# 13 The Kalman Filter

This chapter introduces some very useful tools named for the contributions of R. E. Kalman (1960, 1963). The idea is to express a dynamic system in a particular form called the *state-space representation*. The Kalman filter is an algorithm for sequentially updating a linear projection for the system. Among other benefits, this algorithm provides a way to calculate exact finite-sample forecasts and the exact likelihood function for Gaussian *ARMA* processes, to factor matrix autocovariance-generating functions or spectral densities, and to estimate vector autoregressions with coefficients that change over time.

Section 13.1 describes how a dynamic system can be written in a form that can be analyzed using the Kalman filter. The filter itself is derived in Section 13.2, and its use in forecasting is described in Section 13.3. Section 13.4 explains how to estimate the population parameters by maximum likelihood. Section 13.5 analyzes the properties of the Kalman filter as the sample size grows, and explains how the Kalman filter is related in the limit to the Wold representation and factoring an autocovariance-generating function. Section 13.6 develops a smoothing algorithm, which is a way to use all the information in the sample to form the best inference about the unobserved state of the process at any historical date. Section 13.7 describes standard errors for smoothed inferences and forecasts. The use of the Kalman filter for estimating systems with time-varying parameters is investigated in Section 13.8.

## 13.1. *The State-Space Representation of a Dynamic System*

### *Maintained Assumptions*

Let $\mathbf{y}_t$ denote an $(n \times 1)$ vector of variables observed at date $t$. A rich class of dynamic models for $\mathbf{y}_t$ can be described in terms of a possibly unobserved $(r \times 1)$ vector $\boldsymbol{\xi}_t$ known as the *state vector*. The *state-space representation* of the dynamics of $\mathbf{y}$ is given by the following system of equations:

$$\boldsymbol{\xi}_{t+1} = \mathbf{F}\boldsymbol{\xi}_t + \mathbf{v}_{t+1} \qquad [13.1.1]$$

$$\mathbf{y}_t = \mathbf{A}'\mathbf{x}_t + \mathbf{H}'\boldsymbol{\xi}_t + \mathbf{w}_t, \qquad [13.1.2]$$

where $\mathbf{F}$, $\mathbf{A}'$, and $\mathbf{H}'$ are matrices of parameters of dimension $(r \times r)$, $(n \times k)$, and $(n \times r)$, respectively, and $\mathbf{x}_t$ is a $(k \times 1)$ vector of exogenous or predetermined variables. Equation [13.1.1] is known as the *state equation*, and [13.1.2] is known



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

as the *observation equation*. The $(r \times 1)$ vector $\mathbf{v}_t$ and the $(n \times 1)$ vector $\mathbf{w}_t$ are vector white noise:

$$E(\mathbf{v}_t\mathbf{v}_\tau') = \begin{cases} \mathbf{Q} & \text{for } t = \tau \\ \mathbf{0} & \text{otherwise} \end{cases} \qquad [13.1.3]$$

$$E(\mathbf{w}_t\mathbf{w}_\tau') = \begin{cases} \mathbf{R} & \text{for } t = \tau \\ \mathbf{0} & \text{otherwise,} \end{cases} \qquad [13.1.4]$$

where $\mathbf{Q}$ and $\mathbf{R}$ are $(r \times r)$ and $(n \times n)$ matrices, respectively. The disturbances $\mathbf{v}_t$ and $\mathbf{w}_t$ are assumed to be uncorrelated at all lags:

$$E(\mathbf{v}_t\mathbf{w}_\tau') = \mathbf{0} \qquad \text{for all } t \text{ and } \tau. \qquad [13.1.5]$$

The statement that $\mathbf{x}_t$ is predetermined or exogenous means that $\mathbf{x}_t$ provides no information about $\boldsymbol{\xi}_{t+s}$ or $\mathbf{w}_{t+s}$ for $s = 0, 1, 2, \ldots$ beyond that contained in $\mathbf{y}_{t-1}$, $\mathbf{y}_{t-2}, \ldots, \mathbf{y}_1$. Thus, for example, $\mathbf{x}_t$ could include lagged values of $\mathbf{y}$ or variables that are uncorrelated with $\boldsymbol{\xi}_\tau$ and $\mathbf{w}_\tau$ for all $\tau$.

The system of [13.1.1] through [13.1.5] is typically used to describe a finite series of observations $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T\}$ for which assumptions about the initial value of the state vector $\boldsymbol{\xi}_1$ are needed. We assume that $\boldsymbol{\xi}_1$ is uncorrelated with any realizations of $\mathbf{v}_t$ or $\mathbf{w}_t$:

$$E(\mathbf{v}_t\boldsymbol{\xi}_1') = \mathbf{0} \qquad \text{for } t = 1, 2, \ldots, T \qquad [13.1.6]$$

$$E(\mathbf{w}_t\boldsymbol{\xi}_1') = \mathbf{0} \qquad \text{for } t = 1, 2, \ldots, T. \qquad [13.1.7]$$

The state equation [13.1.1] implies that $\boldsymbol{\xi}_t$ can be written as a linear function of $(\boldsymbol{\xi}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_t)$:

$$\boldsymbol{\xi}_t = \mathbf{v}_t + \mathbf{F}\mathbf{v}_{t-1} + \mathbf{F}^2\mathbf{v}_{t-2} + \cdots + \mathbf{F}^{t-2}\mathbf{v}_2 + \mathbf{F}^{t-1}\boldsymbol{\xi}_1 \qquad [13.1.8]$$
$$\text{for } t = 2, 3, \ldots, T.$$

Thus, [13.1.6] and [13.1.3] imply that $\mathbf{v}_t$ is uncorrelated with lagged values of $\boldsymbol{\xi}$:

$$E(\mathbf{v}_t\boldsymbol{\xi}_\tau') = \mathbf{0} \qquad \text{for } \tau = t-1, t-2, \ldots, 1. \qquad [13.1.9]$$

Similarly,

$$E(\mathbf{w}_t\boldsymbol{\xi}_\tau') = \mathbf{0} \qquad \text{for } \tau = 1, 2, \ldots, T \qquad [13.1.10]$$

$$\begin{aligned} E(\mathbf{w}_t\mathbf{y}_\tau') &= E[\mathbf{w}_t(\mathbf{A}'\mathbf{x}_\tau + \mathbf{H}'\boldsymbol{\xi}_\tau + \mathbf{w}_\tau)'] \\ &= \mathbf{0} \qquad \text{for } \tau = t-1, t-2, \ldots, 1 \end{aligned} \qquad [13.1.11]$$

$$E(\mathbf{v}_t\mathbf{y}_\tau') = \mathbf{0} \qquad \text{for } \tau = t-1, t-2, \ldots, 1. \qquad [13.1.12]$$

The system of [13.1.1] through [13.1.7] is quite flexible, though it is straightforward to generalize the results further to systems in which $\mathbf{v}_t$ is correlated with $\mathbf{w}_t$.[1] The various parameter matrices ($\mathbf{F}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{H}$, or $\mathbf{R}$) could be functions of time, as will be discussed in Section 13.8. The presentation will be clearest, however, if we focus on the basic form in [13.1.1] through [13.1.7].

[1]See, for example, Anderson and Moore (1979, pp. 105–8).



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

## *Examples of State-Space Representations*

Consider a univariate $AR(p)$ process,

$$y_{t+1} - \mu = \phi_1(y_t - \mu) + \phi_2(y_{t-1} - \mu) + \cdots + \phi_p(y_{t-p+1} - \mu) + \varepsilon_{t+1}, \qquad [13.1.13]$$

$$E(\varepsilon_t \varepsilon_\tau) = \begin{cases} \sigma^2 & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases}$$

This could be written in state-space form as follows:

***State Equation*** $(r = p)$:

$$\begin{bmatrix} y_{t+1} - \mu \\ y_t - \mu \\ \vdots \\ y_{t-p+2} - \mu \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} y_t - \mu \\ y_{t-1} - \mu \\ \vdots \\ y_{t-p+1} - \mu \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad [13.1.14]$$

***Observation Equation*** $(n = 1)$:

$$y_t = \mu + [1 \quad 0 \quad \cdots \quad 0] \begin{bmatrix} y_t - \mu \\ y_{t-1} - \mu \\ \vdots \\ y_{t-p+1} - \mu \end{bmatrix}. \qquad [13.1.15]$$

That is, we would specify

$$\boldsymbol{\xi}_t = \begin{bmatrix} y_t - \mu \\ y_{t-1} - \mu \\ \vdots \\ y_{t-p+1} - \mu \end{bmatrix} \qquad \mathbf{F} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

$$\mathbf{v}_{t+1} = \begin{bmatrix} \varepsilon_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \mathbf{Q} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$\mathbf{y}_t = y_t \qquad \mathbf{A}' = \mu \qquad \mathbf{x}_t = 1$$

$$\mathbf{H}' = [1 \quad 0 \quad \cdots \quad 0] \qquad \mathbf{w}_t = 0 \qquad \mathbf{R} = 0.$$

Note that the state equation here is simply the first-order vector difference equation introduced in equation [1.2.5]; $\mathbf{F}$ is the same matrix appearing in equation [1.2.3]. The observation equation here is a trivial identity. Thus, we have already seen that the state-space representation [13.1.14] and [13.1.15] is just another way of summarizing the $AR(p)$ process [13.1.13]. The reason for rewriting an $AR(p)$ process in such a form was to obtain a convenient summary of the system's dynamics, and this is the basic reason to be interested in the state-space representation of any system. The analysis of a vector autoregression using equation [10.1.11] employed a similar state-space representation.

As another example, consider a univariate $MA(1)$ process,

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}. \qquad [13.1.16]$$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

This could be written in state-space form as follows:

***State Equation*** ($r = 2$):

$$\begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1} \\ 0 \end{bmatrix} \qquad [13.1.17]$$

***Observation Equation*** ($n = 1$):

$$y_t = \mu + [1 \quad \theta] \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix}; \qquad [13.1.18]$$

that is,

$$\boldsymbol{\xi}_t = \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix} \qquad \mathbf{F} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \qquad \mathbf{v}_{t+1} = \begin{bmatrix} \varepsilon_{t+1} \\ 0 \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \qquad \mathbf{y}_t = y_t \qquad \mathbf{A}' = \mu \qquad \mathbf{x}_t = 1$$

$$\mathbf{H}' = [1 \quad \theta] \qquad \mathbf{w}_t = 0 \qquad \mathbf{R} = 0.$$

There are many ways to write a given system in state-space form. For example, the $MA(1)$ process [13.1.16] can also be represented in this way:

***State Equation*** ($r = 2$):

$$\begin{bmatrix} \varepsilon_{t+1} + \theta\varepsilon_t \\ \theta\varepsilon_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_t + \theta\varepsilon_{t-1} \\ \theta\varepsilon_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1} \\ \theta\varepsilon_{t+1} \end{bmatrix} \qquad [13.1.19]$$

***Observation Equation*** ($n = 1$):

$$y_t = \mu + [1 \quad 0] \begin{bmatrix} \varepsilon_t + \theta\varepsilon_{t-1} \\ \theta\varepsilon_t \end{bmatrix}. \qquad [13.1.20]$$

Note that the original $MA(1)$ representation of [13.1.16], the first state-space representation of [13.1.17] and [13.1.18], and the second state-space representation of [13.1.19] and [13.1.20] all characterize the same process. We will obtain the identical forecasts of the process or value for the likelihood function from any of the three representations and can feel free to work with whichever is most convenient.

More generally, a univariate $ARMA(p, q)$ process can be written in state-space form by defining $r \equiv \max\{p, q + 1\}$:

$$\begin{aligned} y_t - \mu = {} & \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \cdots + \phi_r(y_{t-r} - \mu) \\ & + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_{r-1}\varepsilon_{t-r+1}, \end{aligned} \qquad [13.1.21]$$

where we interpret $\phi_j = 0$ for $j > p$ and $\theta_j = 0$ for $j > q$. Consider the following state-space representation;

***State Equation*** ($r = \max\{p, q + 1\}$):

$$\boldsymbol{\xi}_{t+1} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{r-1} & \phi_r \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \boldsymbol{\xi}_t + \begin{bmatrix} \varepsilon_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad [13.1.22]$$

***Observation Equation*** ($n = 1$):

$$y_t = \mu + [1 \quad \theta_1 \quad \theta_2 \quad \cdots \quad \theta_{r-1}]\boldsymbol{\xi}_t. \qquad [13.1.23]$$

To verify that [13.1.22] and [13.1.23] describe the same process as [13.1.21], let $\xi_{jt}$ denote the $j$th element of $\boldsymbol{\xi}_t$. Thus, the second row of the state equation asserts



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

that

$$\xi_{2,t+1} = \xi_{1t}.$$

The third row asserts that

$$\xi_{3,t+1} = \xi_{2t} = \xi_{1,t-1},$$

and in general the $j$th row implies that

$$\xi_{j,t+1} = L^{j-1}\xi_{1,t+1}.$$

Thus, the first row of the state equation implies that

$$\xi_{1,t+1} = (\phi_1 + \phi_2 L + \phi_3 L^2 + \cdots + \phi_r L^{r-1})\xi_{1t} + \varepsilon_{t+1}$$

or

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_r L^r)\xi_{1,t+1} = \varepsilon_{t+1}. \qquad [13.1.24]$$

The observation equation states that

$$y_t = \mu + (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_{r-1} L^{r-1})\xi_{1t}. \qquad [13.1.25]$$

Multiplying [13.1.25] by $(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_r L^r)$ and using [13.1.24] gives

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_r L^r)(y_t - \mu)$$
$$= (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_{r-1} L^{r-1})\varepsilon_t,$$

which indeed reproduces [13.1.21].

The state-space form can also be very convenient for modeling sums of stochastic processes or the consequences of measurement error. For example, Fama and Gibbons (1982) wanted to study the behavior of the ex ante real interest rate (the nominal interest rate $i_t$ minus the expected inflation rate $\pi_t^e$). This variable is unobserved, because the econometrician does not have data on the rate of inflation anticipated by the bond market. Thus, the state variable for this application was the scalar $\xi_t = i_t - \pi_t^e - \mu$, where $\mu$ denotes the average ex ante real interest rate. Fama and Gibbons assumed that the ex ante real rate follows an $AR(1)$ process:

$$\xi_{t+1} = \phi\xi_t + v_{t+1}. \qquad [13.1.26]$$

The econometrician has observations on the ex post real rate (the nominal interest rate $i_t$ minus actual inflation $\pi_t$), which can be written as

$$i_t - \pi_t = (i_t - \pi_t^e) + (\pi_t^e - \pi_t) = \mu + \xi_t + w_t, \qquad [13.1.27]$$

where $w_t \equiv (\pi_t^e - \pi_t)$ is the error that people make in forecasting inflation. If people form these forecasts optimally, then $w_t$ should be uncorrelated with its own lagged values or with the ex ante real interest rate. Thus, [13.1.26] and [13.1.27] are the state equation and observation equation for a state-space model with $r = n = 1$, $\mathbf{F} = \phi$, $y_t = i_t - \pi_t$, $\mathbf{A}'\mathbf{x}_t = \mu$, $\mathbf{H} = 1$, and $w_t = (\pi_t^e - \pi_t)$.

In another interesting application of the state-space framework, Stock and Watson (1991) postulated the existence of an unobserved scalar $C_t$ that represents the state of the business cycle. A set of $n$ different observed macroeconomic variables $(y_{1t}, y_{2t}, \ldots, y_{nt})$ are each assumed to be influenced by the business cycle and also to have an idiosyncratic component (denoted $\chi_{it}$) that is unrelated to movements in $y_{jt}$ for $i \neq j$. If the business cycle and each of the idiosyncratic components could be described by univariate $AR(1)$ processes, then the $[(n + 1) \times 1]$ state vector would be

$$\boldsymbol{\xi}_t = \begin{bmatrix} C_t \\ \chi_{1t} \\ \chi_{2t} \\ \vdots \\ \chi_{nt} \end{bmatrix} \qquad [13.1.28]$$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

with state equation

$$\begin{bmatrix} C_{t+1} \\ \chi_{1,t+1} \\ \chi_{2,t+1} \\ \vdots \\ \chi_{n,t+1} \end{bmatrix} = \begin{bmatrix} \phi_C & 0 & 0 & \cdots & 0 \\ 0 & \phi_1 & 0 & \cdots & 0 \\ 0 & 0 & \phi_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \phi_n \end{bmatrix} \begin{bmatrix} C_t \\ \chi_{1t} \\ \chi_{2t} \\ \vdots \\ \chi_{nt} \end{bmatrix} + \begin{bmatrix} v_{C,t+1} \\ v_{1,t+1} \\ v_{2,t+1} \\ \vdots \\ v_{n,t+1} \end{bmatrix} \qquad [13.1.29]$$

and observation equation

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} + \begin{bmatrix} \gamma_1 & 1 & 0 & \cdots & 0 \\ \gamma_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \gamma_n & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} C_t \\ \chi_{1t} \\ \chi_{2t} \\ \vdots \\ \chi_{nt} \end{bmatrix}. \qquad [13.1.30]$$

Thus, $\gamma_i$ is a parameter that describes the sensitivity of the $i$th series to the business cycle. To allow for $p$th-order dynamics, Stock and Watson replaced $C_t$ and $\chi_{it}$ in [13.1.28] with the $(p \times 1)$ vectors $(C_t, C_{t-1}, \ldots, C_{t-p+1})'$ and $(\chi_{it}, \chi_{i,t-1}, \ldots, \chi_{i,t-p+1})'$ so that $\boldsymbol{\xi}_t$ is an $[(n+1)p \times 1]$ vector. The scalars $\phi_i$ in [13.1.29] are then replaced by $(p \times p)$ matrices $\mathbf{F}_i$ with the structure of the matrix $\mathbf{F}$ in [13.1.14], and $[n \times (p-1)]$ blocks of zeros are added between the columns of $\mathbf{H}'$ in the observation equation [13.1.30].

## 13.2. *Derivation of the Kalman Filter*

### *Overview of the Kalman Filter*

Consider the general state-space system [13.1.1] through [13.1.7], whose key equations are reproduced here for convenience:

$$\underset{(r\times 1)}{\boldsymbol{\xi}_{t+1}} = \underset{(r\times r)(r\times 1)}{\mathbf{F}\cdot\boldsymbol{\xi}_t} + \underset{(r\times 1)}{\mathbf{v}_{t+1}} \qquad [13.2.1]$$

$$\underset{(n\times 1)}{\mathbf{y}_t} = \underset{(n\times k)(k\times 1)}{\mathbf{A}'\cdot\mathbf{x}_t} + \underset{(n\times r)(r\times 1)}{\mathbf{H}'\cdot\boldsymbol{\xi}_t} + \underset{(n\times 1)}{\mathbf{w}_t} \qquad [13.2.2]$$

$$E(\mathbf{v}_t\mathbf{v}_\tau') = \begin{cases} \underset{(r\times r)}{\mathbf{Q}} & \text{for } t = \tau \\ \mathbf{0} & \text{otherwise} \end{cases} \qquad [13.2.3]$$

$$E(\mathbf{w}_t\mathbf{w}_\tau') = \begin{cases} \underset{(n\times n)}{\mathbf{R}} & \text{for } t = \tau \\ \mathbf{0} & \text{otherwise.} \end{cases} \qquad [13.2.4]$$

The analyst is presumed to have observed $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$. One of the ultimate objectives may be to estimate the values of any unknown parameters in the system on the basis of these observations. For now, however, we will assume that the particular numerical values of $\mathbf{F}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{R}$ are known with certainty; Section 13.4 will give details on how these parameters can be estimated from the data.

There are many uses of the Kalman filter. It is motivated here as an algorithm for calculating linear least squares forecasts of the state vector on the basis of data observed through date $t$,

$$\hat{\boldsymbol{\xi}}_{t+1|t} \equiv \hat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t),$$

where

$$\mathcal{Y}_t \equiv (\mathbf{y}_t', \mathbf{y}_{t-1}', \ldots, \mathbf{y}_1', \mathbf{x}_t', \mathbf{x}_{t-1}', \ldots, \mathbf{x}_1')' \qquad [13.2.5]$$

and $\hat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t)$ denotes the linear projection of $\boldsymbol{\xi}_{t+1}$ on $\mathcal{Y}_t$ and a constant. The Kalman filter calculates these forecasts recursively, generating $\hat{\boldsymbol{\xi}}_{1|0}, \hat{\boldsymbol{\xi}}_{2|1}, \ldots,$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

$\hat{\boldsymbol{\xi}}_{T|T-1}$ in succession. Associated with each of these forecasts is a mean squared error (*MSE*) matrix, represented by the following $(r \times r)$ matrix:

$$\mathbf{P}_{t+1|t} \equiv E[(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})'].$$ [13.2.6]

## *Starting the Recursion*

The recursion begins with $\hat{\boldsymbol{\xi}}_{1|0}$, which denotes a forecast of $\boldsymbol{\xi}_1$ based on no observations of $\mathbf{y}$ or $\mathbf{x}$. This is just the unconditional mean of $\boldsymbol{\xi}_1$,

$$\hat{\boldsymbol{\xi}}_{1|0} = E(\boldsymbol{\xi}_1),$$

with associated *MSE*

$$\mathbf{P}_{1|0} = E\{[\boldsymbol{\xi}_1 - E(\boldsymbol{\xi}_1)][\boldsymbol{\xi}_1 - E(\boldsymbol{\xi}_1)]'\}.$$

For example, for the state-space representation of the $MA(1)$ system given in [13.1.17] and [13.1.18], the state vector was

$$\boldsymbol{\xi}_t = \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix},$$

for which

$$\hat{\boldsymbol{\xi}}_{1|0} = E\begin{bmatrix} \varepsilon_1 \\ \varepsilon_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$ [13.2.7]

$$\mathbf{P}_{1|0} = E\left(\begin{bmatrix} \varepsilon_1 \\ \varepsilon_0 \end{bmatrix}[\varepsilon_1 \quad \varepsilon_0]\right) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix},$$ [13.2.8]

where $\sigma^2 = E(\varepsilon_t^2)$.

More generally, if eigenvalues of $\mathbf{F}$ are all inside the unit circle, then the process for $\boldsymbol{\xi}_t$ in [13.2.1] is covariance-stationary. The unconditional mean of $\boldsymbol{\xi}_t$ can be found by taking expectations of both sides of [13.2.1], producing

$$E(\boldsymbol{\xi}_{t+1}) = \mathbf{F} \cdot E(\boldsymbol{\xi}_t),$$

or, since $\boldsymbol{\xi}_t$ is covariance-stationary,

$$(\mathbf{I}_r - \mathbf{F}) \cdot E(\boldsymbol{\xi}_t) = \mathbf{0}.$$

Since unity is not an eigenvalue of $\mathbf{F}$, the matrix $(\mathbf{I}_r - \mathbf{F})$ is nonsingular, and this equation has the unique solution $E(\boldsymbol{\xi}_t) = \mathbf{0}$. The unconditional variance of $\boldsymbol{\xi}$ can similarly be found by postmultiplying [13.2.1] by its transpose and taking expectations:

$$E(\boldsymbol{\xi}_{t+1}\boldsymbol{\xi}_{t+1}') = E[(\mathbf{F}\boldsymbol{\xi}_t + \mathbf{v}_{t+1})(\boldsymbol{\xi}_t'\mathbf{F}' + \mathbf{v}_{t+1}')] = \mathbf{F} \cdot E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') \cdot \mathbf{F}' + E(\mathbf{v}_{t+1}\mathbf{v}_{t+1}').$$

Cross-product terms have disappeared in light of [13.1.9]. Letting $\boldsymbol{\Sigma}$ denote the variance-covariance matrix of $\boldsymbol{\xi}$, this equation implies

$$\boldsymbol{\Sigma} = \mathbf{F}\boldsymbol{\Sigma}\mathbf{F}' + \mathbf{Q},$$

whose solution was seen in [10.2.18] to be given by

$$\text{vec}(\boldsymbol{\Sigma}) = [\mathbf{I}_{r^2} - (\mathbf{F} \otimes \mathbf{F})]^{-1} \cdot \text{vec}(\mathbf{Q}).$$

Thus, in general, provided that the eigenvalues of $\mathbf{F}$ are inside the unit circle, the Kalman filter iterations can be started with $\hat{\boldsymbol{\xi}}_{1|0} = \mathbf{0}$ and $\mathbf{P}_{1|0}$ the $(r \times r)$ matrix whose elements expressed as a column vector are given by

$$\text{vec}(\mathbf{P}_{1|0}) = [\mathbf{I}_{r^2} - (\mathbf{F} \otimes \mathbf{F})]^{-1} \cdot \text{vec}(\mathbf{Q}).$$

If instead some eigenvalues of $\mathbf{F}$ are on or outside the unit circle, or if the initial state $\boldsymbol{\xi}_1$ is not regarded as an arbitrary draw from the process implied by [13.2.1], then $\hat{\boldsymbol{\xi}}_{1|0}$ can be replaced with the analyst's best guess as to the initial value of $\boldsymbol{\xi}_1$, where $\mathbf{P}_{1|0}$ is a positive definite matrix summarizing the confidence in



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

this guess. Larger values for the diagonal elements of $\mathbf{P}_{1|0}$ register greater uncertainty about the true value of $\boldsymbol{\xi}_1$.

### *Forecasting* $\mathbf{y}_t$

Given starting values $\hat{\boldsymbol{\xi}}_{1|0}$ and $\mathbf{P}_{1|0}$, the next step is to calculate analogous magnitudes for the following date, $\hat{\boldsymbol{\xi}}_{2|1}$ and $\mathbf{P}_{2|1}$. The calculations for $t = 2, 3, \ldots, T$ all have the same basic form, so we will describe them in general terms for step $t$; given $\hat{\boldsymbol{\xi}}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$, the goal is to calculate $\hat{\boldsymbol{\xi}}_{t+1|t}$ and $\mathbf{P}_{t+1|t}$.

First note that since we have assumed that $\mathbf{x}_t$ contains no information about $\boldsymbol{\xi}_t$ beyond that contained in $\mathcal{Y}_{t-1}$,

$$\hat{E}(\boldsymbol{\xi}_t|\mathbf{x}_t, \mathcal{Y}_{t-1}) = \hat{E}(\boldsymbol{\xi}_t|\mathcal{Y}_{t-1}) = \hat{\boldsymbol{\xi}}_{t|t-1}.$$

Next consider forecasting the value of $\mathbf{y}_t$:

$$\hat{\mathbf{y}}_{t|t-1} \equiv \hat{E}(\mathbf{y}_t|\mathbf{x}_t, \mathcal{Y}_{t-1}).$$

Notice from [13.2.2] that

$$\hat{E}(\mathbf{y}_t|\mathbf{x}_t, \boldsymbol{\xi}_t) = \mathbf{A}'\mathbf{x}_t + \mathbf{H}'\boldsymbol{\xi}_t,$$

and so, from the law of iterated projections,

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{A}'\mathbf{x}_t + \mathbf{H}' \cdot \hat{E}(\boldsymbol{\xi}_t|\mathbf{x}_t, \mathcal{Y}_{t-1}) = \mathbf{A}'\mathbf{x}_t + \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1}. \qquad [13.2.9]$$

From [13.2.2], the error of this forecast is

$$\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1} = \mathbf{A}'\mathbf{x}_t + \mathbf{H}'\boldsymbol{\xi}_t + \mathbf{w}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1} = \mathbf{H}'(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1}) + \mathbf{w}_t$$

with *MSE*

$$\begin{aligned} &E[(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})'] \\ &\qquad = E[\mathbf{H}'(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})'\mathbf{H}] + E[\mathbf{w}_t\mathbf{w}_t']. \end{aligned} \qquad [13.2.10]$$

Cross-product terms have disappeared, since

$$E[\mathbf{w}_t(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})'] = \mathbf{0}. \qquad [13.2.11]$$

To justify [13.2.11], recall from [13.1.10] that $\mathbf{w}_t$ is uncorrelated with $\boldsymbol{\xi}_t$. Furthermore, since $\hat{\boldsymbol{\xi}}_{t|t-1}$ is a linear function of $\mathcal{Y}_{t-1}$, by [13.1.11] it too must be uncorrelated with $\mathbf{w}_t$.

Using [13.2.4] and [13.2.6], equation [13.2.10] can be written

$$E[(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})'] = \mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R}. \qquad [13.2.12]$$

### *Updating the Inference About* $\boldsymbol{\xi}_t$

Next the inference about the current value of $\boldsymbol{\xi}_t$ is updated on the basis of the observation of $\mathbf{y}_t$ to produce

$$\hat{\boldsymbol{\xi}}_{t|t} = \hat{E}(\boldsymbol{\xi}_t|\mathbf{y}_t, \mathbf{x}_t, \mathcal{Y}_{t-1}) = \hat{E}(\boldsymbol{\xi}_t|\mathcal{Y}_t).$$

This can be evaluated using the formula for updating a linear projection, equation [4.5.30]:[2]

$$\begin{aligned} \hat{\boldsymbol{\xi}}_{t|t} = \hat{\boldsymbol{\xi}}_{t|t-1} &+ \{E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})']\} \\ &\times \{E[(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})']\}^{-1} \times (\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}). \end{aligned} \qquad [13.2.13]$$

[2]Here $\boldsymbol{\xi}_t$ corresponds to $\mathbf{Y}_3$, $\mathbf{y}_t$ corresponds to $\mathbf{Y}_2$, and $(\mathbf{x}_t', \mathcal{Y}_{t-1}')'$ corresponds to $\mathbf{Y}_1$ in equation [4.5.30].



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

But

$$\begin{aligned} &E\{(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})'\} \\ &\quad = E\{[\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1}][\mathbf{H}'(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1}) + \mathbf{w}_t]'\} \\ &\quad = E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})'\mathbf{H}] \\ &\quad = \mathbf{P}_{t|t-1}\mathbf{H} \end{aligned} \tag{13.2.14}$$

by virtue of [13.2.11] and [13.2.6]. Substituting [13.2.14], [13.2.12], and [13.2.9] into [13.2.13] gives

$$\hat{\boldsymbol{\xi}}_{t|t} = \hat{\boldsymbol{\xi}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})^{-1}(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1}). \tag{13.2.15}$$

The *MSE* associated with this updated projection, which is denoted $\mathbf{P}_{t|t}$, can be found from [4.5.31]:

$$\begin{aligned} \mathbf{P}_{t|t} &\equiv E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t})(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t})'] \\ &= E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})'] \\ &\quad - \{E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})']\} \\ &\quad \times \{E[(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})']\}^{-1} \\ &\quad \times \{E[(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})']\} \\ &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})^{-1}\mathbf{H}'\mathbf{P}_{t|t-1}. \end{aligned} \tag{13.2.16}$$

## *Producing a Forecast of* $\boldsymbol{\xi}_{t+1}$

Next, the state equation [13.2.1] is used to forecast $\boldsymbol{\xi}_{t+1}$:

$$\begin{aligned} \hat{\boldsymbol{\xi}}_{t+1|t} &= \hat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t) \\ &= \mathbf{F}\cdot\hat{E}(\boldsymbol{\xi}_t|\mathcal{Y}_t) + \hat{E}(\mathbf{v}_{t+1}|\mathcal{Y}_t) \\ &= \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t} + \mathbf{0}. \end{aligned} \tag{13.2.17}$$

Substituting [13.2.15] into [13.2.17],

$$\begin{aligned} \hat{\boldsymbol{\xi}}_{t+1|t} &= \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t-1} \\ &\quad + \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})^{-1}(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1}). \end{aligned} \tag{13.2.18}$$

The coefficient matrix in [13.2.18] is known as the *gain matrix* and is denoted $\mathbf{K}_t$:

$$\mathbf{K}_t \equiv \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})^{-1}, \tag{13.2.19}$$

allowing [13.2.18] to be written

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1}). \tag{13.2.20}$$

The *MSE* of this forecast can be found from [13.2.17] and the state equation [13.2.1]:

$$\begin{aligned} \mathbf{P}_{t+1|t} &= E[(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})'] \\ &= E[(\mathbf{F}\boldsymbol{\xi}_t + \mathbf{v}_{t+1} - \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t})(\mathbf{F}\boldsymbol{\xi}_t + \mathbf{v}_{t+1} - \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t})'] \\ &= \mathbf{F}\cdot E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t})(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t})']\cdot\mathbf{F}' + E[\mathbf{v}_{t+1}\mathbf{v}_{t+1}'] \\ &= \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}' + \mathbf{Q}, \end{aligned} \tag{13.2.21}$$

with cross-product terms again clearly zero. Substituting [13.2.16] into [13.2.21] produces

$$\mathbf{P}_{t+1|t} = \mathbf{F}[\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})^{-1}\mathbf{H}'\mathbf{P}_{t|t-1}]\mathbf{F}' + \mathbf{Q}. \tag{13.2.22}$$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

### *Summary and Remarks*

To summarize, the Kalman filter is started with the unconditional mean and variance of $\boldsymbol{\xi}_1$:

$$\hat{\boldsymbol{\xi}}_{1|0} = E(\boldsymbol{\xi}_1)$$
$$\mathbf{P}_{1|0} = E\{[\boldsymbol{\xi}_1 - E(\boldsymbol{\xi}_1)][\boldsymbol{\xi}_1 - E(\boldsymbol{\xi}_1)]'\}.$$

Typically, these are given by $\hat{\boldsymbol{\xi}}_{1|0} = \mathbf{0}$ and $\text{vec}(\mathbf{P}_{1|0}) = [\mathbf{I}_{r^2} - (\mathbf{F} \otimes \mathbf{F})]^{-1} \cdot \text{vec}(\mathbf{Q})$. We then iterate on

$$\begin{aligned}\hat{\boldsymbol{\xi}}_{t+1|t} &= \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t-1} \\ &\quad + \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})^{-1}(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1})\end{aligned} \qquad [13.2.23]$$

and [13.2.22] for $t = 1, 2, \ldots, T$. The value $\hat{\boldsymbol{\xi}}_{t+1|t}$ denotes the best forecast of $\boldsymbol{\xi}_{t+1}$ based on a constant and a linear function of $(\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1)$. The matrix $\mathbf{P}_{t+1|t}$ gives the *MSE* of this forecast. The forecast of $\mathbf{y}_{t+1}$ is given by

$$\hat{\mathbf{y}}_{t+1|t} \equiv \hat{E}(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \mathcal{Y}_t) = \mathbf{A}'\mathbf{x}_{t+1} + \mathbf{H}'\hat{\boldsymbol{\xi}}_{t+1|t} \qquad [13.2.24]$$

with associated *MSE*

$$E[(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t})(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t})'] = \mathbf{H}'\mathbf{P}_{t+1|t}\mathbf{H} + \mathbf{R}. \qquad [13.2.25]$$

It is worth noting that the recursion in [13.2.22] could be calculated without ever evaluating [13.2.23]. The values for $\mathbf{P}_{t|t-1}$ in [13.2.22] and $\mathbf{K}_t$ in [13.2.19] are not functions of the data, but instead are determined entirely by the population parameters of the process.

An alternative way of writing the recursion for $\mathbf{P}_{t+1|t}$ is sometimes useful. Subtracting the Kalman updating equation [13.2.20] from the state equation [13.2.1] produces

$$\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F}(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1}) - \mathbf{K}_t(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1}) + \mathbf{v}_{t+1}. \qquad [13.2.26]$$

Further substituting the observation equation [13.2.2] into [13.2.26] results in

$$\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t} = (\mathbf{F} - \mathbf{K}_t\mathbf{H}')(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1}) - \mathbf{K}_t\mathbf{w}_t + \mathbf{v}_{t+1}. \qquad [13.2.27]$$

Postmultiplying [13.2.27] by its transpose and taking expectations,

$$\begin{aligned}&E[(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})'] \\ &\quad = (\mathbf{F} - \mathbf{K}_t\mathbf{H}')E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})'](\mathbf{F}' - \mathbf{H}\mathbf{K}_t') + \mathbf{K}_t\mathbf{R}\mathbf{K}_t' + \mathbf{Q};\end{aligned}$$

or, recalling the definition of $\mathbf{P}_{t+1|t}$ in equation [13.2.6],

$$\mathbf{P}_{t+1|t} = (\mathbf{F} - \mathbf{K}_t\mathbf{H}')\mathbf{P}_{t|t-1}(\mathbf{F}' - \mathbf{H}\mathbf{K}_t') + \mathbf{K}_t\mathbf{R}\mathbf{K}_t' + \mathbf{Q}. \qquad [13.2.28]$$

Equation [13.2.28] along with the definition of $\mathbf{K}_t$ in [13.2.19] will produce the same sequence generated by equation [13.2.22].

## 13.3. *Forecasts Based on the State-Space Representation*

The Kalman filter computations in [13.2.22] through [13.2.25] are normally calculated by computer, using the known numerical values of $\mathbf{F}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{R}$ along with the actual data. To help make the ideas more concrete, however, we now explore analytically the outcome of these calculations for a simple example.



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

## *Example—Using the Kalman Filter to Find Exact Finite-Sample Forecasts for an* MA(*1*) *Process*

Consider again a state-space representation for the $MA(1)$ process:

***State Equation*** ($r = 2$):

$$\begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1} \\ 0 \end{bmatrix} \qquad [13.3.1]$$

***Observation Equation*** ($n = 1$):

$$y_t = \mu + [1 \quad \theta] \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix} \qquad [13.3.2]$$

$$\boldsymbol{\xi}_t = \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix} \qquad [13.3.3]$$

$$\mathbf{F} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \qquad [13.3.4]$$

$$\mathbf{v}_{t+1} = \begin{bmatrix} \varepsilon_{t+1} \\ 0 \end{bmatrix} \qquad [13.3.5]$$

$$\mathbf{Q} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \qquad [13.3.6]$$

$$\mathbf{y}_t = y_t \qquad [13.3.7]$$

$$\mathbf{A}' = \mu \qquad [13.3.8]$$

$$\mathbf{x}_t = 1 \qquad [13.3.9]$$

$$\mathbf{H}' = [1 \quad \theta] \qquad [13.3.10]$$

$$\mathbf{w}_t = 0 \qquad [13.3.11]$$

$$\mathbf{R} = 0. \qquad [13.3.12]$$

The starting values for the filter were described in [13.2.7] and [13.2.8]:

$$\hat{\boldsymbol{\xi}}_{1|0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mathbf{P}_{1|0} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}.$$

Thus, from [13.2.24], the period 1 forecast is

$$\hat{y}_{1|0} = \mu + \mathbf{H}'\hat{\boldsymbol{\xi}}_{1|0} = \mu,$$

with $MSE$ given by [13.2.25]:

$$E(y_1 - \hat{y}_{1|0})^2 = \mathbf{H}'\mathbf{P}_{1|0}\mathbf{H} + \mathbf{R} = [1 \quad \theta] \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} 1 \\ \theta \end{bmatrix} + 0 = \sigma^2(1 + \theta^2).$$

These, of course, are just the unconditional mean and variance of $y$.

To see the structure of the recursion for $t = 2, 3, \ldots, T$, consider the basic form of the updating equation [13.2.23]. Notice that since the first row of $\mathbf{F}$ consists entirely of zeros, the first element of the vector $\hat{\boldsymbol{\xi}}_{t+1|t}$ will always equal zero, for all $t$. We see why if we recall the meaning of the state vector in [13.3.3]:

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \begin{bmatrix} \hat{\varepsilon}_{t+1|t} \\ \hat{\varepsilon}_{t|t} \end{bmatrix}. \qquad [13.3.13]$$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

Naturally, the forecast of the future white noise, $\hat{\varepsilon}_{t+1|t}$, is always zero. The forecast of $y_{t+1}$ is given by [13.2.24]:

$$\hat{y}_{t+1|t} = \mu + [1 \quad \theta]\begin{bmatrix}\hat{\varepsilon}_{t+1|t}\\ \hat{\varepsilon}_{t|t}\end{bmatrix} = \mu + \theta\hat{\varepsilon}_{t|t}. \qquad [13.3.14]$$

The Kalman filter updating equation for the *MSE*, equation [13.2.21], for this example becomes

$$\mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}' + \mathbf{Q} = \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}\mathbf{P}_{t|t}\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} + \begin{bmatrix}\sigma^2 & 0\\ 0 & 0\end{bmatrix}. \qquad [13.3.15]$$

Thus, $\mathbf{P}_{t+1|t}$ is a diagonal matrix of the form

$$\mathbf{P}_{t+1|t} = \begin{bmatrix}\sigma^2 & 0\\ 0 & p_{t+1}\end{bmatrix}, \qquad [13.3.16]$$

where the (2, 2) element of $\mathbf{P}_{t+1|t}$ (which we have denoted by $p_{t+1}$) is the same as the (1, 1) element of $\mathbf{P}_{t|t}$. Recalling [13.2.6] and [13.3.13], this term has the interpretation as the *MSE* of $\hat{\varepsilon}_{t|t}$:

$$p_{t+1} = E(\varepsilon_t - \hat{\varepsilon}_{t|t})^2. \qquad [13.3.17]$$

The (1, 1) element of $\mathbf{P}_{t+1|t}$ has the interpretation as the *MSE* of $\hat{\varepsilon}_{t+1|t}$. We have seen that this forecast is always zero, and its *MSE* in [13.3.16] is $\sigma^2$ for all $t$. The fact that $\mathbf{P}_{t+1|t}$ is a diagonal matrix means that the forecast error $(\varepsilon_{t+1} - \hat{\varepsilon}_{t+1|t})$ is uncorrelated with $(\varepsilon_t - \hat{\varepsilon}_{t|t})$.

The *MSE* of the forecast of $y_{t+1}$ is given by [13.2.25]:

$$\begin{aligned} E(y_{t+1} - \hat{y}_{t+1|t})^2 &= \mathbf{H}'\mathbf{P}_{t+1|t}\mathbf{H} + \mathbf{R}\\ &= [1 \quad \theta]\begin{bmatrix}\sigma^2 & 0\\ 0 & p_{t+1}\end{bmatrix}\begin{bmatrix}1\\ \theta\end{bmatrix} + 0 \qquad [13.3.18]\\ &= \sigma^2 + \theta^2 p_{t+1}. \end{aligned}$$

Again, the intuition can be seen from the nature of the forecast in [13.3.14]:

$$\begin{aligned} E(y_{t+1} - \hat{y}_{t+1|t})^2 &= E[(\mu + \varepsilon_{t+1} + \theta\varepsilon_t) - (\mu + \theta\hat{\varepsilon}_{t|t})]^2\\ &= E(\varepsilon_{t+1}^2) + \theta^2 E(\varepsilon_t - \hat{\varepsilon}_{t|t})^2, \end{aligned}$$

which, from [13.3.17], reproduces [13.3.18].

From [13.2.23], the series for $\hat{\varepsilon}_{t|t}$ is generated recursively from

$$\begin{aligned} \begin{bmatrix}0\\ \hat{\varepsilon}_{t|t}\end{bmatrix} &= \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}\begin{bmatrix}0\\ \hat{\varepsilon}_{t-1|t-1}\end{bmatrix}\\ &+ \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}\begin{bmatrix}\sigma^2 & 0\\ 0 & p_t\end{bmatrix}\begin{bmatrix}1\\ \theta\end{bmatrix}\{1/[\sigma^2 + \theta^2 p_t]\}\cdot\{y_t - \mu - \theta\hat{\varepsilon}_{t-1|t-1}\} \end{aligned}$$

or

$$\hat{\varepsilon}_{t|t} = \{\sigma^2/[\sigma^2 + \theta^2 p_t]\}\cdot\{y_t - \mu - \theta\hat{\varepsilon}_{t-1|t-1}\} \qquad [13.3.19]$$

starting from the initial value $\hat{\varepsilon}_{0|0} = 0$. Note that the value for $\hat{\varepsilon}_{t|t}$ differs from the approximation suggested in equations [4.2.36] and [4.3.2],

$$\hat{\varepsilon}_t = y_t - \mu - \theta\hat{\varepsilon}_{t-1} \qquad \hat{\varepsilon}_0 = 0,$$

in that [13.3.19] shrinks the inference $\hat{\varepsilon}_t$ toward zero to take account of the nonzero variance $p_t$ of $\hat{\varepsilon}_{t-1|t-1}$ around the true value $\varepsilon_{t-1}$.

The gain matrix $\mathbf{K}_t$ in equation [13.2.19] is given by

$$\mathbf{K}_t = \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}\begin{bmatrix}\sigma^2 & 0\\ 0 & p_t\end{bmatrix}\begin{bmatrix}1\\ \theta\end{bmatrix}\left(\frac{1}{\sigma^2 + \theta^2 p_t}\right) = \begin{bmatrix}0\\ \sigma^2/[\sigma^2 + \theta^2 p_t]\end{bmatrix}. \qquad [13.3.20]$$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

Finally, notice from [13.2.16] that

$$\mathbf{P}_{t|t} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & p_t \end{bmatrix} - \left(\frac{1}{\sigma^2 + \theta^2 p_t}\right)\begin{bmatrix} \sigma^2 & 0 \\ 0 & p_t \end{bmatrix}\begin{bmatrix} 1 \\ \theta \end{bmatrix}[1 \quad \theta]\begin{bmatrix} \sigma^2 & 0 \\ 0 & p_t \end{bmatrix}.$$

The (1, 1) element of $\mathbf{P}_{t|t}$ (which we saw equals $p_{t+1}$) is thus given by

$$p_{t+1} = \sigma^2 - \{1/[\sigma^2 + \theta^2 p_t]\}\cdot\sigma^4 = \frac{\sigma^2\theta^2 p_t}{\sigma^2 + \theta^2 p_t}. \qquad [13.3.21]$$

The recursion in [13.3.21] is started with $p_1 = \sigma^2$ and thus has the solution

$$p_{t+1} = \frac{\sigma^2\theta^{2t}}{1 + \theta^2 + \theta^4 + \cdots + \theta^{2t}}. \qquad [13.3.22]$$

It is interesting to note what happens to the filter as $t$ becomes large. First consider the case when $|\theta| \leq 1$. Then, from [13.3.22],

$$\lim_{t\to\infty} p_{t+1} = 0,$$

and so, from [13.3.17],

$$\hat{\varepsilon}_{t|t} \xrightarrow{p} \varepsilon_t.$$

Thus, given a sufficient number of observations on $y$, the Kalman filter inference $\hat{\varepsilon}_{t|t}$ converges to the true value $\varepsilon_t$, and the forecast [13.3.14] converges to that of the Wold representation for the process. The Kalman gain in [13.3.20] converges to $(0, 1)'$.

Alternatively, consider the case when $|\theta| > 1$. From [13.3.22], we have

$$p_{t+1} = \frac{\sigma^2\theta^{2t}(1 - \theta^2)}{1 - \theta^{2(t+1)}} = \frac{\sigma^2(1 - \theta^2)}{\theta^{-2t} - \theta^2}$$

and

$$\lim_{t\to\infty} p_{t+1} = \frac{\sigma^2(1 - \theta^2)}{-\theta^2} > 0.$$

No matter how many observations are obtained, it will not be possible to know with certainty the value of the nonfundamental innovation $\varepsilon_t$ associated with date $t$ on the basis of $(y_t, y_{t-1}, \ldots, y_1)$. The gain is given by

$$\frac{\sigma^2}{\sigma^2 + \theta^2 p_t} \to \frac{\sigma^2}{\sigma^2 - \sigma^2(1 - \theta^2)} = \frac{1}{\theta^2},$$

and the recursion [13.3.19] approaches

$$\hat{\varepsilon}_{t|t} = (1/\theta^2)\cdot(y_t - \mu - \theta\hat{\varepsilon}_{t-1|t-1})$$

or

$$\theta\hat{\varepsilon}_{t|t} = (1/\theta)\cdot(y_t - \mu - \theta\hat{\varepsilon}_{t-1|t-1}).$$

Recalling [13.3.14], we thus have

$$\hat{y}_{t+1|t} - \mu = (1/\theta)\cdot[(y_t - \mu) - (\hat{y}_{t|t-1} - \mu)]$$

or

$$\hat{y}_{t+1|t} - \mu = (1/\theta)\cdot(y_t - \mu) - (1/\theta)^2\cdot(y_{t-1} - \mu) + (1/\theta)^3\cdot(y_{t-2} - \mu) - \cdots,$$

which again is the $AR(\infty)$ forecast associated with the invertible $MA(1)$ representation. Indeed, the forecasts of the Kalman filter with $\theta$ replaced by $\theta^{-1}$ and $\sigma^2$ replaced by $\theta^2\sigma^2$ will be identical for any $t$; see Exercise 13.5.

## *Calculating* s*-Period-Ahead Forecasts with the Kalman Filter*

The forecast of $\mathbf{y}_t$ calculated in [13.2.24] is an exact finite-sample forecast of $\mathbf{y}_t$ on the basis of $\mathbf{x}_t$ and $\mathcal{Y}_{t-1} \equiv (\mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \ldots, \mathbf{y}'_1, \mathbf{x}'_{t-1}, \mathbf{x}'_{t-2}, \ldots, \mathbf{x}'_1)'$. If $\mathbf{x}_t$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

is deterministic, it is also easy to use the Kalman filter to calculate exact finite-sample $s$-period-ahead forecasts.

The state equation [13.2.1] can be solved by recursive substitution to yield

$$\boldsymbol{\xi}_{t+s} = \mathbf{F}^s\boldsymbol{\xi}_t + \mathbf{F}^{s-1}\mathbf{v}_{t+1} + \mathbf{F}^{s-2}\mathbf{v}_{t+2} + \cdots + \mathbf{F}^1\mathbf{v}_{t+s-1} + \mathbf{v}_{t+s} \quad \text{for } s = 1, 2, \ldots. \tag{13.3.23}$$

The projection of $\boldsymbol{\xi}_{t+s}$ on $\boldsymbol{\xi}_t$ and $\mathcal{Y}_t$ is given by

$$\hat{E}(\boldsymbol{\xi}_{t+s}|\boldsymbol{\xi}_t, \mathcal{Y}_t) = \mathbf{F}^s\boldsymbol{\xi}_t. \tag{13.3.24}$$

From the law of iterated projections,

$$\hat{\boldsymbol{\xi}}_{t+s|t} \equiv \hat{E}(\boldsymbol{\xi}_{t+s}|\mathcal{Y}_t) = \mathbf{F}^s\hat{\boldsymbol{\xi}}_{t|t}. \tag{13.3.25}$$

Thus, from [13.3.23] the $s$-period-ahead forecast error for the state vector is

$$\boldsymbol{\xi}_{t+s} - \hat{\boldsymbol{\xi}}_{t+s|t} = \mathbf{F}^s(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t}) + \mathbf{F}^{s-1}\mathbf{v}_{t+1} + \mathbf{F}^{s-2}\mathbf{v}_{t+2} + \cdots + \mathbf{F}^1\mathbf{v}_{t+s-1} + \mathbf{v}_{t+s} \tag{13.3.26}$$

with *MSE*

$$\mathbf{P}_{t+s|t} = \mathbf{F}^s\mathbf{P}_{t|t}(\mathbf{F}')^s + \mathbf{F}^{s-1}\mathbf{Q}(\mathbf{F}')^{s-1} + \mathbf{F}^{s-2}\mathbf{Q}(\mathbf{F}')^{s-2} + \cdots + \mathbf{F}\mathbf{Q}\mathbf{F}' + \mathbf{Q}. \tag{13.3.27}$$

To forecast the observed vector $\mathbf{y}_{t+s}$, recall from the observation equation that

$$\mathbf{y}_{t+s} = \mathbf{A}'\mathbf{x}_{t+s} + \mathbf{H}'\boldsymbol{\xi}_{t+s} + \mathbf{w}_{t+s}. \tag{13.3.28}$$

There are advantages if the state vector is defined in such a way that $\mathbf{x}_t$ is deterministic, so that the dynamics of any exogenous variables can be represented through $\boldsymbol{\xi}_t$. If $\mathbf{x}_t$ is deterministic, the $s$-period-ahead forecast of $\mathbf{y}$ is

$$\hat{\mathbf{y}}_{t+s|t} \equiv \hat{E}(\mathbf{y}_{t+s}|\mathcal{Y}_t) = \mathbf{A}'\mathbf{x}_{t+s} + \mathbf{H}'\hat{\boldsymbol{\xi}}_{t+s|t}. \tag{13.3.29}$$

The forecast error is

$$\begin{aligned}\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t} &= (\mathbf{A}'\mathbf{x}_{t+s} + \mathbf{H}'\boldsymbol{\xi}_{t+s} + \mathbf{w}_{t+s}) - (\mathbf{A}'\mathbf{x}_{t+s} + \mathbf{H}'\hat{\boldsymbol{\xi}}_{t+s|t}) \\ &= \mathbf{H}'(\boldsymbol{\xi}_{t+s} - \hat{\boldsymbol{\xi}}_{t+s|t}) + \mathbf{w}_{t+s}\end{aligned}$$

with *MSE*

$$E[(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t})(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t})'] = \mathbf{H}'\mathbf{P}_{t+s|t}\mathbf{H} + \mathbf{R}. \tag{13.3.30}$$

## 13.4. *Maximum Likelihood Estimation of Parameters*

### *Using the Kalman Filter to Evaluate the Likelihood Function*

The Kalman filter was motivated in Section 13.2 in terms of linear projections. The forecasts $\hat{\boldsymbol{\xi}}_{t|t-1}$ and $\hat{\mathbf{y}}_{t|t-1}$ are thus optimal within the set of forecasts that are linear in $(\mathbf{x}_t, \mathcal{Y}_{t-1})$, where $\mathcal{Y}_{t-1} \equiv (\mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \ldots, \mathbf{y}'_1, \mathbf{x}'_{t-1}, \mathbf{x}'_{t-2}, \ldots, \mathbf{x}'_1)'$. If the initial state $\boldsymbol{\xi}_1$ and the innovations $\{\mathbf{w}_t, \mathbf{v}_t\}_{t=1}^T$ are multivariate Gaussian, then we can make the stronger claim that the forecasts $\hat{\boldsymbol{\xi}}_{t|t-1}$ and $\hat{\mathbf{y}}_{t|t-1}$ calculated by the Kalman filter are optimal among any functions of $(\mathbf{x}_t, \mathcal{Y}_{t-1})$. Moreover, if $\boldsymbol{\xi}_1$ and $\{\mathbf{w}_t, \mathbf{v}_t\}_{t=1}^T$ are Gaussian, then the distribution of $\mathbf{y}_t$ conditional on $(\mathbf{x}_t, \mathcal{Y}_{t-1})$ is Gaussian with mean given by [13.2.24] and variance given by [13.2.25]:

$$\mathbf{y}_t|\mathbf{x}_t, \mathcal{Y}_{t-1} \sim N\big((\mathbf{A}'\mathbf{x}_t + \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1}), (\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})\big);$$

that is,

$$\begin{aligned}&f_{\mathbf{Y}_t|\mathbf{X}_t,\mathcal{Y}_{t-1}}(\mathbf{y}_t|\mathbf{x}_t, \mathcal{Y}_{t-1}) \\ &= (2\pi)^{-n/2}|\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R}|^{-1/2} \\ &\quad \times \exp\{-\tfrac{1}{2}(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1})'(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})^{-1} \\ &\quad \times (\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1})\} \qquad \text{for } t = 1, 2, \ldots, T.\end{aligned} \tag{13.4.1}$$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

From [13.4.1], it is a simple matter to construct the sample log likelihood,

$$\sum_{t=1}^{T} \log f_{\mathbf{y}_t|\mathbf{x}_t,\mathcal{Y}_{t-1}}(\mathbf{y}_t|\mathbf{x}_t, \mathcal{Y}_{t-1}). \qquad [13.4.2]$$

Expression [13.4.2] can then be maximized numerically with respect to the unknown parameters in the matrices $\mathbf{F}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{R}$; see Burmeister and Wall (1982) for an illustrative application.

As stressed by Harvey and Phillips (1979), this representation of the likelihood is particularly convenient for estimating regressions involving moving average terms. Moreover, [13.4.2] gives the exact log likelihood function, regardless of whether the moving average representation is invertible.

As an illustrative example, suppose we wanted to estimate a bivariate regression model whose equations were

$$y_{1t} = \mathbf{a}_1'\mathbf{x}_t + u_{1t}$$
$$y_{2t} = \mathbf{a}_2'\mathbf{x}_t + u_{2t},$$

where $\mathbf{x}_t$ is a $(k \times 1)$ vector of exogenous explanatory variables and $\mathbf{a}_1$ and $\mathbf{a}_2$ are $(k \times 1)$ vectors of coefficients; if the two regressions have different explanatory variables, the variables from both regressions are included in $\mathbf{x}_t$ with zeros appropriately imposed on $\mathbf{a}_1$ and $\mathbf{a}_2$. Suppose that the disturbance vector follows a bivariate $MA(1)$ process:

$$\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} + \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix},$$

with $(\varepsilon_{1t}, \varepsilon_{2t})' \sim$ i.i.d. $N(\mathbf{0}, \mathbf{\Omega})$. This model can be written in state-space form by defining

$$\xi_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix} \qquad \mathbf{F} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \qquad \mathbf{v}_{t+1} = \begin{bmatrix} \varepsilon_{1,t+1} \\ \varepsilon_{2,t+1} \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{A}' = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \end{bmatrix}$$

$$\mathbf{H}' = \begin{bmatrix} 1 & 0 & \theta_{11} & \theta_{12} \\ 0 & 1 & \theta_{21} & \theta_{22} \end{bmatrix} \qquad \mathbf{R} = \mathbf{0},$$

where $\sigma_{ij} = E(\varepsilon_{it}\varepsilon_{jt})$. The Kalman filter iteration is started from

$$\hat{\xi}_{1|0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{P}_{1|0} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 & 0 \\ 0 & 0 & \sigma_{11} & \sigma_{12} \\ 0 & 0 & \sigma_{21} & \sigma_{22} \end{bmatrix}.$$

Maximization of [13.4.2] is started by making an initial guess as to the numerical values of the unknown parameters. One obvious way to do this is to regress $y_{1t}$ on the elements of $\mathbf{x}_t$ that appear in the first equation to get an initial guess for $\mathbf{a}_1$. A similar $OLS$ regression for $y_2$ yields a guess for $\mathbf{a}_2$. Setting $\theta_{11} = \theta_{12} = \theta_{21} = \theta_{22} = 0$ initially, a first guess for $\mathbf{\Omega}$ could be the estimated variance-covariance matrix of the residuals from these two $OLS$ regressions. For these initial numerical values for the population parameters, we could construct $\mathbf{F}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{R}$ from the expressions just given and iterate on [13.2.22] through [13.2.25] for $t = 1, 2, \ldots, T - 1$. The sequences $\{\hat{\xi}_{t|t-1}\}_{t=1}^{T}$ and $\{\mathbf{P}_{t|t-1}\}_{t=1}^{T}$ resulting from these iter-



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

ations could then be used in [13.4.1] and [13.4.2] to calculate the value for the log likelihood function that results from these initial parameter values. The numerical optimization methods described in Section 5.7 can then be employed to make better guesses as to the value of the unknown parameters until [13.4.2] is maximized. As noted in Section 5.9, the numerical search will be better behaved if $\mathbf{\Omega}$ is parameterized in terms of its Cholesky factorization.

As a second example, consider a scalar Gaussian $ARMA(1, 1)$ process,

$$y_t - \mu = \phi(y_{t-1} - \mu) + \varepsilon_t + \theta\varepsilon_{t-1},$$

with $\varepsilon_t \sim$ i.i.d. $N(0, \sigma^2)$. This can be written in state-space form as in [13.1.22] and [13.1.23] with $r = 2$ and

$$\mathbf{F} = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix} \qquad \mathbf{v}_{t+1} = \begin{bmatrix} \varepsilon_{t+1} \\ 0 \end{bmatrix} \qquad \mathbf{Q} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{A}' = \mu \qquad \mathbf{x}_t = 1 \qquad \mathbf{H}' = [1 \quad \theta] \qquad \mathbf{R} = 0$$

$$\hat{\boldsymbol{\xi}}_{1|0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \mathbf{P}_{1|0} = \begin{bmatrix} \sigma^2/(1 - \phi^2) & \phi\sigma^2/(1 - \phi^2) \\ \phi\sigma^2/(1 - \phi^2) & \sigma^2/(1 - \phi^2) \end{bmatrix}.$$

This value for $\mathbf{P}_{1|0}$ was obtained by recognizing that the state equation [13.1.22] describes the behavior of $\boldsymbol{\xi}_t = (z_t, z_{t-1}, \ldots, z_{t-r+1})'$, where $z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_r z_{t-r} + \varepsilon_t$ follows an $AR(r)$ process. For this example, $r = 2$, so that $\mathbf{P}_{1|0}$ is the variance-covariance matrix of two consecutive draws from an $AR(2)$ process with parameters $\phi_1 = \phi$ and $\phi_2 = 0$. The expressions just given for $\mathbf{F}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{R}$ are then used in the Kalman filter iterations. Thus, expression [13.4.2] allows easy computation of the exact likelihood function for an $ARMA(p, q)$ process. This computation is valid regardless of whether the moving average parameters satisfy the invertibility condition. Similarly, expression [13.3.29] gives the exact finite-sample $s$-period-ahead forecast for the process and [13.3.30] its $MSE$, again regardless of whether the invertible representation is used.

Typically, numerical search procedures for maximizing [13.4.2] require the derivatives of the log likelihood. These can be calculated numerically or analytically. To characterize the analytical derivatives of [13.4.2], collect the unknown parameters to be estimated in a vector $\boldsymbol{\theta}$, and write $\mathbf{F}(\boldsymbol{\theta})$, $\mathbf{Q}(\boldsymbol{\theta})$, $\mathbf{A}(\boldsymbol{\theta})$, $\mathbf{H}(\boldsymbol{\theta})$, and $\mathbf{R}(\boldsymbol{\theta})$. Implicitly, then, $\hat{\boldsymbol{\xi}}_{t|t-1}(\boldsymbol{\theta})$ and $\mathbf{P}_{t|t-1}(\boldsymbol{\theta})$ will be functions of $\boldsymbol{\theta}$ as well, and the derivative of the log of [13.4.1] with respect to the $i$th element of $\boldsymbol{\theta}$ will involve $\partial\hat{\boldsymbol{\xi}}_{t|t-1}(\boldsymbol{\theta})/\partial\theta_i$ and $\partial\mathbf{P}_{t|t-1}(\boldsymbol{\theta})/\partial\theta_i$. These derivatives can also be generated recursively by differentiating the Kalman filter recursion, [13.2.22] and [13.2.23], with respect to $\theta_i$; see Caines (1988, pp. 585–86) for illustration.

For many state-space models, the $EM$ algorithm of Dempster, Laird, and Rubin (1977) offers a particularly convenient means for maximizing [13.4.2], as developed by Shumway and Stoffer (1982) and Watson and Engle (1983).

## *Identification*

Although the state-space representation gives a very convenient way to calculate the exact likelihood function, a word of caution should be given. In the absence of restrictions on $\mathbf{F}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{R}$, the parameters of the state-space representation are unidentified—more than one set of values for the parameters can give rise to the identical value of the likelihood function, and the data give us no guide for choosing among these. A trivial example is the following system:



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

***State Equation*** ($r = 2$):

$$\boldsymbol{\xi}_{t+1} = \begin{bmatrix} \varepsilon_{1,t+1} \\ \varepsilon_{2,t+1} \end{bmatrix} \qquad [13.4.3]$$

***Observation Equation*** ($n = 1$):

$$y_t = \varepsilon_{1t} + \varepsilon_{2t}. \qquad [13.4.4]$$

Here, $\mathbf{F} = \mathbf{0}$, $\mathbf{Q} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$, $\mathbf{A}' = 0$, $\mathbf{H}' = [1 \quad 1]$, and $\mathbf{R} = 0$. This model asserts that $y_t$ is white noise, with mean zero and variance given by $(\sigma_1^2 + \sigma_2^2)$. The reader is invited to confirm in Exercise 13.4 that the log of the likelihood function from [13.4.1] and [13.4.2] simplifies to

$$\begin{aligned} &\log f_{\mathbf{Y}_T,\mathbf{Y}_{T-1},\ldots,\mathbf{Y}_1}(\mathbf{y}_T, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_1) \\ &= -(T/2)\log(2\pi) - (T/2)\log(\sigma_1^2 + \sigma_2^2) - \sum_{t=1}^{T} y_t^2/[2(\sigma_1^2 + \sigma_2^2)]. \end{aligned} \qquad [13.4.5]$$

Clearly, any values for $\sigma_1^2$ and $\sigma_2^2$ that sum to a given constant will produce the identical value for the likelihood function.

The $MA(1)$ process explored in Section 13.3 provides a second example of an unidentified state-space representation. As the reader may verify in Exercise 13.5, the identical value for the log likelihood function [13.4.2] would result if $\theta$ is replaced by $\theta^{-1}$ and $\sigma^2$ by $\theta^2\sigma^2$.

These two examples illustrate two basic forms in which absence of identification can occur. Following Rothenberg (1971), a model is said to be *globally identified* at a particular parameter value $\boldsymbol{\theta}_0$ if for any value of $\boldsymbol{\theta}$ there exists a possible realization $\mathcal{Y}_T$ for which the value of the likelihood at $\boldsymbol{\theta}$ is different from the value of the likelihood at $\boldsymbol{\theta}_0$. A model is said to be *locally identified* at $\boldsymbol{\theta}_0$ if there exists a $\delta > 0$ such that for any value of $\boldsymbol{\theta}$ satisfying $(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'(\boldsymbol{\theta} - \boldsymbol{\theta}_0) < \delta$, there exists a possible realization of $\mathcal{Y}_T$ for which the value of the likelihood at $\boldsymbol{\theta}$ is different from the value of the likelihood at $\boldsymbol{\theta}_0$. Thus, global identification implies local identification. The first example, [13.4.3] and [13.4.4], is neither globally nor locally identified, while the $MA(1)$ example is locally identified but globally unidentified.

Local identification is much easier to test for than global identification. Rothenberg (1971) showed that a model is locally identified at $\boldsymbol{\theta}_0$ if and only if the information matrix is nonsingular in a neighborhood around $\boldsymbol{\theta}_0$. Thus, a common symptom of trying to estimate an unidentified model is difficulty with inverting the matrix of second derivatives of the log likelihood function. One approach to checking for local identification is to translate the state-space representation back into a vector *ARMA* model and check for satisfaction of the conditions in Hannan (1971); see Hamilton (1985) for an example of this approach. A second approach is to work directly with the state-space representation, as is done in Gevers and Wertz (1984) and Wall (1987). For an illustration of the second approach, see Burmeister, Wall, and Hamilton (1986).

## *Asymptotic Properties of Maximum Likelihood Estimates*

If certain regularity conditions are satisfied, then Caines (1988, Chapter 7) showed that the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_T$ based on a sample of size $T$ is consistent and asymptotically normal. These conditions include the following: (1) the model must be identified; (2) eigenvalues of $\mathbf{F}$ are all inside the unit circle; (3)



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

apart from a constant term, the variables $\mathbf{x}_t$ behave asymptotically like a full-rank linearly indeterministic covariance-stationary process; and (4) the true value of $\boldsymbol{\theta}$ does not fall on a boundary of the allowable parameter space. Pagan (1980, Theorem 4) and Ghosh (1989) examined special cases of state-space models for which

$$\sqrt{T}\mathcal{I}_{2D,T}^{1/2}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{L} N(\mathbf{0}, \mathbf{I}_a), \qquad [13.4.6]$$

where $a$ is the number of elements of $\boldsymbol{\theta}$ and $\mathcal{I}_{2D,T}$ is the $(a \times a)$ information matrix for a sample of size $T$ as calculated from second derivatives of the log likelihood function:

$$\mathcal{I}_{2D,T} = -\frac{1}{T}E\left(\sum_{t=1}^{T} \frac{\partial^2 \log f(\mathbf{y}_t|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right). \qquad [13.4.7]$$

A common practice is to assume that the limit of $\mathcal{I}_{2D,T}$ as $T \to \infty$ is the same as the plim of

$$\hat{\mathcal{I}}_{2D,T} = -\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \log f(\mathbf{y}_t|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_T}, \qquad [13.4.8]$$

which can be calculated analytically or numerically by differentiating [13.4.2]. Reported standard errors for $\hat{\boldsymbol{\theta}}_T$ are then square roots of diagonal elements of $(1/T)(\hat{\mathcal{I}}_{2D,T})^{-1}$.

### *Quasi-Maximum Likelihood Estimation*

Even if the disturbances $\mathbf{v}_t$ and $\mathbf{w}_t$ are non-Gaussian, the Kalman filter can still be used to calculate the linear projection of $\mathbf{y}_{t+s}$ on past observables. Moreover, we can form the function [13.4.2] and maximize it with respect to $\boldsymbol{\theta}$ even for non-Gaussian systems. This procedure will still yield consistent and asymptotically Normal estimates of the elements of $\mathbf{F}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{R}$, with the variance-covariance matrix constructed as described in equation [5.8.7]. Watson (1989, Theorem 2) presented conditions under which the quasi-maximum likelihood estimates satisfy

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{L} N(\mathbf{0}, [\mathcal{I}_{2D}\mathcal{I}_{OP}^{-1}\mathcal{I}_{2D}]^{-1}), \qquad [13.4.9]$$

where $\mathcal{I}_{2D}$ is the plim of [13.4.8] when evaluated at the true value $\boldsymbol{\theta}_0$ and $\mathcal{I}_{OP}$ is the outer-product estimate of the information matrix,

$$\mathcal{I}_{OP} = \text{plim } (1/T)\sum_{t=1}^{T} [\mathbf{h}(\boldsymbol{\theta}_0, \mathcal{Y}_t)][\mathbf{h}(\boldsymbol{\theta}_0, \mathcal{Y}_t)]',$$

where

$$\mathbf{h}(\boldsymbol{\theta}_0, \mathcal{Y}_t) \equiv \frac{\partial \log f(\mathbf{y}_t|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

## 13.5. *The Steady-State Kalman Filter*

### *Convergence Properties of the Kalman Filter*

Section 13.3 applied the Kalman filter to an $MA(1)$ process and found that when $|\theta| \leq 1$,

$$\lim_{t\to\infty} \mathbf{P}_{t+1|t} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\lim_{t\to\infty} \mathbf{K}_t = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

whereas when $|\theta| > 1$,

$$\lim_{t\to\infty} \mathbf{P}_{t+1|t} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2(\theta^2 - 1)/\theta^2 \end{bmatrix}$$

$$\lim_{t\to\infty} \mathbf{K}_t = \begin{bmatrix} 0 \\ 1/\theta^2 \end{bmatrix}.$$

It turns out to be a property of a broad class of state-space models that the sequences $\{\mathbf{P}_{t+1|t}\}_{t=1}^T$ and $\{\mathbf{K}_t\}_{t=1}^T$ converge to fixed matrices, as the following proposition shows.

***Proposition 13.1:*** *Let* $\mathbf{F}$ *be an* $(r \times r)$ *matrix whose eigenvalues are all inside the unit circle, let* $\mathbf{H}'$ *denote an arbitrary* $(n \times r)$ *matrix, and let* $\mathbf{Q}$ *and* $\mathbf{R}$ *be positive semidefinite symmetric* $(r \times r)$ *and* $(n \times n)$ *matrices, respectively. Let* $\{\mathbf{P}_{t+1|t}\}_{t=1}^T$ *be the sequence of MSE matrices calculated by the Kalman filter,*

$$\mathbf{P}_{t+1|t} = \mathbf{F}[\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})^{-1}\mathbf{H}'\mathbf{P}_{t|t-1}]\mathbf{F}' + \mathbf{Q}, \qquad [13.5.1]$$

*where iteration on* [13.5.1] *is initialized by letting* $\mathbf{P}_{1|0}$ *be the positive semidefinite* $(r \times r)$ *matrix satisfying*

$$\text{vec}(\mathbf{P}_{1|0}) = [\mathbf{I}_{r^2} - (\mathbf{F} \otimes \mathbf{F})]^{-1} \cdot \text{vec}(\mathbf{Q}). \qquad [13.5.2]$$

*Then* $\{\mathbf{P}_{t+1|t}\}_{t=1}^T$ *is a monotonically nonincreasing sequence and converges as* $T \to \infty$ *to a steady-state matrix* $\mathbf{P}$ *satisfying*

$$\mathbf{P} = \mathbf{F}[\mathbf{P} - \mathbf{PH}(\mathbf{H}'\mathbf{PH} + \mathbf{R})^{-1}\mathbf{H}'\mathbf{P}]\mathbf{F}' + \mathbf{Q}. \qquad [13.5.3]$$

*Moreover, the steady-state value for the Kalman gain matrix, defined by*

$$\mathbf{K} \equiv \mathbf{FPH}(\mathbf{H}'\mathbf{PH} + \mathbf{R})^{-1}, \qquad [13.5.4]$$

*has the property that the eigenvalues of* $(\mathbf{F} - \mathbf{KH}')$ *all lie on or inside the unit circle.*

The claim in Proposition 13.1 that $\mathbf{P}_{t+1|t} \leq \mathbf{P}_{t|t-1}$ means that for any real $(r \times 1)$ vector $\mathbf{h}$, the scalar inequality $\mathbf{h}'\mathbf{P}_{t+1|t}\mathbf{h} \leq \mathbf{h}'\mathbf{P}_{t|t-1}\mathbf{h}$ holds.

Proposition 13.1 assumes that the Kalman filter is started with $\mathbf{P}_{1|0}$ equal to the unconditional variance-covariance matrix of the state vector $\boldsymbol{\xi}_t$. Although the sequence $\{\mathbf{P}_{t+1|t}\}$ converges to a matrix $\mathbf{P}$, the solution to [13.5.3] need not be unique; a different starting value for $\mathbf{P}_{1|0}$ might produce a sequence that converges to a different matrix $\mathbf{P}$ satisfying [13.5.3]. Under the slightly stronger assumption that either $\mathbf{Q}$ or $\mathbf{R}$ is strictly positive definite, then iteration on [13.5.1] will converge to a unique solution to [13.5.3], where the starting value for the iteration $\mathbf{P}_{1|0}$ can be any positive semidefinite symmetric matrix.

***Proposition 13.2:*** *Let* $\mathbf{F}$ *be an* $(r \times r)$ *matrix whose eigenvalues are all inside the unit circle, let* $\mathbf{H}'$ *denote an arbitrary* $(n \times r)$ *matrix, and let* $\mathbf{Q}$ *and* $\mathbf{R}$ *be positive semidefinite symmetric* $(r \times r)$ *and* $(n \times n)$ *matrices, respectively, with either* $\mathbf{Q}$ *or* $\mathbf{R}$ *strictly positive definite. Then the sequence of Kalman MSE matrices* $\{\mathbf{P}_{t+1|t}\}_{t=1}^T$ *determined by* [*13.5.1*] *converges to a unique positive semidefinite steady-state matrix* $\mathbf{P}$ *satisfying* [*13.5.3*], *where the value of* $\mathbf{P}$ *is the same for any positive semidefinite symmetric starting value for* $\mathbf{P}_{1|0}$. *Moreover, the steady-state value for the Kalman gain matrix* $\mathbf{K}$ *in* [*13.5.4*] *has the property that the eigenvalues of* $(\mathbf{F} - \mathbf{KH}')$ *are all strictly inside the unit circle.*

We next discuss the relevance of the results in Propositions 13.1 and 13.2 concerning the eigenvalues of $(\mathbf{F} - \mathbf{KH}')$.



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

## *Using the Kalman Filter to Find the Wold Representation and Factor an Autocovariance-Generating Function*

Consider a system in which the explanatory variables ($\mathbf{x}_t$) consist solely of a constant term. Without loss of generality, we simplify the notation by assuming that $\mathbf{A}'\mathbf{x}_t \equiv \mathbf{0}$. For such systems, the Kalman filter forecast of the state vector can be written as in [13.2.20]:

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1}). \qquad [13.5.5]$$

The linear projection of $\mathbf{y}_{t+1}$ on the observed finite sample of its own lagged values is then calculated from

$$\hat{\mathbf{y}}_{t+1|t} = \hat{E}(\mathbf{y}_{t+1}|\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1) = \mathbf{H}'\hat{\boldsymbol{\xi}}_{t+1|t}, \qquad [13.5.6]$$

with *MSE* given by [13.2.25]:

$$E[(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t})(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t})'] = \mathbf{H}'\mathbf{P}_{t+1|t}\mathbf{H} + \mathbf{R}. \qquad [13.5.7]$$

Consider the result from applying the Kalman filter to a covariance-stationary process that started up at a time arbitrarily distant in the past. From Proposition 13.1, the difference equation [13.5.5] will converge to

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t-1} + \mathbf{K}(\mathbf{y}_t - \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1}), \qquad [13.5.8]$$

with $\mathbf{K}$ given by [13.5.4]. The forecast [13.5.6] will approach the forecast of $\mathbf{y}_{t+1}$ based on the infinite history of its own lagged values:

$$\hat{E}(\mathbf{y}_{t+1}|\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots) = \mathbf{H}'\hat{\boldsymbol{\xi}}_{t+1|t}. \qquad [13.5.9]$$

The *MSE* of this forecast is given by the limiting value of [13.5.7],

$$E\{[\mathbf{y}_{t+1} - \hat{E}(\mathbf{y}_{t+1}|\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots)][\mathbf{y}_{t+1} - \hat{E}(\mathbf{y}_{t+1}|\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots)]'\} = \mathbf{H}'\mathbf{P}\mathbf{H} + \mathbf{R}, \qquad [13.5.10]$$

where $\mathbf{P}$ is given by [13.5.3].

Equation [13.5.8] can be written

$$\hat{\boldsymbol{\xi}}_{t+1|t} = (\mathbf{F} - \mathbf{K}\mathbf{H}')L\hat{\boldsymbol{\xi}}_{t+1|t} + \mathbf{K}\mathbf{y}_t \qquad [13.5.11]$$

for $L$ the lag operator. Provided that the eigenvalues of $(\mathbf{F} - \mathbf{K}\mathbf{H}')$ are all inside the unit circle, [13.5.11] can be expressed as

$$\begin{aligned}\hat{\boldsymbol{\xi}}_{t+1|t} &= [\mathbf{I}_r - (\mathbf{F} - \mathbf{K}\mathbf{H}')L]^{-1}\mathbf{K}\mathbf{y}_t \\ &= [\mathbf{I}_r + (\mathbf{F} - \mathbf{K}\mathbf{H}')L + (\mathbf{F} - \mathbf{K}\mathbf{H}')^2L^2 + (\mathbf{F} - \mathbf{K}\mathbf{H}')^3L^3 + \cdots]\mathbf{K}\mathbf{y}_t.\end{aligned} \qquad [13.5.12]$$

Substituting [13.5.12] into [13.5.9] gives a steady-state rule for forecasting $\mathbf{y}_{t+1}$ as a linear function of its lagged values:

$$\hat{E}(\mathbf{y}_{t+1}|\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots) = \mathbf{H}'[\mathbf{I}_r - (\mathbf{F} - \mathbf{K}\mathbf{H}')L]^{-1}\mathbf{K}\mathbf{y}_t. \qquad [13.5.13]$$

Expression [13.5.13] implies a $VAR(\infty)$ representation for $\mathbf{y}_t$ of the form

$$\mathbf{y}_{t+1} = \mathbf{H}'[\mathbf{I}_r - (\mathbf{F} - \mathbf{K}\mathbf{H}')L]^{-1}\mathbf{K}\mathbf{y}_t + \boldsymbol{\varepsilon}_{t+1}, \qquad [13.5.14]$$

where

$$\boldsymbol{\varepsilon}_{t+1} \equiv \mathbf{y}_{t+1} - \hat{E}(\mathbf{y}_{t+1}|\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots). \qquad [13.5.15]$$

Thus, $\boldsymbol{\varepsilon}_{t+1}$ is the fundamental innovation for $\mathbf{y}_{t+1}$. Since $\boldsymbol{\varepsilon}_{t+1}$ is uncorrelated with $\mathbf{y}_{t-j}$ for any $j \geq 0$, it is also uncorrelated with $\boldsymbol{\varepsilon}_{t-j} = \mathbf{y}_{t-j} - \hat{E}(\mathbf{y}_{t-j}|\mathbf{y}_{t-j-1}, \mathbf{y}_{t-j-2}, \ldots)$ for any $j \geq 0$. The variance-covariance matrix of $\boldsymbol{\varepsilon}_{t+1}$ can be calculated using [13.5.15] and [13.5.10]):

$$\begin{aligned}E(\boldsymbol{\varepsilon}_{t+1}\boldsymbol{\varepsilon}_{t+1}') &= E\{[\mathbf{y}_{t+1} - \hat{E}(\mathbf{y}_{t+1}|\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots)] \\ &\quad \times [\mathbf{y}_{t+1} - \hat{E}(\mathbf{y}_{t+1}|\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots)]'\} \\ &= \mathbf{H}'\mathbf{P}\mathbf{H} + \mathbf{R}.\end{aligned} \qquad [13.5.16]$$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

Note that [13.5.14] can be written as

$$\{\mathbf{I}_n - \mathbf{H}'[\mathbf{I}_r - (\mathbf{F} - \mathbf{KH}')L]^{-1}\mathbf{K}L\}\mathbf{y}_{t+1} = \boldsymbol{\varepsilon}_{t+1}. \qquad [13.5.17]$$

The following result helps to rewrite the $VAR(\infty)$ representation [13.5.17] in the Wold $MA(\infty)$ form.

***Proposition 13.3:*** *Let* $\mathbf{F}$, $\mathbf{H}'$, *and* $\mathbf{K}$ *be matrices of dimension* $(r \times r)$, $(n \times r)$, *and* $(r \times n)$, *respectively, such that eigenvalues of* $\mathbf{F}$ *and of* $(\mathbf{F} - \mathbf{KH}')$ *are all inside the unit circle, and let* $z$ *be a scalar on the complex unit circle. Then*

$$\{\mathbf{I}_n + \mathbf{H}'(\mathbf{I}_r - \mathbf{F}z)^{-1}\mathbf{K}z\}\{\mathbf{I}_n - \mathbf{H}'[\mathbf{I}_r - (\mathbf{F} - \mathbf{KH}')z]^{-1}\mathbf{K}z\} = \mathbf{I}_n.$$

Applying Proposition 13.3, if both sides of [13.5.17] are premultiplied by $(\mathbf{I}_n + \mathbf{H}'(\mathbf{I}_r - \mathbf{F}L)^{-1}\mathbf{K}L)$, the result is the Wold representation for $\mathbf{y}$:

$$\mathbf{y}_{t+1} = \{\mathbf{I}_n + \mathbf{H}'(\mathbf{I}_r - \mathbf{F}L)^{-1}\mathbf{K}L\}\boldsymbol{\varepsilon}_{t+1}. \qquad [13.5.18]$$

To summarize, the Wold representation can be found by iterating on [13.5.1] until convergence. The steady-state value for $\mathbf{P}$ is then used to construct $\mathbf{K}$ in [13.5.4]. If the eigenvalues of $(\mathbf{F} - \mathbf{KH}')$ are all inside the unit circle, then the Wold representation is given by [13.5.18].

The task of finding the Wold representation is sometimes alternatively posed as the question of factoring the autocovariance-generating function of $\mathbf{y}$. Applying result [10.3.7] to [13.5.16] and [13.5.18], we would anticipate that the autocovariance-generating function of $\mathbf{y}$ can be written in the form

$$\begin{aligned}\mathbf{G}_{\mathbf{Y}}(z) = {} & \{\mathbf{I}_n + \mathbf{H}'(\mathbf{I}_r - \mathbf{F}z)^{-1}\mathbf{K}z\}\{\mathbf{H}'\mathbf{PH} + \mathbf{R}\} \\ & \times \{\mathbf{I}_n + \mathbf{K}'(\mathbf{I}_r - \mathbf{F}'z^{-1})^{-1}\mathbf{H}z^{-1}\}.\end{aligned} \qquad [13.5.19]$$

Compare [13.5.19] with the autocovariance-generating function that we would have written down directly from the structure of the state-space model. From [10.3.5], the autocovariance-generating function of $\boldsymbol{\xi}$ is given by

$$\mathbf{G}_{\boldsymbol{\xi}}(z) = [\mathbf{I}_r - \mathbf{F}z]^{-1}\mathbf{Q}[\mathbf{I}_r - \mathbf{F}'z^{-1}]^{-1},$$

while from [10.3.6] the autocovariance-generating function of $\mathbf{y}_t = \mathbf{H}'\boldsymbol{\xi}_t + \mathbf{w}_t$ is

$$\mathbf{G}_{\mathbf{Y}}(z) = \mathbf{H}'[\mathbf{I}_r - \mathbf{F}z]^{-1}\mathbf{Q}[\mathbf{I}_r - \mathbf{F}'z^{-1}]^{-1}\mathbf{H} + \mathbf{R}. \qquad [13.5.20]$$

Comparing [13.5.19] with [13.5.20] suggests that the limiting values for the Kalman gain and $MSE$ matrices $\mathbf{K}$ and $\mathbf{P}$ can be used to factor an autocovariance-generating function. The following proposition gives a formal statement of this result.

***Proposition 13.4:*** *Let* $\mathbf{F}$ *denote an* $(r \times r)$ *matrix whose eigenvalues are all inside the unit circle; let* $\mathbf{Q}$ *and* $\mathbf{R}$ *denote symmetric positive semidefinite matrices of dimension* $(r \times r)$ *and* $(n \times n)$, *respectively; and let* $\mathbf{H}'$ *denote an arbitrary* $(n \times r)$ *matrix. Let* $\mathbf{P}$ *be a positive semidefinite matrix satisfying [13.5.3] and let* $\mathbf{K}$ *be given by [13.5.4]. Suppose that eigenvalues of* $(\mathbf{F} - \mathbf{KH}')$ *are all inside the unit circle. Then*

$$\begin{aligned}&\mathbf{H}'[\mathbf{I}_r - \mathbf{F}z]^{-1}\mathbf{Q}[\mathbf{I}_r - \mathbf{F}'z^{-1}]^{-1}\mathbf{H} + \mathbf{R} \\ &= \{\mathbf{I}_n + \mathbf{H}'(\mathbf{I}_r - \mathbf{F}z)^{-1}\mathbf{K}z\}\{\mathbf{H}'\mathbf{PH} + \mathbf{R}\}\{\mathbf{I}_n + \mathbf{K}'(\mathbf{I}_r - \mathbf{F}'z^{-1})^{-1}\mathbf{H}z^{-1}\}.\end{aligned} \qquad [13.5.21]$$

A direct demonstration of this claim is provided in Appendix 13.A at the end of this chapter.

As an example of using these results, consider observations on a univariate $AR(1)$ process subject to white noise measurement error, such as the state-space system of [13.1.26] and [13.1.27] with $\mu = 0$. For this system, $\mathbf{F} = \phi$, $\mathbf{Q} = \sigma_V^2$, $\mathbf{A} = \mathbf{0}$, $\mathbf{H} = 1$, and $\mathbf{R} = \sigma_W^2$. The conditions of Proposition 13.2 are satisfied as long as $|\phi| < 1$, establishing that $|F - KH| = |\phi - K| < 1$. From equation



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

[13.5.14], the $AR(\infty)$ representation for this process can be found from

$$y_{t+1} = [1 - (\phi - K)L]^{-1}Ky_t + \varepsilon_{t+1},$$

which can be written

$$[1 - (\phi - K)L]y_{t+1} = Ky_t + [1 - (\phi - K)L]\varepsilon_{t+1}$$

or

$$y_{t+1} = \phi y_t + \varepsilon_{t+1} - (\phi - K)\varepsilon_t. \qquad [13.5.22]$$

This is an $ARMA(1, 1)$ process with $AR$ parameter given by $\phi$ and $MA$ parameter given by $-(\phi - K)$. The variance of the innovation for this process can be calculated from [13.5.16]:

$$E(\varepsilon_{t+1}^2) = \sigma_W^2 + P. \qquad [13.5.23]$$

The value of $P$ can be found by iterating on [13.5.1]:

$$\begin{aligned} P_{t+1|t} &= \phi^2[P_{t|t-1} - P_{t|t-1}^2/(\sigma_W^2 + P_{t|t-1})] + \sigma_V^2 \\ &= \phi^2 P_{t|t-1}\sigma_W^2/(\sigma_W^2 + P_{t|t-1}) + \sigma_V^2, \end{aligned} \qquad [13.5.24]$$

starting from $P_{1|0} = \sigma_V^2/(1 - \phi^2)$, until convergence. The steady-state Kalman gain is given by [13.5.4]:

$$K = \phi P/(\sigma_W^2 + P). \qquad [13.5.25]$$

As a second example, consider adding an $MA(q_1)$ process to an $MA(q_2)$ process with which the first process is uncorrelated at all leads and lags. This could be represented in state-space form as follows:

***State Equation*** $(r = q_1 + q_2 + 2)$**:**

$$\begin{bmatrix} u_{t+1} \\ u_t \\ \vdots \\ u_{t-q_1+1} \\ v_{t+1} \\ v_t \\ \vdots \\ v_{t-q_2+1} \end{bmatrix} = \underset{(q_1+q_2+2)\times(q_1+q_2+2)}{\begin{bmatrix} \mathbf{0}' & 0 & \mathbf{0}' & 0 \\ \mathbf{I}_{q_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & 0 & \mathbf{0}' & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{q_2} & \mathbf{0} \end{bmatrix}} \begin{bmatrix} u_t \\ u_{t-1} \\ \vdots \\ u_{t-q_1} \\ v_t \\ v_{t-1} \\ \vdots \\ v_{t-q_2} \end{bmatrix} + \begin{bmatrix} u_{t+1} \\ 0 \\ \vdots \\ 0 \\ v_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad [13.5.26]$$

***Observation Equation*** $(n = 1)$**:**

$$y_t = [1 \quad \delta_1 \quad \delta_2 \quad \cdots \quad \delta_{q_1} \quad 1 \quad \kappa_1 \quad \kappa_2 \quad \cdots \quad \kappa_{q_2}] \begin{bmatrix} u_t \\ u_{t-1} \\ \vdots \\ u_{t-q_1} \\ v_t \\ v_{t-1} \\ \vdots \\ v_{t-q_2} \end{bmatrix} \qquad [13.5.27]$$

Note that all eigenvalues of $\mathbf{F}$ are equal to zero. Write equation [13.5.18] in the form

$$\begin{aligned} \mathbf{y}_{t+1} &= \{\mathbf{I}_n + \mathbf{H}'(\mathbf{I}_r - \mathbf{F}L)^{-1}\mathbf{K}L\}\varepsilon_{t+1} \\ &= \{\mathbf{I}_n + \mathbf{H}'(\mathbf{I}_r + \mathbf{F}L + \mathbf{F}^2L^2 + \mathbf{F}^3L^3 + \cdots)\mathbf{K}L\}\varepsilon_{t+1}. \end{aligned} \qquad [13.5.28]$$

Let $q \equiv \max\{q_1, q_2\}$, and notice from the structure of $\mathbf{F}$ that $\mathbf{F}^{q+j} = \mathbf{0}$ for $j = 1, 2, \ldots$. Furthermore, from [13.5.4], $\mathbf{F}^q\mathbf{K} = \mathbf{F}^{q+1}\mathbf{PH}(\mathbf{H}'\mathbf{PH} + \mathbf{R})^{-1} = \mathbf{0}$. Thus



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

[13.5.28] takes the form

$$\begin{aligned} y_{t+1} &= \{1 + \mathbf{H}'(\mathbf{I}_r + \mathbf{F}L + \mathbf{F}^2L^2 + \mathbf{F}^3L^3 \\ &\quad + \cdots + \mathbf{F}^{q-1}L^{q-1})\mathbf{K}L\}\varepsilon_{t+1} \qquad [13.5.29] \\ &= \{1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q\}\varepsilon_{t+1}, \end{aligned}$$

where

$$\theta_j = \mathbf{H}'\mathbf{F}^{j-1}\mathbf{K} \qquad \text{for } j = 1, 2, \ldots, q.$$

This provides a constructive demonstration of the claim that an $MA(q_1)$ process plus an $MA(q_2)$ process with which it is uncorrelated can be described as an $MA(\max\{q_1, q_2\})$ process.

The Kalman filter thus provides a general algorithm for finding the Wold representation or factoring an autocovariance-generating function—we simply iterate on [13.5.1] until convergence, and then use the steady-state gain from [13.5.4] either in [13.5.14] (for the $AR(\infty)$ form) or in [13.5.18] (for the $MA(\infty)$ form).

Although the convergent values provide the Wold representation, for any finite $t$ the Kalman filter forecasts have the advantage of calculating the exact optimal forecast of $\mathbf{y}_{t+1}$ based on a linear function of $\{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1\}$.

## 13.6. *Smoothing*

The Kalman filter was motivated in Section 13.2 as an algorithm for calculating a forecast of the state vector $\boldsymbol{\xi}_t$ as a linear function of previous observations,

$$\hat{\boldsymbol{\xi}}_{t|t-1} \equiv \hat{E}(\boldsymbol{\xi}_t | \mathcal{Y}_{t-1}), \qquad [13.6.1]$$

where $\mathcal{Y}_{t-1} \equiv (\mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \ldots, \mathbf{y}'_1, \mathbf{x}'_{t-1}, \mathbf{x}'_{t-2}, \ldots, \mathbf{x}_1)'$. The matrix $\mathbf{P}_{t|t-1}$ represented the MSE of this forecast:

$$\mathbf{P}_{t|t-1} \equiv E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})']. \qquad [13.6.2]$$

For many uses of the Kalman filter these are the natural magnitudes of interest. In some settings, however, the state vector $\boldsymbol{\xi}_t$ is given a structural interpretation, in which case the value of this unobserved variable might be of interest for its own sake. For example, in the model of the business cycle by Stock and Watson, it would be helpful to know the state of the business cycle at any historical date $t$. A goal might then be to form an inference about the value of $\boldsymbol{\xi}_t$ based on the full set of data collected, including observations on $\mathbf{y}_t, \mathbf{y}_{t+1}, \ldots, \mathbf{y}_T, \mathbf{x}_t, \mathbf{x}_{t+1}, \ldots, \mathbf{x}_T$. Such an inference is called the *smoothed* estimate of $\boldsymbol{\xi}_t$, denoted

$$\hat{\boldsymbol{\xi}}_{t|T} \equiv \hat{E}(\boldsymbol{\xi}_t | \mathcal{Y}_T). \qquad [13.6.3]$$

For example, data on GNP from 1954 through 1990 might be used to estimate the value that $\boldsymbol{\xi}$ took on in 1960. The *MSE* of this smoothed estimate is denoted

$$\mathbf{P}_{t|T} \equiv E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|T})(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|T})']. \qquad [13.6.4]$$

In general, $\mathbf{P}_{t|\tau}$ denotes the *MSE* of an estimate of $\boldsymbol{\xi}_t$ that is based on observations of $\mathbf{y}$ and $\mathbf{x}$ through date $\tau$.

For the reader's convenience, we reproduce here the key equations for the Kalman filter:

$$\hat{\boldsymbol{\xi}}_{t|t} = \hat{\boldsymbol{\xi}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})^{-1}(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1}) \qquad [13.6.5]$$

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t} \qquad [13.6.6]$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})^{-1}\mathbf{H}'\mathbf{P}_{t|t-1} \qquad [13.6.7]$$

$$\mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}' + \mathbf{Q}. \qquad [13.6.8]$$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

Consider the estimate of $\boldsymbol{\xi}_t$ based on observations through date $t$, $\hat{\boldsymbol{\xi}}_{t|t}$. Suppose we were subsequently told the true value of $\boldsymbol{\xi}_{t+1}$. From the formula for updating a linear projection, equation [4.5.30], the new estimate of $\boldsymbol{\xi}_t$ could be expressed as[3]

$$\begin{aligned}\hat{E}(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1}, \mathcal{Y}_t) = \hat{\boldsymbol{\xi}}_{t|t} &+ \{E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t})(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})']\} \\ &\times \{E[(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})']\}^{-1} \\ &\times (\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t}).\end{aligned} \tag{13.6.9}$$

The first term in the product on the right side of [13.6.9] can be written

$$E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t})(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})'] = E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t})(\mathbf{F}\boldsymbol{\xi}_t + \mathbf{v}_{t+1} - \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t})'],$$

by virtue of [13.2.1] and [13.6.6]. Furthermore, $\mathbf{v}_{t+1}$ is uncorrelated with $\boldsymbol{\xi}_t$ and $\hat{\boldsymbol{\xi}}_{t|t}$. Thus,

$$E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t})(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})'] = E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t})(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t})'\mathbf{F}'] = \mathbf{P}_{t|t}\mathbf{F}'. \tag{13.6.10}$$

Substituting [13.6.10] and the definition of $\mathbf{P}_{t+1|t}$ into [13.6.9] produces

$$\hat{E}(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1}, \mathcal{Y}_t) = \hat{\boldsymbol{\xi}}_{t|t} + \mathbf{P}_{t|t}\mathbf{F}'\mathbf{P}_{t+1|t}^{-1}(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t}).$$

Defining

$$\mathbf{J}_t \equiv \mathbf{P}_{t|t}\mathbf{F}'\mathbf{P}_{t+1|t}^{-1}, \tag{13.6.11}$$

we have

$$\hat{E}(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1}, \mathcal{Y}_t) = \hat{\boldsymbol{\xi}}_{t|t} + \mathbf{J}_t(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t}). \tag{13.6.12}$$

Now, the linear projection in [13.6.12] turns out to be the same as

$$\hat{E}(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1}, \mathcal{Y}_T); \tag{13.6.13}$$

that is, knowledge of $\mathbf{y}_{t+j}$ or $\mathbf{x}_{t+j}$ for $j > 0$ would be of no added value if we already knew the value of $\boldsymbol{\xi}_{t+1}$. To see this, note that $\mathbf{y}_{t+j}$ can be written as

$$\mathbf{y}_{t+j} = \mathbf{A}'\mathbf{x}_{t+j} + \mathbf{H}'(\mathbf{F}^{j-1}\boldsymbol{\xi}_{t+1} + \mathbf{F}^{j-2}\mathbf{v}_{t+2} + \mathbf{F}^{j-3}\mathbf{v}_{t+3} + \cdots + \mathbf{v}_{t+j}) + \mathbf{w}_{t+j}.$$

But the error

$$\boldsymbol{\xi}_t - \hat{E}(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1}, \mathcal{Y}_t) \tag{13.6.14}$$

is uncorrelated with $\boldsymbol{\xi}_{t+1}$, by the definition of a linear projection, and uncorrelated with $\mathbf{x}_{t+j}$, $\mathbf{w}_{t+j}$, $\mathbf{v}_{t+j}$, $\mathbf{v}_{t+j-1}$, . . . , $\mathbf{v}_{t+2}$ under the maintained assumptions. Thus, the error [13.6.14] is uncorrelated with $\mathbf{y}_{t+j}$ or $\mathbf{x}_{t+j}$ for $j > 0$, meaning that [13.6.13] and [13.6.12] are the same, as claimed:

$$\hat{E}(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1}, \mathcal{Y}_T) = \hat{\boldsymbol{\xi}}_{t|t} + \mathbf{J}_t(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t}). \tag{13.6.15}$$

It follows from the law of iterated projections that the smoothed estimate, $\hat{E}(\boldsymbol{\xi}_t|\mathcal{Y}_T)$, can be obtained by projecting [13.6.15] on $\mathcal{Y}_T$. In calculating this projection, we need to think carefully about the nature of the magnitudes in [13.6.15]. The first term, $\hat{\boldsymbol{\xi}}_{t|t}$, indicates a particular *exact* linear function of $\mathcal{Y}_t$; the coefficients of this function are constructed from population moments, and these coefficients should be viewed as deterministic constants from the point of view of performing a subsequent projection. The projection of $\hat{\boldsymbol{\xi}}_{t|t}$ on $\mathcal{Y}_T$ is thus still $\hat{\boldsymbol{\xi}}_{t|t}$, this same

[3]Here, $\mathbf{Y}_3 = \boldsymbol{\xi}_t$, $\mathbf{Y}_2 = \boldsymbol{\xi}_{t+1}$, and $\mathbf{Y}_1 = \mathcal{Y}_t$.



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

linear function of $\mathcal{Y}_t$—we can't improve on a perfect fit![4] The term $\mathbf{J}_t$ in [13.6.11] is also a function of population moments, and so is again treated as deterministic for purposes of any linear projection. The term $\hat{\boldsymbol{\xi}}_{t+1|t}$ is another exact linear function of $\mathcal{Y}_t$. Thus, projecting [13.6.15] on $\mathcal{Y}_T$ turns out to be trivial:

$$\hat{E}(\boldsymbol{\xi}_t|\mathcal{Y}_T) = \hat{\boldsymbol{\xi}}_{t|t} + \mathbf{J}_t[\hat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_T) - \hat{\boldsymbol{\xi}}_{t+1|t}],$$

or

$$\hat{\boldsymbol{\xi}}_{t|T} = \hat{\boldsymbol{\xi}}_{t|t} + \mathbf{J}_t(\hat{\boldsymbol{\xi}}_{t+1|T} - \hat{\boldsymbol{\xi}}_{t+1|t}). \qquad [13.6.16]$$

Thus, the sequence of smoothed estimates $\{\hat{\boldsymbol{\xi}}_{t|T}\}_{t=1}^{T}$ is calculated as follows. First, the Kalman filter, [13.6.5] to [13.6.8], is calculated and the sequences $\{\hat{\boldsymbol{\xi}}_{t|t}\}_{t=1}^{T}$, $\{\hat{\boldsymbol{\xi}}_{t+1|t}\}_{t=0}^{T-1}$, $\{\mathbf{P}_{t|t}\}_{t=1}^{T}$, and $\{\mathbf{P}_{t+1|t}\}_{t=0}^{T-1}$ are stored. The smoothed estimate for the final date in the sample, $\hat{\boldsymbol{\xi}}_{T|T}$, is just the last entry in $\{\hat{\boldsymbol{\xi}}_{t|t}\}_{t=1}^{T}$. Next, [13.6.11] is used to generate $\{\mathbf{J}_t\}_{t=1}^{T-1}$. From this, [13.6.16] is used for $t = T - 1$ to calculate

$$\hat{\boldsymbol{\xi}}_{T-1|T} = \hat{\boldsymbol{\xi}}_{T-1|T-1} + \mathbf{J}_{T-1}(\hat{\boldsymbol{\xi}}_{T|T} - \hat{\boldsymbol{\xi}}_{T|T-1}).$$

Now that $\hat{\boldsymbol{\xi}}_{T-1|T}$ has been calculated, [13.6.16] can be used for $t = T - 2$ to evaluate

$$\hat{\boldsymbol{\xi}}_{T-2|T} = \hat{\boldsymbol{\xi}}_{T-2|T-2} + \mathbf{J}_{T-2}(\hat{\boldsymbol{\xi}}_{T-1|T} - \hat{\boldsymbol{\xi}}_{T-1|T-2}).$$

Proceeding backward through the sample in this fashion permits calculation of the full set of smoothed estimates, $\{\hat{\boldsymbol{\xi}}_{t|T}\}_{t=1}^{T}$.

Next, consider the mean squared error associated with the smoothed estimate. Subtracting both sides of [13.6.16] from $\boldsymbol{\xi}_t$ produces

$$\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|T} = \boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t} - \mathbf{J}_t\hat{\boldsymbol{\xi}}_{t+1|T} + \mathbf{J}_t\hat{\boldsymbol{\xi}}_{t+1|t}$$

or

$$\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|T} + \mathbf{J}_t\hat{\boldsymbol{\xi}}_{t+1|T} = \boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t} + \mathbf{J}_t\hat{\boldsymbol{\xi}}_{t+1|t}.$$

Multiplying this equation by its transpose and taking expectations,

$$\begin{aligned} &E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|T})(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|T})'] + \mathbf{J}_t E[(\hat{\boldsymbol{\xi}}_{t+1|T}\hat{\boldsymbol{\xi}}'_{t+1|T})]\mathbf{J}'_t \\ &\quad = E[(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t})(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t})'] + \mathbf{J}_t E[(\hat{\boldsymbol{\xi}}_{t+1|t}\hat{\boldsymbol{\xi}}'_{t+1|t})]\mathbf{J}'_t. \end{aligned} \qquad [13.6.17]$$

The cross-product terms have disappeared from the left side because $\hat{\boldsymbol{\xi}}_{t+1|T}$ is a linear function of $\mathcal{Y}_T$ and so is uncorrelated with the projection error $\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|T}$. Similarly, on the right side, $\hat{\boldsymbol{\xi}}_{t+1|t}$ is uncorrelated with $\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t}$.

Equation [13.6.17] states that

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{J}_t\{-E[(\hat{\boldsymbol{\xi}}_{t+1|T}\hat{\boldsymbol{\xi}}'_{t+1|T})] + E[(\hat{\boldsymbol{\xi}}_{t+1|t}\hat{\boldsymbol{\xi}}'_{t+1|t})]\}\mathbf{J}'_t. \qquad [13.6.18]$$

[4]The law of iterated projections states that

$$\hat{E}(\boldsymbol{\xi}_t|\mathcal{Y}_t) = \hat{E}[\hat{E}(\boldsymbol{\xi}_t|\mathcal{Y}_T)|\mathcal{Y}_t].$$

The law of iterated projections thus allows us to go from a larger information set to a smaller. Of course, the same operation does not work in reverse:

$$\hat{E}(\boldsymbol{\xi}_t|\mathcal{Y}_T) \neq \hat{E}[\hat{E}(\boldsymbol{\xi}_t|\mathcal{Y}_t)|\mathcal{Y}_T].$$

We cannot go from a smaller information set to a larger.

An example may clarify this point. Let $y_t$ be an i.i.d. zero-mean sequence with

$$\xi_t = \mu + y_{t+1}.$$

Then

$$\hat{E}(\xi_t|y_t) = \mu$$

and

$$\hat{E}[\hat{E}(\xi_t|y_t)|y_t, y_{t+1}] = \hat{E}[\mu|y_t, y_{t+1}] = \mu.$$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

The bracketed term in [13.6.18] can be expressed as

$$\begin{aligned}
&-E[(\hat{\boldsymbol{\xi}}_{t+1|T}\hat{\boldsymbol{\xi}}'_{t+1|T})] + E[(\hat{\boldsymbol{\xi}}_{t+1|t}\hat{\boldsymbol{\xi}}'_{t+1|t})] \\
&= \{E[(\boldsymbol{\xi}_{t+1}\boldsymbol{\xi}'_{t+1})] - E[(\hat{\boldsymbol{\xi}}_{t+1|T}\hat{\boldsymbol{\xi}}'_{t+1|T})]\} - \{E[(\boldsymbol{\xi}_{t+1}\boldsymbol{\xi}'_{t+1})] - E[(\hat{\boldsymbol{\xi}}_{t+1|t}\hat{\boldsymbol{\xi}}'_{t+1|t})]\} \\
&= \{E[(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|T})(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|T})']\} - \{E[(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})']\} \\
&= \mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}.
\end{aligned} \tag{13.6.19}$$

The second-to-last equality used the fact that

$$\begin{aligned}
E[\boldsymbol{\xi}_{t+1}\hat{\boldsymbol{\xi}}'_{t+1|T}] &= E[(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|T} + \hat{\boldsymbol{\xi}}_{t+1|T})\hat{\boldsymbol{\xi}}'_{t+1|T}] \\
&= E[(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|T})\hat{\boldsymbol{\xi}}'_{t+1|T}] + E[\hat{\boldsymbol{\xi}}_{t+1|T}\hat{\boldsymbol{\xi}}'_{t+1|T}] \\
&= E[\hat{\boldsymbol{\xi}}_{t+1|T}\hat{\boldsymbol{\xi}}'_{t+1|T}],
\end{aligned}$$

since the projection error $(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|T})$ is uncorrelated with $\hat{\boldsymbol{\xi}}_{t+1|T}$. Similarly, $E(\boldsymbol{\xi}_{t+1}\hat{\boldsymbol{\xi}}'_{t+1|t}) = E(\hat{\boldsymbol{\xi}}_{t+1|t}\hat{\boldsymbol{\xi}}'_{t+1|t})$. Substituting [13.6.19] into [13.6.18] establishes that the smoothed estimate $\hat{\boldsymbol{\xi}}_{t|T}$ has *MSE* given by

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{J}_t(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{J}'_t. \tag{13.6.20}$$

Again, this sequence is generated by moving through the sample backward starting with $t = T - 1$.

## 13.7. *Statistical Inference with the Kalman Filter*

The calculation of the mean squared error

$$\mathbf{P}_{\tau|t} = E[(\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|t})(\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|t})']$$

described earlier assumed that the parameters of the matrices $\mathbf{F}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{R}$ were known with certainty. Section 13.4 showed how these parameters could be estimated from the data by maximum likelihood. There would then be some sampling uncertainty about the true values of these parameters, and the calculation of $\mathbf{P}_{\tau|t}$ would need to be modified to obtain the true mean squared errors of the smoothed estimates and forecasts.[5]

Suppose the unknown parameters are collected in a vector $\boldsymbol{\theta}$. For any given value of $\boldsymbol{\theta}$, the matrices $\mathbf{F}(\boldsymbol{\theta})$, $\mathbf{Q}(\boldsymbol{\theta})$, $\mathbf{A}(\boldsymbol{\theta})$, $\mathbf{H}(\boldsymbol{\theta})$, and $\mathbf{R}(\boldsymbol{\theta})$ could be used to construct $\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta})$ and $\mathbf{P}_{\tau|T}(\boldsymbol{\theta})$ in the formulas presented earlier; for $\tau \leq T$, these are the smoothed estimate and *MSE* given in [13.6.16] and [13.6.20], respectively; while for $\tau > T$, these are the forecast and its *MSE* in [13.3.25] and [13.3.27]. Let $\mathcal{Y}_T \equiv (\mathbf{y}'_T, \mathbf{y}'_{T-1}, \ldots, \mathbf{y}'_1, \mathbf{x}'_T, \mathbf{x}'_{T-1}, \ldots, \mathbf{x}'_1)'$ denote the observed data, and let $\boldsymbol{\theta}_0$ denote the true value of $\boldsymbol{\theta}$. The earlier derivations assumed that the true value of $\boldsymbol{\theta}$ was used to construct $\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0)$ and $\mathbf{P}_{\tau|T}(\boldsymbol{\theta}_0)$.

Recall that the formulas for updating a linear projection and its *MSE*, [4.5.30] and [4.5.31], yield the conditional mean and conditional *MSE* when applied to Gaussian vectors; see equation [4.6.7]. Thus, if $\{\mathbf{v}_t\}$, $\{\mathbf{w}_t\}$, and $\boldsymbol{\xi}_1$ are truly Gaussian, then the linear projection $\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0)$ has the interpretation as the expectation of $\boldsymbol{\xi}_\tau$ conditional on the data,

$$\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0) = E(\boldsymbol{\xi}_\tau|\mathcal{Y}_T); \tag{13.7.1}$$

while $\mathbf{P}_{\tau|T}(\boldsymbol{\theta}_0)$ can be described as the conditional *MSE*:

$$\mathbf{P}_{\tau|T}(\boldsymbol{\theta}_0) = E\{[\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0)][\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0)]'|\mathcal{Y}_T\}. \tag{13.7.2}$$

Let $\hat{\boldsymbol{\theta}}$ denote an estimate of $\boldsymbol{\theta}$ based on $\mathcal{Y}_T$, and let $\hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})$ denote the estimate that results from using $\hat{\boldsymbol{\theta}}$ to construct the smoothed inference or forecast in [13.6.16]

[5]This discussion is based on Hamilton (1986).



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

or [13.3.25]. The conditional mean squared error of this estimate is

$$\begin{aligned}
&E\{[\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})][\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})]'|\mathcal{Y}_T\} \\
&\quad = E\{[\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0) + \hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})] \\
&\qquad \times [\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0) + \hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})]'|\mathcal{Y}_T\} \qquad [13.7.3] \\
&\quad = E\{[\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0)][\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0)]'|\mathcal{Y}_T\} \\
&\qquad + E\{[\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})][\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})]'|\mathcal{Y}_T\}.
\end{aligned}$$

Cross-product terms have disappeared from [13.7.3], since

$$\begin{aligned}
&E\{[\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})][\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0)]'|\mathcal{Y}_T\} \\
&\quad = [\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})] \times E\{[\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0)]'|\mathcal{Y}_T\} \\
&\quad = [\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})] \times \mathbf{0}'.
\end{aligned}$$

The first equality follows because $\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0)$ and $\hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})$ are known nonstochastic functions of $\mathcal{Y}_T$, and the second equality is implied by [13.7.1]. Substituting [13.7.2] into [13.7.3] results in

$$\begin{aligned}
&E\{[\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})][\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})]'|\mathcal{Y}_T\} \\
&\quad = \mathbf{P}_{\tau|T}(\boldsymbol{\theta}_0) + E\{[\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})][\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})]'|\mathcal{Y}_T\}. \qquad [13.7.4]
\end{aligned}$$

Equation [13.7.4] decomposes the mean squared error into two components. The first component, $\mathbf{P}_{\tau|T}(\boldsymbol{\theta}_0)$, might be described as the "filter uncertainty." This is the term calculated from the smoothing iteration [13.6.20] or forecast *MSE* [13.3.27] and represents uncertainty about $\boldsymbol{\xi}_\tau$ that would be present even if the true value $\boldsymbol{\theta}_0$ were known with certainty. The second term in [13.7.4],

$$E\{[\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})][\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}_0) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})]'\},$$

might be called "parameter uncertainty." It reflects the fact that in a typical sample, $\hat{\boldsymbol{\theta}}$ will differ from the true value $\boldsymbol{\theta}_0$.

A simple way to estimate the size of each source of uncertainty is by Monte Carlo integration. Suppose we adopt the Bayesian perspective that $\boldsymbol{\theta}$ itself is a random variable. From this perspective, [13.7.4] describes the *MSE* conditional on $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Suppose that the posterior distribution of $\boldsymbol{\theta}$ conditional on the data $\mathcal{Y}_T$ is known; the asymptotic distribution for the *MLE* in [13.4.6] suggests that $\boldsymbol{\theta}|\mathcal{Y}_T$ might be regarded as approximately distributed $N(\hat{\boldsymbol{\theta}}, (1/T)\cdot\mathcal{I}^{-1})$, where $\hat{\boldsymbol{\theta}}$ denotes the *MLE*. We might then generate a large number of values of $\boldsymbol{\theta}$, say, $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(2000)}$, drawn from a $N(\hat{\boldsymbol{\theta}}, (1/T)\cdot\mathcal{I}^{-1})$ distribution. For each draw $(j)$, we could calculate the smoothed estimate or forecast $\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}^{(j)})$. The deviations of these estimates across Monte Carlo draws from the estimate $\hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})$ can be used to describe how sensitive the estimate $\hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})$ is to parameter uncertainty about $\boldsymbol{\theta}$:

$$\frac{1}{2000}\sum_{j=1}^{2000}[\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}^{(j)}) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})][\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}^{(j)}) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})]'. \qquad [13.7.5]$$

This affords an estimate of

$$E\{[\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})][\hat{\boldsymbol{\xi}}_{\tau|T}(\boldsymbol{\theta}) - \hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})]'|\mathcal{Y}_T\},$$

where this expectation is understood to be with respect to the distribution of $\boldsymbol{\theta}$ conditional on $\mathcal{Y}_T$.

For each Monte Carlo realization $\boldsymbol{\theta}^{(j)}$, we can also calculate $\mathbf{P}_{\tau|T}(\boldsymbol{\theta}^{(j)})$ from [13.6.20] or [13.3.27]. Its average value across Monte Carlo draws,

$$\frac{1}{2000}\sum_{j=1}^{2000}\mathbf{P}_{\tau|T}(\boldsymbol{\theta}^{(j)}), \qquad [13.7.6]$$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

provides an estimate of the filter uncertainty in [13.7.4],

$$E[\mathbf{P}_{\tau|T}(\boldsymbol{\theta})|\mathcal{Y}_T].$$

Again, this expectation is with respect to the distribution of $\boldsymbol{\theta}|\mathcal{Y}_T$.

The sum of [13.7.5] and [13.7.6] is then proposed as an *MSE* for the estimate $\hat{\boldsymbol{\xi}}_{\tau|T}(\hat{\boldsymbol{\theta}})$ around the true value $\boldsymbol{\xi}_\tau$.

## 13.8. *Time-Varying Parameters*

### *State-Space Model with Stochastically Varying Coefficients*

Up to this point we have been assuming that the matrices $\mathbf{F}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{R}$ were all constant. The Kalman filter can also be adapted for more general state-space models in which the values of these matrices depend on the exogenous or lagged dependent variables included in the vector $\mathbf{x}_t$. Consider

$$\boldsymbol{\xi}_{t+1} = \mathbf{F}(\mathbf{x}_t)\boldsymbol{\xi}_t + \mathbf{v}_{t+1} \qquad [13.8.1]$$

$$\mathbf{y}_t = \mathbf{a}(\mathbf{x}_t) + [\mathbf{H}(\mathbf{x}_t)]'\boldsymbol{\xi}_t + \mathbf{w}_t. \qquad [13.8.2]$$

Here $\mathbf{F}(\mathbf{x}_t)$ denotes an $(r \times r)$ matrix whose elements are functions of $\mathbf{x}_t$; $\mathbf{a}(\mathbf{x}_t)$ similarly describes an $(n \times 1)$ vector-valued function, and $\mathbf{H}(\mathbf{x}_t)$ an $(r \times n)$ matrix-valued function. It is assumed that conditional on $\mathbf{x}_t$ and on data observed through date $t - 1$, denoted

$$\mathcal{Y}_{t-1} \equiv (\mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \ldots, \mathbf{y}'_1, \mathbf{x}'_{t-1}, \mathbf{x}'_{t-2}, \ldots, \mathbf{x}'_1)',$$

the vector $(\mathbf{v}'_{t+1}, \mathbf{w}'_t)'$ has the Gaussian distribution

$$\left[\begin{matrix}\mathbf{v}_{t+1}\\ \mathbf{w}_t\end{matrix}\middle|\, \mathbf{x}_t, \mathcal{Y}_{t-1}\right] \sim N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{Q}(\mathbf{x}_t) & \mathbf{0}\\ \mathbf{0} & \mathbf{R}(\mathbf{x}_t)\end{bmatrix}\right). \qquad [13.8.3]$$

Note that although [13.8.1] to [13.8.3] generalize the earlier framework by allowing for stochastically varying parameters, it is more restrictive in that a Gaussian distribution is assumed in [13.8.3]; the role of the Gaussian requirement will be explained shortly.

Suppose it is taken as given that $\boldsymbol{\xi}_t|\mathcal{Y}_{t-1} \sim N(\hat{\boldsymbol{\xi}}_{t|t-1}, \mathbf{P}_{t|t-1})$. Assuming as before that $\mathbf{x}_t$ contains only strictly exogenous variables or lagged values of $\mathbf{y}$, this also describes the distribution of $\boldsymbol{\xi}_t|\mathbf{x}_t, \mathcal{Y}_{t-1}$. It follows from the assumptions in [13.8.1] to [13.8.3] that

$$\left[\begin{matrix}\boldsymbol{\xi}_t\\ \mathbf{y}_t\end{matrix}\middle|\, \mathbf{x}_t, \mathcal{Y}_{t-1}\right] \sim N\left(\begin{bmatrix}\hat{\boldsymbol{\xi}}_{t|t-1}\\ \mathbf{a}(\mathbf{x}_t) + [\mathbf{H}(\mathbf{x}_t)]'\hat{\boldsymbol{\xi}}_{t|t-1}\end{bmatrix}, \begin{bmatrix}\mathbf{P}_{t|t-1} & \mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{x}_t)\\ \mathbf{H}'(\mathbf{x}_t)\mathbf{P}_{t|t-1} & [\mathbf{H}(\mathbf{x}_t)]'\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{x}_t) + \mathbf{R}(\mathbf{x}_t)\end{bmatrix}\right). \qquad [13.8.4]$$

Conditional on $\mathbf{x}_t$, the terms $\mathbf{a}(\mathbf{x}_t)$, $\mathbf{H}(\mathbf{x}_t)$, and $\mathbf{R}(\mathbf{x}_t)$ can all be treated as deterministic. Thus, the formula for the conditional distribution of Gaussian vectors [4.6.7] can be used to deduce that[6]

$$\boldsymbol{\xi}_t|\mathbf{y}_t, \mathbf{x}_t, \mathcal{Y}_{t-1} \equiv \boldsymbol{\xi}_t|\mathcal{Y}_t \sim N(\hat{\boldsymbol{\xi}}_{t|t}, \mathbf{P}_{t|t}), \qquad [13.8.5]$$

[6]Here $\mathbf{Y}_1 = \mathbf{y}_t$, $\mathbf{Y}_2 = \boldsymbol{\xi}_t$, $\boldsymbol{\mu}_1 = \mathbf{a}(\mathbf{x}_t) + [\mathbf{H}(\mathbf{x}_t)]'\hat{\boldsymbol{\xi}}_{t|t-1}$, $\boldsymbol{\mu}_2 = \hat{\boldsymbol{\xi}}_{t|t-1}$, $\boldsymbol{\Omega}_{11} = \{[\mathbf{H}(\mathbf{x}_t)]'\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{x}_t) + \mathbf{R}(\mathbf{x}_t)\}$, $\boldsymbol{\Omega}_{22} = \mathbf{P}_{t|t-1}$, and $\boldsymbol{\Omega}_{21} = \mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{x}_t)$.



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

where

$$\hat{\boldsymbol{\xi}}_{t|t} = \hat{\boldsymbol{\xi}}_{t|t-1} + \Big\{\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{x}_t)\big[[\mathbf{H}(\mathbf{x}_t)]'\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{x}_t) + \mathbf{R}(\mathbf{x}_t)\big]^{-1} \times \big[\mathbf{y}_t - \mathbf{a}(\mathbf{x}_t) - [\mathbf{H}(\mathbf{x}_t)]'\hat{\boldsymbol{\xi}}_{t|t-1}\big]\Big\} \qquad [13.8.6]$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \Big\{\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{x}_t) \times \big[[\mathbf{H}(\mathbf{x}_t)]'\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{x}_t) + \mathbf{R}(\mathbf{x}_t)\big]^{-1}[\mathbf{H}(\mathbf{x}_t)]'\mathbf{P}_{t|t-1}\Big\}. \qquad [13.8.7]$$

It then follows from [13.8.1] and [13.8.3] that $\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t \sim N(\hat{\boldsymbol{\xi}}_{t+1|t}, \mathbf{P}_{t+1|t})$, where

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F}(\mathbf{x}_t)\hat{\boldsymbol{\xi}}_{t|t} \qquad [13.8.8]$$

$$\mathbf{P}_{t+1|t} = \mathbf{F}(\mathbf{x}_t)\mathbf{P}_{t|t}[\mathbf{F}(\mathbf{x}_t)]' + \mathbf{Q}(\mathbf{x}_t). \qquad [13.8.9]$$

Equations [13.8.6] through [13.8.9] are just the Kalman filter equations [13.2.15], [13.2.16], [13.2.17], and [13.2.21] with the parameter matrices $\mathbf{F}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{R}$ replaced by their time-varying analogs. Thus, as long as we are willing to treat the initial state $\boldsymbol{\xi}_1$ as $N(\hat{\boldsymbol{\xi}}_{1|0}, \mathbf{P}_{1|0})$, the Kalman filter iterations go through the same as before. The obvious generalization of [13.4.1] can continue to be used to evaluate the likelihood function.

Note, however, that unlike the constant-parameter case, the inference [13.8.6] is a nonlinear function of $\mathbf{x}_t$. This means that although [13.8.6] gives the optimal inference if the disturbances and initial state are Gaussian, it cannot be interpreted as the linear projection of $\boldsymbol{\xi}_t$ on $\mathcal{Y}_t$ with non-Gaussian disturbances.

## *Linear Regression Models with Time-Varying Coefficients*

One important application of the state-space model with stochastically varying parameters is a regression in which the coefficient vector changes over time. Consider

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_t + w_t, \qquad [13.8.10]$$

where $\mathbf{x}_t$ is a $(k \times 1)$ vector that can include lagged values of $y$ or variables that are independent of the regression disturbance $w_\tau$ for all $\tau$. The parameters of the coefficient vector are presumed to evolve over time according to

$$(\boldsymbol{\beta}_{t+1} - \bar{\boldsymbol{\beta}}) = \mathbf{F}(\boldsymbol{\beta}_t - \bar{\boldsymbol{\beta}}) + \mathbf{v}_{t+1}. \qquad [13.8.11]$$

If the eigenvalues of the $(k \times k)$ matrix $\mathbf{F}$ are all inside the unit circle, then $\bar{\boldsymbol{\beta}}$ has the interpretation as the average or steady-state value for the coefficient vector. If it is further assumed that

$$\left[\begin{matrix}\mathbf{v}_{t+1}\\ w_t\end{matrix}\,\middle|\,\mathbf{x}_t, \mathcal{Y}_{t-1}\right] \sim N\left(\begin{bmatrix}\mathbf{0}\\ 0\end{bmatrix}, \begin{bmatrix}\mathbf{Q} & \mathbf{0}\\ \mathbf{0}' & \sigma^2\end{bmatrix}\right), \qquad [13.8.12]$$

then [13.8.10] to [13.8.12] will be recognized as a state-space model of the form of [13.8.1] to [13.8.3] with state vector $\boldsymbol{\xi}_t = \boldsymbol{\beta}_t - \bar{\boldsymbol{\beta}}$. The regression in [13.8.10] can be written as

$$y_t = \mathbf{x}_t'\bar{\boldsymbol{\beta}} + \mathbf{x}_t'\boldsymbol{\xi}_t + w_t, \qquad [13.8.13]$$

which is an observation equation of the form of [13.8.2] with $\mathbf{a}(\mathbf{x}_t) = \mathbf{x}_t'\bar{\boldsymbol{\beta}}$, $\mathbf{H}(\mathbf{x}_t) = \mathbf{x}_t$, and $\mathbf{R}(\mathbf{x}_t) = \sigma^2$. These values are then used in the Kalman filter iterations [13.8.6] to [13.8.9]. A one-period-ahead forecast for [13.8.10] can be calculated from [13.8.4] as

$$E(y_t|\mathbf{x}_t, \mathcal{Y}_{t-1}) = \mathbf{x}_t'\bar{\boldsymbol{\beta}} + \mathbf{x}_t'\hat{\boldsymbol{\xi}}_{t|t-1},$$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

where $\{\hat{\boldsymbol{\xi}}_{t|t-1}\}_{t=1}^{T}$ is calculated from [13.8.6] and [13.8.8]. The *MSE* of this forecast can also be inferred from [13.8.4]:

$$E[(y_t - \mathbf{x}_t'\bar{\boldsymbol{\beta}} - \mathbf{x}_t'\hat{\boldsymbol{\xi}}_{t|t-1})^2|\mathbf{x}_t, \mathcal{Y}_{t-1}] = \mathbf{x}_t'\mathbf{P}_{t|t-1}\mathbf{x}_t + \sigma^2,$$

where $\{\mathbf{P}_{t|t-1}\}_{t=1}^{T}$ is calculated from [13.8.7] and [13.8.9]. The sample log likelihood is therefore

$$\sum_{t=1}^{T} \log f(y_t|\mathbf{x}_t, \mathcal{Y}_{t-1}) = -(T/2)\log(2\pi) - (1/2)\sum_{t=1}^{T}\log(\mathbf{x}_t'\mathbf{P}_{t|t-1}\mathbf{x}_t + \sigma^2)$$
$$- (1/2)\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\bar{\boldsymbol{\beta}} - \mathbf{x}_t'\hat{\boldsymbol{\xi}}_{t|t-1})^2/(\mathbf{x}_t'\mathbf{P}_{t|t-1}\mathbf{x}_t + \sigma^2).$$

The specification in [13.8.11] can easily be generalized to allow for a $p$th-order *VAR* for the coefficient vector $\boldsymbol{\beta}_t$ by defining $\boldsymbol{\xi}_t' \equiv [(\boldsymbol{\beta}_t - \bar{\boldsymbol{\beta}})', (\boldsymbol{\beta}_{t-1} - \bar{\boldsymbol{\beta}})', \ldots, (\boldsymbol{\beta}_{t-p+1} - \bar{\boldsymbol{\beta}})']$ and replacing [13.8.11] with

$$\boldsymbol{\xi}_{t+1} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \cdots & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_p \\ \mathbf{I}_k & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_k & \mathbf{0} \end{bmatrix} \boldsymbol{\xi}_t + \begin{bmatrix} \mathbf{v}_{t+1} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}.$$

## *Estimation of a* VAR *with Time-Varying Coefficients*

Section 12.2 described Litterman's approach to Bayesian estimation of an equation of a vector autoregression with constant but unknown coefficients. A related approach to estimating a *VAR* with time-varying coefficients was developed by Doan, Litterman, and Sims (1984). Although efficiency might be improved by estimating all the equations of the *VAR* jointly, their proposal was to infer the parameters for each equation in isolation from the others.

Suppose for illustration that equation [13.8.10] describes the first equation from a *VAR*, so that the dependent variable $(y_t)$ is $y_{1t}$ and the $(k \times 1)$ vector of explanatory variables is $\mathbf{x}_t = (1, \mathbf{y}_{t-1}', \mathbf{y}_{t-2}', \ldots, \mathbf{y}_{t-p}')'$, where $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ and $k = np + 1$. The coefficient vector is

$$\boldsymbol{\beta}_t = (c_{1,t}, \phi_{11,t}^{(1)}, \phi_{12,t}^{(1)}, \ldots, \phi_{1n,t}^{(1)}, \phi_{11,t}^{(2)}, \phi_{12,t}^{(2)}, \ldots, \phi_{1n,t}^{(2)}, \ldots,$$
$$\phi_{11,t}^{(p)}, \phi_{12,t}^{(p)}, \ldots, \phi_{1n,t}^{(p)})',$$

where $\phi_{1j,t}^{(s)}$ is the coefficient relating $y_{1t}$ to $y_{j,t-s}$. This coefficient is allowed to be different for each date $t$ in the sample.

Doan, Litterman, and Sims specified a Bayesian prior distribution for the initial value of the coefficient vector at date 1:

$$\boldsymbol{\beta}_1 \sim N(\bar{\boldsymbol{\beta}}, \mathbf{P}_{1|0}). \qquad [13.8.14]$$

The prior distribution is independent across coefficients, so that $\mathbf{P}_{1|0}$ is a diagonal matrix. The mean of the prior distribution, $\bar{\boldsymbol{\beta}}$, is that used by Litterman (1986) for a constant-coefficient *VAR*. This prior distribution holds that changes in $y_{1t}$ are probably difficult to forecast, so that the coefficient on $y_{1,t-1}$ is likely to be near unity and all other coefficients are expected to be near zero:

$$\bar{\boldsymbol{\beta}} = (0, 1, 0, 0, \ldots, 0)'. \qquad [13.8.15]$$

As in Section 12.2, let $\gamma$ characterize the analyst's confidence in the prediction that $\phi_{11,1}^{(1)}$ is near unity:

$$\phi_{11,1}^{(1)} \sim N(1, \gamma^2).$$

Smaller values of $\gamma$ imply more confidence in the prior conviction that $\phi_{11,1}^{(1)}$ is near unity.

The coefficient $\phi_{11,1}^{(s)}$ relates the value of variable 1 at date 1 to its own value



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

$s$ periods earlier. Doan, Litterman, and Sims had more confidence in the prior conviction that $\phi_{11,1}^{(s)}$ is zero the greater the lag, or the larger the value of $s$. They represented this with a harmonic series for the variance,

$$\phi_{11,1}^{(s)} \sim N(0, \gamma^2/s) \qquad \text{for } s = 2, 3, \ldots, p.$$

The prior distribution for the coefficient relating variable 1 to lags of other variables was taken to be

$$\phi_{1j,1}^{(s)} \sim N\left(0, \frac{w^2 \cdot \gamma^2 \cdot \hat{\tau}_1^2}{s \cdot \hat{\tau}_j^2}\right) \qquad \begin{matrix} j = 2, 3, \ldots, n \\ s = 1, 2, \ldots, p \end{matrix}. \qquad [13.8.16]$$

As in expression [12.2.4], this includes a correction $(\hat{\tau}_1^2/\hat{\tau}_j^2)$ for the scale of $y_{1t}$ relative to $y_{jt}$, where $\hat{\tau}_j^2$ is the estimated variance of the residuals for a univariate fixed-coefficient $AR(p)$ process fitted to series $j$. The variance in [13.8.16] also includes a factor $w^2 < 1$ representing the prior expectation that lagged values of $y_j$ for $j \neq 1$ are less likely to be of help in forecasting $y_1$ than would be the lagged values of $y_1$ itself; hence, a tighter prior is used to set coefficients on $y_j$ to zero.

Finally, let $g$ describe the variance of the prior distribution for the constant term:

$$c_{1,1} \sim N(0, g \cdot \hat{\tau}_1^2).$$

To summarize, the matrix $\mathbf{P}_{1|0}$ is specified to be

$$\mathbf{P}_{1|0} = \begin{bmatrix} g \cdot \hat{\tau}_1^2 & \mathbf{0}' \\ \mathbf{0} & (\mathbf{B} \otimes \mathbf{C}) \end{bmatrix}, \qquad [13.8.17]$$

where

$$\underset{(p \times p)}{\mathbf{B}} = \begin{bmatrix} \gamma^2 & 0 & 0 & \cdots & 0 \\ 0 & \gamma^2/2 & 0 & \cdots & 0 \\ 0 & 0 & \gamma^2/3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \gamma^2/p \end{bmatrix}$$

$$\underset{(n \times n)}{\mathbf{C}} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & w^2\hat{\tau}_1^2/\hat{\tau}_2^2 & 0 & \cdots & 0 \\ 0 & 0 & w^2\hat{\tau}_1^2/\hat{\tau}_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & w^2\hat{\tau}_1^2/\hat{\tau}_n^2 \end{bmatrix}.$$

For typical economic time series, Doan, Litterman, and Sims recommended using $\gamma^2 = 0.07$, $w^2 = 1/74$, and $g = 630$. This last value ensures that very little weight is given to the prior expectation that the constant term is zero.

Each of the coefficients in the $VAR$ is then presumed to evolve over time according to a first-order autoregression:

$$\boldsymbol{\beta}_{t+1} = \pi_8 \cdot \boldsymbol{\beta}_t + (1 - \pi_8) \cdot \overline{\boldsymbol{\beta}} + \mathbf{v}_{t+1}. \qquad [13.8.18]$$

Thus, the same scalar $\pi_8$ is used to describe a univariate $AR(1)$ process for each element of $\boldsymbol{\beta}_t$; Doan, Litterman, and Sims recommended a value of $\pi_8 = 0.999$. The disturbance $\mathbf{v}_t$ is assumed to have a diagonal variance-covariance matrix:

$$E(\mathbf{v}_t\mathbf{v}_t') = \mathbf{Q}. \qquad [13.8.19]$$

For all coefficients except the constant term, the variance of the $i$th element of $\mathbf{v}_t$ was assumed to be proportional to the corresponding element of $\mathbf{P}_{1|0}$. Thus, for $i = 2, 3, \ldots, k$, the row $i$, column $i$ element of $\mathbf{Q}$ is taken to be $\pi_7$ times the row $i$, column $i$ element of $\mathbf{P}_{1|0}$. The (1, 1) element of $\mathbf{Q}$ is taken to be $\pi_7$ times the (2, 2) element of $\mathbf{P}_{1|0}$. This adjustment is used because the (1, 1) element of $\mathbf{P}_{1|0}$ represents an effectively infinite variance corresponding to prior ignorance about



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

the value for the constant term. Doan, Litterman, and Sims recommended $\pi_7 = 10^{-7}$ as a suitable value for the constant of proportionality.

Equation [13.8.18] can be viewed as a state equation of the form

$$\boldsymbol{\xi}_{t+1} = \mathbf{F}\boldsymbol{\xi}_t + \mathbf{v}_{t+1}, \qquad [13.8.20]$$

where the state vector is given by $\boldsymbol{\xi}_t = (\boldsymbol{\beta}_t - \overline{\boldsymbol{\beta}})$ and $\mathbf{F} = \pi_8 \cdot \mathbf{I}_k$. The observation equation is

$$y_{1t} = \mathbf{x}_t'\overline{\boldsymbol{\beta}} + \mathbf{x}_t'\boldsymbol{\xi}_t + w_{1t}. \qquad [13.8.21]$$

The one parameter yet to be specified is the variance of $w_{1t}$, the residual in the *VAR*. Doan, Litterman, and Sims suggested taking this to be 0.9 times $\hat{\tau}_1^2$.

Thus, the sequence of estimated state vectors $\{\hat{\boldsymbol{\xi}}_{t|t}\}_{t=1}^T$ is found by iterating on [13.8.6] through [13.8.9] starting from $\hat{\boldsymbol{\xi}}_{1|0} = \mathbf{0}$ and $\mathbf{P}_{1|0}$ given by [13.8.17], with $\mathbf{F}(\mathbf{x}_t) = \pi_8 \cdot \mathbf{I}_k$, $\mathbf{Q}(\mathbf{x}_t) = \pi_7 \cdot \mathbf{P}_{1|0}$, $\mathbf{a}(\mathbf{x}_t) = \mathbf{x}_t'\overline{\boldsymbol{\beta}}$ with $\overline{\boldsymbol{\beta}}$ given by [13.8.15], $\mathbf{H}(\mathbf{x}_t) = \mathbf{x}_t$, and $\mathbf{R}(\mathbf{x}_t) = 0.9 \cdot \hat{\tau}_1^2$. The estimated coefficient vector is then $\hat{\boldsymbol{\beta}}_{t|t} = \overline{\boldsymbol{\beta}} + \hat{\boldsymbol{\xi}}_{t|t}$. Optimal one-period-ahead forecasts are given by $\hat{y}_{1,t+1|t} = \mathbf{x}_t'\hat{\boldsymbol{\beta}}_{t|t}$.

Optimal $s$-period-ahead forecasts are difficult to calculate. However, Doan, Litterman, and Sims suggested a simple approximation. The approximation takes the optimal one-period-ahead forecasts for each of the $n$ variables in the *VAR*, $\hat{\mathbf{y}}_{t+1|t}$, and then treats these forecasts as if they were actual observations on $\mathbf{y}_{t+1}$. Then $E(\mathbf{y}_{t+2}|\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1)$ is approximated by $E(\mathbf{y}_{t+2}|\mathbf{y}_{t+1}, \mathbf{y}_t, \ldots, \mathbf{y}_1)$ evaluated at $\mathbf{y}_{t+1} = E(\mathbf{y}_{t+1}|\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1)$. The law of iterated expectations does not apply here, since $E(\mathbf{y}_{t+2}|\mathbf{y}_{t+1}, \mathbf{y}_t, \ldots, \mathbf{y}_1)$ is a nonlinear function of $\mathbf{y}_{t+1}$. However, Doan, Litterman, and Sims argued that this simple approach gives a good approximation to the optimal forecast.

## APPENDIX 13.A. *Proofs of Chapter 13 Propositions*

■ **Proof of Proposition 13.1.**[7] Recall that $\mathbf{P}_{t+1|t}$ has the interpretation as the *MSE* of the linear projection of $\boldsymbol{\xi}_{t+1}$ on $\mathcal{Y}_t \equiv (\mathbf{y}_t', \mathbf{y}_{t-1}', \ldots, \mathbf{y}_1', \mathbf{x}_t', \mathbf{x}_{t-1}', \ldots, \mathbf{x}_1')'$,

$$\mathbf{P}_{t+1|t} = MSE[\hat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t)]. \qquad [13.A.1]$$

Suppose for some reason we instead tried to forecast $\boldsymbol{\xi}_{t+1}$ using only observations $2, 3, \ldots, t$, discarding the observation for date $t = 1$. Thus, define $\mathcal{Y}_t^* \equiv (\mathbf{y}_t', \mathbf{y}_{t-1}', \ldots, \mathbf{y}_2', \mathbf{x}_t', \mathbf{x}_{t-1}', \ldots, \mathbf{x}_2')'$, and let

$$\mathbf{P}_{t+1|t}^* \equiv MSE[\hat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t^*)]. \qquad [13.A.2]$$

Then clearly, [13.A.2] cannot be smaller than [13.A.1], since the linear projection $\hat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t)$ made optimal use of $\mathcal{Y}_t^*$ along with the added information in $(\mathbf{y}_1', \mathbf{x}_1')'$. Specifically, if $\mathbf{h}$ is any $(r \times 1)$ vector, the linear projection of $z_{t+1} \equiv \mathbf{h}'\boldsymbol{\xi}_{t+1}$ on $\mathcal{Y}_t$ has *MSE* given by

$$\begin{aligned} E[z_{t+1} - \hat{E}(z_{t+1}|\mathcal{Y}_t)]^2 &= E[\mathbf{h}'\boldsymbol{\xi}_{t+1} - \mathbf{h}' \cdot \hat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t)]^2 \\ &= \mathbf{h}' \cdot E\{[\boldsymbol{\xi}_{t+1} - \hat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t)][\boldsymbol{\xi}_{t+1} - \hat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t)]'\} \cdot \mathbf{h} \\ &= \mathbf{h}'\mathbf{P}_{t+1|t}\mathbf{h}. \end{aligned}$$

Similarly, the linear projection of $z_{t+1}$ on $\mathcal{Y}_t^*$ has *MSE* $\mathbf{h}'\mathbf{P}_{t+1|t}^*\mathbf{h}$, with

$$\mathbf{h}'\mathbf{P}_{t+1|t}\mathbf{h} \le \mathbf{h}'\mathbf{P}_{t+1|t}^*\mathbf{h}. \qquad [13.A.3]$$

But for a system of the form of [13.2.1] and [13.2.2] with eigenvalues of $\mathbf{F}$ inside the unit circle and time-invariant coefficients, it will be the case that

$$\begin{aligned} &MSE[\hat{E}(\boldsymbol{\xi}_{t+1}|\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_2, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_2)] \\ &\qquad = MSE[\hat{E}(\boldsymbol{\xi}_t|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_1, \mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots, \mathbf{x}_1)], \end{aligned}$$

that is,

$$\mathbf{P}_{t+1|t}^* = \mathbf{P}_{t|t-1}.$$

Hence, [13.A.3] implies that

$$\mathbf{h}'\mathbf{P}_{t+1|t}\mathbf{h} \le \mathbf{h}'\mathbf{P}_{t|t-1}\mathbf{h}$$

[7]The arguments in the proofs of Propositions 13.1 and 13.2 are adapted from Anderson and Moore (1979, pp. 76–82).



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

for any $(r \times 1)$ vector $\mathbf{h}$. The sequence of scalars $\{\mathbf{h}'\mathbf{P}_{t+1|t}\mathbf{h}\}_{t=1}^{T}$ is thus monotonically nonincreasing and is bounded below by zero. It therefore converges to some fixed nonnegative value. Since this is true for any $(r \times 1)$ vector $\mathbf{h}$ and since the matrix $\mathbf{P}_{t+1|t}$ is symmetric, it follows that the sequence $\{\mathbf{P}_{t+1|t}\}_{t=1}^{T}$ converges to some fixed positive semidefinite matrix $\mathbf{P}$.

To verify the claims about the eigenvalues of the matrix $(\mathbf{F} - \mathbf{KH}')$, note that if $\mathbf{P}$ is a fixed point of [13.5.3], then it must also be a fixed point of the equivalent difference equation [13.2.28]:

$$\mathbf{P} = (\mathbf{F} - \mathbf{KH}')\mathbf{P}(\mathbf{F} - \mathbf{KH}')' + \mathbf{KRK}' + \mathbf{Q}. \qquad [13.A.4]$$

Let $\mathbf{x}$ denote an eigenvector of $(\mathbf{F} - \mathbf{KH}')'$ and $\lambda$ its eigenvalue:

$$(\mathbf{F} - \mathbf{KH}')'\mathbf{x} = \lambda\mathbf{x}. \qquad [13.A.5]$$

Although $\mathbf{F}$, $\mathbf{K}$, and $\mathbf{H}$ are all real, the eigenvalue $\lambda$ and eigenvector $\mathbf{x}$ could be complex. If $\mathbf{x}''$ denotes the conjugate transpose of $\mathbf{x}$, then

$$\begin{aligned}\mathbf{x}''(\mathbf{F} - \mathbf{KH}')\mathbf{P}(\mathbf{F} - \mathbf{KH}')'\mathbf{x} &= [(\mathbf{F} - \mathbf{KH}')'\mathbf{x}]''\mathbf{P}[(\mathbf{F} - \mathbf{KH}')'\mathbf{x}] \\ &= [\lambda\mathbf{x}]''\mathbf{P}[\lambda\mathbf{x}] \\ &= |\lambda|^2\mathbf{x}''\mathbf{P}\mathbf{x}.\end{aligned}$$

Thus, if [13.A.4] is premultiplied by $\mathbf{x}''$ and postmultiplied by $\mathbf{x}$, the result is

$$\mathbf{x}''\mathbf{P}\mathbf{x} = |\lambda|^2\mathbf{x}''\mathbf{P}\mathbf{x} + \mathbf{x}''(\mathbf{KRK}' + \mathbf{Q})\mathbf{x},$$

or

$$(1 - |\lambda|^2)\mathbf{x}''\mathbf{P}\mathbf{x} = \mathbf{x}''(\mathbf{KRK}' + \mathbf{Q})\mathbf{x}. \qquad [13.A.6]$$

Now, $(\mathbf{KRK}' + \mathbf{Q})$ is positive semidefinite, so the right side of [13.A.6] is nonnegative. Likewise, $\mathbf{P}$ is positive semidefinite, so $\mathbf{x}''\mathbf{P}\mathbf{x}$ is nonnegative. Expression [13.A.6] then requires $|\lambda| \leq 1$, meaning that any eigenvalue of $(\mathbf{F} - \mathbf{KH}')$ must be on or inside the unit circle, as claimed. ■

■ **Proof of Proposition 13.2.** First we establish the final claim of the proposition, concerning the eigenvalues of $(\mathbf{F} - \mathbf{KH}')$. Let $\mathbf{P}$ denote any positive semidefinite matrix that satisfies [13.A.4], and let $\mathbf{K}$ be given by [13.5.4]. Notice that if $\mathbf{Q}$ is positive definite, then the right side of [13.A.6] is strictly positive for any nonzero $\mathbf{x}$, meaning from the left side of [13.A.6] that any eigenvalue $\lambda$ of $(\mathbf{F} - \mathbf{KH}')$ is strictly inside the unit circle. Alternatively, if $\mathbf{R}$ is positive definite, then the only way that the right side of [13.A.6] could fail to be strictly positive would be if $\mathbf{K}'\mathbf{x} = \mathbf{0}$. But from [13.A.5], this would imply that $\mathbf{F}'\mathbf{x} = \lambda\mathbf{x}$, that is, that $\mathbf{x}$ is an eigenvector and $\lambda$ is an eigenvalue of $\mathbf{F}'$. This, in turn, means that $\lambda$ is an eigenvalue of $\mathbf{F}$, in which case $|\lambda| < 1$, by the assumption of stability of $\mathbf{F}$. Thus, there cannot be an eigenvector $\mathbf{x}$ of $(\mathbf{F} - \mathbf{KH}')'$ associated with an eigenvalue whose modulus is greater than or equal to unity if $\mathbf{R}$ is positive definite.

Turning next to the rest of Proposition 13.2, let $\{\mathbf{P}_{t+1|t}\}$ denote the sequence that results from iterating on [13.5.1] starting from an arbitrary positive semidefinite initial value $\mathbf{P}_{1|0}$. We will show that there exist two other sequences of matrices, to be denoted $\{\underset{\sim}{\mathbf{P}}_{t+1|t}\}$ and $\{\tilde{\mathbf{P}}_{t+1|t}\}$, such that

$$\underset{\sim}{\mathbf{P}}_{t+1|t} \leq \mathbf{P}_{t+1|t} \leq \tilde{\mathbf{P}}_{t+1|t} \qquad \text{for all } t,$$

where

$$\lim_{t\to\infty} \underset{\sim}{\mathbf{P}}_{t+1|t} = \lim \tilde{\mathbf{P}}_{t+1|t} = \mathbf{P}$$

and where $\mathbf{P}$ does not depend on $\mathbf{P}_{1|0}$. The conclusion will then be that $\{\mathbf{P}_{t+1|t}\}$ converges to $\mathbf{P}$ regardless of the value of $\mathbf{P}_{1|0}$.

To construct the matrix $\underset{\sim}{\mathbf{P}}_{t+1|t}$ that is to be offered as a lower bound on $\mathbf{P}_{t+1|t}$, consider the sequence $\{\underset{\sim}{\mathbf{P}}_{t+1|t}\}$ that results from iterating on [13.5.1] starting from the initial value $\underset{\sim}{\mathbf{P}}_{1|0} = \mathbf{0}$. This would correspond to treating the intial state $\boldsymbol{\xi}_1$ as if known with certainty:

$$\underset{\sim}{\mathbf{P}}_{t+1|t} = MSE[\hat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t, \boldsymbol{\xi}_1)]. \qquad [13.A.7]$$

Note that $\mathbf{y}_1$ and $\mathbf{x}_1$ are correlated with $\boldsymbol{\xi}_{t+1}$ for $t = 1, 2, \ldots$ only through the value of $\boldsymbol{\xi}_1$, which means that we could equally well write

$$\underset{\sim}{\mathbf{P}}_{t+1|t} = MSE[\hat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t^*, \boldsymbol{\xi}_1)], \qquad [13.A.8]$$

where $\mathcal{Y}_t^* \equiv (\mathbf{y}_t', \mathbf{y}_{t-1}', \ldots, \mathbf{y}_2', \mathbf{x}_t', \mathbf{x}_{t-1}', \ldots, \mathbf{x}_2')'$. Added knowledge about $\boldsymbol{\xi}_2$ could not hurt the forecast:

$$MSE[\hat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t^*, \boldsymbol{\xi}_2, \boldsymbol{\xi}_1)] \leq MSE[\hat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t^*, \boldsymbol{\xi}_1)], \qquad [13.A.9]$$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

and indeed, $\xi_1$ is correlated with $\xi_{t+1}$ for $t = 2, 3, \ldots$ only through the value of $\xi_2$:

$$MSE[\hat{E}(\xi_{t+1}|\mathcal{Y}_t^*, \xi_2, \xi_1)] = MSE[\hat{E}(\xi_{t+1}|\mathcal{Y}_t^*, \xi_2)]. \qquad [13.A.10]$$

Because coefficients are time-invariant,

$$MSE[\hat{E}(\xi_{t+1}|\mathcal{Y}_t^*, \xi_2)] = MSE[\hat{E}(\xi_t|\mathcal{Y}_{t-1}, \xi_1)] = \underset{\sim}{\mathbf{P}}_{t|t-1}. \qquad [13.A.11]$$

Thus, [13.A.10] and [13.A.11] establish that the left side of [13.A.9] is the same as $\underset{\sim}{\mathbf{P}}_{t|t-1}$, while from [13.A.8] the right side of [13.A.9] is the same as $\underset{\sim}{\mathbf{P}}_{t+1|t}$. Thus, [13.A.9] states that

$$\underset{\sim}{\mathbf{P}}_{t|t-1} \leq \underset{\sim}{\mathbf{P}}_{t+1|t},$$

so that $\{\underset{\sim}{\mathbf{P}}_{t+1|t}\}$ is a monotonically nondecreasing sequence; the farther in the past is the perfect information about $\xi_1$, the less value it is for forecasting $\xi_{t+1}$.

Furthermore, a forecast based on perfect information about $\xi_1$, for which $\underset{\sim}{\mathbf{P}}_{t+1|t}$ gives the *MSE*, must be better than one based on imperfect information about $\xi_1$, for which $\mathbf{P}_{t+1|t}$ gives the *MSE*:

$$\underset{\sim}{\mathbf{P}}_{t+1|t} \leq \mathbf{P}_{t+1|t} \qquad \text{for all } t.$$

Thus, $\underset{\sim}{\mathbf{P}}_{t+1|t}$ puts a lower bound on $\mathbf{P}_{t+1|t}$, as claimed. Moreover, since the sequence $\{\underset{\sim}{\mathbf{P}}_{t+1|t}\}$ is monotonically nondecreasing and bounded from above, it converges to a fixed value $\mathbf{P}$ satisfying [13.5.3] and [13.A.4].

To construct an upper bound on $\mathbf{P}_{t+1|t}$, consider a sequence $\{\tilde{\mathbf{P}}_{t+1|t}\}$ that begins with $\tilde{\mathbf{P}}_{1|0} = \mathbf{P}_{1|0}$, the same starting value that was used to construct $\{\mathbf{P}_{t+1|t}\}$. Recall that $\mathbf{P}_{t+1|t}$ gave the *MSE* of the sequence $\hat{\xi}_{t+1|t}$ described in equation [13.2.20]:

$$\hat{\xi}_{t+1|t} = \mathbf{F}\hat{\xi}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\xi}_{t|t-1}).$$

Imagine instead using a sequence of suboptimal inferences $\{\tilde{\xi}_{t+1|t}\}$ defined by the recursion

$$\tilde{\xi}_{t+1|t} = \mathbf{F}\tilde{\xi}_{t|t-1} + \mathbf{K}(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\tilde{\xi}_{t|t-1}), \qquad [13.A.12]$$

where $\mathbf{K}$ is the value calculated from [13.5.4] in which the steady-state value for $\mathbf{P}$ is taken to be the limit of the sequence $\{\underset{\sim}{\mathbf{P}}_{t+1|t}\}$. Note that the magnitude $\tilde{\xi}_{t+1|t}$ so defined is a linear function of $\mathcal{Y}_t$ and so must have a greater *MSE* than the optimal inference $\hat{\xi}_{t+1|t}$:

$$\tilde{\mathbf{P}}_{t+1|t} \equiv E[(\xi_{t+1} - \tilde{\xi}_{t+1|t})(\xi_{t+1} - \tilde{\xi}_{t+1|t})'] \geq \mathbf{P}_{t+1|t}.$$

Thus, we have established that

$$\underset{\sim}{\mathbf{P}}_{t+1|t} \leq \mathbf{P}_{t+1|t} \leq \tilde{\mathbf{P}}_{t+1|t}$$

and that $\underset{\sim}{\mathbf{P}}_{t+1|t} \to \mathbf{P}$. The proof will be complete if we can further show that $\tilde{\mathbf{P}}_{t+1|t} \to \mathbf{P}$. Parallel calculations to those leading to [13.2.28] reveal that

$$\tilde{\mathbf{P}}_{t+1|t} = (\mathbf{F} - \mathbf{KH}')\tilde{\mathbf{P}}_{t|t-1}(\mathbf{F} - \mathbf{KH}')' + \mathbf{KRK}' + \mathbf{Q}. \qquad [13.A.13]$$

Apply the vec operator to both sides of [13.A.13] and recall Proposition 10.4:

$$\text{vec}(\tilde{\mathbf{P}}_{t+1|t}) = \mathcal{B}\,\text{vec}(\tilde{\mathbf{P}}_{t|t-1}) + c = [\mathbf{I}_{r^2} + \mathcal{B} + \mathcal{B}^2 + \cdots + \mathcal{B}^{t-1}]c + \mathcal{B}^t\,\text{vec}(\tilde{\mathbf{P}}_{1|0}),$$

where

$$\mathcal{B} \equiv (\mathbf{F} - \mathbf{KH}') \otimes (\mathbf{F} - \mathbf{KH}')$$
$$c \equiv \text{vec}(\mathbf{KRK}' + \mathbf{Q}).$$

Recall further that since either $\mathbf{R}$ or $\mathbf{Q}$ is positive definite, the value of $\mathbf{K}$ has the property that all eigenvalues of $(\mathbf{F} - \mathbf{KH}')$ are strictly less than unity in modulus. Thus, all eigenvalues of $\mathcal{B}$ are also strictly less than unity in modulus, implying that

$$\lim_{t\to\infty} \text{vec}(\tilde{\mathbf{P}}_{t+1|t}) = (\mathbf{I}_{r^2} - \mathcal{B})^{-1}c,$$

the same value regardless of the starting value for $\tilde{\mathbf{P}}_{1|0}$. In particular, if the iteration on [13.A.13] is started with $\tilde{\mathbf{P}}_{1|0} = \mathbf{P}$, this being a fixed point of the iteration, the result would be $\tilde{\mathbf{P}}_{t+1|t} = \mathbf{P}$ for all $t$. Thus,

$$\lim_{t\to\infty} \tilde{\mathbf{P}}_{t+1|t} = \mathbf{P},$$

regardless of the value of $\tilde{\mathbf{P}}_{1|0} = \mathbf{P}_{1|0}$ from which the iteration for $\tilde{\mathbf{P}}_{t+1|t}$ is started. ■



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

■ **Proof of Proposition 13.3.** Observe that

$$\begin{aligned}
&\{\mathbf{I}_n + \mathbf{H}'(\mathbf{I}_r - \mathbf{F}z)^{-1}\mathbf{K}z\}\{\mathbf{I}_n - \mathbf{H}'[\mathbf{I}_r - (\mathbf{F} - \mathbf{KH}')z]^{-1}\mathbf{K}z\} \\
&\quad = \mathbf{I}_n - \mathbf{H}'[\mathbf{I}_r - (\mathbf{F} - \mathbf{KH}')z]^{-1}\mathbf{K}z + \mathbf{H}'(\mathbf{I}_r - \mathbf{F}z)^{-1}\mathbf{K}z \\
&\qquad - \{\mathbf{H}'(\mathbf{I}_r - \mathbf{F}z)^{-1}\mathbf{K}z\}\{\mathbf{H}'[\mathbf{I}_r - (\mathbf{F} - \mathbf{KH}')z]^{-1}\mathbf{K}z\} \qquad [13.A.14] \\
&\quad = \mathbf{I}_n + \mathbf{H}'\Big\{-[\mathbf{I}_r - (\mathbf{F} - \mathbf{KH}')z]^{-1} + [\mathbf{I}_r - \mathbf{F}z]^{-1} \\
&\qquad - [\mathbf{I}_r - \mathbf{F}z]^{-1}\mathbf{KH}'z[\mathbf{I}_r - (\mathbf{F} - \mathbf{KH}')z]^{-1}\Big\}\mathbf{K}z.
\end{aligned}$$

The term in curly braces in the last line of [13.A.14] is indeed zero, as may be verified by taking the identity

$$-[\mathbf{I}_r - \mathbf{F}z] + [\mathbf{I}_r - (\mathbf{F} - \mathbf{KH}')z] - \mathbf{KH}'z = \mathbf{0}$$

and premultiplying by $[\mathbf{I}_r - \mathbf{F}z]^{-1}$ and postmultiplying by $[\mathbf{I}_r - (\mathbf{F} - \mathbf{KH}')z]^{-1}$:

$$\begin{aligned}
&-[\mathbf{I}_r - (\mathbf{F} - \mathbf{KH}')z]^{-1} + [\mathbf{I}_r - \mathbf{F}z]^{-1} \\
&\quad - [\mathbf{I}_r - \mathbf{F}z]^{-1}\mathbf{KH}'z[\mathbf{I}_r - (\mathbf{F} - \mathbf{KH}')z]^{-1} = \mathbf{0}. \quad ■ \qquad [13.A.15]
\end{aligned}$$

■ **Proof of Proposition 13.4.** Notice that

$$\begin{aligned}
&\{\mathbf{I}_n + \mathbf{H}'(\mathbf{I}_r - \mathbf{F}z)^{-1}\mathbf{K}z\}\{\mathbf{H}'\mathbf{PH} + \mathbf{R}\}\{\mathbf{I}_n + \mathbf{K}'(\mathbf{I}_r - \mathbf{F}'z^{-1})^{-1}\mathbf{H}z^{-1}\} \\
&\quad = \{\mathbf{H}'\mathbf{PH} + \mathbf{R}\} + \mathbf{H}'(\mathbf{I}_r - \mathbf{F}z)^{-1}\mathbf{K}\{\mathbf{H}'\mathbf{PH} + \mathbf{R}\}z \qquad [13.A.16] \\
&\qquad + \{\mathbf{H}'\mathbf{PH} + \mathbf{R}\}\mathbf{K}'(\mathbf{I}_r - \mathbf{F}'z^{-1})^{-1}\mathbf{H}z^{-1} \\
&\qquad + \mathbf{H}'(\mathbf{I}_r - \mathbf{F}z)^{-1}\mathbf{K}\{\mathbf{H}'\mathbf{PH} + \mathbf{R}\}\mathbf{K}'(\mathbf{I}_r - \mathbf{F}'z^{-1})^{-1}\mathbf{H}.
\end{aligned}$$

Now [13.5.4] requires that

$$\mathbf{K}\{\mathbf{H}'\mathbf{PH} + \mathbf{R}\} = \mathbf{FPH} \qquad [13.A.17]$$

$$\{\mathbf{H}'\mathbf{PH} + \mathbf{R}\}\mathbf{K}' = \mathbf{H}'\mathbf{PF}' \qquad [13.A.18]$$

$$\begin{aligned}
\mathbf{K}\{\mathbf{H}'\mathbf{PH} + \mathbf{R}\}\mathbf{K}' &= \mathbf{FPH}\{\mathbf{H}'\mathbf{PH} + \mathbf{R}\}^{-1}\mathbf{H}'\mathbf{PF}' \\
&= \mathbf{FPF}' - \mathbf{P} + \mathbf{Q}, \qquad [13.A.19]
\end{aligned}$$

with the last equality following from [13.5.3]. Substituting [13.A.17] through [13.A.19] into [13.A.16] results in

$$\begin{aligned}
&\{\mathbf{I}_n + \mathbf{H}'(\mathbf{I}_r - \mathbf{F}z)^{-1}\mathbf{K}z\}\{\mathbf{H}'\mathbf{PH} + \mathbf{R}\}\{\mathbf{I}_n + \mathbf{K}'(\mathbf{I}_r - \mathbf{F}'z^{-1})^{-1}\mathbf{H}z^{-1}\} \\
&\quad = \{\mathbf{H}'\mathbf{PH} + \mathbf{R}\} + \mathbf{H}'(\mathbf{I}_r - \mathbf{F}z)^{-1}\mathbf{FPH}z + \mathbf{H}'\mathbf{PF}'(\mathbf{I}_r - \mathbf{F}'z^{-1})^{-1}\mathbf{H}z^{-1} \\
&\qquad + \mathbf{H}'(\mathbf{I}_r - \mathbf{F}z)^{-1}\{\mathbf{FPF}' - \mathbf{P} + \mathbf{Q}\}(\mathbf{I}_r - \mathbf{F}'z^{-1})^{-1}\mathbf{H} \qquad [13.A.20] \\
&\quad = \mathbf{R} + \mathbf{H}'\Big\{\mathbf{P} + (\mathbf{I}_r - \mathbf{F}z)^{-1}\mathbf{FP}z + \mathbf{PF}'(\mathbf{I}_r - \mathbf{F}'z^{-1})^{-1}z^{-1} \\
&\qquad + (\mathbf{I}_r - \mathbf{F}z)^{-1}\{\mathbf{FPF}' - \mathbf{P} + \mathbf{Q}\}(\mathbf{I}_r - \mathbf{F}'z^{-1})^{-1}\Big\}\mathbf{H}.
\end{aligned}$$

The result in Proposition 13.4 follows provided that

$$\begin{aligned}
&\mathbf{P} + (\mathbf{I}_r - \mathbf{F}z)^{-1}\mathbf{FP}z + \mathbf{PF}'(\mathbf{I}_r - \mathbf{F}'z^{-1})^{-1}z^{-1} \qquad [13.A.21] \\
&\qquad + (\mathbf{I}_r - \mathbf{F}z)^{-1}\{\mathbf{FPF}' - \mathbf{P}\}(\mathbf{I}_r - \mathbf{F}'z^{-1})^{-1} = \mathbf{0}.
\end{aligned}$$

To verify that [13.A.21] is true, start from the identity

$$\begin{aligned}
&(\mathbf{I}_r - \mathbf{F}z)\mathbf{P}(\mathbf{I}_r - \mathbf{F}'z^{-1}) + \mathbf{FP}z(\mathbf{I}_r - \mathbf{F}'z^{-1}) \qquad [13.A.22] \\
&\qquad + (\mathbf{I}_r - \mathbf{F}z)\mathbf{PF}'z^{-1} + \mathbf{FPF}' - \mathbf{P} = \mathbf{0}.
\end{aligned}$$

Premultiplying [13.A.22] by $(\mathbf{I}_r - \mathbf{F}z)^{-1}$ and postmultiplying by $(\mathbf{I}_r - \mathbf{F}'z^{-1})^{-1}$ confirms [13.A.21]. Substituting [13.A.21] into [13.A.20] produces the claim in Proposition 13.4. ■

## *Chapter 13 Exercises*

13.1. Suppose we have a noisy indicator $y$ on an underlying unobserved random variable $\xi$:

$$y = \xi + \varepsilon.$$



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

Suppose moreover that the measurement error ($\varepsilon$) is $N(0, \tau^2)$, while the true value $\xi$ is $N(\mu, \sigma^2)$, with $\varepsilon$ uncorrelated with $\xi$. Show that the optimal estimate of $\xi$ is given by

$$E(\xi \mid y) = \mu + \frac{\sigma^2}{\tau^2 + \sigma^2}(y - \mu)$$

with associated *MSE*

$$E[\xi - E(\xi \mid y)]^2 = \frac{\sigma^2\tau^2}{\tau^2 + \sigma^2}.$$

Discuss the intuition for these results as $\tau^2 \to \infty$ and $\tau^2 \to 0$.

13.2. Deduce the state-space representation for an $AR(p)$ model in [13.1.14] and [13.1.15] and the state-space representation for an $MA(1)$ model given in [13.1.17] and [13.1.18] as special cases of that for the $ARMA(r, r - 1)$ model of [13.1.22] and [13.1.23].

13.3. Is the following a valid state-space representation of an $MA(1)$ process?

***State Equation*:**

$$\begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_t \end{bmatrix}$$

***Observation Equation*:**

$$y_t - \mu = [1 \quad \theta] \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix}$$

13.4. Derive equation [13.4.5] as a special case of [13.4.1] and [13.4.2] for the model specified in [13.4.3] and [13.4.4] by analysis of the Kalman filter recursion for this case.

13.5. Consider a particular $MA(1)$ representation of the form of [13.3.1] through [13.3.12] parameterized by $(\theta, \sigma^2)$ with $|\theta| < 1$. The noninvertible representation for the same process is parameterized by $(\tilde{\theta}, \tilde{\sigma}^2)$ with $\tilde{\theta} = 1/\theta$ and $\tilde{\sigma}^2 = \theta^2\sigma^2$. The forecast generated by the Kalman filter using the noninvertible representation satisfies

$$\hat{y}_{t+1|t} = \mathbf{A}'\mathbf{x}_{t+1} + \tilde{\mathbf{H}}'\hat{\boldsymbol{\xi}}_{t+1|t} = \mu + \tilde{\theta}\tilde{\varepsilon}_{t|t},$$

where $\tilde{\varepsilon}_{t|t} = \{\tilde{\sigma}^2/[\tilde{\sigma}^2 + \tilde{\theta}^2\tilde{p}_t]\}\cdot\{y_t - \mu - \tilde{\theta}\tilde{\varepsilon}_{t-1|t-1}\}$. The *MSE* of this forecast is

$$E(y_{t+1} - \hat{y}_{t+1|t})^2 = \tilde{\mathbf{H}}'\tilde{\mathbf{P}}_{t+1|t}\tilde{\mathbf{H}} + \tilde{\mathbf{R}} = \tilde{\sigma}^2 + \tilde{\theta}^2\tilde{p}_{t+1},$$

where $\tilde{p}_{t+1} = (\tilde{\sigma}^2\tilde{\theta}^{2t})/(1 + \tilde{\theta}^2 + \tilde{\theta}^4 + \cdots + \tilde{\theta}^{2t})$. Show that this forecast and *MSE* are identical to those for the process as parameterized using the invertible representation $(\theta, \sigma^2)$. Deduce that the likelihood function given by [13.4.1] and [13.4.2] takes on the same value at $(\theta, \sigma^2)$ as it does at $(\tilde{\theta}, \tilde{\sigma}^2)$.

13.6. Show that $\varepsilon_t$ in equation [13.5.22] is fundamental for $y_t$. What principle of the Kalman filter ensures that this will be the case? Show that the first autocovariance of the implied $MA(1)$ error process is given by

$$-(\phi - K)E(\varepsilon_t^2) = -\phi\sigma_W^2$$

while the variance is

$$[1 + (\phi - K)^2]E(\varepsilon_t^2) = (1 + \phi^2)\sigma_W^2 + \sigma_V^2.$$

Derive these expressions independently, using the approach to sums of *ARMA* processes in Section 4.7.

13.7. Consider again the invertible $MA(1)$ of equations [13.3.1] to [13.3.12]. We found that the steady-state value of $\mathbf{P}_{t|t-1}$ is given by

$$\mathbf{P} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

From this, deduce that the steady-state value of $\mathbf{P}_{t|t+s} = \mathbf{0}$ for $s = 0, 1, 2, \ldots$. Give the intuition for this result.

## *Chapter 13 References*


Anderson, Brian D. O., and John B. Moore. 1979. *Optimal Filtering*. Englewood Cliffs, N.J.: Prentice-Hall.

Burmeister, Edwin, and Kent D. Wall. 1982. "Kalman Filtering Estimation of Unobserved Rational Expectations with an Application to the German Hyperinflation." *Journal of Econometrics* 20:255–84.





This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms

———, ———, and James D. Hamilton. 1986. "Estimation of Unobserved Expected Monthly Inflation Using Kalman Filtering." *Journal of Business and Economic Statistics* 4:147–60.

Caines, Peter E. 1988. *Linear Stochastic Systems*. New York: Wiley.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society* Series B, 39:1–38.

Doan, Thomas, Robert B. Litterman, and Christopher A. Sims. 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews* 3:1–100.

Fama, Eugene F., and Michael R. Gibbons. 1982. "Inflation, Real Returns, and Capital Investment." *Journal of Monetary Economics* 9:297–323.

Gevers, M., and V. Wertz. 1984. "Uniquely Identifiable State-Space and ARMA Parameterizations for Multivariable Linear Systems." *Automatica* 20:333–47.

Ghosh, Damayanti. 1989. "Maximum Likelihood Estimation of the Dynamic Shock-Error Model." *Journal of Econometrics* 41:121–43.

Hamilton, James D. 1985. "Uncovering Financial Market Expectations of Inflation." *Journal of Political Economy* 93:1224–41.

———. 1986. "A Standard Error for the Estimated State Vector of a State-Space Model." *Journal of Econometrics* 33:387–97.

Hannan, E. J. 1971. "The Identification Problem for Multiple Equation Systems with Moving Average Errors." *Econometrica* 39:751–65.

Harvey, Andrew, and G. D. A. Phillips. 1979. "Maximum Likelihood Estimation of Regression Models with Autoregressive–Moving Average Disturbances." *Biometrika* 66:49–58.

Kalman, R. E. 1960. "A New Approach to Linear Filtering and Prediction Problems." *Journal of Basic Engineering, Transactions of the ASME* Series D, 82:35–45.

———. 1963. "New Methods in Wiener Filtering Theory," in John L. Bogdanoff and Frank Kozin, eds., *Proceedings of the First Symposium of Engineering Applications of Random Function Theory and Probability*, 270–388. New York: Wiley.

Litterman, Robert B. 1986. "Forecasting with Bayesian Vector Autoregressions—Five Years of Experience." *Journal of Business and Economic Statistics* 4:25–38.

Meinhold, Richard J., and Nozer D. Singpurwalla. 1983. "Understanding the Kalman Filter." *American Statistician* 37:123–27.

Nicholls, D. F., and A. R. Pagan. 1985. "Varying Coefficient Regression," in E. J. Hannan, P. R. Krishnaiah, and M. M. Rao, eds., *Handbook of Statistics*, Vol. 5. Amsterdam: North-Holland.

Pagan, Adrian. 1980. "Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients." *Journal of Econometrics* 13:341–63.

Rothenberg, Thomas J. 1971. "Identification in Parametric Models." *Econometrica* 39:577–91.

Shumway, R. H., and D. S. Stoffer. 1982. "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm." *Journal of Time Series Analysis* 3:253–64.

Sims, Christopher A. 1982. "Policy Analysis with Econometric Models." *Brookings Papers on Economic Activity* 1:107–52.

Stock, James H., and Mark W. Watson. 1991. "A Probability Model of the Coincident Economic Indicators," in Kajal Lahiri and Geoffrey H. Moore, eds., *Leading Economic Indicators: New Approaches and Forecasting Records*. Cambridge, England: Cambridge University Press.

Tanaka, Katsuto. 1983. "Non-Normality of the Lagrange Multiplier Statistic for Testing the Constancy of Regression Coefficients." *Econometrica* 51:1577–82.

Wall, Kent D. 1987. "Identification Theory for Varying Coefficient Regression Models." *Journal of Time Series Analysis* 8:359–71.

Watson, Mark W. 1989. "Recursive Solution Methods for Dynamic Linear Rational Expectations Models." *Journal of Econometrics* 41:65–89.

——— and Robert F. Engle. 1983. "Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC, and Varying Coefficient Regression Models." *Journal of Econometrics* 23:385–400.

White, Halbert. 1982. "Maximum Likelihood Estimation of Misspecified Models." *Econometrica* 50:1–25.



This content downloaded from 130.237.245.8 on Fri, 31 Mar 2023 11:48:42 UTC
All use subject to https://about.jstor.org/terms