FORECASTING AND SIGNAL EXTRACTION IN AUTOREGRESSIVE-MOVING AVERAGE MODELS

by

Peter Burridge and Kenneth F. Wallis

Number 274

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

<u>FORECASTING AND SIGNAL EXTRACTION IN</u>

<u>AUTOREGRESSIVE-MOVING AVERAGE MODELS</u>

by

Peter Burridge and Kenneth F. Wallis

Number 274

October 1986

This paper is a complete revision and considerable extension of "Signal extraction in nonstationary series", Warwick Economic Research Paper No.234, August 1983. Early versions of parts of the work were presented at European Meetings of the Econometric Society, Pisa, 1983 and Madrid, 1984. This research was supported by a grant from the Economic and Social Research Council.

1. INTRODUCTION AND SUMMARY

Prediction theory for stationary stochastic processes as developed by Wiener and Kolmogorov and exposited by Whittle (1963) has a wide range of applications in theoretical and empirical economics. Results for the forecasting or pure prediction problem underlie theoretical models of optimal behaviour in dynamic and uncertain environments, and support empirical time series studies in considerable number. The signal extraction problem arises when the variable to be predicted or estimated is jointly stationary with the observed series, specifically as a signal observed with a superimposed error, or more generally as one component of a process comprising several unobserved components; applications range from permanent income theory to the seasonal adjustment of economic time series. The Wiener-Kolmogorov theory is heavily utilized by such textbook authors as Sargent (1979) in macroeconomic theory and Nerlove et al. (1979) in time series analysis.

The basic requirement of the Wiener-Kolmogorov theory for purely non-deterministic stationary processes is knowledge of the autocovariances (or equivalently the spectral density) of the variable to be predicted, or knowledge of the auto- and cross-covariances of the unobserved components, and hence of the observed variable, in the signal extraction problem. Both in theoretical work and in practical implementation, this requirement has been met by postulating models for the relevant processes and expressing the autocovariances as functions of model parameters. Most often, these models are linear autoregressive-moving average (ARMA) models. Attention is usually restricted to linear least squares (l.l.s.) prediction and estimation, in which the unknown variable is predicted or estimated by a linear function of observed values chosen to minimize the mean square error. The unrestricted least squares predictor is the conditional expectation, but if the process is normally distributed this coincides with the l.l.s. predictor. The set of observed values is usually assumed to extend from the indefinite past, that is, to comprise a semi-infinite sample.

Whereas the classical theory is limited to stationary processes, applications often relate to non-stationary processes variously termed accumulated, integrated, or difference-stationary processes. Thus theoretical models may deliver propositions that certain variables follow random walks, and in empirical analysis of economic time series it is a common finding that series must be differenced at least once before appearing stationary. In such circumstances the usual approach is to apply to the differenced series the theory relevant to stationary processes. The assumption of a semi-infinite sample causes difficulties that have not been satisfactorily resolved, for example, the observed series then has unbounded variance. Of course in practice only a finite record is available, and we see below that a relevant approach can be developed which, by paying proper attention to initial conditions, satisfactorily accommodates difference-stationary processes.

In this paper we present the theory of l.l.s. forecasting and signal extraction for autoregressive-moving average models that are free of any parameter restriction, such as stationarity or invertibility restrictions. The approach we adopt is recursive, in which estimates or forecasts formed from information available at a particular point in time are sequentially updated as each successive observation becomes available. The techniques are those of Kalman filtering, studied more extensively in the control theory literature, and much of the paper is concerned with extending and adapting results from that literature to the present statistical time series problems. The basic requirement of the theory remains the covariance structure of relevant variables, and this is provided by the ARMA model, whose parameters are assumed known and time-invariant, together with appropriate initial conditions.

In Section 2 we present two familiar examples of the general problems, namely forecasting the first-order ARMA process and extracting a first-order autoregressive signal from noise-contaminated observations. In each case we develop recursions for the l.l.s. forecasts or signal estimates from first principles, and the treatment is accessible to a reader with no prior knowledge of the Kalman filter. The equivalence to Wiener-Kolmogorov theory in the case that this is applicable is described, but it is also seen that the l.l.s. recursions are applicable to non-stationary and non-invertible cases of the same underlying models.

In Section 3 we review relevant material from the control theory literature. The state space representation of a linear dynamic system is presented, together with the l.l.s. recursions more familiarly known in that context as the Kalman filter. An important question is whether these time-varying recursions eventually achieve a form which has constant coefficients, or a steady state, and we record a number of theorems dealing with the convergence of a key quantity in the Kalman filter equations, namely the error variance of the one-step-ahead estimate of the state variable.

The forecasting and signal extraction problems for general ARMA models are dealt with in turn in Sections 4 and 5. In each case we first present a convenient state space representation of the underlying time series model, and then investigate the applicability of the apparatus of

Section 3. The l.l.s. recursions are shown to converge to a steady state from given initial conditions in circumstances more general than those considered in the time series literature. As with the simple examples of Section 2, the results cover non-stationary and non-invertible cases of the underlying models. For stationary models, to which the Wiener-Kolmogorov theory applies, we show explicitly that the steady state of the Kalman filter coincides with the classical time-series formulae.

A particular circumstance not covered by these general results is one in which two components of an unobserved-component ARMA model have autoregressive operators sharing an unstable common factor. For example, in the seasonal time series context, the seasonal component might contain a seasonal difference operator and the non-seasonal component a simple difference operator. In Section 6 we return to direct consideration of a simple example, namely one in which a single explosive common factor is present, and show that the l.l.s. recursions lead to a steady state of the expression for the estimate of the components, while its error variance diverges. Because convergence of the error variance is sufficient for convergence of the state estimate recursion, the control theory literature has concentrated on circumstances in which the former is achieved. Our example, however, provides a practical demonstration that it is not a necessary condition.

In Section 7 we discuss the relationship between the results of this paper and other treatments of the same, or closely-related, problems in the time series literature, and present some concluding comments.

2. TWO LEADING EXAMPLES

2.1 Forecasting the ARMA(1,1) process

In this section we develop from first principles recursive methods of forecasting a variable y which obeys the ARMA(1,1) process

$$y_t = \phi y_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} \tag{2.1}$$

where ε_t is white noise with constant variance σ_ε^2. It is assumed that $\phi \neq \theta$, but these parameters are otherwise unrestricted. At time t the observed history is $\{y_\tau;\ 0 \leq \tau \leq t\}$, and the problem is to construct the l.l.s. forecast of y_{t+1}, denoted $\hat{y}_{t+1,t}$, together with quantities such as the mean square error or innovation variance $E\{(y_{t+1} - \hat{y}_{t+1,t})^2\} = \Sigma_{t+1}$. It is convenient to express y as the sum of two random variables, one of which, x, is forecastable while the other, ε, is not (other than at a value of zero). Thus we write (2.1) as

$$\begin{aligned} x_t &= \phi y_{t-1} - \theta\varepsilon_{t-1} \\ y_t &= x_t + \varepsilon_t . \end{aligned} \tag{2.2}$$

Then x obeys the stochastic difference equation

$$x_t = \phi x_{t-1} + (\phi-\theta)\varepsilon_{t-1} , \tag{2.3}$$

and forecasts of x and y coincide.

The recursive procedure is started off at time 0. Before any observations are made, our state of knowledge about x_0 is assumed to comprise its expectation $\hat{x}_{0,-1}$ and variance $p_{0,-1} = E\{(x_0 - \hat{x}_{0,-1})^2\}$. Since ε is serially uncorrelated, the l.l.s. forecast of y_0 is simply $\hat{x}_{0,-1}$, with variance $\Sigma_0 = p_{0,-1} + \sigma_\varepsilon^2$, and the covariance between y_0 and

x_0 is equal to $p_{0,-1}$. On observing y_0 we first update our initial estimate of x_0 by projection on y_0, giving the l.l.s. estimate

$$\hat{x}_{0,0} = \hat{x}_{0,-1} + p_{0,-1}\Sigma_0^{-1}(y_0 - \hat{y}_{0,-1}) \tag{2.4}$$

$$= \hat{x}_{0,-1} + p_{0,-1}\Sigma_0^{-1}\tilde{y}_0 ,$$

where $\tilde{y}_0$ is the innovation in y, and $p_{0,-1}\Sigma_0^{-1}$ is the least squares regression coefficient. The corresponding estimate of the unobserved ε_0 is

$$\hat{\varepsilon}_{0,0} = y_0 - \hat{x}_{0,0} = \sigma_\varepsilon^2\Sigma_0^{-1}\tilde{y}_0 , \tag{2.5}$$

using $\Sigma_0 - p_{0,-1} = \sigma_\varepsilon^2$. Since $\varepsilon_0 - \hat{\varepsilon}_{0,0} = -(x_0 - \hat{x}_{0,0})$ their variances are equal, to $p_{0,0}$ say, which from the projection (2.4) is given as

$$p_{0,0} = p_{0,-1} - p_{0,-1}^2\Sigma_0^{-1} = \sigma_\varepsilon^2 p_{0,-1}\Sigma_0^{-1} . \tag{2.6}$$

Turning to the one-step-ahead forecast, from (2.2) we have

$$\hat{x}_{1,0} = \phi y_0 - \theta\hat{\varepsilon}_{0,0} , \tag{2.7}$$

with variance $p_{1,0} = E\{(x_1 - \hat{x}_{1,0})^2\} = \theta^2 p_{0,0}$, so from (2.6) we have the recursion

$$p_{1,0} = \theta^2\sigma_\varepsilon^2 p_{0,-1}\Sigma_0^{-1} . \tag{2.8}$$

Finally, since the forecasts of x and y coincide, from (2.5) and (2.7) the l.l.s. forecast recursion for y is obtained as

$$\hat{y}_{1,0} = \phi y_0 - \theta\sigma_\varepsilon^2\Sigma_0^{-1}(y_0 - \hat{y}_{0,-1}) , \tag{2.9}$$

with mean square error $\Sigma_1 = p_{1,0} + \sigma_\varepsilon^2$.

The sequence of steps leading from (2.4) to (2.9) may be repeated as each new observation arrives, yielding the general recursions, valid for t=0,1,2,...,

$$(2.10)\,(a)\qquad \hat{y}_{t+1,t} = \phi y_t - \theta\sigma_\varepsilon^2\Sigma_t^{-1}(y_t - \hat{y}_{t,t-1})$$

$$(b)\qquad \Sigma_t = p_{t,t-1} + \sigma_\varepsilon^2$$

$$(c)\qquad p_{t+1,t} = \theta^2\sigma_\varepsilon^2 p_{t,t-1}\Sigma_t^{-1}\ .$$

This system of equations provides the l.l.s. forecasts of y, irrespective of the stationarity or otherwise of the process, or of the invertibility of its moving average component, no assumption about the values of ϕ and θ having been made, except that they are distinct.

The coefficient of the innovation in the forecasting rule (2.10a) varies over time, and it is of interest to study its evolution, as described by equations (b) and (c). We note, however, that these are independent of the data, so there is no necessary connection between the coefficient sequence and the observation sequence: the coefficients can be calculated "off-line". Combining equations (2.10b,c) their evolution is described by the nonlinear difference equation

$$(2.11)\qquad p_{t+1,t} = \theta^2\sigma_\varepsilon^2 p_{t,t-1}(\sigma_\varepsilon^2 + p_{t,t-1})^{-1} = h(p_{t,t-1})\ ,\ \text{say.}$$

This has fixed points at $p=0$ and $p = (\theta^2 - 1)\sigma_\varepsilon^2$, furthermore $h'(p_{t,t-1}) > 0$ and $h''(p_{t,t-1}) < 0$ for $p_{t,t-1} \geq 0$. In consequence

(i) if $|\theta| \leq 1$, then $p_{t+1,t}$ converges monotonically to zero and the innovation variance to σ_ε^2 for all $p_{0,-1} \geq 0$, the second fixed point being non-positive,

(ii) if $|\theta| > 1$, then $p_{t+1,t}$ converges monotonically to $(\theta^2 - 1)\sigma_\varepsilon^2$ and the innovation variance to $\theta^2\sigma_\varepsilon^2$ for all $p_{0,-1} > 0$; if, on the other hand, $p_{0,-1} = 0$, then $p_{t+1,t} = 0$ for all t.

Thus in all cases, again irrespective of the value of ϕ, the recursions (2.10) deliver a steady-state forecasting rule. In case (i), this is

$$(2.12) \qquad \hat{y}_{t+1,t} = \phi y_t - \theta(y_t - \hat{y}_{t,t-1})$$

whereas in case (ii), with $|\theta| > 1$, unless $p_{0,-1} = 0$, we have

$$(2.13) \qquad \hat{y}_{t+1,t} = \phi y_t - \theta^{-1}(y_t - \hat{y}_{t,t-1}) .$$

With respect to invertibility, we note that this is immaterial to the existence of the steady-state forecasting rule; indeed, if the process is non-invertible, with $|\theta| > 1$, the recursions deliver an "invertible" steady-state rule (2.13), in which the coefficient of the innovation $\tilde{y}_t$ is θ^{-1}, corresponding to the moving average coefficient in the observationally equivalent invertible process. The only case in which this is not so is $|\theta| > 1$ and $p_{0,-1} = 0$. The latter requirement represents perfect knowledge of x_0, and is unrealistic: in practical situations it is customary to assign a relatively large value to $p_{0,-1}$, amounting to a "diffuse prior" on x_0, and in the remainder of this section it is assumed that $p_{0,-1} > 0$.

The forecast error associated with the application of the steady-state forecasting rule from the beginning can be readily described. First, if $|\theta| \leq 1$, repeated substitutions in (2.12) yield the forecast as a function of the data and initial condition:

$$(2.14) \qquad \hat{y}_{t+1,t} = (\phi-\theta) \sum_{j=0}^{t} \theta^j y_{t-j} + \theta^{t+1}\hat{y}_{0,-1} .$$

Similarly, lagging and repeatedly substituting the relation $\varepsilon_t = y_t - \phi y_{t-1} + \theta\varepsilon_{t-1}$ gives an autoregressive expression for y_{t+1}:

$$y_{t+1} = \varepsilon_{t+1} + (\phi-\theta) \sum_{j=0}^{t} \theta^j y_{t-j} + \theta^{t+1} y_0 - \theta^{t+1}\varepsilon_0 .$$

Subtracting, the forecast error is obtained as

$$y_{t+1} - \hat{y}_{t+1,t} = \varepsilon_{t+1} + \theta^{t+1}(x_0 - \hat{x}_{0,-1}) ,$$

which approaches the white noise random variable ε_{t+1} if $|\theta| < 1$; the initial error $(x_0 - \hat{x}_{0,-1})$ has a persistent effect if $|\theta| = 1$, reflecting the continuing cost of using the steady-state rule rather than the general l.l.s. recursions (2.10) in this case. Secondly, when $|\theta| > 1$, from (2.13) we similarly obtain

$$\hat{y}_{t+1,t} = (\phi-\theta^{-1}) \sum_{j=0}^{t} \theta^{-j} y_{t-j} + \theta^{-(t+1)} \hat{y}_{0,-1} ,$$

but now to obtain a comparable autoregressive representation for y_{t+1} we utilize the invertible ARMA(1,1) form

$$y_t = \phi y_{t-1} + \eta_t - \theta^{-1} \eta_{t-1}$$

which is observationally equivalent to the original specification in the second-moment sense. Repeated substitutions yield

$$y_{t+1} = \eta_{t+1} + (\phi-\theta^{-1}) \sum_{j=0}^{t} \theta^{-j} y_{t-j} + \theta^{-(t+1)} y_0 - \theta^{-(t+1)} \eta_0 .$$

On subtracting, the forecast error is seen to be

$$y_{t+1} - \hat{y}_{t+1,t} = \eta_{t+1} + \theta^{-(t+1)} (y_0 - \hat{y}_{0,-1} - \eta_0) ,$$

which approaches η_{t+1}, the driving random variable in the invertible representation, with variance $\sigma_\eta^2 = \theta^2 \sigma_\varepsilon^2$. This representation has been used simply to obtain a closed form for the forecast error, and we again emphasize that it is not necessary to work with the invertible representation in developing l.l.s. forecasts, the correct steady-state rule (2.13) being given automatically by the recursions. The choice of invertible or non-invertible representation is more apparent than real, since whatever choice is made, the result is the same steady-state forecasting rule with the same forecast mean square error.[1]

The equivalence to the Wiener-Kolmogorov theory can be easily analysed. Attention is usually restricted to stationary invertible processes, and the forecast is expressed as a function of observations extending from the indefinite past. Accordingly we consider the limit of the steady-state forecasting rule (2.12), which either directly or via (2.14) can be written as the linear filter (with lag operator L)

$$\hat{y}_{t+1,t} = f_1(L)y_t \tag{2.15}$$

where the generating function of the coefficients of the observations is

$$f_1(z) = \frac{\phi - \theta}{1 - \theta z} . \tag{2.16}$$

Simple manipulation of this function yields

$$f_1(z) = \frac{\phi - \theta}{1 - \theta z} = \frac{z^{-1}(1-\theta z) - z^{-1}(1-\phi z)}{1 - \theta z}$$

$$= \frac{1 - \phi z}{1 - \theta z}\left\{\frac{z^{-1}(1-\theta z)}{1 - \phi z} - z^{-1}\right\} = \frac{1-\phi z}{1-\theta z}\left[\frac{z^{-1}(1-\theta z)}{1 - \phi z}\right]_+$$

where $[a(z)]_+$ denotes that part of the polynomial $a(z)$ containing non-negative powers of z, the operator $[\cdot]_+$ being sometimes termed an "annihilation" operator. This last expression is the form in which the Wiener-Kolmogorov predictor is usually given (Whittle, 1963, Ch.3; Nerlove et al., 1979, Ch.V). It is derived under the assumption that $|\phi| < 1$, and there are clearly difficulties with the term $(1 - \phi z)^{-1}$ if this is not so. But this assumption has no bearing on the first form in which $f_1(z)$ is given, which thus represents a more convenient device for calculating the coefficients, nor has it any bearing on our derivation. So in addition to an equivalence to the Wiener-Kolmogorov predictor in the case for which it is defined, we see that the same expression, interpreted as the limit of our l.l.s. forecast recursions, applies to non-stationary cases of the same process. Equally, when the Wiener-Kolmogorov predictor is expressed in recursive form,[2] it coincides with the steady-state l.l.s. recursion (2.12), again valid more generally. The particular case $\phi=1$ has received much attention; here (2.12) can be rearranged to give the familiar "adaptive expectations" formula

$$\hat{y}_{t+1,t} - \hat{y}_{t,t-1} = (1-\theta)(y_t - \hat{y}_{t,t-1}) .$$

Forecasting j steps ahead ($j>1$) is straightforward in the present framework, since $\hat{y}_{t+j,t}$ is simply $\hat{x}_{t+j,t}$. Returning to the general case, this together with its forecast error is obtained from (2.3) as

$$\hat{x}_{t+j,t} = \phi^{j-1}\hat{x}_{t+1,t}, \qquad j=2,3,\ldots$$

$$x_{t+j} - \hat{x}_{t+j,t} = (\phi-\theta)(\varepsilon_{t+j-1} + \phi\varepsilon_{t+j-2} + \ldots + \phi^{j-2}\varepsilon_{t+1}) + \phi^{j-1}(x_{t+1} - \hat{x}_{t+1,t}) .$$

Thus the j-step-ahead forecast $\hat{y}_{t+j,t} = \phi^{j-1}\hat{y}_{t+1,t}$ has mean square error

$$E\{(y_{t+j} - \hat{y}_{t+j,t})^2\} = (\phi-\theta)^2 \sum_{i=0}^{j-2} \phi^{2i}\sigma_\varepsilon^2 + \phi^{2j-2}P_{t+1,t} + \sigma_\varepsilon^2 .$$

If $|\phi| < 1$, we see that as j increases the forecast $\hat{y}_{t+j,t}$ tends to zero, the unconditional mean, the observed history summarized in $\hat{x}_{t+1,t}$ becoming less informative; similarly the forecast error variance tends to the variance of y, namely $\sigma_\varepsilon^2(1-2\phi\theta+\theta^2)/(1-\phi^2)$. Conversely, if $|\phi| > 1$, neither the forecast nor its error variance approach limits, while if $\phi=1$ the forecast is constant, with error variance which increases linearly in j, as is well known. In all cases, the steady-state forecasting rule can be expressed as a function of past observations, corresponding to (2.15) and (2.16), giving

$$\hat{y}_{t+j,t} = f_j(L)y_t \quad ; \quad f_j(z) = \phi^{j-1}f_1(z) = \frac{\phi^{j-1}(\phi-\theta)}{1-\theta z} \; .$$

It is then easily seen that this again corresponds to the usual result for the stationary case, namely (Nerlove et al., 1979, p.92)

$$f_j(z) = \frac{1-\phi z}{1-\theta z}\left[\frac{(1-\theta z)z^{-j}}{1-\phi z}\right]_+ ,$$

sustaining the same interpretation as in the previous paragraph.

2.2 Extracting an AR(1) signal masked by white noise

In this section we consider a simple signal extraction problem and again develop from first principles recursive methods of calculating the l.l.s. signal estimate. The signal s is assumed to follow an AR(1) process, and represents the variable of principal interest, but it is observed subject to a superimposed white noise error, η. Thus the model is

$$\begin{aligned} y_t &= s_t + \eta_t \\ s_{t+1} &= \phi s_t + \varepsilon_{t+1} \end{aligned} \qquad t=0,1,2,\ldots \tag{2.17}$$

where η and ε are contemporaneously and serially uncorrelated, with variances σ_η^2 and σ_ε^2 respectively, and the problem is to estimate s_t from observations on y.

In advance of the first observation, y_0, knowledge about s_0 is summarized in its expectation $\hat{s}_{0,-1}$ and variance $p_{0,-1} \geq 0$. The corresponding mean and variance of y_0 are then

$$\hat{y}_{0,-1} = \hat{s}_{0,-1} \quad , \quad \Sigma_0 = p_{0,-1} + \sigma_\eta^2$$

and the covariance between y_0 and s_0 is $p_{0,-1}$. Once the observation y_0 is available, the signal estimate is updated by the projection

$$\hat{s}_{0,0} = \hat{s}_{0,-1} + p_{0,-1}\Sigma_0^{-1}(y_0 - \hat{y}_{0,-1}),$$

and its mean square error correspondingly updated as

$$E\{(s_0 - \hat{s}_{0,0})^2\} = p_{0,-1} - p_{0,-1}^2\Sigma_0^{-1} .$$

The l.l.s. forecast of the next value s_1 is now simply

$$\hat{s}_{1,0} = \phi\hat{s}_{0,0} ,$$

and since the error in this forecast is

$$s_1 - \hat{s}_{1,0} = \phi(s_0 - \hat{s}_{0,0}) + \varepsilon_1$$

its mean square error is given as

$$p_{1,0} = E\{(s_1 - \hat{s}_{1,0})^2\} = \phi^2(p_{0,-1} - p_{0,-1}^2\Sigma_0^{-1}) + \sigma_\varepsilon^2 .$$

This procedure may be repeated as the observations $y_1, y_2, \ldots$ in turn become available, giving the general recursions

(2.18)(a) $$\hat{s}_{t,t} = \phi\hat{s}_{t-1,t-1} + p_{t,t-1}\Sigma_t^{-1}(y_t - \phi\hat{s}_{t-1,t-1})$$

(b) $$\Sigma_t = p_{t,t-1} + \sigma_\eta^2$$

(c) $$p_{t+1,t} = \phi^2(p_{t,t-1} - p_{t,t-1}^2\Sigma_t^{-1}) + \sigma_\varepsilon^2 .$$

Equation (2.18a) gives a recursion for the current estimate of the signal, with coefficients that vary over time, but again independently of the observations.

As in the previous example, we are interested in the possibility that the error variance $p_{t,t-1}$ approaches a limit. From (2.18b,c) we obtain the nonlinear difference equation

$$(2.19) \qquad p_{t+1,t} = \phi^2 p_{t,t-1}\sigma_\eta^2(p_{t,t-1} + \sigma_\eta^2)^{-1} + \sigma_\varepsilon^2 = h(p_{t,t-1}) \text{ , say,}$$

and on setting $p = h(p)$ we see that the fixed points are the solutions of the quadratic equation

$$(2.20) \qquad f(p) = p^2 + \{\sigma_\eta^2(1-\phi^2) - \sigma_\varepsilon^2\}p - \sigma_\varepsilon^2\sigma_\eta^2 = 0 .$$

This has real solutions of opposite sign, since $f(0) < 0$, and 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 (2.19) to p for any $p_{0,-1} \geq 0$. As in the first example, convergence of the error variance does not depend on stationarity, that is, on the value of ϕ.

The steady-state signal estimate recursion obtained by setting $p_{t,t-1}$ equal to p in (2.18) can be rearranged as

$$\hat{s}_{t,t} = b\hat{s}_{t-1,t-1} + \left(1 - \frac{b}{\phi}\right) y_t$$

where $b = \phi\sigma_\eta^2/(p + \sigma_\eta^2)$. Repeated substitutions then give the current estimate as a function of the observed history and the initial estimate, namely

$$(2.21) \qquad \hat{s}_{t,t} = \left(1 - \frac{b}{\phi}\right) \sum_{j=0}^{t} b^j y_{t-j} + b^{t+1}\hat{s}_{0,-1}/\phi .$$

The steady state of the variance recursion (2.19) gives an alternative expression for the coefficient b, namely

$$b = \frac{\phi\sigma_\eta^2}{p+\sigma_\eta^2} = \frac{p-\sigma_\varepsilon^2}{\phi p} ,$$

and from one or the other we see that $|b| < 1$ irrespective of the value of ϕ. The expression (2.21) may be compared with the classical formula obtained when $|\phi| < 1$ and a semi-infinite sample is available, which is

(Whittle, 1963, §6.3; Nerlove et al., 1979, §V.5)

$$(2.22)\qquad \hat{s}_{t,t} = (1 - \beta/\phi)(1 - \beta L)^{-1} y_t .$$

Here β is found from the unique invertible factorization of the covariance generating function of $(y_t - \phi y_{t-1})$; that is, β satisfies

$$(2.23)\qquad g(z) = \sigma_\varepsilon^2 + \sigma_\eta^2 (1-\phi z)(1-\phi z^{-1}) = \sigma^2 (1-\beta z)(1-\beta z^{-1}) ,$$

and is the moving average coefficient in the ARMA(1,1) representation of y. The quadratic equation to be solved for β, which has a reciprocal pair of solutions, is obtained by setting $z=\beta$ in (2.23), whence

$$(2.24)\qquad \sigma_\varepsilon^2 + \sigma_\eta^2 (1-\phi\beta)(1-\phi/\beta) = 0 .$$

Making a change of variable from β to p by substituting $\phi\sigma_\eta^2/(p+\sigma_\eta^2)$ for β in (2.24) and rearranging yields the quadratic equation (2.20), and it is clear that the positive solution for p corresponds to the invertible solution for β, so b and β in (2.21) and (2.22) coincide. Thus the classical result, derived under a stationarity assumption, again coincides with the limiting case of the l.l.s. recursion, which is valid more generally.

The principal case of this model discussed in the literature occurs when $\phi=1$, so that the signal follows a random walk. For example, s_t might then represent the "persisting" or "permanent" component of income, and observed income y_t comprises this together with a purely random "transitory" component, η_t. Permanent income is then estimated from current and past observed income via (2.22), which in this case gives the familiar exponentially weighted moving average (Muth, 1960). Its classical derivation is unsatisfactory, however, since apparatus developed for stationary processes is being applied to a nonstationary process.[3] Nevertheless we have seen that l.l.s. recursions can be readily developed in this case, with proper attention to initial conditions, and the sense in which the same apparatus can be applied is made precise.

In some situations it may be desirable to update the estimate of s_t as further data arrive, that is, to obtain a sequence $\hat{s}_{t,t}$, $\hat{s}_{t,t+1}$, $\hat{s}_{t,t+2}$, ... of increasingly accurate l.l.s. estimates of s_t. This can be easily handled once the present approach has been suitably generalized, and we return to this problem, known as the "smoothing" problem, below.

Summarizing the lessons of the two examples, we first note the similar roles of x_t in the first example and s_t in the second. Both variables follow first order autoregressions, and neither is observed directly. Secondly, the fundamental quantity required to calculate successive estimates is in each case the error variance of the l.l.s. estimate of this variable, $p_{t,t-1}$. In effect, each of these examples has been analysed in state-space form and, as discussed below, the recursions developed from first principles are scalar versions of the Kalman filter. The ease with which the behaviour of $p_{t,t-1}$ is analysed results from the fact that the relevant state variable (x_t or s_t) is a scalar. To consider more general forecasting and signal extraction problems we introduce representations in which the state variable is a vector, with the error covariance matrix $P_{t,t-1}$ replacing the scalar variance. These representations, and the appropriate generalizations of the l.l.s. recursions, are presented next.

3. STATE-SPACE METHODS AND CONVERGENCE CONDITIONS

3.1 The state-space form and the Kalman filter

We consider linear dynamic systems in which a first-order vector autoregression describes the evolution of intermediate "state" or "process" variables x_t and a contemporaneous equation expresses the observations y_t as a linear combination of state variables, possibly with measurement error. The state transition and measurement equations are, respectively,

$$x_{t+1} = Fx_t + Gw_t \tag{3.1}$$

$$y_t = H'x_t + v_t \,. \tag{3.2}$$

The random input w_t and the observation noise v_t are serially uncorrelated random variables with zero mean and finite covariance matrix

$$\mathrm{cov}\begin{pmatrix} w_t \\ v_t \end{pmatrix} = \begin{pmatrix} Q & S \\ S' & R \end{pmatrix} ;$$

this matrix, together with the coefficient matrices F, G and H are assumed known and time-invariant. In our use of this model to analyse univariate ARMA processes, y_t is a scalar.

The representation (3.1)-(3.2) (or 'realization' in control theory parlance) of the relation between w_t, v_t and the output, y_t, is not unique since this relation is invariant to non-singular transformations of the states. Indeed, in many situations the order of the state vector is not uniquely determined either, and these features of the set-up can be exploited to improve the tractability of the analysis in particular situations. In some circumstances it is convenient to date the input, w, at t+1 rather than t, but this makes no essential difference to the Kalman filtering equations. In this paper we are concerned with the filtering problem, that is, recovery of x_t from observations on y. A mathematically dual problem is that of controlling x_t by varying the input, w_t, and a number of the concepts we employ take their names from this dual; the duality is discussed by numerous authors including, in an economic context, Preston and Pagan (1982).

We seek the l.l.s. estimator of x_t, say, given information up to and including y_{t+k}. To solve this problem we need the covariances of x_t and y_τ $(\tau=0,1,\ldots,t+k)$. These are rendered well-defined by the model (3.1)-(3.2) and the assumed initial conditions that, in advance of any observations, x_0 is known to be randomly distributed with mean $\hat{x}_{0,-1}$ and variance $P_{0,-1}$. Thus the information available at time τ is $\Omega_\tau \equiv \{\hat{x}_{0,-1}, P_{0,-1}, y_0, \ldots, y_\tau\}$. We write the l.l.s. estimate of a random variable u_t given Ω_τ as $\hat{u}_{t,\tau}$, and the innovation in u_t is defined as the error in the one-step-ahead l.l.s. forecast: $\tilde{u}_t = u_t - \hat{u}_{t,t-1}$. In passing, we note that the information sets Ω_τ and $\tilde{\Omega}_\tau \equiv \{\hat{x}_{0,-1}, P_{0,-1}, \tilde{y}_0, \ldots, \tilde{y}_\tau\}$ are identical, since each can be constructed from the other. In the statistical time series literature it has been commonplace not to indicate explicitly the dependence of Ω_τ on initial conditions, since the lack of dependence is implicit in stationarity assumptions. In the present context, where stationarity is not assumed, the nature of the initial conditions is important not only practically, but also theoretically, to define the underlying sigma fields (cf. Florens and Mouchart, 1982).

The l.l.s. estimates of the state vector and their covariances

$$P_{t+j,t} = E\{(x_{t+j} - \hat{x}_{t+j,t})(x_{t+j} - \hat{x}_{t+j,t})'\}, \qquad j=0,1$$

may be obtained recursively from the following Kalman filter equations:

$$(3.3)\quad (a)\quad \hat{x}_{t,t} = \hat{x}_{t,t-1} + C_{t,t}\tilde{y}_t$$

$$(b)\quad \hat{x}_{t+1,t} = F\hat{x}_{t,t} + GS\Sigma_t^{-1}\tilde{y}_t$$

$$(c)\quad C_{t,t} = P_{t,t-1}H\Sigma_t^{-1}$$

$$(d)\quad \Sigma_t = E\{\tilde{y}_t\tilde{y}_t'\} = H'P_{t,t-1}H + R\ .$$

Equation (a) expresses the fact that the innovation in x_t, which is not observed, is estimated by its orthogonal projection on that in y_t. Similarly, the second term on the right-hand side of equation (b) is the projection of Gw_t on $\tilde{y}_t$; equations (c) and (d) are definitional. The covariance update is the Riccati difference equation

$$(e)\quad P_{t+1,t} = FP_{t,t-1}F' + GQG' - K_t\Sigma_tK_t'$$

$$(f)\quad K_t = FC_{t,t} + GS\Sigma_t^{-1}\ .$$

The final term of equation (e) is the reduction in the conditional variance of x_{t+1} attributable to the information in $\tilde{y}_t$, and K_t is the Kalman gain, so called because it gives the amount by which the new information affects the one-step-ahead forecast of the state: combining (a) and (b) we have

$$(g)\quad \hat{x}_{t+1,t} = F\hat{x}_{t,t-1} + K_t\tilde{y}_t\ .$$

Finally, some intermediate quantities are defined as follows:

$$(h)\quad P_{t,t} = P_{t,t-1}(I - HC_{t,t}')$$

$$(i)\quad \tilde{y}_t = y_t - H'\hat{x}_{t,t-1}$$

$$(j)\quad \tilde{F}_t = F - K_tH'\ .$$

Equation (j) defines the 'closed loop system matrix' $\tilde{F}_t$, which is important in the filtering context because it determines the properties of the error process, $e_t = x_t - \hat{x}_{t,t-1}$. Subtracting (3.3g) from the state transition equation and substituting (3.3i) and the measurement equation gives a first order vector difference equation for e, namely

$$(3.4) \qquad e_{t+1} = \tilde{F}_t e_t - K_t v_t + G w_t .$$

Thus if the error variance is to remain bounded as t increases then it is necessary in general that there exist gain matrices, K_t, such that the closed loop system matrix is stable (has eigenvalues inside the unit circle). That this condition is also sufficient to guarantee that the error variance, $P_{t+1,t}$, converges to a steady state, P, given as a fixed point of (3.3e), is less obvious, however, and is discussed below in the context of our applications. Equation (3.4) indicates that, unlike the one-step-ahead forecast errors for the observed series, y_t, which form an innovation sequence, the errors in the one-step-ahead estimates of the unobserved state vector follow a first-order autoregression. The difference is essentially that observing y_t allows the previous estimate to be fully corrected before the forecast of y_{t+1} is made, but only partial correction to the state forecast is possible, resulting in the persistence of errors.

Forecasting the state vector more than one period ahead is very simple in the present framework. Since w_{t+j}, $j > 0$, is uncorrelated with all variables in Ω_t, we have $\hat{w}_{t+j,t} = 0$, so that

$$(3.5) \quad (a) \qquad \hat{x}_{t+j,t} = F\hat{x}_{t+j-1,t}, \qquad j=2,3,\ldots$$

with error variance given by the auxiliary recursion

$$(b) \qquad P_{t+j,t} = FP_{t+j-1,t}F' + GQG' .$$

Forecasts of y are then obtained as

$$\text{(c)} \qquad \hat{y}_{t+j,t} = H'\hat{x}_{t+j,t}$$

with error variance $H'P_{t+j,t}H + R$, which reduces to the innovation variance Σ_{t+1} when j=1. Notice that the sequence of j-step-ahead forecast errors $\ldots,(y_{t+j} - \hat{y}_{t+j,t}), (y_{t+j+1} - \hat{y}_{t+j+1,t+1}),\ldots$ is not an innovation sequence when j >1, but exhibits autocorrelation of order j-1.

Once the concurrent estimate of the state is available at time t, say, it may be improved as further observations arrive. This is achieved by the smoothing recursions[4]

$$\text{(3.6) (a)} \quad \hat{x}_{t,t+k} = \hat{x}_{t,t+k-1} + C_{t,t+k}\tilde{y}_{t+k}, \qquad k=0, 1,\ldots$$

$$\text{(b)} \quad P_{t,t+k} = P_{t,t+k-1} - C_{t,t+k}\Sigma_{t+k}C'_{t,t+k}$$

$$\text{(c)} \quad C_{t,t+k} = \bar{P}'_{t,t+k-1}H\Sigma^{-1}_{t+k}$$

$$\text{(d)} \quad \bar{P}_{t,t+k} = \tilde{F}_{t+k}\bar{P}_{t,t+k-1}$$

initializing (3.6d) by $\bar{P}_{t,t-1} = P_{t,t-1}$ from (3.3), with the first three equations corresponding at k=0 to (3.3a,c,h). Here $\bar{P}_{t,t+k}$ is the covariance of the error in the current one-step forecast of the state and that in the estimate of x_t:

$$\bar{P}_{t,t+k} = E\{(x_{t+k+1} - \hat{x}_{t+k+1,t+k})(x_t - \hat{x}_{t,t+k})'\}$$

so that $C_{t,t+k}\tilde{y}_{t+k}$ appearing in (3.6a) has a similar interpretation to $C_{t,t}\tilde{y}_t$ of equation (3.3a). Furthermore, $P_{t,t+k}$, the error variance of the smoothed estimate, is monotonically non-increasing in k (that is, $P_{t,t+k} - P_{t,t+k+1}$ is positive semi-definite) and bounded below by zero, so that $C_{t,t+k}$ tends to zero and revisions to the state estimate eventually die out. Finally, (3.6b) reflects the fact that the revision,

$(\hat{x}_{t,t+k} - \hat{x}_{t,t+k-1})$, is a function of the (t+k)th innovation and is thus orthogonal to the error in $\hat{x}_{t,t+k-1}$(were this not so, $\hat{x}_{t,t+k}$ would not be the l.l.s. estimate).

That the forecasting and signal extraction problems considered in Section 2 are examples of the application of the Kalman filter is clear once notational equivalences are established. First, with x_t scalar and $F = \phi$, $G = \phi - \theta$, $H = 1$, $w_t = \varepsilon_t$, $v_t = \varepsilon_t$ and $Q = S = R = \sigma_\varepsilon^2$ the state-space form gives the ARMA(1,1) model (2.2) and (2.3); the l.l.s. recursions (2.10) then correspond to (the relevant parts of) the Kalman filter equations (3.3). Secondly, with scalar state variable $x_t = s_t$ and $F = \phi$, $G = 1$, $H = 1$, $w_t = \varepsilon_{t+1}$, $v_t = \eta_t$, $Q = \sigma_\varepsilon^2$, $R = \sigma_\eta^2$ and $S = 0$, the state-space form gives the AR(1)-plus-noise model (2.17), and the l.l.s. recursions (2.18) may again be obtained as a special case of (3.3).

In our discussion of these examples and in consideration of the Kalman filter more generally an important aspect of the recursions is their time-varying nature, and an important question concerns their possible convergence to a steady state. In applied work it is clearly more convenient computationally if the recursions have fixed coefficients, and in studying the relations with the statistical time series literature this is a central question, since in that literature most attention is given to time-invariant forecasting or filtering formulae. In the examples and more generally, a key role is played by the forecast error covariance matrix $P_{t+1,t}$ (written $p_{t+1,t}$ in the scalar examples): if this converges, possibly rapidly, to a fixed point P of the recursion (3.3e), or takes such a value from the beginning, then the complete recursions (3.3) are in steady state (are time-invariant), as are the extended forecast recursions (3.5) and the smoothing recursions (3.6). Conditions under which this occurs are described in the next section.

3.2 Conditions for convergence of the covariance sequence

In this section we state without proof a number of theorems concerning the behaviour of the covariance of the one-step-ahead state-estimation error, $P_{t+1,t}$, associated with the system (3.1)-(3.2). In subsequent sections these results are specialized and extended to cover forecasting and signal extraction in ARMA processes. Proofs of the theorems, and an indication of their antecedents in the control literature, may be found in the papers by Caines and Mayne (1970, 1971), Hager and Horowitz (1976) and Chan, Goodwin and Sin (1984) to which we refer below. The behaviour of the filter covariance is determined by two properties of the system, first identified by Kalman (1960), which relate to the extent to which changes in the state vector, x_t, affect the measurements or output, y_t, and the extent to which variations in the input, w_t, affect the state vector. We begin by defining these system properties.

Definition 1

The pair (F,H) is said to be detectable if F has no eigenvalue λ with corresponding non-zero eigenvector b such that $|\lambda| \geq 1$ and $H'b = 0$. If $H'b \neq 0$ for all λ the pair is said to be observable.

Definition 2

The pair $(F, GQ^{\frac{1}{2}})$ is said to be stabilizable if F has no eigenvalue λ with corresponding non-zero left eigenvector a' such that $|\lambda| \geq 1$ and

$a'GQ^{\frac{1}{2}} = 0$. If $a'GQ^{\frac{1}{2}} \neq 0$ for all λ the pair is said to be controllable.

Observability and controllability are frequently expressed as the condition that the range space of F^n is contained in those of the matrices

$$O(F,H) \equiv \left(H \;\vdots\; FH \;\vdots \ldots \vdots\; F^{n-1}H\right),$$

$$C(F,GQ^{\frac{1}{2}}) \equiv \left(GQ^{\frac{1}{2}} \;\vdots\; FGQ^{\frac{1}{2}} \;\vdots \ldots \vdots\; F^{n-1}GQ^{\frac{1}{2}}\right).$$

These conditions are implied by definitions 1 and 2 (Kailath, 1980, p.135). In the context of ARMA models the eigenvalues and eigenvectors of F are easily obtained, and so it is convenient to work with the definitions in the form given. As noted above, these concepts take their names from the dual control problem (see Preston and Pagan, 1982, chs.5,6). Controllability is the ability to move the state in finite time from an arbitrary initial point to an arbitrary terminal point by manipulation of the input (control instruments); stabilizability, or asymptotic controllability is the asymptotic version of this capability. Whereas controllability is an existence property, observability is a uniqueness property. Interpreted as a property of the dynamic structure of the model, it specifies that all the natural modes of the state dynamics are contained in, or observed by, the output dynamics. When observability fails, detectability rules out instability of the unobserved modes. In the filtering problem, detectability holds when movements in the state vector along directions which do not affect the measurement have bounded variance whenever the input, w, has bounded variance. An important implication of detectability is that there exists a column vector $\bar{K}$ such that the matrix $\{F-\bar{K}H'\}$ has eigenvalues inside the unit circle.

If the covariance possesses a steady state, P, then there exists a corresponding steady-state closed-loop system matrix, $\tilde{F}$, which from (3.3) is a function of P, F, G, H, S and R. Chan, Goodwin and Sin (1984) define a <u>strong</u> steady-state solution, P, of (3.3e) as one for which the eigenvalues of $\tilde{F}$ lie <u>inside or on</u> the unit circle, and a <u>stabilizing</u> solution as one for which these eigenvalues lie <u>strictly inside</u> the unit circle. Usually no other steady state exists, as is clear from the form of the error process (3.4), but exceptions are possible, as discussed in Section 4.4 below.

Before giving the theorem statements, it is helpful to show that a system in which the input and measurement noises are correlated may be converted to one in which they are not, with considerable notational savings in the ensuing analysis of the covariance recursions. Consider the system (3.1)-(3.2), which we rewrite here for convenience:

$$x_{t+1} = Fx_t + Gw_t$$

$$y_t = H'x_t + v_t$$

where

$$E\{(w_t', v_t')'(w_r', v_r')\} = \begin{pmatrix} Q & S \\ S' & R \end{pmatrix} \delta_{tr}$$

and $Q \geq 0$, $R > 0$ (i.e. positive semi-definite and positive definite, respectively). Replacing w_t by the residual in its projection on v_t, and incorporating the matrix G into the definition, so that $w_t^* = G(w_t - SR^{-1}v_t)$, the state transition equation may be rearranged as

$$x_{t+1} = (F - GSR^{-1}H')x_t + w_t^* + GSR^{-1}y_t$$

$$= F^*x_t + w_t^* + GSR^{-1}y_t ,$$

also using the measurement equation. The covariance of w_t^* and v_t is

$$E\{(w_t^{*\prime}, v_t')'(w_t^{*\prime}, v_t')\} = \begin{pmatrix} GQG' - GSR^{-1}S'G' & 0 \\ 0 & R \end{pmatrix} .$$

Since at time t, $GSR^{-1}y_t$ is known, the covariance $P_{t+1,t}$ is unaffected by the presence of this term, and its properties may be obtained by reference to those of the system

$$
\begin{aligned}
x_{t+1} &= F^*x_t + w_t^* \\
y_t &= H'x_t + v_t
\end{aligned}
\qquad (3.7)
$$

with the covariance matrix given above, in particular with w_t^* and v_t uncorrelated. For convenience we factorize the positive semi-definite covariance of w_t^* in the form $DD' \equiv G(Q - SR^{-1}S')G'$. The following results now relate to the model (3.7).

<u>Theorem 3.1</u> (Chan, Goodwin and Sin, 1984)

If the pair (F^*,H) is detectable and F^* is non-singular, then

(i) there exists a unique strong solution

(ii) if the pair (F^*,D) is stabilizable, then the strong solution is the only non-negative definite solution

(iii) if the pair (F^*,D) has no uncontrollable eigenvalues on the unit circle, then the strong solution is also stabilizing

(iv) if the pair (F^*,D) has an uncontrollable eigenvalue on the unit circle, then there is no stabilizing solution

(v) if the pair (F^*,D) has an uncontrollable eigenvalue inside or on the unit circle, then the strong solution is not positive definite

(vi) if the pair (F^*,D) has an uncontrollable eigenvalue outside the unit circle, then there is at least one other solution besides the strong solution.

Sufficient conditions for convergence of the covariance $P_{t+1,t}$ to the strong solution are given in the following theorems. The first, due to Hagar and Horowitz (1976), generalizes a result for controllable systems, obtained by Caines and Mayne (1970), to the stabilizable case.[5]

Theorem 3.2 (after Hager and Horowitz, 1976, who give the dual control result)

If (F^*,D) is stabilizable and (F^*,H) is detectable, then the sequence of covariance matrices $P_{0,-1}, P_{1,0},\ldots, P_{t+1,t},\ldots$ converges (exponentially fast) to the unique stabilizing solution, P, from all initial conditions $P_{0,-1} \geq 0$.

Theorem 3.3 (Chan, Goodwin and Sin, 1984)

If (F^*H) is observable and $P_{0,-1} - P \geq 0$, then the sequence of covariance matrices $P_{0,-1}, P_{1,0},\ldots, P_{t+1,t},\ldots$ converges to the strong solution, P.

Theorem 3.4 (Caines and Mayne, 1970, 1971)

If $R>0$, $DD'\geq 0$, (F^*,D) is controllable and (F^*,H) is detectable, then the sequence $P_{0,-1}, P_{1,0},\ldots, P_{t+1,t},\ldots$ converges to the unique positive semi-definite stabilizing solution, P, from all $P_{0,-1} \geq 0$.

4. FORECASTING THE ARMA(p,q) PROCESS

4.1 Setting up the problem

As is apparent from the previous discussion, once the problem in hand has been cast in state-space form corresponding to (3.1)-(3.2), the construction of l.l.s. forecasts is straightforward. Our principal objective in this section is to show how the Kalman filter framework may be used to extend the well-known theory of forecasting stationary series to more general cases. That this is possible has often been stated; the recent results of Chan, Goodwin and Sin (1984) are required, however, to treat the non-invertible moving average case in any generality.

We consider a scalar variable y which is generated by the ARMA(p,q) process

(4.1) $$\phi(L)\, y_t = \theta(L)\, \varepsilon_t$$

where $\phi(L)$ and $\theta(L)$ are polynomials in the lag operator L of degrees p and q respectively, and ε_t is white noise. A state-space representation in which the first element of the state vector is the forecastable part of y_t, that is $x(1)_t = y_t - \varepsilon_t$, is obtained from the following equivalences. Define $r = \max(p,q)$, and $\phi_{p+1} = \ldots = \phi_r = 0$ if $p<q$, or $\theta_{q+1} = \ldots = \theta_r = 0$ if $p>q$. Let

$$F = \begin{pmatrix} \phi_1 & 1 & 0 & \ldots & 0 \\ \phi_2 & 0 & & & \vdots \\ \vdots & \vdots & & 1 & 0 \\ & & & 0 & 1 \\ \phi_r & 0 & \ldots & 0 & 0 \end{pmatrix}, \quad F_\theta = \begin{pmatrix} \theta_1 & 1 & 0 & \ldots & 0 \\ \theta_2 & 0 & & & \vdots \\ \vdots & \vdots & & 1 & 0 \\ & & & 0 & 1 \\ \theta_r & 0 & \ldots & 0 & 0 \end{pmatrix}$$

$$G = (\phi_1 - \theta_1, \phi_2 - \theta_2, \ldots, \phi_r - \theta_r)'$$

$$H = (1, 0, \ldots, 0)' .$$

Then defining the state transition and measurement equations as

(4.2) $$x_t = Fx_{t-1} + G\varepsilon_{t-1}$$

(4.3) $$y_t = H'x_t + \varepsilon_t ,$$

the latter gives $x(1)_t = y_t - \varepsilon_t$ as above. Using this repeatedly, and substituting in turn from the rth, (r-1)th,... rows of (4.2) to the first gives

$$x(1)_t = \sum_{j=1}^{r} \phi_j y_{t-j} - \sum_{k=1}^{r} \theta_k \varepsilon_{t-k}$$

as required. A simple relation between the transition matrices and the ARMA coefficient polynomials is $|I-Fz| = \phi(z)$, $|I-F_\theta z| = \theta(z)$. The system (4.2)-(4.3) corresponds to (3.1)-(3.2), with $w_t = v_t = \varepsilon_t$, and $Q = S = R = \sigma_\varepsilon^2$ (all scalars). In the alternative form (3.7) the matrix $F^* = F - GSR^{-1}H$ is equal to F_θ, defined above, in this case, and $\text{cov}(w_t^*) = G(Q-SR^{-1}S')G' = 0$.

To express the forecast as a function of the observations and to consider the steady-state filter we examine the form taken by equations (3.3) in the present case. First, the covariance recursion can be greatly simplified. From (3.3c,e,f) applied to the alternative form (3.7), as specialized in the preceding paragraph, we obtain

(4.4) $$P_{t+1,t} = F_\theta\{P_{t,t-1} - P_{t,t-1}H\Sigma_t^{-1}H'P_{t,t-1}\}F_\theta' .$$

Furthermore, the closed-loop system matrix is now

(4.5) $$\tilde{F}_t = F_\theta\{I - P_{t,t-1}H\Sigma_t^{-1}H'\} .$$

We note that these expressions do not involve the autoregressive parameters so that, for example, in computing the one-step-ahead forecast error variance we need only consider the initial conditions and the moving average operator; specifically, the stationarity or otherwise of the process is irrelevant. The limiting behaviour of (4.4) depends on the location of the eigenvalues of F_θ, which are the roots of $\theta(L)$. We consider the various possibilities case by case, using the convergence and existence theorems given in section 3.2. To apply these to the ARMA(p,q) form (4.2)-(4.3) we note first that (F^*,H) is observable since for any vector b, $F^*b = \lambda b$ and $H'b = 0$ together imply $b = 0$. Secondly, (F^*,D) is not controllable since $D = 0$; it is stabilizable if and only if all roots of $\theta(L)$, the eigenvalues of F^*, lie inside the unit circle.

In the statement of Theorem 3.1, F^* is restricted to be non-singular, which in this context requires $r=q\geq p$. In fact the restriction to non-singularity is inessential in the general theorem, but the details of a proof for the singular case have not yet appeared. In the present problem we can obtain the result we require for the case $p>q$, as follows. Let $m = p - q$, and factor the autoregressive operator as $\phi(L) = \phi_1(L)\phi_2(L)$, where $\phi_2(L)$ is of order m. Then defining the quasi-differenced series $y_t^* = \phi_2(L)y_t$, we have

$$\phi_1(L)y_t^* = \theta(L)\varepsilon_t ,$$

and in the corresponding reduced order system $r=q$ so that F_θ is non-singular. The forecast of y_{t+1} is then obtained as

$$\hat{y}_{t+1,t} = \hat{y}^*_{t+1,t} + \phi_{21}y_t + \ldots + \phi_{2m}y_{t-m+1}$$

from which it is clear that the variance of $\hat{y}_{t+1,t}$ is just that of $\hat{y}^*_{t+1,t}$, once $t\geq m-1$. Thus we need only consider the case $q\geq p$ in what follows.

4.2 The invertible moving average case

If the roots of $\theta(L)$ lie inside the unit circle, it follows from Theorems 3.1 and 3.2 that the sequence of covariance matrices $P_{0,-1},\ldots, P_{t,t-1}, P_{t+1,t},\ldots$ converges to the unique non-negative definite fixed point P of (4.4) from all bounded non-negative definite $P_{0,-1}$. Since, as is clear by inspection, P=0 is such a fixed point, and the associated steady-state gain (K), innovation variance (Σ) and closed loop system matrix ($\tilde{F}$) are G, R and F_θ respectively, the steady-state filter takes the simple form

$$\hat{x}_{t+1,t} = F_\theta \hat{x}_{t,t-1} + Gy_t .$$

Repeated substitution in this equation gives

$$\hat{x}_{t+1,t} = \sum_{i=0}^{t} F_\theta^i G y_{t-i} + F_\theta^t \hat{x}_{0,-1}$$

and from (3.5) the j-step ahead forecast of y is given as

$$\hat{y}_{t+j,t} = H'F^{j-1}\hat{x}_{t+1,t} \qquad j=1,2,\ldots .$$

Combining these two equations, we obtain a generating function for the weight on y_{t-i} in the forecast of y_{t+j} as

$$(4.6) \qquad f_j(z) = H'F^{j-1}\{I - zF_\theta\}^{-1}G \qquad j=1,2,\ldots .$$

The forecast error variance is

$$V_j = H'P_{t+j,t}H + R$$

where, using (3.5b) and noting that Q=R,

$$P_{t+j,t} = \sum_{i=0}^{j-2} F^i GRG'(F')^i .$$

If the eigenvalues of F (that is, the roots of $\phi(L)$) lie inside the unit circle, we see from (4.6) that as the forecast horizon, j, increases so the observed history ceases to be informative, and the forecast $\hat{y}_{t+j,t}$ tends to the unconditional mean, zero. Similarly, V_j tends to the variance of y. Conversely, if $\phi(L)$ is not invertible, neither the forecast, nor its error variance in general approach limits as j increases. We note further that since P=O, we must have $e_t = 0$, and $\hat{x}_{t+1,t} = x_{t+1}$, hence $\tilde{y}_{t+1}$, the innovation is equal to ε_{t+1}; comparing (3.4) we have $Kv_t = Gw_t$ since K=G and $v_t = w_t = \varepsilon_t$.

Finally, we can relate (4.6) to the conventional formulae as follows. First, an expression appearing in (4.6) can be written in terms of the ARMA polynomial operators as

$$H'(I - zF_\theta)^{-1}G = z^{-1}\theta^{-1}(z)\{\theta(z) - \phi(z)\} ,$$

hence we also have

$$\left(1 - zH'(I - zF_\theta)^{-1}G\right) = \phi(z)\theta^{-1}(z) .$$

Next, since $F_\theta = F - GH'$ we may write

$$I - zF = \{I - zGH'(I - zF_\theta)^{-1}\}(I - zF_\theta).$$

If $\phi(L)$ is invertible, so that $(I - zF)^{-1}$ exists for $|z|=1$, this gives

$$\begin{aligned}(I - zF_\theta)^{-1}G &= (I - zF)^{-1}\{I - zGH'(I - zF_\theta)^{-1}\}G \\ &= (I - zF)^{-1}G\{I - zH'(I - zF_\theta)^{-1}G\}\end{aligned}$$

so that on substituting into (4.6) we obtain

$$f_j(z) = H'F^{j-1}(I - zF)^{-1}G \,.\, \phi(z)\theta^{-1}(z) .$$

Since $(I - zF)^{-1}$ may be expanded as a power series, we may write the right-hand side as

$$H'\left(z^{-(j-1)}(I - zF)^{-1}G\right)_+ \phi(z)\theta^{-1}(z)$$

which using results above may be written

$$\left(z^{-j}\ \phi^{-1}(z)\{\theta(z) - \phi(z)\}\right)_+ \phi(z)\theta^{-1}(z)$$

to give, finally,

$$(4.7) \qquad f_j(z) = \left(z^{-j}\frac{\theta(z)}{\phi(z)}\right)_+ \phi(z)\theta^{-1}(z)$$

which is the formula to be found in Whittle (1963, ch.3). It is clear from its definition at (4.6) that $f_j(z)$ has a convergent power series expansion even when $(I - zF)^{-1}$ does not exist, and for this reason (4.6) is to be preferred to (4.7), in general, as a basis for calculating the coefficients.

4.3 Moving average with roots on the unit circle

If $\theta(L)$ has roots inside and on the unit circle, then by Theorems 3.1 and 3.3, the covariance sequence again converges to P=0, the unique non-negative definite fixed point of (4.4). In this case, however, the closed loop system matrix, $\tilde{F}=F_\theta$, has at least one unit eigenvalue if $\theta(L)$ has any unit roots. Then (4.6) no longer holds in the usual sense, that is $\sum_{i=0}^{\infty} z^i F_\theta^i \neq (I-zF_\theta)^{-1}$ for $|z| = 1$, but if it is regarded primarily as shorthand for the power series in F_θ, we may think of taking $|z| < 1$ and deriving (4.7) as before. To avoid such heuristic devices, however, it is clearly preferable to write the generating function of the coefficients on past observations in the forecast of y_{t+j} in the explicit form

$$(4.8) \qquad f_j(z) = H'F^{j-1}\left(\sum_{i=0}^{t} z^i F_\theta^i\right)G\ ,$$

which indicates how its coefficients should be calculated.

4.4 Moving average with roots outside the unit circle

If $\theta(L)$ has any roots outside the unit circle, it follows from Theorem 3.1 that there are at least two steady states, but that only one of these yields a closed-loop system matrix with eigenvalues inside or on the unit circle. As shown below, the matrix $\tilde{F}$ then has eigenvalues which are the roots of $\theta(L)$ inside or on the unit circle together with the complex conjugate inverses of those roots outside $|z| = 1$. Furthermore, if $P_{0,-1}$ is positive definite, which is the case of practical relevance, it follows from Theorem 3.3 that the sequence of covariance matrices tends to this 'strong' solution.

Obviously, P=0 is again a solution of (4.4) with a filter z-transform again given by (4.8), but it is not the strong solution. To find the strong solution we first note that $\tilde{F} = F_\theta\{I - PH\Sigma^{-1}H'\}$ is of the same form as F_θ. To see this, note that the matrix $PH\Sigma^{-1}H'$ has a single non-zero column followed by r-1 columns of zeros, so that $\tilde{F}$ differs from F_θ only in the first column. Thus if we let $\theta^*(L)$ denote the lag operator which corresponds to $\tilde{F}$ in the same way that $\theta(L)$ corresponds to F_θ, the steady-state filter in question has z-transform

(4.9) $$f_j(z) = H'\left(F^{j-1}\sum_{i=0}^{t} z^i F_{\theta*}^i\right)G .$$

This filter has one-step-ahead variance Σ (the innovation variance) which is given by the solution to the invertible factorization of the c.g.f. of $\phi(L)y_t$, namely

(4.10) $$\Sigma\theta^*(z)\theta^*(z^{-1}) = R\theta(z)\theta(z^{-1}) .$$

To establish this result, first observe that since Σ is independent of the autoregressive operator, we can take this to be invertible without loss of generality. Since Q=R=S, the c.g.f. of y_t is then

$$(4.11) \qquad g_{yy}(z) = \{1 + zH'(I-zF)^{-1}G\}R\{G'(I-z^{-1}F')^{-1}Hz^{-1} + 1\} ,$$

and, as shown in the Appendix, this may be written

$$(4.12) \qquad g_{yy}(z) = \{1 + zH'(I-zF)^{-1}K\}\Sigma\{K'(I-z^{-1}F')^{-1}Hz^{-1} + 1\} .$$

Now $g_{yy}(z)$ is a scalar, and is thus equal to its determinant, and using the result that if T_1 is n×m, T_2 m×n, then $|I_n + T_1T_2| = |I_m + T_2T_1|$, we have, for example

$$\begin{aligned} |1 + zH'(I-zF)^{-1}G| &= |I + z(I-zF)^{-1}GH'| \\ &= |(I-zF)^{-1}| \,.\, |I - zF + zGH'| \\ &= |(I-zF)^{-1}| \,.\, |I - zF_\theta| . \end{aligned}$$

Similarly, $|1 + zH'(I-zF)^{-1}K| = |(I-zF)^{-1}| \,.\, |I - z\tilde{F}|$; hence equating (4.11) and (4.12) and cancelling common factors gives

$$|I - zF_\theta|.R.|I - z^{-1}F'_\theta| = |I - z\tilde{F}|.\Sigma.|I - z^{-1}\tilde{F}'| .$$

But $|I - zF_\theta| = \theta(z)$, and $|I - z\tilde{F}| = \theta^*(z)$, so this gives (4.10), as asserted.

This treatment provides a generalization of the discussion of the ARMA(1,1) example in Section 2.1, and the same general considerations apply. As with the example in equation (2.13), and seen to be true more generally, for a non-invertible moving average process the l.l.s. recursions deliver a steady-state forecasting rule of an "invertible" form, automatically parameterized in terms of the coefficients of the observationally equivalent invertible moving average. This coincides with the Wiener-Kolmogorov predictor in the case in which the latter is defined, but is also valid if the process is non-stationary.

5. SIGNAL EXTRACTION IN UNOBSERVED-COMPONENT ARMA MODELS

5.1 Setting up the problem

We now assume that the observations, y_t, are the sum of two or more unobserved components, each of which is an ARMA process of known form. Such specifications are frequently employed in studies of seasonal adjustment, where we find decompositions of time series into either two components (one seasonal, the other not) or three components (seasonal, cyclical, irregular), one of which is white noise. In either case the problem of interest is to estimate the seasonal component, in order to subtract it from the observed series to give a "seasonally adjusted" series. A three-component specification can always be reduced to two components by absorbing the white noise into one of the other components. Since the converse is not true unless at least one of the (non-white) processes has MA order greater than or equal to its AR order, the two-component specification is more general. From the point of view of state-space representations, however, it is more natural to think of the two-component specification as a special case of three components that arises when the noise in the measurement equation, v_t, has zero variance ($R=0$). Since results for two-component processes can be recovered from those for three components by taking $R=0$, we prefer to work with the three-component specification.

In this section we apply the general results of Section 3 to indicate the circumstances under which the well-known theory of signal extraction for stationary components may be extended to cover difference-stationary or explosively non-stationary cases. Some of the results for difference-stationary processes have appeared elsewhere (Cleveland and Tiao, 1976; Pierce, 1979) or been foreshadowed by results for similar, but not identical, specifications (Hannan, 1967; Sobel, 1967), but have not, in our view, rested on firm foundations. The treatment of explosively non-stationary processes that the Kalman filter makes possible appears to be new, at least in the present context. Again we seek equivalences between the Kalman filter and classical results, in the cases in which the latter are defined. A non-detectable example is reserved until Section 6.

We consider the following three-component model

$$(5.1) \qquad y_t = S_t + C_t + I_t \qquad t=0,1,2,...$$

where S_t and C_t are ARMA processes, and I_t is white noise. That is, $\phi^s(L)S_t = \theta^s(L)w_{1t}$, $\phi^c(L)C_t = \theta^c(L)w_{2t}$, and $I_t = v_t$, where the four lag polynomials are of degree m, n, p and q respectively, and the uncorrelated white noise variables w_{1t}, w_{2t} and v_t have variances $\sigma^2_{w_1}$, $\sigma^2_{w_2}$ and σ^2_v. We seek the l.l.s. estimate of S_t given observations on y and appropriate initial conditions. To do this we employ the general results in Section 3, having first cast the problem in a suitable state-space form. The representation of an ARMA process employed in the previous section is no longer appropriate, and instead we employ the following equivalences, which have the incidental merit of retaining the original variables of (5.1) in the state vector:

$$x_t = (x_{1t}', x_{2t}')'$$

$$x_{1t} = (S_t, S_{t-1}, \ldots, S_{t-m+1}, w_{1,t}\ w_{1,t-1}, \ldots, w_{1,t-n+1})'$$

$$x_{2t} = (C_t, C_{t-1}, \ldots, C_{t-p+1}, w_{2,t}\ w_{2,t-1}, \ldots, w_{2,t-q+1})'$$

$$F = \text{block diagonal } (F_1, F_2)$$

(5.2)
$$F_1 = \left[\begin{array}{ccccc|cccc} \phi_{s,1} & \phi_{s,2} & \cdots & \phi_{s,m-1} & \phi_{s,m} & -\theta_{s,1} & -\theta_{s,2} & \cdots & -\theta_{s,n} \\ 1 & 0 & \ldots & 0 & 0 & & & & \\ 0 & 1 & \ldots & 0 & 0 & & & 0 & \\ \vdots & & & & \vdots & & & & \\ 0 & 0 & \ldots & 1 & 0 & & & & \\ \hline & & & & & 0 & 0 & \ldots & 0 \\ & & & & & 1 & 0 & \ldots & 0 \\ & & 0 & & & 0 & 1 & \ldots & 0 \\ & & & & & \vdots & & & \vdots \\ & & & & & 0 & 0 & \ldots\ 1 & 0 \end{array}\right]$$

F_2 is similarly defined by matching coefficients in the model for C_t to elements of x_{2t},

$$G' = \begin{bmatrix} 1\ 0 \ldots 0\ 1\ 0 \ldots 0\ 0\ 0 \ldots 0\ 0\ 0 \ldots 0 \\ 0\ 0 \ldots 0\ 0\ 0 \ldots 0\ 1\ 0 \ldots 0\ 1\ 0 \ldots 0 \end{bmatrix}$$

$$H' = (1\ 0 \ldots \qquad 0\ 1\ 0 \ldots \qquad 0)$$

$$w_t = (w_{1t}, w_{2t})' \qquad Q = \text{diag}\left(\sigma^2_{w_1}, \sigma^2_{w_2}\right)$$

$$R = \sigma^2_v \text{ (scalar)} \qquad S = 0 .$$

Although there are other representations, similar to that used in the previous section, which are of smaller order, the representation (5.2) is convenient to use, given our objective of connecting the Kalman filter results to the usual results for stationary ARMA processes; an empirical application of this representation is given by Burridge and Wallis (1985).

Our first concern is to establish conditions under which the covariance generated by (3.3e) goes to a steady state. Before applying relevant theorems from Section 3.2 we need to check the detectability and controllability of the system (5.2). Since S=0, the matrices defined in (3.7) are $F^* = F$ and $DD' = GQG'$. The eigenvalues of F are the union of those of F_1 and F_2, which are obtained from the characteristic equations $|F_1 - \lambda I| = (-\lambda)^{n+m}\phi^s(\lambda^{-1})$ and $|F_2 - \lambda I| = (-\lambda)^{p+q}\phi^c(\lambda^{-1})$. Thus the eigenvalues of F are $\{\lambda_{si}\}$ and $\{\lambda_{cj}\}$ where $\phi^s(L) = \prod_{i=1}^{m}(1 - \lambda_{si}L)$ and $\phi^c(L) = \prod_{j=1}^{p}(1 - \lambda_{cj}L)$, together with n+q zeros.

(i) Detectability: in the system (5.2), the pair (F^*,H) is detectable if and only if the polynomials $\phi^s(L)$ and $\phi^c(L)$ contain no unstable common factor.

Proof. Suppose b is a right eigenvector of F. Now $H'b = b_1 + b_{m+n+1}$, and if λ is an eigenvalue of either F_1 or F_2, but not of both, then either b_{m+n+1} or b_1 must be zero. However, it is easy to check that since the top left blocks of F_1 and F_2 are companion matrices, their corresponding non-trivial eigenvectors cannot have zero first elements. Thus if F_1 and F_2 have no eigenvalues in common, then $H'b \neq 0$ for all λ and the pair (F,H) is observable. Conversely, if λ is an eigenvalue of <u>both</u> F_1 and F_2, the vector $b = \{1, \lambda^{-1}, \ldots, \lambda^{-(m-1)}, 0, \ldots, 0, -1, -\lambda^{-1}, \ldots, -\lambda^{-(p-1)}, 0, \ldots, 0\}'$ is a right eigenvector of F, and satisfies $H'b = 0$. Finally, F_1 and F_2 have common eigenvalue λ if and only if $(1-\lambda L)$ is a factor of both $\phi^s(L)$ and $\phi^c(L)$.

(ii) Controllability: in the system (5.2), the pair (F^*,D) is controllable.

Proof. Noting that $D = GQ^{\frac{1}{2}}$ has the form

$$D = \begin{bmatrix} \sigma_{w_1} & 0 \dots 0 & \sigma_{w_1} & 0 \dots 0 & 0 & 0 \dots 0 & 0 & 0 \dots 0 \\ 0 & 0 \dots 0 & 0 & 0 \dots 0 & \sigma_{w_2} & 0 \dots 0 & \sigma_{w_2} & 0 \dots 0 \end{bmatrix}'$$

we see that $a'D = 0$ implies that $a_1 + a_{m+1} = 0$ and $a_{m+n+1} + a_{m+n+p+1} = 0$; but then a' cannot be a non-trivial left eigenvector of F, as is clear by inspection.

From Theorem 3.4 it thus follows that the covariance, $P_{t+1,t}$, converges to the unique positive semi-definite stabilizing solution to (3.3) from all $P_{0,-1} \geq 0$ if and only if $\phi^S(L)$ and $\phi^C(L)$ have no common factor $(1 - \lambda L)$ with $|\lambda| \geq 1$. Furthermore, this convergence is exponentially fast, by Theorem 3.2, and so the result is of practical use in the non-stationary case, as in the examples of Burridge and Wallis (1985). The ensuing discussion falls naturally into three parts, dealing with first the case in which $\phi^S(L)$ and $\phi^C(L)$ are both invertible, then the general detectable case, and finally the undetectable case in which a common factor of $(1-L)$ is present.

5.2 The stationary case

If S_t and C_t are stationary processes, the classical l.l.s. theory presented, for example, by Whittle (1963, ch.6) may be applied to obtain the estimate of S_t given observations up to y_{t+k} in the form

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

where the coefficient of y_{t-j}, $j \geq -k$, is the coefficient of z^j in the first element of the power series expansion of

$$(5.3) \qquad f_k(z) = \left[g_{xy}(z) \{g_{yy}^-(z)\}^{-1} \right]_{-k} \{g_{yy}^+(z)\}^{-1} .$$

In this expression the covariance generating function of y, obtained directly from (3.1)-(3.2) with S=0 as

$$(5.4) \qquad g_{yy}(z) = R + H'(I - zF)^{-1} GQG'(I - z^{-1}F')^{-1} H$$

is to be factored in the form

$$g_{yy}(z) = g^{+}_{yy}(z)g^{-}_{yy}(z) .$$

In the classical approach this factorization problem is solved by expressing the c.g.f. of $\phi^{s}(L)\phi^{c}(L)y_t$, which is

$$\sigma_v^2 \phi^{s}(z)\phi^{c}(z)\phi^{s}(z^{-1})\phi^{c}(z^{-1}) + \sigma_{w_1}^2 \phi^{c}(z)\theta^{s}(z)\phi^{c}(z^{-1})\theta^{s}(z^{-1}) + \sigma_{w_2}^2 \phi^{s}(z)\theta^{c}(z)\phi^{s}(z^{-1})\theta^{c}(z^{-1}) ,$$

in the invertible form $\sigma^2\beta(z)\beta(z^{-1})$. In the present framework the corresponding result is obtained via the steady-state filter and the innovations, $\tilde{y}_t$. The steady state version of (3.3g) is

$$\hat{x}_{t+1,t} = F\hat{x}_{t,t-1} + K\tilde{y}_t ,$$

and repeated substitution, and the definition (3.3i), yields y_t as a linear combination of the innovations

$$y_t = \tilde{y}_t + \sum_{i=0}^{t-1} H'F^{i}K\tilde{y}_{t-1-i} . \tag{5.5}$$

Since $\tilde{y}_t$ is a white noise process with constant variance Σ, and F has eigenvalues inside the unit circle, we obtain the c.g.f. of y_t directly from (5.5) as

$$g_{yy}(z) = \{1 + zH'(I-zF)^{-1}K\}\Sigma\{1 + z^{-1}K'(I-z^{-1}F')^{-1}H\} \tag{5.6}$$

which is the required factorization. We show that (5.4) and (5.6) are identical by an algebraic argument in the Appendix.

The remaining quantity in (5.3), the covariance of x_t and y_t, is again obtained directly from (3.1)-(3.2) as

$$(5.7) \qquad g_{xy}(z) = (I - zF)^{-1}GQG'(I - z^{-1}F')^{-1}H .$$

Since $\bar{g}_{yy}(z)$ is a scalar, attention can be restricted to the first element of $g_{xy}(z)$. To express this in a more familiar lag polynomial form we first note that the inverse of $(I - zF)$ is block diagonal; denoting these blocks by A_1 and A_2, and those of $(I - z^{-1}F')^{-1}$ by B_1 and B_2 it then follows from the form of G, H and Q in (5.2) that the first element of (5.7) is

$$(5.8) \qquad g_{sy}(z) = \sigma^2_{w_1}\{A_1(1,1) + A_1(1,m+1)\}\{B_1(1,1) + B_1(m+1,1)\} .$$

Writing F_1 in the partitioned form $F_1 = \begin{pmatrix} M_{11} & M_{12} \\ 0 & M_{22} \end{pmatrix}$, we see that

$$A_1 = (I-zF_1)^{-1} = \begin{pmatrix} (I-zM_{11})^{-1} & -(I-zM_{11})^{-1}(-zM_{12})(I-zM_{22})^{-1} \\ 0 & (I-zM_{22})^{-1} \end{pmatrix} .$$

Now $A_1(1,1) = \{\phi^s(z)\}^{-1}$, and a little further algebra shows that $A_1(1,m+1) = \{\phi^s(z)\}^{-1}\{\theta^s(z)-1\}$ so that, with a symmetric treatment of B_1, the right-hand side of (5.8) is $\sigma^2_{w_1}\{\phi^s(z)\phi^s(z^{-1})\}^{-1}\theta^s(z)\theta^s(z^{-1})$. Using this expression in (5.3), and writing (5.6) as $\sigma^2\beta(z)\beta(z^{-1})$, we obtain (5.3) in the form of Whittle's (6.1.13), and equation (4.5) of Pierce (1979), whose treatment admits difference-stationary components. On the other hand, substituting from (5.6) and (5.7) gives (5.3) in terms of the system matrices as

$$(5.9) \qquad \begin{aligned} f_k(z) &= \left[(I-zF)^{-1}GQG'(I-z^{-1}F')^{-1}H\{1 + z^{-1}K'(I-z^{-1}F')^{-1}H\}^{-1}\right]_{-k} \\ &\quad \times \Sigma^{-1}\{1 + zH'(I-zF)^{-1}K\}^{-1} , \end{aligned}$$

and to complete our account of the equivalences we show that the steady-state Kalman filter does indeed lead to this expression.

Let us suppose, then, that $P_{0,-1} = P$, a fixed point of (3.3e), so that all the resulting quantities in (3.3) and (3.6) are time-invariant, that is $C_{t,t+k} = C_k$ for all t and k=0,1,2,..., $\Sigma_t = \Sigma$, $K_t = K$, and $\tilde{F}_t = \tilde{F}$, with the eigenvalues of $\tilde{F}$ inside the unit circle. The one-step-ahead estimates of the state, $\hat{x}_{t+1,t}$, may then be expressed, by making repeated substitutions in (3.3g), in the non-recursive form

$$(5.10) \qquad \hat{x}_{t+1,t} = \sum_{i=0}^{t} \tilde{F}^i K y_{t-i} + \tilde{F}^t \hat{x}_{0,-1} ,$$

This estimator is, strictly speaking, affine, rather than linear in y, because of the presence of the initial conditions. In the classical theory, $\hat{x}_{0,-1}$ would be set to its unconditional value, zero, resulting in a linear form, as it were fortuitously. Since $\tilde{F}^t$ converges to zero, however, the effect of $\hat{x}_{0,-1}$ on the estimate of x_{t+1} dies out as t increases and for simplicity we take $\hat{x}_{0,-1} = 0$ from here on. The coefficient of y_{t-i} in (5.10) is given by the coefficient of z^i in the generating function

$$(5.11) \qquad f_{-1}(z) = \{I - z\tilde{F}\}^{-1} K ,$$

from which it follows that the innovation, $\tilde{y}_t$, may be expressed as a linear combination of the observations with generating function

$$(5.12) \qquad \tilde{w}(z) = 1 - zH'\{I - z\tilde{F}\}^{-1} K .$$

In steady state the smoothed estimate of x_t given Ω_{t+k} takes the form (cf. 3.6a)

$$(5.13) \qquad \hat{x}_{t,t+k} = \hat{x}_{t,t-1} + \sum_{i=0}^{k} C_i \tilde{y}_{t+i} ,$$

which may be expressed in terms of the observations by substituting for $\hat{x}_{t,t-1}$ and $\tilde{y}_{t+i}$ using (5.10)-(5.12), from which we obtain the generating function, centred at t:

(5.14) $$f_k(z) = zf_{-1}(z) + \sum_{i=0}^{k} z^{-i} C_i \tilde{w}(z) .$$

Here the coefficient of y_{t-j}, $j \geq -k$, is given by the coefficient of z^j, and $C_i = P(\tilde{F}')^i H\Sigma^{-1}$. Again all that is required for (5.14) to be a convergent power series in z is that the eigenvalues of $\tilde{F}$ have modulus less than unity, which is the case. Substituting for $f_{-1}(z)$ and C_i gives

(5.15) $$f_k(z) = z\{I - z\tilde{F}\}^{-1}K + P\left[\{I - z^{-1}\tilde{F}'\}^{-1}H\Sigma^{-1}\right]_{-k} \tilde{w}(z) .$$

As y_t is stationary, $\tilde{w}(z)$ is invertible, and this can be written as

(5.16) $$f_k(z) = \left[z\{I - z\tilde{F}\}^{-1}K\tilde{w}^{-1}(z)\Sigma + P\{I - z^{-1}\tilde{F}'\}^{-1}H\right]_{-k} \Sigma^{-1}\tilde{w}(z) .$$

We wish to show that the coefficient of z^j in (5.16) is the same as that in (5.9). First we note that their last factors coincide. From (5.12) we have

$$\tilde{w}(z) = |1 - zH'\{I - z\tilde{F}\}^{-1}K| = |I - z\{I - z\tilde{F}\}^{-1}KH'|$$

$$= |I - z\tilde{F}|^{-1} . |I - zF|$$

since $\tilde{F} = F - KH'$, again using $|I_n + T_1T_2| = |I_m + T_2T_1|$. Similar operations on the last term in (5.9) give

$$\{1 + zH'(I - zF)^{-1}K\}^{-1} = |I - zF| . |I - z\tilde{F}|^{-1}$$

as required. Likewise $\{1 + z^{-1}K'(I - z^{-1}F')^{-1}H\}^{-1} = \{1 - z^{-1}K'(I - z^{-1}\tilde{F}')^{-1}H\}$, so that

$$H\{1 + z^{-1}K'(I - z^{-1}F')^{-1}H\}^{-1} = \{I - z^{-1}HK'(I - z^{-1}\tilde{F}')^{-1}\}H$$

$$= \{I - z^{-1}F'\}\{I - z^{-1}\tilde{F}'\}^{-1}H .$$

Substituting this gives the expression inside the annihilation operator in (5.9) as

$$(5.17) \qquad (I - zF)^{-1}GQG'(I - z^{-1}\tilde{F}')^{-1}H .$$

Turning finally to (5.15), in a similar fashion we have

$$\begin{aligned} z\{I - z\tilde{F}\}^{-1}K\tilde{w}^{-1}(z)\Sigma &= z\{I - z\tilde{F}\}^{-1}K\left(1 + zH'(I-zF)^{-1}K\right)\Sigma \\ &= z\{I - z\tilde{F}\}^{-1}(I + zKH'\{I-zF\}^{-1})K\Sigma \\ &= z(I - zF)^{-1}K\Sigma . \end{aligned}$$

Taking out a left factor of $(I - zF)^{-1}$, a right factor of $(I - z^{-1}\tilde{F}')^{-1}H$ and noting that $K\Sigma = FPH$, the expression in the annihilation operator in (5.16) becomes

$$(5.18) \qquad (I - zF)^{-1}\{zFP(I-z^{-1}\tilde{F}') + (I-zF)P\}(I - z^{-1}\tilde{F}')^{-1}H .$$

The middle term in this expression is then just $P - FP\tilde{F}'$, which is equal to GQG' (see 3.3e). Thus (5.17) and (5.18) coincide, and we have shown explicitly that the classical and steady-state Kalman filters are identical. Other authors have considered this equivalance, but so far as we are aware the precise connection between expressions such as (5.9) and (5.16) has not been made explicit (cf. Anderson and Moore, 1979, p.257; Whittle, 1983, p.151). The discussion in Section 2.2 concerns the same equivalence in respect of the filter $f_0(z)$, in the present notation, in a simple two-component example of the unobserved-component ARMA model.

5.3 The non-stationary detectable case

If $\phi^c(L)$ and/or $\phi^s(L)$ are not invertible, but contain no non-invertible common factor, then the covariance $P_{t+1,t}$ attains a steady state, P, and $\tilde{F}$ has all eigenvalues inside the unit circle. In this case, therefore, there is still a steady-state filter of the form (5.15), and the question that then arises is how much of the analysis in the preceding section survives? In particular, is there an expression corresponding to (5.9) in this case? To make progress, it is necessary to be clear about the practical significance of the various z-transforms manipulated above. Although we replace the series $I + z\tilde{F} + z^2\tilde{F}^2 + \ldots$ by its limit $\{I - z\tilde{F}\}^{-1}$, at (5.11), and also $I + zF + z^2F^2 + \ldots$ by $(I - zF)^{-1}$ at various points, the practical reality is that only a finite record is ever analysed, so that we could equally proceed with, for example, (5.11) written as

$$f_{-1,t}(z) = \sum_{i=0}^{t} z^i\tilde{F}^iK$$

and (5.12) as

$$\tilde{w}_t(z) = 1 - zH'\left\{\sum_{i=0}^{t} z^i\tilde{F}^iK\right\} .$$

Now an 'inverse' of $\tilde{w}_t(z)$ could be defined as

$$w(z) = 1 + zH'\left\{\sum_{i=0}^{\infty} z^iF^iK\right\} .$$

This is the z-transform for the coefficient on $\tilde{y}_{t-i}$ when y_t is written as a linear combination of the innovations, as in (5.5). It is the inverse of $\tilde{w}_t(z)$ in the sense that

$$w(z)\tilde{w}_t(z) = 1 + \text{terms resulting from truncation at any finite power of } z.$$

However, in order to apply (5.15) in practice, all we need ensure is that coefficients on powers of z from -k to t are correct. With this in mind (5.15) can be written as

$$(5.19) \qquad f_k(z) = \left\{z\{I+z\tilde{F}+z^2\tilde{F}^2+\ldots\}Kw(z) + P\{\sum_{j=0}^{k} z^{-j}(\tilde{F}')^jH\Sigma^{-1}\}\right\}\tilde{w}(z) + U(z)$$

where U(z) is a remainder involving powers of z greater than t (it is assumed that the first term inside the braces is expanded far enough to give the first term of (5.15) exactly, before the remainder is taken). Now by direct calculation the first term inside the braces of (5.19) can be seen to be $z(I + zF + z^2F^2 + \ldots)K$ plus a remainder due to truncation; the same argument applies, however, and so we can write

$$(5.20) \qquad f_k(z) = \left\{ z(I+zF+z^2F^2+\ldots)FPH + P\left(\sum_{i=0}^{k} z^{-i}(\tilde{F}')^i H\right)\right\} \Sigma^{-1}\tilde{w}(z) + U^*(z) .$$

Pursuing the same line of reasoning, we might seek to obtain $f_k(z)$ by expanding the following z-transform, obtained from (5.9) by replacing $(I - zF)^{-1}$ and $(I - z^{-1}F')^{-1}$ by the divergent power series $(I + zF + z^2F^2 + \ldots)$ and $(I + z^{-1}F' + z^{-2}(F')^2 + \ldots)$ respectively:

$$(5.21) \qquad f_k^*(z) = \left[(I+zF+z^2F^2+\ldots)GQG'(I+z^{-1}F'+z^{-2}(F')^2+\ldots)H\tilde{w}'(z^{-1})\right]_{-k} \Sigma^{-1}\tilde{w}(z) .$$

The requirement here is that the polynomial in negative powers of z inside the annihilation operator converges more rapidly than F^j diverges. Recall that

$$\tilde{w}'(z^{-1}) = 1 - z^{-1}K'(I + z^{-1}\tilde{F}' + z^{-2}(\tilde{F}')^2 + \ldots)H .$$

By direct calculation $f_k^*(z)$ can be written as

$$(5.22) \qquad f_k^*(z) = \left[(I+zF+z^2F^2+\ldots)GQG'(I+z^{-1}\tilde{F}'+z^{-2}(\tilde{F}')^2+\ldots)H\right]_{-k} \Sigma^{-1}\tilde{w}(z)$$

in which the coefficients on finite powers of z inside the annihilation operator are as follows:

coefficient of z^0 : $\sum_{i=0}^{\infty} F^i GQG'(\tilde{F}')^i H$

coefficient of z^{-j} : $\sum_{i=0}^{\infty} F^i GQG'(\tilde{F}')^{i+j} H$

coefficient of z^j : $\sum_{i=0}^{\infty} F^{i+j} GQG'(\tilde{F}')^i H$.

That these coefficients are bounded if F has eigenvalues on or inside the unit circle follows from the fact that those of $\tilde{F}$ are then strictly inside the unit circle. A simple rearrangement of (3.3e) in steady state gives $GQG' = P - F P\tilde{F}'$, so the coefficient of z^0 is then PH, and so on. We thus see that $f_k^*(z)$ coincides with $f_k(z)$ in (5.20) up to any desired power of z. The interest in this result is that formulae analogous to (5.21) for the case in which F has eigenvalues on or inside the unit circle have appeared in the literature, and although these provide a means of correctly calculating the coefficients, as we have shown, they have not been derived directly, as here. As a basis for calculating the coefficients, (5.15) is obviously to be preferred, however, since the intervention of rounding errors is thereby minimized. Whether any meaning can be attached to (5.22) in the explosive case remains an open question.

5.4 The non-detectable case

The model of (5.1) and (5.2) fails the detectability condition if the polynomials $\phi^S(L)$ and $\phi^C(L)$ have a common factor $(1-\lambda L)$ with $|\lambda| \geq 1$. Although cases with $|\lambda| > 1$ might be considered to be coincidental situations of little practical relevance, in the next section we consider a particular example whose main interest is in showing that Σ_t and K_t may converge to finite limits even when $P_{t+1,t}$ increases without limit. In the seasonal adjustment literature cases with $\lambda=1$ arise, through specifications in which $\phi^S(L)$ has a factor $(1-L^d)$ where d is the seasonal period, and $\phi^C(L)$ has a factor (1-L), and we briefly consider this case here.

Burridge and Hall (1986) show that convergence of the filter gain to a steady state, K, does occur when a common factor of (1-L) is present. The matrix $\tilde{F}$ has an eigenvalue of unity, however, and so we must consider carefully what meaning, if any, is to be attached to expressions such as (5.20) and (5.22) in these circumstances. The time-invariant Kalman filter can still be written as

$$(5.23) \qquad f_{k,t}(z) = zf_{-1,t}(z) + \sum_{i=0}^{k} z^i C^i \tilde{w}_{t+k}(z) ,$$

with the understanding that power series expansions are to be retained:

$$f_{-1,t}(z) = \sum_{i=0}^{t} z^i \tilde{F}^i K , \quad \tilde{w}_{t+k}(z) = 1 - zH' \sum_{i=0}^{t+k} z^i \tilde{F}^i K$$

and $C_i = P_{t,t-1}(\tilde{F})^i H$. That C_i is time invariant when K is follows from the convergence proof of Burridge and Hall (1986). Now exactly the same argument can be applied as in the previous section, except that $GQG' = P_{t+1,t} - FP_{t,t-1}\tilde{F}'$ and a time subscript on P needs to be retained in obtaining the equivalence of (5.22) and (5.20), since $P_{t+1,t} \neq P_{t,t-1}$ here. Again it is clear that (5.23) provides a better means of calculating the coefficients than (5.22). In the limit this filter coincides with that considered by Pierce (1979), who shows that if this filter delivers the conditional expectation of S_t (in the infinite-sample case) then the estimation error is non-stationary in the presence of a common factor (1-L).

6. A NON-DETECTABLE EXAMPLE

In this section we consider a particular example of the unobserved-component ARMA model analysed in the previous section. By direct argument we show that it is possible that K_t and $\tilde{F}_t$ converge to constants, even when $P_{t+1,t}$ does not, a case that has received relatively little attention in the control theory literature.

In the three-component model of Section 5, namely

$$y_t = S_t + C_t + I_t ,$$

it is now assumed that S_t and C_t are first-order autoregressive processes, with the parameters ϕ_s and ϕ_c respectively. Thus in the general representation

$$x_{t+1} = Fx_t + Gw_{t+1}$$

$$y_t = H'x_t + v_t$$

we have $F = \begin{pmatrix} \phi_s & 0 \\ 0 & \phi_c \end{pmatrix}$, $G = I_2$, $H' = \begin{pmatrix} 1 & 1 \end{pmatrix}$ and $Q = \mathrm{diag}(\sigma_s^2, \sigma_c^2)$. The previous discussion deals with the cases in which $\phi_s \neq \phi_c$ or $\phi_s = \phi_c = \phi$, say, with $|\phi| < 1$. We now assume that $\phi_s = \phi_c = \phi$ with $|\phi| \geq 1$, thus the system fails the detectability condition. For a two-component vector $K^* = (k_1^*, k_2^*)'$, it is easily verified that the eigenvalues of $\{F - K^*H'\}$ are ϕ and $\phi - k_1^* - k_2^*$, thus if $|\phi| \geq 1$ there is no K^* which renders this matrix stable. As a result, with $Q>0$, $P_{t+1,t}$ increases without limit. It is of interest to answer two questions, however: (i) What functions of the state vector can be estimated with bounded variance; (ii) Does the quantity K_t still approach a limit so that a time invariant rule of the usual form

$$\hat{x}_{t+1,t} = F\hat{x}_{t,t-1} + K\{y_t - H'\hat{x}_{t,t-1}\}$$

may eventually be applied? So far as we are aware, the latter possibility has not previously been formally investigated in the Kalman filter literature.

Both questions are most easily addressed by applying a non-singular transformation to the state vector which, as noted in Section 3.1, leaves the y process unaltered. Consider, then, the transformation $x_t^* = T^{-1}x_t$ with $T = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} = T^{-1}$, so that the system becomes

(6.1)
$$x_{t+1}^* = F^*x_t^* + w_{t+1}^*$$

$$y_t = H^{*\prime}x_t^* + v_t$$

with $F^* = T^{-1}FT = F$, $H^{*\prime} = H'T = (1, 0)$, and $E\{w_t^* w_t^{*\prime}\} = Q^* = \begin{pmatrix} \sigma_s^2+\sigma_c^2 & -\sigma_c^2 \\ -\sigma_c^2 & \sigma_c^2 \end{pmatrix}$.
Writing the variance of $x_{t+1}^* - \hat{x}_{t+1,t}^*$ as $P_{t+1,t}^*$ the recursions (3.3) applied to the transformed system give

$$(6.2) \qquad P_{t+1,t}^* = FP_{t,t-1}^*\tilde{F}_t^{*\prime} + Q^* ,$$

$$\tilde{F}_t^* = F - K_t^*H^{*\prime} = F - FP_{t,t-1}^*H^*\Sigma_t^{*-1}H^{*\prime} .$$

Noting the particular form of H^* in the present case we have

$$\Sigma_t^* = P_{t,t-1}^*(1,1) + R, \qquad P_{t,t-1}^*H^* = \begin{pmatrix} P_{t,t-1}^*(1,1) \\ P_{t,t-1}^*(2,1) \end{pmatrix} ,$$

so that (6.2) may be written out element-by-element to give

$$(6.3)\,(a) \quad P_{t+1,t}^*(1,1) = \phi P_{t,t-1}^*(1,1)\left(\phi - \phi P_{t,t-1}^*(1,1)\{P_{t,t-1}^*(1,1) + R\}^{-1}\right) + Q^*(1,1)$$

$$(b) \quad P_{t+1,t}^*(2,1) = \phi P_{t,t-1}^*(2,1)\left(\phi - \phi P_{t,t-1}^*(1,1)\{P_{t,t-1}^*(1,1) + R\}^{-1}\right) + Q^*(2,1)$$

$$(c) \quad P_{t+1,t}^*(2,2) = \phi\left(P_{t,t-1}^*(2,2) - P_{t,t-1}^*(2,1)\{P_{t,t-1}^*(1,1) + R\}^{-1}P_{t,t-1}^*(2,1)\right)\phi + Q^*(2,2) .$$

Now (6.3a) is of the same form as equation (2.19) of Section 2.2, so $P_{t+1,t}^*(1,1)$ converges monotonically to a positive constant $P^*(1,1)$, say, from all $P_{0,-1}^*(1,1) \geq 0$ by the argument used there. In the more general setting of Section 3, the transformation gives a decomposition of the original system into two sub-systems, the first of which is observable (cf. Preston and Pagan, 1982, §6.3). By the application of Theorem 3.2 to this sub-system, $P_{t,t+1}^*(1,1)$ converges, exponentially fast.

Turning to the second element, we first condense the notation by writing $\tilde{\phi}_t = \phi\left(1 - P_{t,t-1}^*(1,1)\{P_{t,t-1}^*(1,1) + R\}^{-1}\right)$, and $\tilde{\phi}$ as its steady-state value. Then (6.3b) possesses a fixed point if and only if $\phi\tilde{\phi} \neq 1$. The steady state of equation (6.3a) and the fact that $P^*(1,1) > 0$ implies that $0 < \phi\tilde{\phi} < 1$, hence (6.3b) possesses a unique fixed point for any ϕ. If $|\phi| = 1$, then $|\phi\tilde{\phi}_t| = |\tilde{\phi}_t| < 1$ for $t \geq 1$, and the recursion (6.3b) converges to this fixed point. If $|\phi| > 1$, the same conclusion, less obviously, can be established as follows. The condition for $|\phi\tilde{\phi}_t|$ to be less than unity is simply that $P^*_{t,t-1}(1,1) > (\phi^2 - 1)R$. From (6.3a) we have $P^*(1,1) > (\phi^2 - 1)R$, so even if $P^*_{0,-1}(1,1) < (\phi^2 - 1)R$, the convergence properties of $P^*_{t+1,t}(1,1)$ are such that we must have $P^*_{t+1,t}(1,1) > (\phi^2 - 1)R$ after a finite number of steps. It thus follows that $P^*_{t+1,t}(2,1)$ converges, since either $0 < \phi\tilde{\phi}_t < 1$ for all t, or $\phi\tilde{\phi}_t \geq 1$ for at most a finite number of steps.

Finally we see that $P^*_{t+1,t}(2,2)$ diverges, with a coefficient $\phi^2 \geq 1$ in the recursion (6.3c), the second term in parentheses on the right-hand side converging to a finite limit, as established above.

Returning to the original problem we note that $P_{t,t-1}H = TP^*_{t,t-1}T'(T')^{-1}H^* = TP^*_{t,t-1}H^*$ converges to a unique steady state by virtue of the preceding results. Clearly, then, $K_t = FP_{t,t-1}H\Sigma_t^{-1}$ converges to a limit in all cases, and we have an example in which K_t converges even though $P_{t+1,t}$ does not. That is, there is a well-behaved l.l.s. recursion, producing estimates whose variance is diverging. However, as the discussion of the transformed model makes clear, the sum $S_t + C_t$ can be estimated with bounded error variance. Since both S_t and C_t have the same autoregressive structure in this example, so does their sum, and considering only this composite signal reduces the problem to the case considered in Section 2.2, which is well-behaved whatever the value of the autoregressive parameter. Despite having an estimate of $S_t + C_t$ with bounded error variance, however, l.l.s. estimation of the separate components in the nonstationary case is subject to unbounded errors, which are offsetting in the sense that their sum is bounded, or their correlation coefficient approaches -1.

7. DISCUSSION

In the preceding sections we have shown how the theory of forecasting and signal extraction for stationary ARMA time series can be extended to cover non-stationary processes. While some new results have been obtained, it has been equally important to place some long-standing results on rigorous foundations. The techniques we employ have been hitherto most widely used in the control theory literature, and a parallel objective has been to relate that literature to the statistical time series literature. In particular, the connections between the Kalman filter apparatus and Wiener-Kolmogorov filters have been made explicit. It remains to comment briefly on a number of earlier treatments of non-stationary processes in the time series literature.

The prediction of a difference-stationary process, possibly masked by stationary noise, is discussed by Whittle (1963, ch.8.5). He obtains the z-transform of the l.l.s. predictor in the case where the AR operator contains a (possibly repeated) factor of $(1 - \phi L)$, with $|\phi| < 1$, and then allows $\phi \to 1$ in this z-transform. Whittle (p.95) does not claim that the resulting predictor is the l.l.s. predictor, but that this is true, with appropriate assumptions on initial conditions, is shown above.

The extraction of non-stationary signals with roots on the unit circle observed in stationary noise where an infinite sample is available is also considered by Hannan (1967) and Sobel (1967). These two papers are closely related in that they address the same problem, but they use different methods. Hannan finds that filter in the class of linear filters which pass a polynomial trend generated by the same m^{th} order difference equation as the signal, which minimizes the m.s.e. His result is a straightforward generalization of Whittle's. Sobel, on the other hand, proves by Hilbert space arguments that in this class of process the projection of S_t on $\{y_t = S_t + N_t\}$ tends to a limit as given by Hannan. Sobel's argument requires the noise to be stationary, and the variance of S_t to be bounded at time τ, say, by taking as 'initial' conditions $S_{\tau-m}, \ldots, S_{\tau-2}, S_{\tau-1}$ with finite variance. This device, which is used also by Bell (1984), makes the resulting estimates sensitive to the choice of τ, a fact which neither of these authors discuss. Sobel was aware of Kalman's work, but seems to have dismissed it on the (mistaken) grounds that only autoregressive processes could be handled - with the benefit of hindsight this seems to have been unfortunate, perhaps contributing to the slow diffusion of state-space methods into statistical time series analysis.

Cleveland and Tiao (1976) consider the extension of Whittle's result to situations in which signal and noise processes may have shared and repeated unit roots. That is, $\phi^s(L)$ may have a factor $(1-L^d)$ and $\phi^n(L)$ a factor of $(1-L)$, giving a non-detectable model as discussed in Section 5.4. They assert that the l.l.s. estimator (in their case, the conditional expectation) takes the same form as in the stationary case. However, their argument allows no proof of convergence of the filter to steady state, and no initial conditions which would bound the variance of S_t, N_t or y_t at some point in the sample are given.[6]

More recently, Bell (1984) extends Sobel's result to situations in which the signal and noise processes have unrestricted autoregressive operators of finite order. As noted above, his results depend on the arbitrary treatment of some part of the record as 'initial' conditions. In particular, his claim (pp.660-661) that the signal extraction error in the non-detectable case discussed by Pierce (see Section 5.4) has finite variance, contrary to Pierce's result, rests on the use of initial conditions in the middle of an infinite record, which seems to conflict with what is usually intended when writers in this field discuss such a record. Bell suggests that the Kalman filter, which he does not use, is a convenient device for solving the data processing problem in finite samples, to which his results do not apply.

That a variety of signal extraction problems may be usefully cast in the state-space framework is recognized by several authors. For example, Pagan (1975) displays the steady-state filter for a simple unobserved-components model with autoregressive components, and Engle (1978) and Kitagawa (1981) discuss both practical signal extraction and parameter estimation for various ARMA components models. In fact, the most common applications of Kalman filtering in econometrics are to be found in the estimation context, where the prediction error decomposition facilitates evaluation of the likelihood function for models with state structure (Harvey, 1984; Engle and Watson, 1985). Parameter estimation problems are not discussed in the present paper; rather the emphasis is on the use of state space methods in prediction theory. In that context, the techniques applied above have been seen to be powerful and direct, and their application to further problems, such as those of specification error and parameter uncertainty, will undoubtedly be equally rewarding.

APPENDIX

Alternative forms for the covariance generating function of y_t

In Sections 4.4 and 5.2 the c.g.f. of y_t, defined only in the stationary case, is expressed in two different forms. In this Appendix we show their equivalence by a direct algebraic argument.

First, from the general system (3.1)-(3.2) we obtain

$$\text{(A.1)} \quad g_{yy}(z) = R + H'(I-zF)^{-1}GQG'(I-z^{-1}F')^{-1}H + zH'(I-zF)^{-1}GS + z^{-1}S'G'(I-z^{-1}F')^{-1}H .$$

The second form is obtained from the steady-state relationship between y_t and current and past innovations as

$$\text{(A.2)} \quad g_{yy}(z) = \{1 + zH'(I-zF)^{-1}K\}\Sigma\{1 + z^{-1}K'(I-z^{-1}F')^{-1}H\} .$$

In this expression we have

$$\Sigma = H'PH + R, \quad K = (FPH + GS)\Sigma^{-1}, \quad P = FPF' + GQG' - K\Sigma K' .$$

Expanding (A.2) and substituting for K and Σ then gives

$$\begin{aligned} \text{(A.3)} \quad g_{yy}(z) = {} & H'PH + R + zH'(I - zF)^{-1}(FPH + GS) \\ & + z^{-1}(H'PF' + S'G')(I - z^{-1}F')^{-1}H \\ & + H'(I - zF)^{-1}(FPF' + GQG' - P)(I - z^{-1}F')^{-1}H . \end{aligned}$$

Subtracting (A.1) from (A.3) leaves the remainder

$$\begin{aligned} & H'PH + zH'(I - zF)^{-1}FPH + z^{-1}H'PF'(I - z^{-1}F')^{-1}H \\ & + H'(I - zF)^{-1}(FPF' - P)(I - z^{-1}F')^{-1}H , \end{aligned}$$

and on rearranging this in the form $H'(I - zF)^{-1}\Delta(I - z^{-1}F')^{-1}H$ it is readily seen that $\Delta=0$, as required. This algebra both generalizes and simplifies the discussion of Anderson and Moore (1979, §4.5), who treat the case $S=0$.

In the state-space representation of the ARMA model that is used in Section 4 we have $Q=S=R$, and using this in (A.1) gives equation (4.11) of the main text. In the representation used in Section 5 we have $S=0$, and again specializing (A.1) gives equation (5.4); the alternative form (A.2) is obtained as (5.6) of the main text.

REFERENCES

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

BELL, W.R. (1984). Signal extraction for nonstationary time series. Annals of Statistics, 12, 646-664.

BURRIDGE, P. and HALL, A.R. (1986). Convergence of the Kalman filter gain for a class of non-detectable signal extraction problems. Unpublished paper, University of Birmingham.

BURRIDGE, P. and WALLIS, K.F. (1985). Calculating the variance of seasonally adjusted series. Journal of the American Statistical Association, 80, 541-552.

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.

——— (1971). On the discrete time matrix Riccati equation of optimal control: a correction. International Journal of Control, 14, 205-207.

CHAN, S.W., GOODWIN, G.C. and SIN, K.S. (1984). Convergence properties of the Riccati difference equation in optimal filtering of non-stabilizable systems. IEEE Transactions on Automatic Control, 29, 110-118.

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.

ENGLE, R.F. (1978). Estimating structural models of seasonality. In Seasonal Analysis of Economic Time Series (A. Zellner, ed.), pp.281-297. Washington, DC: Bureau of the Census.

ENGLE, R.F. and WATSON, M. (1985). Applications of Kalman filtering in econometrics. Presented at the Econometric Society World Congress, Cambridge, Mass., August 1985.

FLORENS, J.P. and MOUCHART, M. (1982). A note on non-causality. Econometrica, 50, 583-591.

GRANGER, C.W.J. and NEWBOLD, P. (1977). Forecasting Economic Time Series. New York: Academic Press.

HAGER, W.W. and HOROWITZ, L.L. (1976). Convergence and stability properties of the discrete Riccati operator equation and the associated optimal control and filtering problems. SIAM Journal of Control and Optimization, 14, 295-312.

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

HARVEY, A.C. (1981). Time Series Models. Oxford: Philip Allan.

——— (1984). Dynamic models, the prediction error decomposition and state space. In Econometrics and Quantitative Economics (D.F. Hendry and K.F. Wallis, eds.), pp.37-59. Oxford: Basil Blackwell.

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

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

KALMAN, R.E. (1960). A new approach to linear filtering and prediction problems. Transactions ASME, Journal of Basic Engineering, 85, 35-45.

KITAGAWA, G. (1981). A nonstationary time series model and its fitting by a recursive filter. Journal of Time Series Analysis, 2, 103-116.

MUTH, J.F. (1960). Optimal properties of exponentially weighted forecasts. Journal of the American Statistical Association, 55, 299-305.

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.

PRESTON, A.J. and PAGAN, A.R. (1982). The Theory of Economic Policy. Cambridge: Cambridge University Press.

SARGENT, T.J. (1979). Macroeconomic Theory. New York: Academic Press.

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

WHITTLE, P. (1963). Prediction and Regulation by Linear Least Square Methods. London: English Universities Press. Second edition, Minneapolis: University of Minnesota Press, 1983.

FOOTNOTES

1. The suggestion of Harvey (1981, pp.37, 129, 161) that forecasts using a non-invertible MA are unattractive because they are inefficient seems to rest on the mistaken assumption that σ_ε^2 and σ_η^2 are equal. Granger and Newbold (1977, pp.144-145), working with conditional expectations, emphasize the equivalence of forecasts from the two representations, and comment that "the invertible form is inevitably used in practice since the computation of coefficient estimates and of forecasts from the fitted model is considerably easier for invertible than for non-invertible models." The computation of forecasts through the recursions (2.10) is unaffected by this choice.

2. For the present model this is given (with a typographical error) as Exercise 3.3.12 of Whittle (1963, p.35).

3. In his presentation and discussion of Muth's result Sargent (1979, p.310) notes "... a technical difficulty that arises because our $\{Y_i\}$ process is (borderline) nonstationary. In particular, the variance of Y is not finite, making application of least squares projection theory a touchy matter."

4. For a derivation see Anderson and Moore (1979, ch.7) or, for the Gaussian case from a slightly different perspective, Jazwinski (1970, pp.215-218).

5. This result is sometimes incorrectly attributed to Caines and Mayne, whose original proof of their theorem was in error. The confusion arises because in a subsequent correction (Caines and Mayne, 1971) the condition that (F*,D) be stabilizable had to be strengthened to controllability, and this change has been overlooked by some authors.

6. Burridge and Hall (1986) show that the filter coefficients converge to a steady state in this case using an argument along the lines of that in Section 6.