldsymbol{\theta}}_T) \rightarrow \mathbf{U}(\boldsymbol{\theta}_0)$ as $T \rightarrow \infty$ [cf. (A.5)], it is clear that $U_T(a, b)$ is asymptotically equivalent to

$$T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\{E(Z_{ta}Z_{sb}) - E(Z_{ta})E(Z_{sb})\}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

But the true innovations, $\mathbf{e}(t, \boldsymbol{\theta}_0)$ and $\mathbf{e}(s, \boldsymbol{\theta}_0)$, are uncorrelated, $s \neq t$, so that (A.13) is asymptotically equivalent to

$$T^{-1}\sum_{t=1}^{T}\{E(Z_{ta}Z_{tb}) - E(Z_{ta})E(Z_{tb})\}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \quad (A.14)$$

By the stability of the filter, as $T \rightarrow \infty$, we have

$$T^{-1}\sum_{t=1}^{T} E\{Z_{ta}(\boldsymbol{\theta}_0)Z_{tb}(\boldsymbol{\theta}_0)\} \rightarrow \mu_4(a, b) \quad (A.15)$$

and

$$T^{-1}\sum_{t=1}^{T} E\{Z_{ta}(\boldsymbol{\theta}_0)\}E\{Z_{tb}(\boldsymbol{\theta}_0)\} \rightarrow \mu_2(a)\mu_2(b), \quad (A.16)$$

where

$$\mu_4(a, b) = E\{\partial[\mathbf{u}'(t)\boldsymbol{\Sigma}^{-1}\mathbf{u}(t)]/\partial\theta_a\ \partial[\mathbf{u}'(t)\boldsymbol{\Sigma}^{-1}\mathbf{u}(t)]/\partial\theta_b\}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

and

$$\mu_2(a) = E\{\partial[\mathbf{u}'(t)\boldsymbol{\Sigma}^{-1}\mathbf{u}(t)]/\partial\theta_a\}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

where $\{\mathbf{u}(t)\}$ denotes the steady-state innovation sequence and $\boldsymbol{\Sigma}$ denotes the steady-state error covariance matrix, $\boldsymbol{\Sigma} = E\{\mathbf{u}(t)\mathbf{u}'(t)\}$ (Anderson and Moore 1979, sec. 4.4).

In view of (A.12), it remains to show that $T^{-1}\Sigma_{t=1}^{T} Z_{ta}(\hat{\boldsymbol{\theta}}_T)Z_{tb}(\hat{\boldsymbol{\theta}}_T) \rightarrow \mu_4(a, b)$ and $T^{-1}\Sigma_{t=1}^{T} Z_{ta}(\hat{\boldsymbol{\theta}}_T) \rightarrow \mu_2(a)$ almost surely as $T \rightarrow \infty$. By Watanabe (1984, theorem 1) we have $T^{-1}\Sigma_{t=1}^{T}\mathbf{e}(t, \hat{\boldsymbol{\theta}}_T) = T^{-1}\Sigma_{t=1}^{T}\mathbf{e}(t, \boldsymbol{\theta}_0) + \mathbf{o}_{\text{a.s.}}(1)$, from which it follows by the differentiability and boundedness conditions on the $Z_{ta}(\boldsymbol{\theta})$ (Ljung and Caines 1979, assumption 2.7) that $T^{-1}\Sigma_{t=1}^{T} Z_{ta}(\hat{\boldsymbol{\theta}}_T) = T^{-1}\Sigma_{t=1}^{T} Z_{ta}(\boldsymbol{\theta}_0) + o_{\text{a.s.}}(1)$, and $T^{-1}\Sigma_{t=1}^{T} Z_{ta}(\hat{\boldsymbol{\theta}}_T)Z_{tb}(\hat{\boldsymbol{\theta}}_T) = T^{-1}\Sigma_{t=1}^{T} Z_{ta}(\boldsymbol{\theta}_0)Z_{tb}(\boldsymbol{\theta}_0) + o_{\text{a.s.}}(1)$. But, by (A.6), we may write

$$T^{-1}\sum_{t=1}^{T} Z_{ta}(\boldsymbol{\theta}_0) = T^{-1}\sum_{t=1}^{T} E\{Z_{ta}(\boldsymbol{\theta}_0)\} + o_{a.s.}(1) \quad (A.17)$$

and

$$T^{-1}\sum_{t=1}^{T} Z_{ta}(\boldsymbol{\theta}_0)Z_{tb}(\boldsymbol{\theta}_0) = T^{-1}\sum_{t=1}^{T} E\{Z_{ta}(\boldsymbol{\theta}_0)Z_{tb}(\boldsymbol{\theta}_0)\} + o_{\text{a.s.}}(1). \quad (A.18)$$

The lemma now follows from (A.17) and (A.18) by the stability of the innovation sequence—the same argument used to establish (A.15) and (A.16).

[*Received June 1984. Revised March 1991.*]

REFERENCES

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

Ansley, C. F., and Newbold, P. (1980), "Finite Sample Properties of Estimators for Autoregressive Moving Average Processes," *Journal of Econometrics*, 13, 159–183.

Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control* (rev. ed.), San Francisco: Holden-Day.

Burmeister, E., and Wall, K. D. (1982), "Kalman Filtering Estimation of Unobserved Rational Expectations With an Application to the German Hyperinflation," *Journal of Econometrics*, 20, 255–284.

Burmeister, E., Wall, K. D., and Hamilton, J. D. (1986), "Estimation of Unobserved Expected Monthly Inflation Using Kalman Filtering," *Journal of Business and Economic Statistics*, 4, 147–160.

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

Dennis, J., and Schnabel, R. B. (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Englewood Cliffs: Prentice-Hall.

Dent, W., and Min, A.-S. (1978), "A Monte Carlo Study of Autoregressive-Integrated-Moving Average Processes," *Journal of Econometrics*, 7, 23–55.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 7, 1–26.

Freedman, D. A. (1981), "Bootstrapping Regression Models," *The Annals of Statistics*, 9, 1218–1228.

Freedman,, D. A., and Peters, S. C. (1984a), "Bootstrapping a Regression Equation: Some Empirical Results," *Journal of the American Statistical Association*, 79, 97–106.

——— (1984b), "Bootstrapping an Econometric Model: Some Empirical Results," *Journal of Business and Economic Statistics*, 2, 150–158.

Findley, D. F. (1985), "On Bootstrap Estimates of Forecast Mean Square Errors for Autoregressive Processes," *17th Symposium on the Interface* (March 17–19, 1985), ed. D. M. Allen, Amsterdam: North-Holland.

Gardner, G., Harvey, A. C., and Phillips, G. D. A. (1980), "An Algorithm for Exact Maximum Likelihood Estimation of Autoregressive-Moving Average Models by Means of Kalman Filtering," *Applied Statistics*, 29, 311–322.

Goodrich, R. L., and Caines, P. E. (1979), "Linear System Identification from Nonstationary Cross-Sectional Data," *IEEE Transactions on Automatic Control*, 24, 403–410.

Harvey, A. C. (1981), *Time Series Models*, New York: Halsted Press.

Harvey, A. C., and Durbin, J. (1986), "The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling," (with discussion) *Journal of the Royal Statistical Society*, Ser. A, 149, 187–227.

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

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

Harvey, A. C., and Todd, P. H. J. (1983), "Forecasting Economic Time Series With Structural and Box–Jenkins Models: A Case Study" (with comments), *Journal of Business and Economic Statistics*, 1, 299–315.

Jones, R. H. (1980), "Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations," *Technometrics*, 22, 389–395.

Kitagawa, G., and Gersch, W. (1984), "A Smoothness Priors Modeling of Time Series With Trend and Seasonality," *Journal of the American Statistical Association*, 79, 378–389.

Ljung, L., and Caines, P. E. (1979), "Asymptotic Normality of Prediction Error Estimators for Approximate System Models," *Stochastics*, 3, 29–46.

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

Pagan, A. (1980), "Some Identification and Estimation Results for Regression Models With Stochastically Varying Coefficients," *Journal of Econometrics*, 13, 341–363.

Shumway, R. H. (1988), *Applied Statistical Time Series Analysis*, Englewood Cliffs, NJ: Prentice-Hall.

Shumway, R. H., and Stoffer, D. S. (1982), "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," *Journal of Time Series Analysis*, 3, 253–264.

Spall, J. C., and Wall, K. D. (1984), "Asymptotic Distribution Theory for the Kalman Filter State Estimator," *Communications in Statistics—Theory and Methods*, 13, 1981–2003.

Stine, R. A. (1985), "Bootstrap Prediction Intervals for Regression," *Journal of the American Statistical Association*, 80, 1026–1031.

Wall, K. D. (1987), "Identification Theory for Time-Varying Coefficient Regression Models," *Journal of Time Series Analysis*, 8, 359–371.

Watanabe, N. (1985), "Note on the Kalman Filter With Estimated Parameters," *Journal of Time Series Analysis*, 6, 269–278.