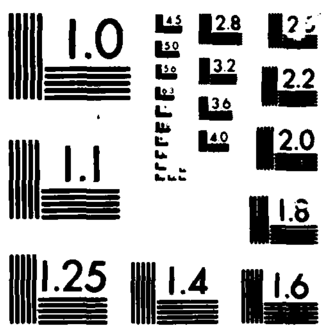
AD-A167 099 BOOTSTRAPPING NONLINEAR LEAST SQUARES ESTIMATES IN THE KALMAN FILTER MODE..(U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS D S STOFFER JAN 86 TR-86-02 1/1

UNCLASSIFIED AFOSR-TR-86-0252 F49620-85-C-0008 F/G 12/1 NL



MICROCOP CHART

(2)

AFOSR-TR- 86-0252

BOOTSTRAPPING NONLINEAR LEAST SQUARES
ESTIMATES IN THE KALMAN FILTER MODEL

David S. Stoffer*

Center for Multivariate Analysis
University of Pittsburgh

Center for Multivariate Analysis

University of Pittsburgh

BOOTSTRAPPING NONLINEAR LEAST SQUARES
ESTIMATES IN THE KALMAN FILTER MODEL

David S. Stoffer*

Center for Multivariate Analysis
University of Pittsburgh

Technical Report No. 86-02

January 1986

Center for Multivariate Analysis
University of Pittsburgh
515 Thackeray Hall
Pittsburgh, PA 15260

*Part of the work of this author was supported by the Air Force Office of Scientific Research under Contracts F49620-85-C-0008 and [redacted]R-84-0113. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

1. Introduction

The Kalman filter (KF) has become an important and powerful tool for the statistician. Recently, many authors have exploited the state-space model and KF recursions for estimation and prediction of time series. For example, Jones (1980) and Harvey and Pierse (1984) use the KF to obtain maximum likelihood estimates of the parameters of ARMA processes when observations are missing. It has been suggested by Morrison and Pike (1977) and others (cf. Kendall (1973)) that the KF model provides an appropriate setting within which to parametrize smoothing and forecasting problems.

To be specific, we suppose that a p×1 vector time series $\{y_t;\ t = 0, \pm 1, \pm 2, \ldots\}$ is being generated by the following dynamic system

$$y_t = x_t + v_t \tag{1.1}$$

where x_t is an unobservable zero mean, p×1 vector stationary stochastic signal, and v_t is p×1 Gaussian white noise, $v_t \sim N(0,R)$. The dynamics of the stationary signal is given by

$$x_t = \Phi x_{t-1} + w_t \tag{1.2}$$

where Φ is the p×p transition matrix and w_t is p×1 Gaussian white noise, $w_t \sim N(0,Q)$. Furthermore, $\{v_t\}$ and $\{w_t\}$ are mutually independent and we assume that the system and the filter have reached steady state. We remark that the superficially more general model in which (1.1) is replaced by

$$y_t = Mx_t + v_t$$

where M is a nonsingular known design matrix may be reduced to (1.1) by an appropriate change of bases.

Given the parameters of the model, namely, Φ, Q and R, one may obtain the minimum mean square error filter and forecasts for the system via the KF recursions.

However, the parameters are rarely known and hence must be estimated. Moreover, since the forecasts are based on the estimate of the state transition matrix Φ, the precision of the estimate must be evaluated. We propose the bootstrap as a method to evaluate the precision of the transition parameter estimates, in particular, to provide robustness against departure from normality in the Gaussian state and observation errors, and to assist in estimating forecast errors.

In most cases, parameter estimation for the KF model has been accomplished by maximum likelihood techniques involving the use of scoring or Newton-Raphson techniques to solve the nonlinear equations which result from differentiating the log-likelihood function (cf. Gupta and Mehra (1974)). Several examples have been given, notably by Ledolter (1979) and Goodrich and Caines (1979), which demonstrate the feasibility of these methods for several specific cases. Maximum likelihood estimation of parameters in the autoregressive moving average (ARMA) model expressed in state-space form has been considered by Harvey and Phillips (1979) and Jones (1980). The methods in the above references typically involve using a set of recursions for the derivatives of the log-likelihood and require that one invert a matrix of partial derivatives at each step. When the matrix of partials (or its expectation) is well behaved, the Newton-Raphson and scoring procedures enjoy quadratic convergence in the neighborhood of the maximum and one has a ready-made estimator for the covariance matrix of the parameters. We discuss the Newton-Raphson procedure for the KF model in Section 4.

Another maximum likelihood technique uses the EM algorithm to estimate the parameters of the KF model (cf. Shumway and Stoffer (1982)). Although this procedure is relatively simple and always increases the likelihood, the matrix of partials is never computed so that it is not available for providing estimates of the standard errors. However, the bootstrap may be able to augment this procedure by providing an approximation to the distribution of the parameter estimates.

The maximum likelihood techniques mentioned above require that one supply initial estimates (or starting values) which are sufficiently close to the true parameters. As will be seen, we shall require initial consistent estimates of Φ, Q and R, which converge faster that $n^{-1/4}$. Such estimates have been given by Anderson et. al. (1969). Their estimates, which are computationally simple to obtain, are discussed in Section 4. Further in Section 4, it is shown that when the aforementioned initial estimates are used, the one-step Newton-Raphson yields an efficient estimate of the transition parameter Φ when the noise processes are Gaussian. We make bootstrapping the Newton-Raphson estimate of Φ appealing by showing, in Section 5, that the bootstrap gives the right answers with large samples. That is, the bootstrap is at least as sound as the conventional asymptotics.

Finally, in Section 6, we give emperical evidence of the bootstrap's importance in Kalman filtering by comparing the bootstrap to the Newton-Raphson in the cases when the likelihood is Gaussian and when the likelihood is contaminated Gaussian.

Our goal is to estimate the precision of the parameter estimate of Φ as well as the precision of the forecasts $\hat{x}_{n+1}, \hat{x}_{n+2}, \ldots, \hat{x}_{n+k}$. The techniques used here are based on the bootstrap (cf. Efron (1979)) and the methods used in bootstrapping least squeares estimates discussed in Bickel and Freedman (1981), Freedman (1981) and Freedman and Peters (1984). It is noted in the above references that in regression models (static or dynamic), it is appropriate to resample the centered residuals after estimating the parameters. This is not possible in the present model (1.1) and (1.2) since the signal is not observable. However, we may base the procedure on the innovations which are obtained by taking the conditional expectation of the signal given the data. The bootstrap procedure will involve the resampling of the innovation sequence

$$r_t^{t-1} = y_t - x_t^{t-1} \qquad (1.3)$$

where by x_t^{t-1} we mean $E(x_t|y_{t-1},y_{t-2},\ldots)$. Of coures x_t^{t-1} will be obtained recursively via the KF.

Under the conditions stated in the next section we will be able to put this problem into the nonlinear regression context as discussed in Efron (1979, Section 7). That is, we may write

$$y_t = g_t(\Phi,Q,R,x_t,y_{t-1},y_{t-2},\ldots) + \varepsilon_t$$

where ε_t are iid zero mean random vectors (namely, the innovations) and $g_t(\cdot)$ is a particularly complicated, but known, nonlinear function of the parameters Φ, Q, and R, the signal x_t, and the data $y_{t-1},y_{t-2},\ldots$. In particular, $g_t(\cdot) = x_t^{t-1}$, the filtered value of the signal.

In the next section we give conditions under which we are able to bootstrap the innovations, (1.3). The bootstrap procedure is given in Section 3.

2. The Steady-State Innovation Sequence

Throughout the remainder of this paper we make the following assumptions on the p×p parameter matrices: (A1) Q and R are positive definite, and (A2) Φ is nonsingular with spectal norm, $\rho(\Phi)$, less than unity. These conditions ensure the asymptotic global stability of the KF (cf. Deyst and Price (1968)).

The steady-state KF recursions are given by (cf. Jazwinski (1970))

$$K = P(P+R)^{-1}, \tag{2.1a}$$

$$P = \Phi[P-P(P+R)^{-1}P]\Phi' + Q, \tag{2.1b}$$

$$x_{t+1}^{t} = \Phi x_t^t, \tag{2.1c}$$

$$x_t^t = x_t^{t-1} + K(y_t - x_t^{t-1}). \tag{2.1d}$$

In the KF above, K is the steady-state gain matrix, P is the steady-state prediction error, $P = E\{(x_t-x_t^{t-1})(x_t-x_t^{t-1})'\}$, and $x_t^{t-1} = E(x_t|y_{t-1},y_{t-2},\ldots)$ is the steady-state filter estimate of x_t based on the data $y_{t-1}, y_{t-2}, \ldots$.

Lemma 2.1 Under steady-state and optimal filtering, the p×1 vector innovation sequence

$$r_t^{t-1} = y_t - x_t^{t-1} \tag{2.2}$$

is a zero-mean, white Gaussian sequence with covariance matrix P+R.

Proof Write $r_t^{t-1} = e_t + v_t$ where $e_t = x_t - x_t^{t-1}$ and note that $E(e_t) = E(v_t) = 0$. The r_t^{t-1} are Gaussian since they are linear combinations of Gaussian random vectors. To establish the orthogonality of the innovations, it is easy to see that while $r_t^{t-1}(= y_t - x_t^{t-1})$ is in the linear space spanned by $\{y_t, y_{t-1}, \ldots\}$, $r_t^{t-1}(= e_t + v_t)$ is orthogonal to the linear space spanned by $\{y_{t-1}, y_{t-2}, \ldots\}$. Hence, for s < t,

$$E(r_s^{s-1} r_t^{t-1'}) = E\{r_s^{s-1} E(r_t^{t-1'} | y_s, y_{s-1}, \ldots)\} = 0.$$

Also, since e_t and v_t are uncorrelated we have that

$$\mathrm{Cov}(r_t^{t-1}) = \mathrm{Cov}(e_t) + \mathrm{Cov}(v_t) = P+R. \qquad \square$$

As a final remark, we note that via (2.1c), (2.1d), and (2.2) we may write y_t in terms of the steady-state innovations as

$$\begin{aligned} y_t &= \textstyle\sum_{j=1}^{\infty} \Phi^j K\, r_{t-j}^{t-j-1} + r_t^{t-1} \\ &= \textstyle\sum_{j=1}^{t-1} \Phi^j K\, r_{t-j}^{t-j-1} + r_t^{t-1} + \Phi^t x_o^o \end{aligned} \tag{2.3}$$

which follows from the fact that $||\Phi^j|| \to 0$ exponentially fast as $j \to \infty$ since $\rho(\Phi) < 1$, where $||\Phi||^2 = \mathrm{trace}(\Phi'\Phi)$. This result will be useful in establishing the bootstrap procedure.

3. The Bootstrap Estimate of Precision

As previously mentioned, the bootstrap technique will be employed by resampling the steady-state innovation sequence. Recall that under optimal filtering the innovation sequence r_t^{t-1}, $t=1,\ldots,n$ is $p\times 1$ Gaussian white noise, $r_t^{t-1} \sim N_p(0,P+R)$ where P is the steady-state error covariance matrix given in (2.1b).

The bootstrap procedure begins by estimating the parameters $\Theta = \{\Phi,Q,R\}$ of the model (1.1), (1.2) by the procedures mentioned in the Introduction. We shall discuss the particulars in Section 4. Call these estimates $\tilde{\Theta} = \{\tilde{\Phi},\tilde{Q},\tilde{R}\}$.

From these preliminary estimates obtain a suboptimal innovation sequence by filtering (cf. 2.1) under $\tilde{\Theta}$. Call this innovation sequence $\tilde{r}_t^{t-1}$. Make the sequence $\{\tilde{r}_t^{t-1}\}_{t=1}^n$ independent and identically distributed with distribution equal to the emperical distribution by putting mass n^{-1} on each innovation $\tilde{r}_t^{t-1}$, $t=1,\ldots,n$.

Next, draw a "bootstrap sample" of innovations, r_t^{*t-1}, $t=1,\ldots,n$ by independent random sampling of the residuals $\tilde{r}_t^{t-1}$. That is, sample the $\tilde{r}_t^{t-1}$, n times, with replacement from $\{\tilde{r}_1^0,\tilde{r}_2^1,\ldots,\tilde{r}_n^{n-1}\}$. From this we obtain a "bootstrap sample" of data $y_1^*,\ldots,y_n^*$ by setting (cf. 2.3)

$$y_t^* = \tilde{r}_t^{*t-1} + \sum_{j=1}^{t-1} \tilde{\Phi}^j \tilde{K} \tilde{r}_{t-j}^{*t-j-1}, \qquad t=1,\ldots,n \tag{3.1}$$

where $\tilde{K}$ is the estimated gain matrix obtained via filtering under parameters $\tilde{\Theta}$.

We make the following suggestions before proceeding with step (3.1). First, as suggested in Freedman (1981), one should center the residuals $\tilde{r}_t^{t-1}$ before resampling them so that the emperical distriubtion puts mass n^{-1} on $\tilde{r}_t^{t-1} - \tilde{\mu}_n$ where $\tilde{\mu}_n = n^{-1} \sum_{t=1}^n \tilde{r}_t^{t-1}$. Second, we suggest checking whether the innovations are nearly white. It is known that a suboptimal filter produces correlated innovations (see, for example, Mehra (1970)) and hence this is a check on the "goodness" of the estimates. Various methods are available for testing the whiteness of the innovations many of which are listed in Mehra (1970).

Now, suppose that the bootstrap data $\{y_1^*,\ldots,y_n^*\}$ come from the model

$$y_t^* = x_t + v_t^*, \qquad t \geq 1, \tag{3.2a}$$

$$x_t = \Phi^* x_{t-1} + w_t^*, \quad t \geq 1, \tag{3.2b}$$

where v_t^* is p×1 Gaussian white noise $v_t^* \sim N(0,R^*)$ and is independent of w_t^* which is p×1 Gaussian white noise $w_t^* \sim N(0,Q^*)$. Assume the parameters $\Theta^* = \{\Phi^*,Q^*,R^*\}$ are unknown and to be estimated.

The parameters Θ^* are then estimated by the initial optimal procedure to produce estimates $\tilde{\Theta}_1^* = \{\tilde{\Phi}_1^*,\tilde{Q}_1^*,\tilde{R}_1^*\}$. Then, the suboptimal innovation sequence is resampled and the bootstrap procedure is reiterated.

The entire process is repeated some large number "L" of times obtaining L bootstrap replications $\tilde{\Theta}_1^*,\tilde{\Theta}_2^*,\ldots,\tilde{\Theta}_L^*$. The distribution of the errors

$$\tilde{\Phi}^* - \tilde{\Phi} \tag{3.3}$$

are then computed to give an approximation as to the distribution of

$$\tilde{\Phi} - \Phi. \tag{3.4}$$

The bootstrap distribution of the errors (3.3) may then be used to obtain confidence regions and tests of hypotheses about the parameters Φ. Justification of this procedure is given in Section 5.

Forecasting k steps into the future, say $x_{n+j}^n = E(x_{n+j}|y_n,y_{n-1},\ldots)$, $j=1,2,\ldots,k$ is easily accomplished via the filter equations (2.1), namely

$$x_{n+j}^n = \Phi^j x_n^n, \qquad j=1,\ldots,k. \tag{3.5}$$

The suboptimal forecasts will be obtained via the KF under parameter estimates $\tilde{\Theta}$ so that

$$\tilde{x}_{n+j}^n = \tilde{\Phi}^j \tilde{x}_n^n, \qquad j=1,\ldots,k \tag{3.6}$$

will be the actual forecasts. If at each bootstrap replication we obtain

$$\{\tilde{x}^{n*}_{n+1},\ldots,\tilde{x}^{n*}_{n+k}\}^{(1)},\ldots,\{\tilde{x}^{n*}_{n+1},\ldots,\tilde{x}^{n*}_{n+k}\}^{(L)} \tag{3.7}$$

we may extract the emperical distribution of the forecast residuals

$$\tilde{x}^{n*}_{n+j} - \tilde{x}^{n}_{n+j}, \qquad j=1,\ldots,k \tag{3.8}$$

which can then be used to approximate the distribution of the actual forecast errors

$$\tilde{x}^{n}_{n+j} - x^{n}_{n+j}, \qquad j=1,\ldots,k. \tag{3.9}$$

From the distributions of (3.8) we may obtain prediction regions for the forecasts (3.5).

4. Parameter Estimation

In this section we give the details of the consistent and efficient estimation of the parameters of the KF model (1.1), (1.2). Recall that the system is in steady-state and the parameters $\Theta = \{\Phi,Q,R\}$ satisfy the conditions (A1) and (A2) given in Section 2. First, we discuss the initial consistent estimates given in Anderson et. al. (1969) and give related results. Second, we discuss the Newton-Raphson procedure for the KF model and in particular we show that the procedure is sound for the given model. We note that the assumption of normality of the error processes is not needed to establish the results of this section.

4.1 Initial Consistent Estimates

The following estimates are given in Anderson et. al. (1969). Let

$$\hat{\Phi}_n = (\textstyle\sum_{t=3}^{n} y_t y'_{t-2})(\sum_{t=3}^{n} y_{t-1} y'_{t-2})^{+}, \qquad n \geq 3 \tag{4.1}$$

where by + we mean generalized inverse. Further, define

$$\hat{B}_n(i) = n^{-1} \sum_{t=3}^{n}(y_t - \hat{\Phi}_n^i y_{t-i})(y_t - \hat{\Phi}_n^i y_{t-i})', \qquad n \geq 3, \quad i=1,2$$

and set

$$\hat{R}_n = \frac{1}{2}\{\hat{B}_n(1) + \hat{\Phi}_n^{-1}[\hat{B}_n(1) - \hat{B}_n(2)]\hat{\Phi}_n^{-1}{}'\} \tag{4.2}$$

and

$$\hat{Q}_n = \hat{B}_n(1) - \hat{R}_n - \hat{\Phi}_n \hat{R}_n \hat{\Phi}_n' \tag{4.3}$$

provided that $\hat{\Phi}_n$ is invertible.

Anderson et. al. (1969) show the strong consistency ($n \to \infty$) of $\hat{\Phi}_n$, $\hat{Q}_n$, and $\hat{R}_n$ for Φ, Q and R, respectively, under the model assumptions (1.1), (1.2). To establish the bootstrap principle in Section 5, we need the following results which exhibit the behavior of the suboptimal filter and forecasts (see Anderson et. al. (1969), Theorems 2.4, 2.5 and Corollary 2.4). Denote positive (semi)-definite by p.(s.)d.

Result 4.1. If Q is p.d. and if $\hat{\Phi}_n, \hat{Q}_n, \hat{R}_n$ are strongly consistent estimates of Φ,Q,R, respectively, for which $\hat{Q}_n$ is p.d. and R_n is p.s.d. for all $n \geq 1$, then $\hat{P}_n \to P$ and $\hat{K}_n \to K$ a.s. as $n \to \infty$ where $\hat{P}_n$ and $\hat{K}_n$ are the estimates of the steady-state filter covariance and gain matrices, respectively.

Result 4.2. Let the hypotheses of Result 4.1 be satisfied, and suppose $\rho(\Phi) < 1$. If in addition, $E\{|v_t|^k\} < \infty$ and $E\{|w_t|^k\} < \infty$ for some $k \geq 1$, then

$$\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} |x_t^t - \hat{x}_t^t|^k = 0 \qquad \text{a.s.}$$

and

$$\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} |x_{t+\ell}^t - \hat{\Phi}_n^\ell \hat{x}_t^t|^k = 0 \qquad \text{a.s.}$$

for any integer $\ell \geq 1$.

Anderson et. al. (1969) do not establish the asymptotic normality of the estimates given in (4.1), (4.2) and (4.3). This, however, is easily accomplished via the following theorem which may be found in parts in Hannan (1970, Chapter 4). First, we need some definitions. If $u_t = \sum_{j=-\infty}^{\infty} A_j \varepsilon_{t-j}$, $\sum_{-\infty}^{\infty} ||A_j|| < \infty$, and the ε_t are independent and identically distributed with mean vector zero and finite covariance matrix G, then we say that u_t is generated by a linear process. Define the sample autocovariance function of u_t from a sample of length n to be

$$C_n(h) = n^{-1} \sum_{t=h+1}^{n} u_t u_{t-h}' \; ; \; h=0,1,2,... \qquad (4.4)$$

and the autocavariance function of u_t to be $\Gamma(h) = E\{u_t u_{t-h}'\}$.

Theorem 4.1 Let u_t be generated by a linear process and suppose the fourth cumulant of ε_t is finite. Let $c_{ij}(h)$ and $\gamma_{ij}(h)$ denote the ij^{th} element of $C_n(h)$ and $\Gamma(h)$, respectively. Then (a) $C_n(h) \to \Gamma(h)$ a.s. as $n \to \infty$ for any h, and (b) for any integer $H > 0$ and integers $\ell(h)$, the joint law of

$$\sqrt{n}\, \{c_{ij}(\ell(h)) - \gamma_{ij}(\ell(h))\} \qquad i,j=1,...,p; \quad h=1,...,H$$

converges $(n \to \infty)$ to that of a zero-mean normal with asymptotic covariances

$$n\,\mathrm{Cov}(c_{ij}(m), c_{k\ell}(h))$$
$$\to \sum_{r=-\infty}^{\infty} \{\gamma_{ik}(r)\gamma_{j\ell}(r+h-m) + \gamma_{i\ell}(r+h)\gamma_{ij}(r-m) + \kappa_{ijk\ell}(0,m,r,r+h)\} \qquad (4.5)$$

where $\kappa_{ijk\ell}$ is the fourth cumulant function of u_t; m, h ε $\{\ell(1),...,\ell(H)\}$. The fact that $\kappa_{ijk\ell}$ is absolutely summable follows from the finiteness of the fourth cumulants of ε_t (cf. Hannan (1970) p. 211 for details). Note that if u_t is Gaussian the fourth cumulants vanish.

It is clear from Theorem 4.1 that since y_t given by (1.1) is generated by a linear process (cf. 2.3), the estimates given in (4.1), (4.2), and (4.3) are in fact strongly consistent: Simply note that $\Gamma(h) = \Phi^h \sum_{j=0}^{\infty} \Phi^j Q \Phi^{j'} + \delta_o^h R$ where δ is the Kronecker δ. Moreover, since the estimates are linear combinations of the asymptotically jointly normal variates $C_n(h)$, it is clear that $\sqrt{n}\,(\hat{\Phi}_n - \Phi)$, $\sqrt{n}\,(\hat{Q}_n - Q)$, and $\sqrt{n}\,(\hat{R}_n - R)$ are asymptotically normal with covariance matrices determined via (4.5). This establishes the desired rate of convergence needed for the Newton-Raphson. That is, $(\hat{\Phi}_n - \Phi)$, $(\hat{Q}_n - Q)$ and $(\hat{R}_n - R)$ are all $o_p(n^{-1/4})$. We note that we shall use the same order notation, o_p and O_p, for matrix as well as vector variates, no confusion should arise from this.

4.2 The Newton-Raphson Procedure

In this subsection we demonstrate the soundness of the Newton-Raphson procedure for the given KF model. The techniques used in this section will also help us establish the bootstrap principle in the next section.

So that we may explicitly exhibit a Newton-Raphson iteration we reparametrize the problem. Let P be as defined in (2.1b) and let $W = (P+R)^{-1}$ be the inverse of the covariance matrix of an innovation. We then consider the problem of estimating (Φ,P,W) via Newton-Raphson. Note that our original parameters (Φ,Q,R) are easily identified from (Φ,P,W), namely $Q = P - \Phi[P-PWP]\Phi'$ (cf. 2.1b) and $R = W^{-1} - P$. In this manner we may write (2.3) as

$$\begin{aligned} y_t &= \sum_{j=1}^{\infty} \Phi^j PWr_{t-j} + r_t \\ &= \sum_{j=1}^{t-1} \Phi^j PWr_{t-j} + r_t + \Phi^t x_o^o \end{aligned} \tag{4.6}$$

where we have dropped the superscript t-1 from the r_t's.

Let θ^o be the $k\times 1$ vector containing the distinct parameters of (Φ,P,W) and note that $k = p(2p+1)$ since Φ contains p^2, P contains $p(p+1)/2$, and W contains $p(p+1)/2$ distinct parameters. The Newton-Raphson procedure considers minimizing

$$\hat{Q}(\theta) = n^{-1} \sum_{t=1}^{n} \{\hat{W}^{\frac{1}{2}} [\hat{r}_t - \hat{G}_t(\theta^o-\hat{\theta})]\}'\{\hat{W}^{\frac{1}{2}} [\hat{r}_t - \hat{G}_t(\theta^o-\hat{\theta})]\} \tag{4.7}$$

where $\hat{\theta}$ is an initial estimate of θ^o, $\hat{r}_t = y_t - \hat{x}_t^{t-1}$ where $\hat{x}_t^{t-1}$ is obtained by running a KF under parameter estimates $\hat{\theta}$, and $\hat{G}_t$ is the $p\times k$ matrix of partials

$\partial r_t/\partial\theta$ evaluated at $\theta = \hat{\theta}$. $\hat{G}_t$ may be explicitly obtained via equation (4.6) by considering the model in canonical form. Specifically, let E be the nonsingular matrix for which $E^{-1}\Phi E$ is block diagonal. Then, let $s_t = E^{-1}x_t$, $e_t = E^{-1}w_t$, $z_t = E^{-1}y_t$ and $\eta_t = E^{-1}v_t$ in which case we transform the model (1.1), (1.2) to

$$z_t = s_t + \eta_t$$

and

$$s_t = \Lambda s_{t-1} + e_t$$

where $\Lambda = E^{-1}\Phi E$ is block diagonal. Then we may consider writing z_t in the form of (4.6) in which case Λ^j has a nice form (see Fuller (1976) p. 49).

In view of (4.7), the one-step Newton-Raphson estimate of θ^o is given by

$$\tilde{\theta} = \hat{\theta} + [n^{-1} \textstyle\sum_{t=1}^{n} \hat{G}_t'\hat{W}\hat{G}_t]^{-1}[n^{-1} \textstyle\sum_{t=1}^{n} \hat{G}_t'\hat{W}\hat{r}_t]. \qquad (4.8)$$

In the examples of Section 6, we shall consider the univariate case, p=1. Thus, it is worthwhile to give the explicit Newton-Raphson procedure for that case here:

1. Estimate Φ, Q and R via (4.1), (4.2) and (4.3), respectively. Call the estimates $\hat{\Phi}$, $\hat{Q}$ and $\hat{R}$.
2. Run a KF under $\hat{\Phi}$, $\hat{Q}$, $\hat{R}$, and $\hat{x}_o^o = 0$ to obtain $\hat{P}$, $\hat{K}$, and $\hat{x}_1^o,\ldots,\hat{x}_n^{n-1}$. Obtain the innovations $\hat{r}_t = y_t - \hat{x}_t^{t-1}$; t=1,...,n.
3. Calculate the partials via (4.6):

$$\hat{G}_t = [-\textstyle\sum_{j=1}^{t-1} j\hat{\Phi}^{j-1}\hat{P}\hat{W}\hat{r}_{t-j},\ -\textstyle\sum_{j=1}^{t-1} \hat{\Phi}^j\hat{W}\hat{r}_{t-j},\ -\textstyle\sum_{j=1}^{t-1} \hat{\Phi}^j\hat{P}\hat{r}_{t-j}]$$

   for t=2,...,n, $\hat{G}_1 = [0,0,0]$; where $\hat{W} = (\hat{P}+\hat{R})^{-1}$.
4. Update via (4.8):

$$\begin{bmatrix}\tilde{\Phi}\\ \tilde{P}\\ \tilde{W}\end{bmatrix} = \begin{bmatrix}\hat{\Phi}\\ \hat{P}\\ \hat{W}\end{bmatrix} + [\textstyle\sum_{t=2}^{n} \hat{G}_t'\hat{G}_t]^{-1}[\textstyle\sum_{t=2}^{n} \hat{G}_t'\hat{r}_t] .$$

We now establish the asymptotic properties of the Newton-Raphson estimate $\tilde{\theta}$ given in (4.8).

<u>Theorem 4.2</u>. Let Φ, Q, and R satisfy assumptions (A1) and (A2) given in Section 2. Let $\hat{\Phi}_n$, $\hat{Q}_n$, and $\hat{R}_n$ be the initial consistent estimates given in (4.1), (4.2), and (4.3), respectively, and let $\hat{\theta}$, as defined above, consist of these estimates. Further, let $\hat{x}_o^o$ be an estimator of x_o^o which is bounded in probability, and assume that r_t has finite fourth cumulant. Then

$$\sqrt{n}\,(\tilde{\theta} - \theta^o) \xrightarrow{\mathcal{L}} N_k(0, B^{-1}(\theta^o))$$

where $\tilde{\theta}$ is defined in (4.8), $k = p(2p+1)$, and $B(\theta^o) = \text{plim}_{n\to\infty}\, n^{-1} \sum_{t=1}^{n} \hat{G}_t' \hat{W} \hat{G}_t$ where $\hat{G}_t$ and $\hat{W}$ are defined in (4.7).

<u>Proof</u>. The proof parallels the proof of Fuller (1976, Theorem 8.3.1). See also Fuller (1976, Theorem 5.5.1 and Corollary 5.5.1). One must simply note that the elements of the matrices of first, second and third order partials of r_t with respect to θ converge to linear processes as $t \to \infty$ and Theorem 4.1a applies. One may then show that

$$(\tilde{\theta} - \theta^o) = [n^{-1} \sum_{t=1}^{n} G_t^{o\prime} W G_t^o]^{-1} \, [n^{-1} \sum_{t=1}^{n} G_t^{o\prime} W r_t] + o_p(n^{-1/2}),$$

where G_t^o is the $p\times k$ matrix $\partial r_t/\partial\theta$ evaluated at $\theta = \theta^o$, and that

$$\sqrt{n}\,[n^{-1} \sum_{t=1}^{n} G_t^{o\prime} W r_t] \xrightarrow{\mathcal{L}} N_k(0, B(\theta^o)).$$

The result of the theorem then follows. □

5. The Bootstrap Principle

In this section we justify the techniques established in Section 3. Throughout this section we replace the normality assumptions with the assumption that the noise processes w_t and v_t have finite fourth moments so that the observations y_t satisfy the conditions of Theorem 4.1. As in the previous section we drop the superscript t-1 from the innovations r_t^{t-1}.

Before proving the bootstrap principle given in Section 3, we state the following useful lemmas. First some notation is needed. If R^p is a p-dimensional space equipped with the Euclidean norm $|\cdot|$ and $\alpha \geq 1$, then $d_\alpha^p(\mu,\nu)$ is the distance between probability measures μ and ν in R^p defined as the infimum of $E\{|U-V|^\alpha\}^{1/\alpha}$ over all pairs of random vectors U with law μ and V with law ν (cf. Bickel and Freedman (1981)).

Lemma 5.1 Let $\Theta = (\Phi,Q,R)$ and $\hat{\Theta}_n = (\hat{\Phi}_n,\hat{Q}_n,\hat{R}_n)$ satisfy the conditions of Result 4.2. Let $\hat{F}_n$ be the emperical distribution function (e.d.f.) of the suboptimal innovations $\hat{r}_t$, t=1,...,n generated by $\hat{\Theta}_n$ and let F_n be the e.d.f. of the optimal innovations r_t t=1,...,n generated by Θ. Then $d_4^p(\hat{F}_n,F_n) \to 0$ almost surely (a.s.) as $n \to \infty$.

Proof Noting that $r_t = y_t - x_t^{t-1}$ and $\hat{r}_t = y_t - \hat{x}_t^{t-1}$, in view of Result 4.2, we have

$$d_4^p(\hat{F}_n,F_n)^4 \leq n^{-1} \sum_{t=1}^n |\hat{r}_t - r_t|^4$$

$$= n^{-1} \sum_{t=1}^n |x_t^{t-1} - \hat{x}_t^{t-1}|^4 \to 0 \quad \text{a.s.}$$

as $n \to \infty$. □

Lemma 5.2 Let F_n be the e.d.f. of the optimal innovations, r_t, t=1,...,n and let F be the common distribution of r_t. Then $d_4^p(F_n,F) \to 0$ a.s. as $n \to \infty$.

Proof Since the optimal steady-state innovations are iid (cf. Proposition 2.1) with finite fourth moments, this follows from Lemma 8.4 of Bickel and Freedman (1981). □

Now, let $C_n(h)$ be the sample autocovariance function of the observations y_t (cf. 4.4). Let $\psi_{n,h}(F)$ be the law of $C_n(h)$ when the law of r_t is F. Metrize the ψ's by $d_1^{p\times p}$ and the F's by d_2^p. Then we have the following lemma which is similar to Freedman (1984, Lemma 6.3), however, for the sake of completeness, we provide a proof.

Lemma 5.3 The $\psi_{n,h}(F)$ are equiuniformly continuous functions of F on

$$S = \{F: \int_{R^p} |r|^2 \, dF(r) \leq a^2 < \infty\}.$$

Proof. Fix F and F* in S. Construct iid random vectors r_t, r_t^*; t=1,...,n, so that r_t has law F and r_t^* has law F*, and

$$d_2^p \, (F,F^*)^2 = E\{|r_t - r_t^*|^2\}.$$

See Bickel and Freedman (1981, Lemma 8.1). Build y_t from the r_t and y_t^* from the r_t^* as in (2.3). Then, for $h \geq 0$

$$\begin{aligned} d_1^{p\times p}[\psi_{n,h}(F),\psi_{n,h}(F^*)] &\leq E\{|n^{-1} \textstyle\sum_{t=h+1}^{n} (y_t y'_{t-h} - y_t^* y^{*\prime}_{t-h})|\} \\ &\leq E\{|y_t y'_{t-h} - y_t^* y^{*\prime}_{t-h}|\} \\ &\leq E\{|y_t|\cdot|y_{t-h} - y^*_{t-h}|\} + E\{|y_t - y_t^*|\cdot|y^*_{t-h}|\}. \quad (5.1) \end{aligned}$$

We concentrate on the first term in (5.1), the second being treated similarly.

Now, by the Cauchy-Schwartz inequality and the fact that y_t is a linear process,

$$\begin{aligned} E\{|y_t|\cdot|y_s - y_s^*|\}^2 &\leq E\{|y_t|^2\} \, E\{|y_s - y_s^*|^2\} \\ &\leq c^2 \, E\{|y_s - y_s^*|^2\}. \end{aligned}$$

Using the fact that if U_j are independent random vectors, then

$$E\{|\Sigma_j U_j|^2\} \leq \Sigma_j \; E\{|U_j|^2\} + |\Sigma_j \; E\{U_j\}|^2$$

we have that in view of equation (2.3)

$$E\{|y_s - y_s^*|^2\} = E\{|\textstyle\sum_{j=1}^{\infty} \Phi^j \; K(r_{s-j} - r_{s-j}^*) + (r_s - r_s^*)|^2\}$$

$$\leq E\{|\delta\;|^2\} + E\{|\delta\;|^2\} \frac{\rho k}{1-\rho^2} + |E\{\delta\;\}|^2 \; [(\frac{\rho k}{1-\rho}) + 1]^2 \qquad (5.2)$$

where

$$\delta = r_1 - r_1^*, \quad \rho = ||\Phi|| < 1, \quad \text{and} \quad k = ||K||.$$

It is clear that (5.2) is small if F and F^* are close in d_2. □

Now, let starred variables denote those obtained via the bootstrap sample $\{y_1^*, \ldots, y_n^*\}$. In this manner we denote

$$C_n^*(h) = n^{-1} \textstyle\sum_{t=h+1}^{n} y_t^* y_{t-h}^{*\prime}, \quad h \geq 0$$

as the bootstrap counterpart of $C_n(h)$. Furthermore, let E, E^* denote expectation under F, $\hat{F}_n$, respectively. Let $Z_{ij}(t,h) = y_{ti} y_{t-h,j} - E(y_{ti} y_{t-h,j})$ and let $Z_{ij}^*(t,h) = y_{ti}^* y_{t-h,j}^* - E^*(y_{ti}^* y_{t-h,j}^*)$, $i,j=1,\ldots,p$; $h=0,1,\ldots$. Then $\bar{Z}_{ij}(h) = n^{-1} \sum_{t=1}^{n} Z_{ij}(t,h)$ will have the same ergodic properties as $\{c_{ij}(h) - \gamma_{ij}(h)\}$ (Hannan (1970) calls the matrix whose elements are $\bar{Z}_{ij}(h)$, $\tilde{C}_n(h) - \Gamma(h)$, and shows that $\tilde{C}_n(h)$ and $C_n(h)$ as we've defined it have the same limiting properties). For details see Hannan (1970, p. 208 and p. 228).

We now state the following theorem.

Theorem 5.1. Let y_t satisfy the conditions of Theorem 4.1. Then, along almost all sample sequences, as $n \to \infty$, conditionally on the data, for any integer $H > 0$ and integers $\ell(h)$,

(1) $C_n^*(\ell(h)) \to \Gamma(\ell(h))$ in conditional probability, and

(2) the joint conditional law of $\sqrt{n}\ \bar{Z}_{ij}^*(\ell(h))$ merges with the joint unconditional law of $\sqrt{n}\ \bar{Z}_{ij}(\ell(h))$, $i,j=1,\ldots,p$; $h=1,2,\ldots,H$.

Proof. The proof of part (1) follows from Lemmas 5.1, 5.2, and 5.3. That is, the conditional law of $C_n^*(h)$ given the data differs little in the sense of $d_1^{p\times p}$ from the unconditional law of $C_n(h)$ by Lemma 5.3, because the e.d.f. of the suboptimal innovations, $\hat{F}_n$, differs little in the sense of d_2^p from the law of the optimal innovations, F, by the combination of Lemmas 5.1 and 5.2.

We prove part (2) by showing that as $n \to \infty$, the joint conditional law of $\sqrt{n}\ \bar{Z}_{ij}^*(\ell(h))$ is the joint law described in Theorem 4.1.

Let r_t^* be iid $\hat{F}_n$ (appropriately centered) and let y_t^* be generated by (cf. 2.3)

$$y_t^* = \sum_{j=1}^{\infty} \hat{\Phi}_n^j \hat{K}_n\, r_{t-j}^* + r_t^*$$

where $\hat{\Phi}_n$ and $\hat{K}_n$ are the consistent estimates of Φ and K described in Section 4 such that $\rho(\hat{\Phi}_n) < 1$. For convenience, define p×p matrices $\hat{A}(j) = \hat{\Phi}_n^j \hat{K}_n$, $j=1,2,\ldots$, and $\hat{A}(0) = I$. Then for all h,

$$E^*\{y_t^* y_{t-h}^{*\prime}\} = E^*(\sum_{j=0}^{\infty} \hat{A}(j) r_{t-j}^*)(\sum_{k=0}^{\infty} \hat{A}(k) r_{t-h-k}^*)'$$

$$= \sum_{j=0}^{\infty} \hat{A}(j+h)\ E^*(r_t^* r_t^{*\prime})\ \hat{A}'(j)$$

$$= \sum_{j=0}^{\infty} \hat{A}(j+h)\ (n^{-1} \sum_{m=1}^{n} \hat{r}_m \hat{r}_m')\ \hat{A}'(j). \qquad (5.3)$$

Since by Lemmas 5.1 and 5.2, $d_2^p(\hat{F}_n,F) \to 0$ a.s. as $n \to \infty$, it follows that given the data, $E^*(r_t^* r_t^{*\prime}) \to Cov(r_t) = P+R$ a.s. as $n \to \infty$ (cf. Bickel and Freedman (1981), Lemma 8.3). Hence, we conclude from (5.3) that conditional on the data, as $n \to \infty$, $E^*(y_t^* y_{t-h}^{*\prime}) \to \Gamma(h)$, all h.

Note that for a,b,c,d, $i,j,k,\ell=1,\ldots,p$, and $t,u,v,w=0,\pm 1,\pm 2,\ldots$,

$$E^*(r^*_{ti} r^*_{uj} r^*_{vk} r^*_{w\ell}) = \begin{cases} n^{-1} \sum_{m=1}^{n} \hat{r}_{mi}\hat{r}_{mj}\hat{r}_{mk}\hat{r}_{m\ell} & t = u = v = w \\ (n^{-1} \sum_{m=1}^{n} \hat{r}_{ma}\hat{r}_{mb})(n^{-1} \sum_{m=1}^{n} \hat{r}_{mc}\hat{r}_{md}) & t,u,v,w \text{ equal in pairs but not all equal} \\ 0 & \text{otherwise} \end{cases}$$

where a,b and c,d correspond to the pairs of subscripts which are equal (e.g., if $t = u$ and $v = w$, then $a = i$, $b = j$, $c = k$, $d = \ell$). It is clear that the fourth cumulants of the r^*_t are finite and hence, so is the fourth cumulant function of the y^*_t. Thus, the y^*_t satisfy the conditions of Theorem 4.1. That is, if $y^*_1,\ldots,y^*_m$ is a (bootstrap) sample of size m, the joint law of $\sqrt{m}\,\bar{Z}^*_{ij}(\ell(h))$, $i,j=1,\ldots,p$; $h=1,\ldots,H$, $H > 0$ integer, converges ($m \to \infty$) to a zero-mean normal law with asymptoric covariances evaluated as in (4.5). Hence, part (2) follows if the conditional moments $E^*(y^*_{ti}y^*_{sj}) \equiv \hat{\gamma}_{ij}(t-s)$ and fourth cumulants $\hat{\kappa}_{ijk\ell}(0,t,u,v) \equiv E^*(y^*_{si}y^*_{s+t,j}y^*_{s+u,k}y^*_{s+v,\ell}) - \hat{\gamma}_{ij}(t)\hat{\gamma}_{k\ell}(v-u) - \hat{\gamma}_{ik}(u)\hat{\gamma}_{j\ell}(v-t) - \hat{\gamma}_{i\ell}(v)\hat{\gamma}_{jk}(u-t)$, converge ($n \to \infty$) a.s. to the unconditional values $\gamma_{ij}(t-s)$ and $\kappa_{ijk\ell}(0,t,u,v)$, respectively. We have already seen that $\hat{\gamma}_{ij}(t-s) \to \gamma_{ij}(t-s)$ a.s. as $n \to \infty$. Also, by Lemmas 5.1 and 5.2, $d^p_4(\hat{F}_n,F) \to 0$ a.s. as $n \to \infty$ from which it follows that the fourth conditional moments of y^*_t converges ($n \to \infty$) a.s. to the fourth moments of y_t (cf. Bickel and Freedman (1981), Lemma 8.3) which completes the proof. □

Let $\hat{\Phi}^*_n$, $\hat{Q}^*_n$, and $\hat{R}^*_n$ denote the bootstrap initial estimates of Φ, Q and R, respectively, obtained by evaluating (4.1), (4.2), and (4.3), respectively, with y_t replaced by y^*_t. In view of Theorem 5.1, we have that $\hat{\Phi}^*_n$, $\hat{Q}^*_n$, and $\hat{R}^*_n$ are consistent for Φ, Q, and R, respectively, in conditional probability with the desired convergence rate of $o^*_p(n^{-1/4})$. By $o^*_p(n^{-1/4})$ we mean a variate which is of smaller order in conditional probability than $n^{-1/4}$. These facts, of course, parallel those given in Section 4.1 for the original data y_t.

Next, we establish the appropriate asymptotics for the Newton-Raphson procedure involving the bootstrap data y_t^*; paralleling the results of Section 4.2 for the original data y_t. Recall that in the realm of the bootstrap the data y_t^* are generated via (cf. 3.1)

$$y_t^* = \sum_{j=1}^{\infty} \tilde{\Phi}^j \tilde{K} \tilde{r}_{t-j}^* + \tilde{r}_t^* \tag{5.4}$$

where $\tilde{\Phi}$ and $\tilde{K} = \tilde{P}\tilde{W}$ are the Newton-Raphson estimates obtained via (4.8); $\tilde{r}_t = y_t - \tilde{x}_t^{t-1}$, where $\tilde{x}_t^{t-1}$ is obtained via the KF under parameters $\tilde{\theta}$, and $\tilde{r}_t^*$ is obtained, as described in Section 3, by resampling the $\tilde{r}_t$.

Let $\hat{\theta}^*$ be the k×1 vector containing the appropriate elements of $\hat{\Phi}_n^*$, $\hat{P}_n^*$, and $\hat{W}_n^*$. The Newton-Raphson estimate of $\tilde{\theta}$ based on a bootstrap sample of size n is thus

$$\tilde{\theta}^* = \hat{\theta}^* + [n^{-1} \sum_{t=1}^{n} \hat{G}_t^{*'} \hat{W}_n^* \hat{G}_t^*]^{-1} [n^{-1} \sum_{t=1}^{n} \hat{G}_t^* \hat{W}_n^* \hat{r}_t^*] \tag{5.5}$$

where $\hat{r}_t^* = y_t^* - \hat{x}_t^{t-1*}$, $\hat{x}_t^{t-1*}$ is obtained via filtering under parameter estimates $\hat{\theta}^*$, and $\hat{G}_t^*$ is the p×k matrix of partials $\partial \tilde{r}_t^*/\partial \tilde{\theta}$ evaluated at $\tilde{\theta} = \hat{\theta}^*$. We now state the bootstrap principle in the following theorem.

<u>Theorem 5.2</u> Let $\tilde{\theta}^*$ be the k×1 bootstrap estimate given in (5.5) and let $\tilde{\theta}$ be given by (4.8). Then, along almost all sample sequences, as $n \to \infty$, conditional on the data, the law of $\sqrt{n}\,(\tilde{\theta}^* - \tilde{\theta})$ merges with the law of $\sqrt{n}\,(\tilde{\theta} - \theta^o)$ as given by Theorem 4.2.

<u>Proof</u>. The proof of this theorem will parallel that of Theorem 4.2. That is, in view of (5.4) the elements of the matrices of the first, second, and third order partials of $\tilde{r}_t^*$ with respect to $\tilde{\theta}$ are linear processes and we may show that conditional on the data

$$(\tilde{\theta}^* - \tilde{\theta}) = [n^{-1} \sum_{t=1}^{n} \tilde{G}_t' \tilde{W} \tilde{G}_t]^{-1} [n^{-1} \sum_{t=1}^{n} \tilde{G}_t' \tilde{W} \tilde{r}_t^*] + o_p^*(n^{-1/2}) \tag{5.6}$$

since $\hat{\Phi}_n^*$, $\hat{Q}_n^*$, and $\hat{R}_n^*$ have the desired convergence rate previously established. In (5.6), $\tilde{G}_t$ is the p×k matrix $\partial r_t/\partial\theta$ evaluated at $\theta = \tilde{\theta}$. Moreover, since $(\tilde{\theta} - \theta^o)$ is $o_p(n^{-1/4})$ and $n^{-1} \sum_{t=1}^n \tilde{G}_t'\tilde{W}\tilde{G}_t = O_p(1)$, we have that conditional on the data

$$(\tilde{\theta}^* - \tilde{\theta}) = [n^{-1} \sum_{t=1}^n G_t^{o'} W G_t^o]^{-1} [n^{-1} \sum_{t=1}^n G_t^{o'} W \tilde{r}_t^*] + o_p^*(n^{-1/2}) + o_p(n^{-1/2}). \quad (5.7)$$

Next, following Lemmas 5.1 and 5.2 we may show that $d_2^p(\tilde{F}_n, F) \to 0$ a.s. as $n \to \infty$, where $\tilde{F}_n$ is the e.d.f. of the $\tilde{r}_t^*$ and F is the common distribution of r_t. Hence we may easily establish (by paralleling Theorem 4.2) that conditional on the data, as $n \to \infty$

$$\sqrt{n}\, [n^{-1} \sum_{t=1}^n G_t^{o'} W \tilde{r}_t^*] \xrightarrow{\mathcal{L}} N_k(0, B(\theta^o))$$

from which, in view of (5.7), the theorem follows. □

6. Examples

In this section we submit two examples. The first example considers the bootstrap for the univariate KF model when the noise processes are Gaussian. In the second example, we consider the case when the noise processes are contaminated normals. For the sake of clarity, we first provide the step-by-step bootstrap procedure for the KF:

Given the data $y_1, \ldots, y_n$

1. Obtain the initial consistent estimates of Φ, Q, and R via equations (4.1), (4.2) and (4.3), respectively.
2. Filter (cf. 2.1) under $\hat{\Phi}_n$, $\hat{Q}_n$, and $\hat{R}_n$ to obtain $\hat{P}_n$, $\hat{K}_n$, and $\hat{r}_t = y_t - \hat{x}_t^{t-1}$; $t=1,\ldots,n$.
3. Obtain the Newton-Raphson estimates $\tilde{\theta}$ via (4.8). Also, see the discussion below (4.8).
4. Filter under $\tilde{\theta}$ and obtain the innovations $\tilde{r}_t = y_t - \tilde{x}_t^{t-1}$; $t=1,\ldots,n$. Center the $\tilde{r}_t$.

5. Sample with replacement, n times from $\{\tilde{r}_1,\ldots,\tilde{r}_n\}$ to obtain $\{\tilde{r}_1^*,\ldots,\tilde{r}_n^*\}$.

6. Obtain the bootstrap data $y_1^*,\ldots,y_n^*$ via (3.1).

7. Repeat steps 1, 2 and 3 using the bootstrap data yielding $\tilde{\theta}_1^*$, the first bootstrap estimate of $\tilde{\theta}$.

8. Repeat steps 5, 6, and 7 a large number "L" of times to obtain $\tilde{\theta}_1^*,\ldots,\tilde{\theta}_L^*$.

<u>Example 6.1</u> In this example we generated n=250 Gaussian observations from the KF model (1.1), (1.2) with parameters $\Phi = 0.8$, $Q = 4.0$, and $R = 1.0$. The one-step Newton-Raphson estimates were then bootstrapped L = 250 times and we compared the Newton-Raphson estimate of the standard error of $\tilde{\Phi}$, based on the asymptotics of Theorem 4.2, to the bootstrap estimate of the standard error of $\tilde{\Phi}$. The summary results of 30 such runs are given in Table 6.1. Also included in Table 6.1 is the emperical standard error of the Newton-Raphson estimate of Φ obtained from 2000 generated samples of length n = 250 observations from the model.

<u>TABLE 6.1</u> Standard Error of $\tilde{\Phi}$

	Mean	Standard Deviation	Bias
Bootstrap[a]	3.933×10^{-3} [b]	0.651×10^{-3} [b]	-0.166×10^{-3} [b,d]
Newton-Rapshon	1.380×10^{-3} [b]	0.479×10^{-3} [b]	———
Emperical	3.605×10^{-3} [c]	———	-0.026×10^{-3} [c,e]

Table 6.1: Summary of the estimates of the standard error of the Newton-Raphson estimate of Φ in the KF model with Gaussian noise for samples of length n = 250.

a: Based on L = 250 replications

b: Based on 30 runs

c: Based on 2000 runs

d: Average bias relative to the corresponding Newton-Raphson estimate

e: Bias relative to the true value of Φ.

Example 6.2 In this example we generated n = 250 contaminated normal observations from the KF model (1.1), (1.2) with parameters $\Phi = 0.8$, $Q = 4.0\ (90\%) + 16.0\ (10\%) = 5.2$ and $R = 1.0\ (90\%) + 9.0\ (10\%) = 1.8$. That is, the state noise is $N(0,4)$ with probability 90% and $N(0,16)$ with probability 10%, while the observation noise is $N(0,1)$ with probability 90% and $N(0,9)$ with probability 10%. The one-step Newton-Raphson estimates were then bootstrapped L = 250 times and we compared the estimates of the standard error of the state transition parameter estimate as in Example 6.1. Table 6.2 gives the summary of 30 runs and compares these with the emperical standard error based on 2000 runs.

TABLE 6.2 Standard Error of $\tilde{\Phi}$

	Mean	Standard Deviation	Bias
Bootstrap[a]	4.662×10^{-3} [b]	1.044×10^{-3} [b]	-0.464×10^{-3} [b,d]
Newton-Raphson	1.871×10^{-3} [b]	0.778×10^{-3} [b]	——
Emperical	4.197×10^{-3} [c]	——	-0.068×10^{-3} [c,e]

Table 6.2: Summary of the estimates of the standard error of the Newton-Raphson estimate of Φ in the KF model with contaminated Gaussian noise for samples of length n = 250

a: Based on L = 250 replications

b: Based on 30 runs

c: Based on 2000 runs

d: Average bias relative to corresponding Newton-Raphson estimate

e: Bias relative to the true value of Φ.

In each example, the advantage of the bootstrap is clear. In both examples, the bootstrap estimate of the standard error of $\tilde{\Phi}$ tended to be slightly larger than the emperical standard error, whereas the standard error of $\tilde{\Phi}$ obtained via the larger sample theory of the Newton-Raphson was always considerably smaller than the emperical value. Thus, the bootstrap has the desired property that the confidence and prediction regions obtained via the bootstrap will tend to be conservative. The bootstrap is clearly a perfect complement to the Newton-Raphson procedure.

References

Anderson, T.W. (1971). The Statistical Analysis of Time Series, Wiley, New York.

Anderson, W.N., Kleindorfer, G.B., and Kleindorfer, P.R., and Woodroofe, M.B. (1969). Consistent estimates of the parameters of a linear system. Ann. Math. Stat., 6, 2064-2075.

Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. Annals of Statistics, 6, 1196-1217.

Deyst, J.J. and Price, C.F. (1968). Conditions for asymptotic stability of the discrete minimum variance linear estimator, IEEE Trans. on Aut. Control, AC-13, 702-705.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife, Annals of Statistics, 7, 1-26.

Freedman, D.A. (1981). Bootstrapping regression models, Annals of Statistics, 9, 1218-1228.

Freedman, D.A. (1984). On bootstrapping two-stage least square estimates in stationary linear models, Annals of Statistics, 12, 827-842.

Freedman, D.A. and Peters, S.C. (1984). Bootstrapping a regression equation: some empirical results, Journal of the American Statistical Association, 79, 97-106.

Fuller, W.A. (1976). Introduction to Statistical Time Series, Wiley, New York.

Goodrich, R.L. and Caines, P.E. (1979). Linear system identification from non-stationary cross-sectional data. IEEE Trans. on Aut. Control AC-24, 403-411.

Gupta, N.K. and Mehra, R.K. (1974). Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations. IEEE Trans. Aut. Cont., AC-19, 774-783.

Hannan, E.J. (1970). Multiple Time Series, Wiley, New York.

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 Phillips, G.D.A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. Biometrika, 66, 49-58.

Jazwinski, A.H. (1970). Stochastic Processes and Filtering Theory. Academic Press, New York.

Jones, R.H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. Technometrics, 22, 389-395.

Kendall, M.G. (1973). Time Series, Hafner Pressm New York.

Ledolter, J. (1979). A recursive approach to parameter estimation in regression and time series problems. Comm. Statist. Theor. Meth., A8, 1227-1245.

Mehra, R.K. (1970). On the identification of variances and adaptive Kalman filtering, IEEE Trans. on Aut. Control, AC-15, 175-184.

Morrison, G.W. and Pike, D.H. (1977). Kalman filtering applied to statistical forecasting, Management Science, 23, 768-774.

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

Wu, C.F. (1983). On the convergence of the EM algorithm. Annals of Statistics, 11, 95-103.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

AD-A167 099

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 86-0252
6a. NAME OF PERFORMING ORGANIZATION: University of Pittsburgh; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code): Dept Mathmatics and Statistics, Pittsburgh, PA 15260	7b. ADDRESS (City, State and ZIP Code): Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR; 8b. OFFICE SYMBOL (If applicable): NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F49620-85-C-0008
8c. ADDRESS (City, State and ZIP Code): Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NOS.: PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2304; TASK NO. A5; WORK UNIT NO.

11. TITLE (Include Security Classification): BOOTSTRAPPING NONLINEAR LEAST SQUARES ESTIMATES IN THE KALMAN FILTER MODEL

12. PERSONAL AUTHOR(S): David S. Stoffer

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Technical	FROM ______ TO ______	86 January	25

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

ABSTRACT (Continue on reverse side if necessary and identify by block number)

The bootstrap is proposed as a method for estimating the precision of forecasts and maximum likelihood estimates of the transition parameters of the Kalman filter model when the estimates are obtained via Newton-Raphson. It is shown that when the system and the filter are in steady-state, the bootstrap applied to the Gaussian innovations yields asymptotically consistant standard errors. That the bootstrap works well with moderate sample sizes and supplies robustness against departures from normality is substantiated by emperical evidence.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL: Major Brian W. Woodruff	22b. TELEPHONE NUMBER (Include Area Code): (202) 767-5027; 22c. OFFICE SYMBOL: NM

DD FORM 1473, 83 APR   EDITION OF 1 JAN 73 IS OBSOLETE.   UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

END
FILMED

5-86

DTIC