SIGNAL EXTRACTION IN NONSTATIONARY SERIES

Peter Burridge and Kenneth F. Wallis

University of Warwick

NUMBER 234

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

SIGNAL EXTRACTION IN NONSTATIONARY SERIES

Peter Burridge and Kenneth F. Wallis

University of Warwick

NUMBER 234

August 1983

AMS 1980 subject classification. 90A20, 93E11.
Key words and phrases. Time series, signal extraction, nonstationarity, autoregressive models, Kalman filter, controllability, detectability, initial conditions, seasonal adjustment.

This research was supported by a grant from the Social Science Research Council.
The helpful comments of William P. Cleveland, Peter Crouch, Chris Rogers and Mark Salmon are gratefully acknowledged.

SUMMARY

SIGNAL EXTRACTION IN NONSTATIONARY SERIES

The state-space method is applied to the problem of separating an autoregressive (AR) signal from composite AR and white normal noise. In the stationary case, for which the Wiener filter exists, we show explicitly its equivalence to the steady-state Kalman filter. Existing results for difference-stationary processes are generalized to the explosive AR case, with careful attention paid to initial conditions; the limiting filter is shown to be stable. Conditions are given for convergence of the signal extraction error variance, and these are seen to exclude the existence of an unstable common factor in signal and noise autoregressions, but not nonstationarity. The general argument is illustrated with simple examples and the role of controllability and detectability is explored in an appendix.

1. INTRODUCTION

In the statistical analysis of economic time series, models in which the observed sequence is generated as the sum of two or more unobserved variables have a long history, recently surveyed by Nerlove et al. (1979, Ch.I). For example, in the early studies of economic cycles it was assumed that some pre-processing of the data was necessary, to rid the observed series of unwanted variation that would otherwise obscure the feature of principal interest. Also the traditional model for a variable observed at regular intervals each year decomposes the series into trend, cycle, seasonal and irregular components, and this model still underlies the seasonal adjustment procedures used by many statistical agencies. The signal extraction problem is to "estimate" the component of interest (the "signal") by a function of the observed data, and so separate it from the remainder, regarded as "noise". In addition to seasonal adjustment, a further application is in models that assume that an economic agent's behaviour depends on an unobserved component, such as "normal" price or "permanent" income, which is estimated as a function of current and past observed values of price or income, yielding a distributed lag relation among the observed variables.

This paper considers the extension of existing signal extraction theory, where the unobserved components are stationary stochastic processes, to a specific class of nonstationary models, namely those containing explosive autoregressions. It extends the work of Cleveland and Tiao (1976) and Pierce (1979), who consider series that can be reduced to stationarity by differencing. A motivation arises in studies of the linear filters implicit in the Census Bureau X-11 seasonal adjustment program. In seeking autoregressive models for which the one-sided filter for the adjustment of current data minimises the mean square of first annual revisions, no non-explosive solution is observed (Wallis, 1981). Likewise, explosive models are found in subsequent searches for models for which the X-11 program approximates the optimal signal extraction procedure, prompting the question of whether the standard algebraic formulae are indeed valid in these circumstances. More generally, given the recent interest in testing for unit roots in time series models, it seems appropriate to consider the extension of existing theory to cases that arise if a two-sided alternative hypothesis is postulated.

The basic set-up of the problem and the general approach we adopt are described in Section 2. Following Pagan's (1975) suggestion that unobserved-components models may be usefully analysed within the state-space framework, our approach uses the Kalman filter (Kalman, 1960), which is developed in Section 3. The steady-state Kalman filter is shown to be equivalent to the existing results for the stationary case, presented for example by Whittle (1963), but, in the context of a simple example of a series that comprises a first-order autoregressive signal and a purely random noise, it is shown to be applicable in the nonstationary case too. The convergence of the Kalman algorithm to a steady state is considered in Section 4, and conditions which guarantee this are presented. Under these conditions, the implied linear filter of the observations again corresponds to the existing results for the stationary case, but stationarity is not a necessary assumption. In the "nondetectable" case in which these conditions are not met, namely that in which the component autoregressions have an unstable common factor, the covariance of the signal extraction error diverges, and although the Kalman gain converges in numerical examples, a general proof is not yet available. Discussion of our results and their relation to the existing literature is contained in Section 5. In common with this literature it is assumed that the forms of the models and their parameter values are known, and problems of identification and estimation of unobserved-components models are not discussed.

2. SETTING UP THE PROBLEM

The observed variable y_t is given as the sum of two unobserved uncorrelated random processes respectively termed signal, s_t, and noise. The signal is the component of principal interest, and the noise is simply taken to be the remainder, $y_t - s_t$; the noise may in turn comprise more than one contributory component. The signal extraction or filtering problem is to estimate s_t from observations on y. The theory of the linear least squares (l.l.s.) approach to this problem for stationary time series, developed by Wiener and Kolmogorov, is treated extensively by Whittle (1963); see also Nerlove et al. (1979, ch. V) and Priestley (1971, ch. 10). Here the approach is to construct a linear function of the observations, or linear filter,

$$\hat{s}_t = f(L)y_t \tag{2.1}$$

where $f(L)$ is a polynomial in the lag operator L, so as to minimise the mean square error

$$E(\delta_t^2) = E(s_t - \hat{s}_t)^2. \tag{2.2}$$

Different filters result from different informational assumptions: one might consider only the finite sample $(y_o, y_1, \ldots, y_{t+k})$, or the semi-infinite sample $(y_\tau;\ \tau \leq t + k)$, or indeed the infinite sample $(y_\tau;\ -\infty < \tau < \infty)$.

As Whittle observes, "the l.l.s approach is theoretically elegant, in that it fits in naturally with the representation theory of stationary processes; concerned with canonical linear representations calculated from the covariances" of the observed process. However, given an unrestricted choice of function of the observations then the estimate of s_t which minimises the mean square error is the conditional expectation

$$\tilde{s}_t = E(s_t | y_o, \ldots, y_{t+k}). \tag{2.3}$$

This coincides with the l.l.s. estimate $\hat{s}_t$ if the random variables $s_t, y_o, \ldots, y_{t+k}$ are jointly normally distributed. In much of the time series literature attention is restricted to linear functions of the data irrespective of the joint distribution, the argument being that if this distribution is normal then the linear estimator is optimal in the mean square error sense, whereas if it is not "then, in general, we would be unable to evaluate the expression (2.3) and so we might as well seek the 'best' linear predictor", in Priestley's words. While it is easy to produce examples in which a linear estimator is absurd, these usually rest on explicit distributional assumptions that one would hope to be able to exploit in practical situations to construct the conditional expectation. In what follows it is often convenient to work with the conditional expectation, yet we seek equivalences to the existing l.l.s. results, and so we explicitly assume that the relevant random variables have a joint normal distribution.

We consider autoregressive models in which the stationarity condition on the roots of the autoregressive operator is relaxed. Thus we remain in the class of models with time-invariant parameters, and the particular evolutionary feature of the model is that, beginning from given initial conditions, the variance increases without limit. It is then inappropriate to consider a sample of observations extending from the infinite past to the present, and we take the relevant information set, denoted Ω_τ, to be the finite sample $y_o, y_1, \ldots, y_\tau$ together with some description of the initial conditions.

Initially we consider one-sided filters, that is, estimates of s_t based on Ω_t. As time goes by, and more y-observations become available, the estimate of the signal at time t can clearly be improved, and an asymmetric two-sided filter results. In the seasonal context this is the problem of preliminary adjustment and subsequent revision of seasonally adjusted data, and Wallis (1982) has shown that a set of (finite) linear filters can be constructed to represent the operation of the X-11 seasonal adjustment method. These are of the form

$$y_t^{(i)} = a_i(L)y_t = \sum_{j=-i}^{m} a_{i,j}\, y_{t-j}, \qquad i = 0,1,\ldots,m$$

where $y_t^{(i)}$ denotes the seasonally adjusted value of y_t based on observations up to time $t+i$, thus $y_t^{(o)}$ is the first-announced or preliminary adjusted value, and $y_t^{(m)}$ is the final or historical adjusted value. The filters are said to be "time-varying" because, on running the seasonal adjustment program at time τ, the m most recent values of the output series are $y_t^{(\tau-t)}$, $t = \tau-m,\ldots,\tau$, each a different linear filter of the input series.

If we think of the intermediate adjustment problem as that of obtaining the best (l.l.s.) estimate of $y_t^{(m)}$ given Ω_{t+i}, that is, data only up to time $t+i$, then the solution (Geweke, 1978; Pierce, 1980) is that the filter $a_i(L)$ that minimises the mean square of the revision

$y_t^{(m)} - y_t^{(i)}$ is given by the application of $a_m(L)$ to the series $y_{t-m},\ldots,y_t,\ldots,y_{t+i},\hat{y}_{t+i+1},\ldots,\hat{y}_{t+m}$, where the last $m-i$ values are the l.l.s. forecasts of the missing future y-values. Since these are linear filters of the observed y-values, the net effect is again a linear filter, namely $a_i(L)$. A set of filters constructed in this way is said to be internally consistent with respect to the given y-process (Wallis, 1981), and it minimises revisions throughout the whole sequence of adjustments. Thus if the filters $a_i(L)$ and $a_k(L)$ minimise the mean square of the revisions $y_t^{(m)} - y_t^{(i)}$ and $y_t^{(m)} - y_t^{(k)}$ respectively, being identical to the application of $a_m(L)$ to a series of observations augmented by $m-i$ and $m-k$ forecasts respectively, then $a_i(L)$, $i < k$, also minimises the mean square of the revision $y_t^{(k)} - y_t^{(i)}$ and is identical to the application of $a_k(L)$ to a series extended by $k-i$ forecasts. If the linear filters represent conditional expectations, then internal consistency is simply an implication of the "tower property" of conditional expectations, that is,

$$E(s_t|\Omega_{t+i}) = E\,\{E(s_t|\Omega_{t+k})|\Omega_{t+i}\}. \tag{2.4}$$

By regarding the signal, s_t, as an unobserved 'state', and the observations, y, as 'outputs', one may translate the signal extraction problem into a state-space estimation problem. The state estimation problem can be solved using the recursive formulae derived in Section 3, namely the Kalman filter. The recursions are derived by exploiting the fact that all the information about s_t available at time $t+i$ is contained in the state estimate $\hat{s}_{t,t+i}$ and the covariance $P_{t,t+i} = E\{(\hat{s}_{t,t+i} - s_t)^2\,|\Omega_{t+i}\}$, so that as new observations arrive, new estimates may be formed as a function of the old estimates and the new observations:

repeated filtering of the past history of the series is not necessary. We show in Section 3.1 that the state estimate may be updated by a recursion of the form

$$(2.5) \qquad E(s_t|\Omega_{t+i}) = E(s_t|\Omega_{t+i-1}) + K_{t+i}\{y_{t+i} - E(y_{t+i}|\Omega_{t+i-1})\}$$

where K_{t+i} is a function only of model parameters. Thus the update in the state estimate is a function of the "unanticipated" element in the new observation. Indeed, $E(s_t|\Omega_{t+i})$ may be written as a linear combination of such "innovations" in the observed series, and since this plays a crucial role in obtaining equivalences between the Kalman and Wiener filters, we give some further preliminaries.

The <u>innovation</u> in y_t is defined as

$$(2.6) \qquad \tilde{y}_t = y_t - E(y_t|\Omega_{t-1}), \; t \geqslant 0.$$

The sequence is initialized by assuming a distribution for y_o, thus Ω_{-1} comprises the information that y_o is normally distributed with mean μ_o and given variance, and $\tilde{y}_o = y_o - \mu_o$. The information sets

$$\Omega_{t+1} \equiv \{\Omega_{-1}, y_o, \ldots, y_{t+i}\}, \; \tilde{\Omega}_{t+i} \equiv \{\Omega_{-1}, \tilde{y}_o, \ldots, \tilde{y}_{t+i}\}$$

are identical, with the consequence that

$$(2.7) \qquad E(x|\tilde{\Omega}_{t+i}) = E(x|\Omega_{t+i})$$

for any random variable x for which either side of (2.7) exists. This follows immediately by repeated application of (2.6) to prove that $\Omega_{t+i} \subseteq \tilde{\Omega}_{t+i} \subseteq \Omega_{t+i}$. The innovations $\{\tilde{y}_o, \tilde{y}_1, \ldots, \tilde{y}_{t+i}\}$ form an uncorrelated sequence: if this were not so then Ω_{t-1}, which contains $\{\tilde{y}_o, \tilde{y}_1, \ldots, \tilde{y}_{t-1}\}$, would contain information about $\tilde{y}_t$, but it is immediate from (2.6) that $E(\tilde{y}_t | \Omega_{t-1}) = 0$. Further, if x and y_t are jointly normally distributed, then so are x and $\tilde{y}_t$, so that $E(x|\tilde{\Omega}_{t+i})$ is linear in $\tilde{y}_o, \ldots, \tilde{y}_{t+i}$.

As noted above, we are concerned to relax the usual stationarity assumption on unobserved-component models, and to relate the resulting treatment to existing results for the stationary case. Since the stationarity of an ARMA model is determined by its autoregressive operator, we consider autoregressive models in detail. Most of the features encountered in practice can be captured by a three-component model in which the "noise" is the sum of an autoregression and an independent white noise. Thus we consider the model

(2.8) $$y_t = s_t + n_t + \eta_t$$

where $\phi(L)s_t = \varepsilon_t, \quad \psi(L)n_t = \nu_t,$

the lag polynomials $\phi(L)$ and $\psi(L)$ being of degree p and r respectively, and the uncorrelated white noise variables ε_t, ν_t and η_t having variances σ^2_ε, σ^2_ν and σ^2_η respectively. We principally analyse this model in state-space form, writing the state transition equation and the measurement equation as

(2.9a) $$x_{t+1} = Fx_t + Gw_{t+1}$$

(2.9b) $y_t = H^T x_t + v_t$.

In general x_t and y_t denote the state vector and the output vector respectively, and w_t and v_t are independent serially uncorrelated normal random variables with zero means and covariance matrices Q and R respectively. In the present case y_t is a scalar, and the model (2.8) is expressed in the form (2.9) through the following definitions and equivalences:

$$x_t = (s_t, s_{t-1}, \ldots, s_{t-p+1}, n_t, n_{t-1}, \ldots, n_{t-r+1})^T$$

$$F = \left[\begin{array}{ccccc|cccc} \phi_1 & \phi_2 & \ldots\phi_{p-1} & \phi_p & & & & & \\ 1 & 0 & \ldots 0 & 0 & & & & & \\ 0 & 1 & \ldots 0 & 0 & & & 0 & & \\ \vdots & & & \vdots & & & & & \\ 0 & 0 & \ldots 1 & 0 & & & & & \\ \hline & & & & & \psi_1 & \ldots\psi_{r-1} & \psi_r & \\ & & & & & 1 & \ldots 0 & 0 & \\ & & 0 & & & \vdots & & \vdots & \\ & & & & & 0 & \ldots 1 & 0 & \end{array}\right] \qquad (2.10)$$

$$G^T = \begin{bmatrix} 1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \end{bmatrix}$$

$$H^T = \begin{bmatrix} 1 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \end{bmatrix}$$

$$w_t \equiv (w_{1t}, w_{2t})^T = (\varepsilon_t, \nu_t)^T, \; Q = \text{diag}\,(\sigma_\varepsilon^2, \sigma_\nu^2)$$

$$v_t = \eta_t, \; R = \sigma_\eta^2 \text{ (scalar)}.$$

Some slight redundancy in notation results from our wish to reconcile two traditions.

3. THE KALMAN FILTER, ITS RELATION TO THE WIENER FILTER, AND A SIMPLE EXAMPLE

3.1 First principles: the Kalman filter

In this section we give a derivation of the Kalman filter. The treatment is quite standard, and is given for completeness and to help fix ideas. We consider the state-space model (2.9), specializing to (2.10) in later sections.

It is assumed that in advance of any observations, the initial state vector x_o is known to be normally distributed with mean $\bar{x}_o$ and covariance $0 < P_o < \infty$. We derive the filter equations in four stages (cf. Anderson and Moore, 1979, § 3.1).

I. The random variable $(x_o^T, y_o^T)^T$ has mean $(\bar{x}_o^T, \bar{x}_o^T H)^T$ and covariance matrix

$$(3.1) \qquad \begin{bmatrix} P_o & P_o H \\ H^T P_o & H^T P_o H + R \end{bmatrix} .$$

Hence, conditional on y_o, x_o is normally distributed with mean and covariance given by

$$(3.2) \qquad \begin{aligned} \hat{x}_{o,o} &= \bar{x}_o + P_o H (H^T P_o H + R)^{-1} (y_o - H^T \bar{x}_o), \\ P_{o,o} &= P_o - P_o H (H^T P_o H + R)^{-1} H^T P_o . \end{aligned}$$

II. Conditional on y_o, x_1 is normal with mean and covariance

(3.3)
$$\hat{x}_{1,o} = F\,\hat{x}_{o,o}$$

$$P_{1,o} = F\,P_{o,o}F^T + G\,Q\,G^T\,.$$

III. It follows that conditional on y_o, the random variable $(x_1^T, y_1^T)^T$ is normally distributed with mean and covariance

(3.4)
$$\begin{bmatrix} \hat{x}_{1,o} \\ H^T\hat{x}_{1,o} \end{bmatrix}, \begin{bmatrix} P_{1,o} & P_{1,o}H \\ H^TP_{1,o} & H^TP_{1,o}H + R \end{bmatrix}.$$

IV. Thus conditional on (y_o,y_1), x_1 is normal with mean and covariance

(3.5)
$$\hat{x}_{1,1} = \hat{x}_{1,o} + P_{1,o}H(H^TP_{1,o}H + R)^{-1}\,(y_1 - H^T\,\hat{x}_{1,o})$$

$$P_{1,1} = P_{1,o} - P_{1,o}H(H^TP_{1,o}H + R)^{-1}\,H^T\,P_{1,o}\,.$$

We can now repeat steps II through IV and obtain the following recursions, valid for $t > 0$:

(3.6a) $$\hat{x}_{t,t} = \hat{x}_{t,t-1} + K_t\,(y_t - H^T\hat{x}_{t,t-1})$$

(3.6b) $$\hat{x}_{t+1,t} = F\,\hat{x}_{t,t}$$

(3.6c) $$P_{t,t} = P_{t,t-1} - K_t\,H^TP_{t,t-1}$$

(3.6d) $$P_{t+1,t} = F\,P_{t,t}F^T + G\,Q\,G^T$$

where $$K_t = P_{t,t-1} H(H^T P_{t,t-1} H + R)^{-1}.$$

These four equations constitute the Kalman filter. Equations (3.6a) and (3.6c) are sometimes called the updating equations, and (3.6b) and (3.6d) the prediction equations. Note that the final term in parentheses in (3.6a) is the innovation $\tilde{y}_t$, so that the equation is of the form of (2.5), with $i = 0$. It is useful to simplify (3.6c) by using the following identity, valid for nonsingular A,

$$A - AH(H^T AH + R)^{-1} H^T A = (A^{-1} + HR^{-1} H^T)^{-1},$$

and this yields the alternative form

$$P_{t,t} = (P_{t,t-1}^{-1} + HR^{-1}H^T)^{-1}. \tag{3.7}$$

The sequences of coefficients $\{K_t\}$, $\{P_{t,t}\}$ and $\{P_{t+1,t}\}$, are independent of the data, and their index has no necessary connection with the observation sequence. That is, they can be calculated "off line". Conditions under which these sequences have a limit or "steady state" as $t \to \infty$ are discussed in Section 4. Convergence to the steady state is conveniently studied by rearranging (3.6d) and (3.7) into two of the many possible forms of the <u>Riccati equation</u>:

$$P_{t+1,t+1} = \{(FP_{t,t}F^T + GQG^T)^{-1} + HR^{-1}H^T\}^{-1} \tag{3.8a}$$

$$P_{t+1,t} = F(P_{t,t-1}^{-1} + HR^{-1}H^T)^{-1}F^T + GQG^T. \tag{3.8b}$$

The choice of working with (3.6a) and (3.6b) depends on the particular form of the various parameter matrices, and for our purposes (3.8b) is more useful. Setting $P_{t,t-1} = P_{t+1,t} = P$, say, assuming that a steady state exists, gives the following expression for the steady-state one-step-ahead covariance

$$(3.9) \qquad P = F(P^{-1} + HR^{-1}H^T)^{-1}F^T + GQG^T.$$

Rearranging (3.6a) and (3.6b) then gives the steady-state filter recursion

$$(3.10) \qquad \hat{x}_{t,t} = F\hat{x}_{t-1,t-1} + K(y_t - H^T F\hat{x}_{t-1,t-1})$$

where $\quad K = PH(H^TPH + R)^{-1}.$

A further problem of interest is the "smoothing" problem, that of constructing $\hat{x}_{t,t+j}$. Two possible approaches to this problem are (a) to use a "fixed lag" smoother, obtaining a recursion for the sequence $\hat{x}_{t,t+j}, \hat{x}_{t+1,t+j+1}, \ldots$ in which j is constant, (b) to use a "fixed point" smoother, obtaining a recursion for the sequence $\hat{x}_{t,t+j}, \hat{x}_{t,t+j+1}, \ldots$ in which t is constant. These methods are discussed at length by Anderson and Moore (1979, Ch.7). The latter approach corresponds to the problem of revising a given seasonally adjusted figure as time goes by and more data become available, and best suits the purposes of Section 4. As shown in Appendix A, the steady-state recursion is

$$(3.11) \qquad \hat{x}_{t,t+j} = \hat{x}_{t,t+j-1} + K_j\tilde{y}_{t+j}$$

with $\quad K_j = P\{(I-HK^T)F^T\}^j H(H^TPH + R)^{-1}$, and P and K as defined at (3.9)-(3.10).

3.2 Equivalence of the Kalman filter and the Wiener filter in the stationary case

It has long been recognized that since the Kalman filter and the Wiener filter both yield the linear least squares solution to the signal extraction problem, they "must" be equivalent in situations in which both are defined. However explicit demonstrations of this equivalence in familiar time-series settings are not commonplace, and one is provided here. Moreover, this is not without interest to readers already familiar with state-space methods, since we see in Section 4 that the complete treatment of a class of models routinely employed in the statistical literature requires some extensions of existing state-space results.

We consider the model (2.8) in which the autoregressive components s_t and n_t are now assumed to be stationary processes, and in which the information set comprises the semi-infinite sample $(y_\tau; -\infty < \tau \leqslant t+k)$. With given initial conditions this is equivalent to the information set $(\tilde{y}_\tau; -\infty < \tau \leqslant t+k)$, and the stationarity assumption allows conditions in the infinite past to be neglected. In order to show that the l.l.s. estimate of s_t as presented by Whittle (1963) exactly corresponds to that delivered by the Kalman filter in steady state, it is convenient to work with the innovations $\tilde{y}_\tau$; developments of the l.l.s. results in the statistical literature are similarly based on the moving average representation of a stationary process. Equivalent formulations in terms of the observations, y_τ, for both stationary and nonstationary series are given in Section 4.2.

First, the Kalman filter for $\hat{x}_{t,t+k}$ (and hence $\hat{s}_{t,t+k}$) is expressed in terms of the innovations $\tilde{y}_T$. By using (3.11) recursively we obtain

(3.12) $$\hat{x}_{t,t+k} = \hat{x}_{t,t} + \sum_{j=1}^{k} K_j \tilde{y}_{t+j}.$$

Treating (3.10) as a difference equation and making repeated substitutions yields

(3.13) $$\hat{x}_{t,t} = \sum_{j=o}^{\infty} F^j K \tilde{y}_{t-j}.$$

Combining these equations and using the definition of K_j at (3.11) gives

(3.14) $$\hat{x}_{t,t+k} = \sum_{j=o}^{\infty} F^j K\tilde{y}_{t-j} + \sum_{j=1}^{k} P\{(I-HK^T)F^T\}^j H(H^T PH + R)^{-1}\tilde{y}_{t+j} .$$

Stationarity ensures that the eigenvalues of F lie inside the unit circle; as we see in Section 4, those of $(I-HK^T)F^T$ also lie inside the unit circle in circumstances in which the steady-state Kalman filter exists. Hence the coefficient of $\tilde{y}_{t+j}$, $-\infty < j \leq k$ in (3.14) may be obtained as the coefficient of z^{-j} in the generating function

(3.15) $$C(z) = \left(P\{I-z^{-1}(I-HK^T)F^T\}^{-1} + zF(I-zF)^{-1}P\right)H(H^T PH+R)^{-1}.$$

Secondly, to obtain the Wiener filter we introduce the covariance generating function of y, namely

$$g_{yy}(z) = g_{ss}(z) + g_{nn}(z) + \sigma_\eta^2$$

$$= \sigma_\varepsilon^2 \phi^{-1}(z)\phi^{-1}(z^{-1}) + \sigma_\nu^2 \psi^{-1}(z)\psi^{-1}(z^{-1}) + \sigma_\eta^2.$$

As shown in Appendix B, by virtue of the definitions and equivalences (2.10), this can be written as

$$(3.16) \qquad g_{yy}(z) = R + H^T(I-zF)^{-1}\, GQG^T(I-z^{-1}F^T)^{-1}H$$

which has the canonical factorization $g_{yy}(z) = W(z)W^T(z^{-1})$ where

$$(3.17) \qquad W(z) = \{1 + zH^T(I-zF)^{-1}FK\}\,(H^TPH + R)^{\frac{1}{2}}$$

$$= \sum_{j=o}^{\infty} w_j z^j, \quad \text{say.}$$

Since $H^TPH + R$ is the steady-state innovation variance $\sigma^2_{\tilde{y}}$ (compare (3.4)), this corresponds to the innovations representation

$$y_t = (H^TPH + R)^{-\frac{1}{2}} \sum_{j=o}^{\infty} w_j \tilde{y}_{t-j}.$$

Following Whittle (1963, Ch.6), the generating function for the coefficients on the innovations in the l.l.s. estimate of s_t given $\tilde{\Omega}_{t+k}$ is

$$C_s(z) = g_{ss}(z)\; W^{-T}(z^{-1})\; \sigma^{-1}_{\tilde{y}}$$

$$= g_{ss}(z)\; \{1 + z^{-1}K^TF^T(I-z^{-1}F^T)^{-1}H\}^{-1}(H^TPH+R)^{-1}$$

where the expansion is taken over powers of z from $-k$ to $+\infty$. Since s_t is the first element of x_t, to show that this gives the same result as the Kalman filter we need to show that

$$c_s(z) = s^T c(z),$$

where s^T is a row vector that selects the first element of the vector $c(z)$ defined in (3.15). In practice, it is more convenient to show that

$$g_{ss}(z) = s^T c(z)\sigma_{\tilde{y}} w^T(z^{-1}). \tag{3.18}$$

The polynomial on the right-hand side of (3.18) is

$$\left(P\{I-z^{-1}(I-HK^T)F^T\}^{-1} + zF(I-zF)^{-1}P\right)H\{1 + z^{-1}K^T F^T(I-z^{-1}F^T)^{-1}H\}$$

and on employing the identity

$$\{I-z^{-1}(I-HK^T)F^T\}^{-1} = (I-z^{-1}F^T)^{-1}\{I + z^{-1}HK^T F^T(I-z^{-1}F^T)^{-1}\}^{-1}$$

this reduces to

$$\{P(I-z^{-1}F^T)^{-1} + zF(I-zF)^{-1}P + F(I-zF)^{-1}\ PHK^T F^T(I-z^{-1}F^T)^{-1}\}H. \tag{3.19}$$

From (3.6c), (3.6d) and (3.11) we have

$$PHK^T = P - F^{-1}(P-GQG^T)F^{-T}$$

and substitution in (3.19) and rearrangement yields

$$s^T c(z)\sigma_{\tilde{y}} w^T(z^{-1}) = s^T(I-zF)^{-1}GQG^T(I-z^{-1}F^T)^{-1}H,$$

which is equal to $g_{ss}(z)$ as required (compare (3.16) and Appendix B).

The large amount of algebraic manipulation that is necessary to translate the state-space results into their classical form is mainly due to the fact that the Kalman filter yields only an implicit solution for the steady-state covariance matrices via the algebraic Riccati equations. However such equations may sometimes be solved analytically, by obtaining quadratic equations for individual elements of P, and this is the case in the example presented in the next section, in which the state vector is a scalar.

3.3 A scalar example

This example not only illustrates the previous equivalence of the two approaches, but also illustrates the Kalman filter's ability to handle nonstationary processes, and so motivates the general discussion of filter convergence in Section 4.

We consider the AR(1) signal plus white noise case:

$$s_t = \phi s_{t-1} + \varepsilon_t \qquad (3.20)$$
$$y_t = s_t + \eta_t$$

specializing the general state-space form (2.9) by setting $x_t = s_t$, $F = \phi$, $G = 1$, $w_t = \varepsilon_t$, $H^T = 1$, $v_t = \eta_t$, $Q = \sigma_\varepsilon^2$ and $R = \sigma_\eta^2$. The Riccati equation for the one-step-ahead covariance, (3.8b), now becomes

$$P_{t+1,t} = \phi^2(P_{t,t-1}^{-1} + \sigma_\eta^{-2})^{-1} + \sigma_\varepsilon^2 = h(P_{t,t-1}), \text{ say,} \qquad (3.21)$$

and (3.6a,b) yield the signal estimate recursion

$$\hat{s}_{t,t} = \phi\, \hat{s}_{t-1,t-1} + K_t \;(y_t - \phi\, \hat{s}_{t-1,t-1})$$

(3.22)
$$= \frac{\phi\, \sigma_\eta^2}{P_{t,t-1} + \sigma_\eta^2}\; \hat{s}_{t-1,t-1} + K_t y_t$$

$$= a_t\, \hat{s}_{t-1,t-1} + K_t\, y_t \;, \quad \text{say}\;,$$

with $K_t = P_{t,t-1}/(P_{t,t-1} + \sigma_\eta^2)$. The coefficient sequences $\{P_{t,t-1}\}$, $\{a_t\}$ and $\{K_t\}$ are independent of the data, and can be calculated off line. To consider the limits of these sequences we set $P_{t+1,t} = P_{t,t-1} = P$ in (3.21) to obtain the quadratic equation

(3.23)
$$f(P) = P^2 + \{\sigma_\eta^2\,(1-\phi^2) - \sigma_\varepsilon^2\}\, P - \sigma_\varepsilon^2\, \sigma_\eta^2 = 0.$$

This has real solutions of opposite sign since $f(0) < 0$. Choosing the positive solution, we also have $P > \sigma_\varepsilon^2$ since $f(\sigma_\varepsilon^2) < 0$. Moreover, $h'(P_{t,t-1}) > 0$ and $h''(P_{t,t-1}) < 0$ guarantee convergence of the iteration (3.21) to P for any finite $P_0 \geqslant 0$. The associated limits of the sequences $\{a_t\}$ and $\{K_t\}$ are, for any value of ϕ,

$$a = \frac{\phi\sigma_\eta^2}{P + \sigma_\eta^2}\;, \qquad K = \frac{P}{P+\sigma_\eta^2}\;,$$

and these coefficients give the steady-state Kalman filter as

(3.24)
$$s^*_{t,t} = a\, s^*_{t-1,t-1} + K\, y_t.$$

This signal extraction recursion can be "solved" to express $s^*_{t,t}$ as a linear filter of current and past y-observations with coefficients tending to zero provided that a is less than one in absolute value. This is obviously true for $|\phi| \leq 1$. To see that it is also true for $|\phi| > 1$ we notice that the steady state of the variance recursion (3.21) gives

$$a = \frac{P-\sigma_\varepsilon^2}{\phi P} ,$$

which is, again, obviously less than one in absolute value.

We now consider the Wiener filter in the stationary case. The process $y_t - \phi\, y_{t-1}$ has covariance generating function

$$(3.25) \qquad g(z) = \sigma_\varepsilon^2 + \sigma_\eta^2\,(1-\phi z)(1-\phi z^{-1})$$

$$= \sigma^2\,(1-\beta z)(1-\beta z^{-1}), \text{ say,}$$

where in the latter canonical factorization we choose σ^2 so that $|\beta| < 1$. The quadratic equation to be solved for β, which has a reciprocal pair of solutions, is obtained by setting $z = \beta$ in (3.25), whence

$$(3.26) \qquad \sigma_\varepsilon^2 + \sigma_\eta^2(1-\phi\beta)(1 - \frac{\phi}{\beta}) = 0 ,$$

while setting $z = \phi$ gives a relation from which the normalizing variance is then obtained:

$$(3.27) \qquad \sigma^2(1 - \beta\phi)(1 - \frac{\beta}{\phi}) = \sigma_\varepsilon^2 .$$

For this problem, the l.l.s. signal extraction filter based on current and past data is then given (Whittle, 1963, §6.3; Nerlove et al., 1979, § V.5) as

$$(3.28)\qquad \hat{s}_{t,t} = f_o(L)y_t = \frac{1-\beta/\phi}{1-\beta L}\,y_t \quad .$$

The associated mean square error is as follows, the first expression being obtained on simplifying Whittle's (6.3.7), and the second on using (3.26) above:

$$(3.29)\qquad E(s_t-\hat{s}_{t,t})^2 = \sigma_\varepsilon^2\,\frac{\beta/\phi}{1-\phi\beta} = \sigma_\eta^2\,(1-\frac{\beta}{\phi}) .$$

The invertibility condition $|\beta| < 1$ ensures that (3.28) has a convergent power series expansion and that the right-hand side of (3.29) is positive.

To check the equivalence of the steady-state Kalman filter to the Wiener filter (3.28) we first show that the quadratic equations solved in each case are equivalent, and then show that the filters coincide. For the first part, equating the steady-state current covariance, given by (3.7) as $P\sigma_\eta^2/(P+\sigma_\eta^2)$, to $\sigma_\eta^2(1-\beta/\phi)$ from (3.29) suggests a change of variable from β to P. Accordingly, on substituting for β in (3.26) and rearranging, the quadratic (3.23) is then obtained. Since the implied transformation is $P = \sigma_\eta^2(\phi/\beta-1)$, the positive solution for P clearly corresponds to the invertible solution for β. Finally, making this substitution in the coefficients of (3.24) gives

$$(3.30)\qquad s^*_{t,t} = \beta s^*_{t-1,t-1} + (1-\beta/\phi)y_t .$$

Thus the Kalman filter coincides with the Wiener filter (3.28) for $|\phi| < 1$, but is also defined, and takes the same form, for $|\phi| \geq 1$. Hence the Wiener filter apparatus is also applicable in nonstationary cases of this model, if reinterpreted as the steady-state Kalman filter under appropriate assumptions on initial conditions.

As more information arrives, we can continue to update the estimate of s_t by employing the steady-state filter

$$(3.31) \qquad s^*_{t,t+j} = s^*_{t,t+j-1} + K_j \tilde{y}_{t+j}$$

where K_j is obtained by specializing (3.11):

$$(3.32) \qquad K_j = P\,\{(1-\frac{P}{P+\sigma^2_\eta})\phi\}^j/(P+\sigma^2_\eta) = \beta^j(1-\beta/\phi),$$

using once more the above substitution for P.

In the stationary case the Wiener filter for "projection on the semi-infinite sample" is given, on translating Whittle's equation (6.3.8) into our present notation, as

$$(3.33) \qquad f_j(z) = \frac{1-\beta/\phi}{1-\beta z}\left[1 + \frac{\beta z^{-1}(1-\beta^j z^{-j})}{1-\beta z^{-1}}(1-\phi z)\right].$$

To obtain a recursion we observe that

$$f_j(z) - f_{j-1}(z) = \frac{1-\beta/\phi}{1-\beta z}(1-\phi z)\beta^j z^{-j},$$

and noting that the innovation $\tilde{y}_t$ is given as

$$\tilde{y}_t = y_t - E(y_t|\Omega_{t-1}) = \frac{1 - \phi L}{1 - \beta L} y_t$$

we then have

$$(3.34) \qquad \hat{s}_{t,t+j} - \hat{s}_{t,t+j-1} = (1 - \frac{\beta}{\phi}) \beta^j \tilde{y}_{t+j} ,$$

which coincides with the steady-state Kalman recursion (3.31). Again the standard expression for the l.l.s. filter in the stationary case is seen to be applicable in the nonstationary case under our assumptions.

4. EXISTENCE AND GENERAL CHARACTERISTICS OF THE STEADY-STATE KALMAN FILTER

4.1 Introduction

In Section 3.1 we presented the steady-state Kalman filter under the assumption that the sequence of one-step-ahead covariance matrices $\{P_{t,t-1}\}$ converges to a limit, and when signal and noise are both stationary this filter coincides with the Wiener filter, as was shown explicitly in Section 3.2. In the scalar example of Section 3.3 we saw that convergence to a steady state does not depend on stationarity, and that the filter weights could be obtained from the same polynomial quotient in the nonstationary as in the stationary case. In this section we give sufficient conditions for the covariance sequence to converge, and indicate the form of the steady-state filter in both stationary and nonstationary cases. In the literature on seasonal adjustment of economic time series, models in which the latter conditions do not hold are frequently encountered. The difficulties that this raises for the analysis of the limiting behaviour of the Kalman filter are discussed in Section 4.3.

To proceed we require the concepts, from linear state-space theory, of observability and controllability. Loosely speaking, the state-space realization (2.10) is said to be completely controllable if any initial state x_o may be driven to an arbitrary chosen state x_c in a finite number of steps by an appropriate sequence of inputs $\{w_t\}$, here regarded as control variables rather than random variables. Similarly the realization is completely observable if all possible movements in the state vector can eventually affect the observed process y_t. More formally, we have the following definitions (cf. Kailath, 1980a, p.135).

<u>Definition 4.1</u> λ is an <u>uncontrollable eigenvalue</u> of the pair (F,G) if there exists a row vector, $a^T \neq 0$, such that $a^T F = \lambda a^T$, and $a^T G = 0$.

<u>Definition 4.2</u> λ is an <u>unobservable eigenvalue</u> of the pair (F,H) if there exists a column vector, $b \neq 0$, such that $Fb = \lambda b$, and $H^T b = 0$.

<u>Definition 4.3</u> The pair (F,G) is <u>stabilizable</u> if either all eigenvalues of F are controllable, or its uncontrollable eigenvalues lie inside the unit circle.

<u>Definition 4.4</u> The pair (F,H) is <u>detectable</u> if either all eigenvalues of F are observable, or its unobservable eigenvalues lie inside the unit circle.

The major result we require, proved by Caines and Mayne (1970), is the following:

<u>Theorem</u> If R is positive definite, GQG^T is non-negative definite, the pair (F,G) is completely controllable, and the pair (F,H) is detectable, then for any positive semi-definite initial condition, P_o, the sequence $\{P_{t,t-1}\}$ converges to the unique positive definite solution, P, of the algebraic Riccati equation (3.10); furthermore, the eigenvalues of $F(I-KH^T)$ lie inside the unit circle.

As stated this is slightly less general than Caines and Mayne's theorems 2.2 and 2.3 which allow for correlation between w_t and v_t, but in the signal extraction problem these are usually taken to be independent. The general role of the controllability and detectability conditions in this theorem is examined by worked examples in Appendix C. In the next section we apply the preceding definitions and Caines and Mayne's theorem to establish sufficient conditions for convergence in the present problem.

4.2 The Steady-state Kalman filter: detectable case

Recalling the forms of G, H and F from (2.10), and writing $a^T = (a_1,\ldots,a_p,a_{p+1},\ldots,a_{p+r})$ we see that $a^T G = 0$ implies $a_1 = a_{p+1} = 0$, but then a^T cannot be a left eigenvector of F, as is clear by inspection. Thus, since F is of full rank and has no uncontrollable eigenvalues, in our problem (F,G) is completely controllable. Defining b in similar fashion, we see that $H^T b = 0$ implies $b_1 + b_{p+1} = 0$. Partitioning F in the obvious way, let

$$Fb = \begin{bmatrix} F_1 b^1 \\ F_2 b^2 \end{bmatrix} = \lambda \begin{bmatrix} b^1 \\ b^2 \end{bmatrix}$$

Then either λ is an eigenvalue of both F_1 and F_2, or one of b^1 and b^2 must be zero, but this contradicts $H^T b = 0$. Simple calculation shows that the eigenvalues of F_1 and F_2 are the solutions of $\phi(\lambda^{-1}) = 0$ and $\psi(\lambda^{-1}) = 0$ respectively. Thus (F,H) has unobservable eigenvalue λ if and only if $\phi(L)$ and $\psi(L)$ have common factor $(1-\lambda L)$; referring to definition 4.4 we see that (F,H) is detectable if and only if $\phi(\cdot)$ and $\Psi(\cdot)$ have no common roots on or outside the unit circle. Finally, our assumptions on Q and R clearly meet the requirements of Caines and Mayne's theorem. Thus the Kalman filter covariance converges to a steady state provided only that the signal and noise AR operators have no unstable common factor.

In the general detectable case (stationary or not) we now consider the linear filter of the observations, namely

$$\hat{x}_{t,t+k} = f_k(L)y_t,$$

that is implied by the steady-state Kalman filter. First, the one-sided filter $f_o(z)$ is obtained by repeated substitution in (3.10) as

$$(4.1) \qquad f_o(z) = \sum_{j=0}^{\infty} z^j\{(I-KH^T)F\}^j K.$$

Since the eigenvalues of $F(I-KH^T)$ lie inside the unit circle, by the Caines and Mayne theorem, the right-hand side is a convergent power series, and so we may write

$$(4.2) \qquad f_o(z) = \{I-z(I-KH^T)F\}^{-1}K.$$

To obtain $f_k(z)$ in general, we note that the innovation $\tilde{y}_{t+i}$ is given as

$$\tilde{y}_{t+i} = y_{t+i} - H^T F x_{t+i-1,\ t+i-1},$$

so that if this is regarded as a linear filter of the data, $\tilde{y}_{t+i} = h_i(L)y_t$, (4.3) $$h_i(z) = z^{-i} - H^T F z^{-i+1} f_o(z)$$

$$= z^{-i}\left(1 - zH^T F\{I-z(I-KH^T)F\}^{-1}K\right).$$

The second factor on the right-hand side, equal to $z^i h_i(z)$, may be simplified as follows, the second rearrangement resulting from application of the well-known formula for $(I+ab^T)^{-1}$:

$$z^i h_i(z) = 1 - zH^T F(I-zF)^{-1}\{I+zKH^T F(I-zF)^{-1}\}^{-1}K$$

$$= 1 - zH^T F(I-zF)^{-1}K\{1+zH^T F(I-zF)^{-1}K\}^{-1}$$

$$=\{1 + zH^T F(I-zF)^{-1}K\}^{-1}$$

$$=(H^T PH + R)^{\frac{1}{2}} W^{-1}(z),$$

which may be compared with (3.17), obtained in the stationary case. Since the right-hand side of (4.3) has a convergent power series expansion, so has $W^{-1}(z)$, and the canonical factorization (see Appendix B) is always invertible.

From (3.12) we then have

(4.4) $$f_k(z) = f_o(z) + \sum_{j=1}^{k} K_j z^{-j} (H^T PH+R)^{\frac{1}{2}} W^{-1}(z)$$

which, upon writing $f_o(z)$ as

$$(I-zF)^{-1}K\ (H^T PH+R)^{\frac{1}{2}}\ W^{-1}(z)$$

may be rearranged to yield

$$f_k(z) = \left((I-zF)^{-1}K + \sum_{j=1}^{k} K_j z^{-j}\right)(H^T PH+R)^{\frac{1}{2}} W^{-1}(z)$$

However, the first term in braces is just that part of $C(z)$, defined in (3.15), involving powers of z greater than or equal to $-k$, and so we may write

$$(4.5) \qquad f_k(z) = \left((I-zF)^{-1}GQG^T(I-z^{-1}F^T)^{-1}H\, W^{-T}(z^{-1})\right)_{-k} W^{-1}(z)$$

This is a generalization of Whittle's (6.1.13) in the sense that attention is no longer restricted to the stationary case. Also, $f_k(z)$ is a vector whose elements are the filters yielding $\hat{s}_{t,t+k}$, $\hat{s}_{t-1,t+k}, \ldots,$ $\hat{n}_{t,t+k}$, $\hat{n}_{t-1,t+k}, \ldots$. The first element also coincides with equation (4.5) of Pierce (1979), whose treatment admits difference-stationary signal and noise processes. However, Pierce also allows $\phi(L)$ and $\psi(L)$, in our notation, to have a common unit root, which has been excluded in the preceding discussion. Such a nondetectable case is discussed in the next section.

For computational purposes, (4.4) is to be recommended, since it presents no difficulties that might be caused by divergent power series even in explosive cases, as already noted. On the other hand, calculating (4.5) by convoluting the power series expansions of the four factors will produce numerical difficulties, since the first two factors are individually divergent in nonstationary cases.

4.3 The undetectable case

The results of the last section rested on the convergence of the covariance sequence $\{P_{t,t-1}\}$ to a limit P. The conditions given in the Caines and Mayne theorem are sufficient, but not, in general, necessary to ensure this. It appears from numerical calculations, however, that in the signal extraction context the covariance does diverge in undetectable cases, and the counterexample given in Appendix C has an undetectable state which is constant over time (corresponding to a 'noise' with zero variance).

As shown in Section 4.1, the pair (F,H) has an undetectable eigenvalue λ when the factor $(1-\lambda L)$ appears in both $\phi(L)$ and $\psi(L)$, and $|\lambda| \geq 1$; such "coincidental situations", in Granger's sense, are not of practical significance, except in the case $\lambda = \pm 1$ which does arise in the analysis of seasonal time series. That is, where the AR operator of the signal process contains the factor $(1-L^D)$, D being the seasonal periodicity, and that of the noise component, the factor $(1-L)$, which is thus common to both, giving the matrix F an undetectable unit root. Cleveland and Tiao (1976), give a formula for this case which corresponds to (4.5) with $k \to \infty$, and, as noted above, Pierce (1979) gives the formula for finite k which results from simplification of (4.5) when a common unit root is present. In order to derive their result, the former authors work explicitly with the infinite sample $\{y_\tau, -\infty < \tau < \infty\}$, rather than with the sequences $\{P_{t,t-1}\}$ and $\{K_t\}$ used in the present paper. Their method thus relies on the possibility that such a sample could be observed in nonstationary cases, and we prefer not to utilise this assumption for the reason given in Section 2. Pierce's approach, which mirrors that adopted by Whittle (1963, §8.5) for a difference-stationary signal in stationary noise, is to apply the formula appropriate to the stationary case, and then argue that the polynomial quotient so obtained yields a well defined stable filter. As suggested by Masani (1966), this line of argument, although leading to an intuitively plausible result, does not constitute a proof that the resulting filter delivers the l.l.s. estimate.

In numerical analysis of examples in which a common unit root is present, but all other roots of $\phi(L)$ and $\psi(L)$ lie on or <u>inside</u> the unit circle, we find that although the covariance $P_{t,t-1}$ increases without limit, the gain K_t nevertheless goes to a steady state, and the formula (4.5) still characterizes the filter weights. We note that in such a case the formula (4.5) is numerically stable, while (4.4), which is generally to be preferred, cannot be used since it requires the steady-state covariance, P, which no longer exists. It seems, then, that the formulae which have appeared elsewhere in the literature do indeed apply in the undetectable case, but a satisfactory proof does not yet appear to be available.

5. DISCUSSION

The state-space approach adopted in this paper has, in the past, found little favour in statistical time series analysis even though it has long been recognized that it is the most natural framework within which to handle many types of nonstationarity. In part this is the legacy of Wiener's work in the 1940's and 50's, in which the properties of stationary processes could readily be exploited to derive large-sample results. Thus the apparent dependence of Kalman's (1960) recursive method on initial conditions was seen as a disadvantage, as was the restriction to autoregressive representations (Sobel, 1967). While the fact that Kalman's approach can accommodate non-stationary processes has been a commonplace in control theory for twenty years, seen from the traditional perspective of statistical time series analysis the simplest approach has been to obtain results for stationary models and then allow the process parameters to approach suitable limits. Whatever merits this latter strategy might possess, a major shortcoming is that nothing can be said about processes whose parameters are neither in the stationary domain nor on its boundary. Thus the results for explosive autoregressive processes contained in Section 4.2 could not have been obtained by such means.

More rigorous derivations of large-sample extraction filters for signals with unit roots corrupted by stationary noise were offered by Hannan (1967) and Sobel (1967). Hannan argued in the frequency domain, seeking a filter which minimized the spectrum of the error process, while Sobel used projections in Hilbert space to directly obtain the l.l.s. estimate. Both approaches required the manipulation of quantities with unbounded variance, and the initial conditions imposed by Sobel were less natural than those required by the Kalman theory. More recently, Pierce (1979) has shown that application of a filter corresponding to (4.5) to a series in which signal and noise processes share a common unit root yields a signal estimate with unbounded variance. This result is in accordance with the discussion in Section 4.2 in which we show that such a common factor yields an undetectable unit eigenvalue of F. Unlike these earlier authors we have not treated moving average processes in detail, because nonstationarity has been our main concern; this omission is one which might usefully be remedied in future work.

None of the papers cited above discusses numerical matters, but the computational advantages of various representations of the basic filter (3.6) have attracted a great deal of attention in the state-space literature. If the eigenvalues of F are unfavourably distributed (a fortiori if one or more are undetectable) then the recursions (3.6) may become numerically unstable, while if F, G, H, Q and R are time invariant (as here) then many fewer calculations need to be performed than are required for direct implementation of (3.6). Such considerations are examined at length by Kailath (1980b). In the seasonal adjustment context the extreme sparseness of F, G and H should also be exploited to save computing time. It might appear that asymptotic filters for explosive processes are unlikely to be of use in handling a finite record, but numerical experience indicates that provided the roots of $\phi(z)$ and $\psi(z)$ are not too close together, $P_{t,t-1}$ (and hence K_t) settle down to a steady state very quickly, so that little is to be lost by employing the steady-state filter over the whole series.

The formal similarity of (A.3) to (2.9) may be exploited to obtain the following recursions, analogous to (3.6):

(A.4)

$$\text{(a)}\quad \hat{x}^{\dagger}_{r+1,r} = F^{\dagger}\hat{x}^{\dagger}_{r,r}$$

$$\text{(b)}\quad \hat{x}^{\dagger}_{r,r} = \hat{x}^{\dagger}_{r,r-1} + K^{\dagger}_{r}(y_r - H^{\dagger^T}\hat{x}^{\dagger}_{r,r-1})$$

$$\text{(c)}\quad P^{\dagger}_{r,r-1} = F^{\dagger}P^{\dagger}_{r-1,r-1}F^{\dagger^T} + G^{\dagger}QG^{\dagger^T}$$

$$\text{(d)}\quad P^{\dagger}_{r-1,r-1} = P^{\dagger}_{r-1,r-2}(I - H^{\dagger}K^{\dagger^T}_{r-1})$$

where $r = t+k$, and the augmented gain is defined as

$$K^{\dagger}_{r} = P^{\dagger}_{r,r-1}(H^{\dagger^T}P^{\dagger}_{r,r-1}H^{\dagger}+R)^{-1}. \quad \text{(A.5)}$$

The recursions (A.4) are defined for $r-k > 0$ with initial conditions obtained by running the standard filter (3.6) for k periods. We see from (A.1) that $\hat{x}_{t,t+k}$ is the bottom column block of $\hat{x}^{\dagger}_{r+1,r}$, and from (A.4a,b) and the structure of $F^{\dagger}$ that this may be written

$$\hat{x}_{t,r} = \hat{x}_{t,r-1} + K_r\tilde{y}_r \quad \text{(A.6)}$$

where K_r is the bottom block of $K^{\dagger}_{r}$, and is given by

$$K_r = P^{\dagger}_{r,r-1}(k+1,1)\ H(H^T P^{\dagger}_{r,r-1}(1,1)H+R)^{-1} \quad \text{(A.7)}$$

with $P^{\dagger}_{r,r-1}(1,1) = P_{r,r-1}$. The final term in parentheses in (A.7) is, of course, the innovations variance, which is unaffected by the stacking

The formal similarity of (A.3) to (2.9) may be exploited to obtain the following recursions, analogous to (3.6):

(A.4)

(a) $\hat{x}^{\dagger}_{r+1,r} = F^{\dagger}\hat{x}^{\dagger}_{r,r}$

(b) $\hat{x}^{\dagger}_{r,r} = \hat{x}^{\dagger}_{r,r-1} + K^{\dagger}_{r}(y_r - H^{\dagger^T}\hat{x}^{\dagger}_{r,r-1})$

(c) $P^{\dagger}_{r,r-1} = F^{\dagger}P^{\dagger}_{r-1,r-1}F^{\dagger^T} + G^{\dagger}QG^{\dagger^T}$

(d) $P^{\dagger}_{r-1,r-1} = P^{\dagger}_{r-1,r-2}(I - H^{\dagger}K^{\dagger^T}_{r-1})$

where $r = t+k$, and the augmented gain is defined as

(A.5) $$K^{\dagger}_{r} = P^{\dagger}_{r,r-1}(H^{\dagger^T}P^{\dagger}_{r,r-1}H^{\dagger}+R)^{-1}.$$

The recursions (A.4) are defined for $r-k > 0$ with initial conditions obtained by running the standard filter (3.6) for k periods. We see from (A.1) that $\hat{x}_{t,t+k}$ is the bottom column block of $\hat{x}^{\dagger}_{r+1,r}$, and from (A.4a,b) and the structure of $F^{\dagger}$ that this may be written

(A.6) $$\hat{x}_{t,r} = \hat{x}_{t,r-1} + K_r\tilde{y}_r$$

where K_r is the bottom block of $K^{\dagger}_{r}$, and is given by

(A.7) $$K_r = P^{\dagger}_{r,r-1}(k+1,1)\ H(H^T P^{\dagger}_{r,r-1}(1,1)H+R)^{-1}$$

with $P^{\dagger}_{r,r-1}(1,1) = P_{r,r-1}$. The final term in parentheses in (A.7) is, of course, the innovations variance, which is unaffected by the stacking of the model, and $P^{\dagger}_{r,r-1}(k+1,1) = \text{cov}\,\{x_t, x_{t+k} | \Omega_{t+k-1}\}$. Thus, in accordance with the projection theorem for conditional expectations which was used repeatedly to derive the basic filter in Section 3.1, we may write

$$\text{(A.8)} \qquad \hat{x}_{t,r} = \hat{x}_{t,r-1} + \frac{\text{cov}\{x_t, y_r | \Omega_{r-1}\}}{\text{var}\{y_r | \Omega_{r-1}\}} \tilde{y}_r \quad .$$

Embedded in the covariance recursions of (A.4) is a sub-sequence which yields $P^{\dagger}_{r,r-1}(k+1,1)$ in terms of $k-1$ successive one-step-ahead covariance matrices $P_{t+1,t}, \ldots, P_{r-1,r-2}$ together with $P_{t,t}$. These recursions are, for $t,j > 0$, and $q = t+j$:

$$\text{(A.9)} \qquad \begin{aligned} P^{\dagger}_{q,q-1}(j+1,1) &= P^{\dagger}_{q-1,q-1}(j,1)F^T \\ P^{\dagger}_{q-1,q-1}(j,1) &= P^{\dagger}_{q-1,q-2}(j,1)\ (I-HK^T_{q-1}). \end{aligned}$$

By making repeated substitutions in (A.9), all quantities not derived directly from the basic filter, (3.6), may be eliminated, terminating with $P^{\dagger}_{t+1,t}(2,1) = P_{t,t}F^T$. In steady state the necessary substitutions yield, as required,

$$K_j = P\{(I-HK^T)\ F^T\}^j\ H(H^TPH+R)^{-1}.$$

APPENDIX B SOME COVARIANCE GENERATING FUNCTION AND z-TRANSFORM IDENTITIES

(i) The c.g.f. of y_t

We wish to prove that

$$\text{(B.1)} \qquad g_{yy}(z) = R + H^T(I-zF)^{-1} GQG^T(I-z^{-1}F^T)^{-1}H .$$

Partitioning the two inverses conformably as

$$(I-zF)^{-1} = \begin{bmatrix} A & O \\ O & B \end{bmatrix}, \quad (I-z^{-1}F^T)^{-1} = \begin{bmatrix} C & O \\ O & D \end{bmatrix},$$

and noting the forms of H, G, and Q from (2.10), the right hand side of (B.1) can be written

$$W(z)W^T(z^{-1}) = \sigma_\eta^2 + \sigma_\varepsilon^2 a_{11}c_{11} + \sigma_\nu^2 b_{11}d_{11} .$$

Now $|A^{-1}| = \phi(z)$, and the top left term of adj $[A^{-1}]$ is easily seen to be unity, so that, by symmetry, as asserted,

$$a_{11}c_{11} = \frac{1}{\phi(z)\phi(z^{-1})}, \quad b_{11}d_{11} = \frac{1}{\psi(z)\psi(z^{-1})}$$

(ii) The symmetric factorization, $W(z)$

In steady state, we have, substituting (3.6c) into (3.6d)

$$\text{(B.2)} \qquad GQG^T = P - F\{P-PH(H^TPH+R)^{-1}H^TP\}F^T$$

and from the definition of K, therefore

$$(B.3) \qquad GQG^T = P - F(P-K(H^TPH+R)K^T)F^T .$$

Substitution of (B.3) into the right hand side of (B.1) gives

$$W(z)W^T(z^{-1}) = R + H^T(I-zF)^{-1}\left(P-F\{P-K(H^TPH+R)K^T\}F^T\right)(I-z^{-1}F^T)^{-1}H;$$

$H^T(I-zF)^{-1}(P-FPF^T)(I-z^{-1}F^T)^{-1}H$ may be rearranged, using

$(I-zF)^{-1} = I+z(I-zF)^{-1}F$, to the form

$$(B.4) \qquad H^TPH + zH^T(I-zF)^{-1}FPH + z^{-1}H^TPF^T(I-z^{-1}F^T)^{-1}H ,$$

and substitution of $K(H^TPH+R)$ for PH in the second and third terms of (B.4) gives, putting $M = (R+H^TPH)$:

$$(B.5) \qquad W(z)W^T(z^{-1}) = M + zH^T(I-zF)^{-1}FKM + z^{-1}MK^TF^T(I-z^{-1}F^T)^{-1}H$$

$$+ H^T(I-zF)^{-1}FKMK^TF^T(I-z^{-1}F^T)^{-1}H$$

$$= \{I+zH^T(I-zF)^{-1}FK\}M\{I+z^{-1}K^TF^T(I-z^{-1}F^T)^{-1}H\}$$

as required.

APPENDIX C CONTROLLABILITY AND DETECTABILITY

In this appendix we illustrate the role of some fundamental state-space concepts in determining convergence of the signal extraction error covariance to a steady state. We thus examine the questions of existence, uniqueness, convergence and independence from initial conditions in the context of a simple example. The theorem of Caines and Mayne employed in the body of the paper provides sufficient but not necessary conditions for convergence, and thus a case of general statistical interest (cf. Section 4.3) could not be treated fully. However, the pursuit of necessary conditions is beset with difficulties, as the examples show. We note that despite the importance of the topic, the standard texts on state-space estimation do not discuss such details.

We consider a simple 2-dimensional realization:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + G \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix}$$

$$y_t = H^T \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} + v_t$$

where $w_{1,t}$, $w_{2,t}$ and v_t are independent white noises with variances σ_1^2, σ_2^2, and R. Where no confusion may arise we denote Diag $\{\lambda_1, \lambda_2\}$ by F, and we consider various possibilities for G and H.

Case 1

$$G = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad H = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Applying definitions 4.1-4.4 we see that the pair (F,G) has uncontrollable eigenvalue λ_2, and is thus stabilizable if and only if $|\lambda_2| < 1$, while the pair (F,H) is completely observable if and only if $\lambda_1 \neq \lambda_2$, and if $\lambda_1 = \lambda_2 = \lambda$ it is detectable if and only if $|\lambda| < 1$.

(i) Existence. In this case $w_{2,t}$ plays no role, and $x_{2,t} = \lambda_2^t x_{2,0}$, so that if $x_{2,0}$ is known exactly we can transform to $y_t^* = y_t - x_{2,t}$ irrespective of the value of λ_2, thus reducing the problem to the scalar case of Section 3.3. Thus there exists a steady state of the form $P^* = \begin{bmatrix} p_{11}^* & 0 \\ 0 & 0 \end{bmatrix}$.

(ii) Uniqueness. Substituting (3.6c) into (3.6d) yields a form of the Riccati equation valid for singular P, from which we obtain the following three equations which must hold in steady state:

(C.2) $$p_{11} = \lambda_1^2 \ (p_{11} - \sigma^{-2}(p_{11} + p_{12})^2) + \sigma_1^2$$

(C.3) $$p_{12} = \lambda_1\lambda_2(p_{12} - \sigma^{-2}(p_{11} + p_{12})(p_{12} + p_{22}))$$

(C.4) $$p_{22} = \lambda_2^2 \ (p_{22} - \sigma^{-2}(p_{12} + p_{22})^2)$$

where $\sigma^2 = H^TPH + R$ and $p_{21} = p_{12}$. It follows that

(a) If (F,G) is stabilizable ($|\lambda_2| < 1$) then P^* is the unique solution: this is obvious, since $x_{2,t} \to 0$ as $t \to \infty$. Furthermore, $|\lambda_2| < 1$ implies that (F,H) is detectable so this case departs from the conditions of the Caines and Mayne theorem by virtue of the <u>stabilizability</u> rather than controllability of (F,G), and in consequence although the steady-state covariance is unique, it is non-negative rather than positive-definite.

(b) If (F,G) is not stabilizable, but $\lambda_1 \neq \lambda_2$ so that (F,H) is detectable, we have two possibilities: if $|\lambda_2| = 1$, then P^* is again the unique solution; if $|\lambda_2| > 1$ there always exists a second steady state which is positive definite. This illustrates the difficulty that any search for necessary and sufficient conditions for uniqueness must encounter.

(c) If (F,H) is not detectable and (F,G) is not stabilizable $(\lambda_1 = \lambda_2 = \lambda, |\lambda| \geqslant 1)$ we again have different situations if $\lambda = \pm 1$ or $|\lambda| > 1$. If $\lambda = \pm 1$, then $p_{12} = -p_{22}$, from (C.4), and if we let $p_{22} = d \geqslant 0$ then (C.2) gives p_{11} as the positive solution of the quadratic

$$\text{(C.5)} \qquad p_{11} = d + \tfrac{1}{2}\,(\sigma_1^2 \pm \sigma_1(\sigma_1^2 + 4R)^{\frac{1}{2}})$$

There are thus arbitrarily many such steady states. Conversely, if $|\lambda| > 1$, then P^* is again unique.

(iii) Convergence. Establishing convergence is complicated because of the three dimensional time path of $P_{t,t-1}$, and even in this simple example the powerful general methods of Caines and Mayne seem to be required. They show that detectability is sufficient for convergence For the undetectable case in which $|\lambda| > 1$, but $x_{2,0}$ is not known exactly, transforming to the sum and difference of x_1 and x_2 shows that the difference is unobservable but explodes over time, and this results in the divergence of $P_{t,t-1}$. When $|\lambda| = 1$, we can show that the variance of $\hat{x}_{2t,t-1}$ is monotonically decreasing in t, and hence that it attains some limit, d. We can also show that p_{11} then converges to a solution of (C.5). The value of P attained depends on the initial conditions, however.

Case 2

$$G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Here (F,G) is controllable, and (F,H) has unobservable eigenvalue λ_2, and is thus detectable if and only if $|\lambda_2| < 1$.

(i) Existence. That a steady state exists for p_{11} follows from the fact that $y_t = x_{1,t} + v_t$, and $x_{1,t}$ does not depend on $x_{2,t}$ so the analysis of Section 3.3 again applies. If $|\lambda_2| < 1$, the steady state of p_{22} is easily seen to be $p_{22} = \sigma_2^2/(1-\lambda_2^2)$, the unconditional variance of a stationary AR(1) process. If $|\lambda_2| > 1$ we find the only solution of the Riccati equation is negative definite, while if $|\lambda_2| = 1$ it has no real solution.

(ii) Uniqueness, convergence and initial conditions. In this case, with (F,G) controllable, we find that $P_{t,t-1}$ converges to a unique positive-definite steady state if and <u>only if</u> (F,H) is detectable. If (F,H) has an undetectable eigenvalue, then that part of $P_{t,t-1}$ associated with the detectable subspace still converges.

REFERENCES

ANDERSON, B.D.O. and MOORE, J.B. (1979), Optimal Filtering. Englewood Cliffs, NJ : Prentice-Hall.

CAINES, P.E. and MAYNE, D.Q. (1970). On the discrete time matrix Riccati equation of optimal control. International Journal of Control, 12, 785-794.

CLEVELAND, W.P. and TIAO, G.C. (1976). Decomposition of seasonal time series: a model for the Census X-11 program. Journal of the American Statistical Association, 71, 581-587.

GEWEKE, J. (1978). The revision of seasonally adjusted time series. Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp. 320-325.

HANNAN, E.J. (1967). Measurement of a wandering signal amid noise. Journal of Applied Probability, 4, 90-102.

KAILATH, T. (1980a). Linear Systems. Englewood Cliffs, NJ : Prentice-Hall.

KAILATH, T. (1980b). Some alternatives in recursive estimation. International Journal of Control, 32, 311-328.

KALMAN, R.E. (1960). A new approach to linear filtering and prediction problems. Transactions ASME, Journal of Basic Engineering, 82, 35-45.

MASANI, P. (1966). Review of Whittle (1963). Journal of the American Statistical Association, 61, 268-273.

NERLOVE, M., GRETHER, D.M. and CARVALHO, J.L. (1979). Analysis of Economic Time Series. New York : Academic Press.

PAGAN, A.R. (1975). A note on the extraction of components from time series. Econometrica, 43, 163-168.

PIERCE, D.A. (1979). Signal extraction error in nonstationary time series. Annals of Statistics, 7, 1303-1320.

PIERCE, D.A. (1980). Data revisions with moving average seasonal adjustment procedures. Journal of Econometrics, 14, 95-114.

PRIESTLEY, M.B. (1981). Spectral Analysis and Time Series. London : Academic Press.

SOBEL, E. (1967). Prediction of a noise-distorted, multivariate, non-stationary signal. Journal of Applied Probability, 4, 330-342.

WALLIS, K.F. (1981). Models for X-11 and 'X-11-FORECAST' procedures for preliminary and revised seasonal adjustments. Presented at the ASA-CENSUS-NBER Conference on Applied Time Series Analysis of Economic Data, Washington DC, October 1981.

WALLIS, K.F. (1982). Seasonal adjustment and revision of current data : linear filters for the X-11 method. Journal of the Royal Statistical Society A, 145, 74-85.

WHITTLE, P. (1963). Prediction and Regulation by Linear Least-Square Methods. London : English Universities Press.