State-Space Model in Linear Case

Hisashi Tanizaki
Faculty of Economics, Kobe University

Chapter 1 of *Nonlinear Filters: Estimation and Applications* (Springer-Verlag, 1996)

1 Introduction

There is a great deal of literature about applications of the Kalman filter model. Kalman (1960) and Kalman and Bucy (1961) developed this new approach to linear filtering and prediction, which has also been developed by engineers. Economic applications of the Kalman filter include a time-varying parameter model, an autoregressive-moving average process, estimation of seasonal components, prediction of final data given preliminary data and so on, which are introduced later in this paper. Thus, the Kalman filter can be applied to various models, especially, the models that include unobservable variables.

As it is well known, there are many approaches to derive the Kalman filter algorithm. Here, three derivations are discussed; the first is based on normality assumption for error terms, the second is related to the mixed estimation approach, so-called Goldberger-Theil estimator, where we do not have to impose normality assumption for error terms, and the third is interpreted as minimum mean square linear estimator. All the derivations result in the same linear recursive algorithm, although the derivation procedure is different.

Finally, in this paper, an estimation problem of unknown parameters is taken. In the case where measurement and transition equations include unknown parameters, the maximum likelihood method is usually used for estimation. The maximum likelihood estimation procedure is to maximize an innovation form of the likelihood function, which is very simple and requires no extra computation because all of the values used in the maximization procedure are obtained in the filtering algorithm.

Thus, we aim to introduce past research on the Kalman filter by way of a preface to this paper.

In general, the model represented by the following two equations is called the state-space model.

$$\text{(Measurement equation)} \qquad y_t = Z_t\alpha_t + d_t + S_t\epsilon_t, \tag{1}$$

$$\text{(Transition equation)} \qquad \alpha_t = T_t\alpha_{t-1} + c_t + R_t\eta_t, \tag{2}$$

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_t & 0 \\ 0 & Q_t \end{pmatrix} \right),$$

$$\begin{array}{lllll} y_t: g\times 1, & Z_t: g\times k, & d_t: g\times 1, & S_t: g\times g, & \epsilon_t: g\times 1, \\ \alpha_t: k\times 1 & T_t: k\times k, & c_t: k\times 1, & R_t: k\times k, & \eta_t: k\times 1, \end{array}$$

where ϵ_t is a serially uncorrelated disturbance with mean zero and covariance matrix H_t, and α_t is a $k \times 1$ vector which is not observable and is assumed to be generated by the first-order Markov process, as shown in equation (2). T_t is a transition matrix and η_t is a random vector of serially uncorrelated disturbances with mean zero and covariance matrix Q_t. And also Z_t, d_t, S_t, T_t, c_t and R_t may depend on information available at time $t-1$ if we assume normality for the errors ϵ_t and η_t (see Harvey (1989)). Equation (1) is known as the measurement equation while equation (2) is called the transition equation. The state-space model requires the following two additional assumptions:

(i) The initial vector α_0 has a mean of a_0 and a covariance matrix of Σ_0, i.e., $\mathrm{E}(\alpha_0) = a_0$ and $\mathrm{Var}(\alpha_0) = \Sigma_0$.

(ii) The disturbances ϵ_t and η_s are uncorrelated with each other for all time periods, and uncorrelated with the initial state-variable, i.e., $\mathrm{E}(\epsilon_t\eta_s') = 0$ for all t and all s, and $\mathrm{E}(\epsilon_t\alpha_0') = \mathrm{E}(\eta_t\alpha_0') = 0$ for $t = 1, \cdots, T$.

Notable points are as follows:

(1) The assumption (ii) guarantees no correlation between ϵ_t and α_t, and no correlation between η_t and α_{t-1}, i.e., $\mathrm{E}(\epsilon_t\alpha_t')=0$ and $\mathrm{E}(\eta_t\alpha_{t-1}')=0$ for all t.

(2) Z_t, d_t, S_t, T_t, c_t and R_t may depend on an unknown parameter vector, say θ. In this case, θ has to be estimated. We consider this estimation problem later in this paper.

(3) The error terms ϵ_t and η_t are usually assumed to be normal, but we do not necessarily need normality assumption, depending on the derivation procedure of the linear recursive algorithm. In the case of the derivations by mixed estimation and minimum mean square linear estimator, we do not need to assume any distribution for the errors ϵ_t and η_t. However, if we derive the standard linear recursive algorithm based on density functions, normality assumption is required. This is discussed in Section 3.

2 Applications of Kalman Filter

Some applications of the state-space model are shown below. In this section, we take the following examples to Kalman filter; Time-Varying Parameter Model (Section 2.1), Autoregressive-Moving Average Process (Section 2.2), Seasonal Component Model (Section 2.3), Prediction of Final Data Based on Preliminary Data (Section 2.4) and Estimation of Permanent Consumption (Section 2.5).

2.1 Time-Varying Parameter Model

In the case where we deal with time series data, we can usually write the regression model as follows:

$$y_t = x_t\beta + u_t,$$

for $t = 1,\cdots,T$, where y_t is a dependent variable, x_t denotes a $1\times k$ vector of the explanatory variables, a $k\times 1$ vector of unknown parameters to be estimated is given by β, and u_t is the error term distributed with mean zero and variance σ^2. There are some methods to estimate the equation above, for example, ordinary least squares, generalized least squares, instrumental variable estimation, etc. In any case, the estimated parameters are constant over time. This model is known as the fixed-parameter model. However, structural changes (for example, the first- and the second-oil shocks), specification errors, nonlinearities, proxy variables and aggregation are all the sources of parameter variation; see Sarris (1973), Belsley (1973), Belsley and Kuh (1973) and Cooley and Prescott (1976). Therefore, we need to consider the model such that the parameter is a function of time, which is called the time-varying parameter model. Using the state-space form, the model is represented as the following two equations:

$$\text{(Measurement equation)} \qquad y_t = x_t\beta_t + u_t, \tag{3}$$
$$\text{(Transition equation)} \qquad \beta_t = \Psi\beta_{t-1} + v_t, \tag{4}$$

where the movement of the parameter is generally assumed to be the first-order autoregressive process. The error term v_t, independent of u_t, is a white noise with mean zero and variance R. Here, equations (3) and (4) are referred to as the measurement equation and the transition equation, respectively. Note that it is possible to consider that β_t follows an AR(p) model for the transition equation. The time-varying parameter β_t is unobservable, which is estimated using the observed data y_t and x_t. Clearly, we have $Z_t = x_t$, $d_t = c_t = 0$, $S_t = R_t = I_k$, $\alpha_t = \beta_t$, $T_t = \Psi$ and $g = 1$ in equations (1) and (2).

As an example of the time-varying parameter model, Tanizaki (1987, 1993b) estimated a small macro-econometric model of Japan, using the Kalman filtering technique, where a consumption function, an investment function, an export function, an import function, a money demand function and so on are estimated equation-by-equation, and the movements of each coefficient are analyzed. There are numerous other papers which deal with the time-varying parameter model, for example, Cooley (1977), Cooper (1973), Cooley, Rosenberg and Wall (1977), Dziechciarz (1989), Johnston (1984), Nicholls and Pagan (1985), Laumas and Mehra (1976), Garbade (1976), Pagan (1980), Rosenberg (1977), Tanizaki (1989, 1993a) and Watanabe (1985).

2.2 Autoregressive-Moving Average Process

It is well-known that any autoregressive-moving average process (ARMA) can be written as the state-space form. See, for example, Aoki (1987), Gardner, Harvey and Phillips (1980), Hannan and Deistler (1988), Harvey (1981, 1989), Kirchen (1988) and Burridge and Wallis (1988). First of all, consider the following ARMA(p,q) process.

$$y_t = a_1 y_{t-1} + \cdots + a_p y_{t-p} + \epsilon_t + b_1 \epsilon_{t-1} + \cdots + b_q \epsilon_{t-q},$$

where ϵ_t is a white noise. The model above is rewritten as:

$$y_t = a_1 y_{t-1} + \cdots + a_m y_{t-m} + \epsilon_t + b_1 \epsilon_{t-1} + \cdots + b_{m-1} \epsilon_{t-m+1},$$

where $m = \max(p, q+1)$ and some of the coefficients $a_1, \cdots, a_m$, $b_1, \cdots, b_{m-1}$ can be zero. As it is well known, the ARMA process above is represented as:

$$\begin{aligned} &\text{(Measurement equation)} && y_t = z\alpha_t, \\ &\text{(Transition equation)} && \alpha_t = A\alpha_{t-1} + B\epsilon_t, \end{aligned} \qquad (5)$$

where z, A and B are defined as:

$$\underset{1\times m}{z = (1, 0, \cdots, 0)}, \qquad A = \underset{m \times m}{\left(\begin{array}{c|c} a_1 & \\ \vdots & I_{m-1} \\ a_{m-1} & \\ \hline a_m & 0 \end{array}\right)}, \qquad B = \underset{m \times 1}{\begin{pmatrix} 1 \\ b_1 \\ \vdots \\ b_{m-1} \end{pmatrix}},$$

Thus, the state-space form is obtained by setting $Z_t = z$, $d_t = S_t = c_t = 0$, $T_t = A$, $R_t = B$, $g = 1$ and $k = m$ in equations (1) and (2).

As an extension of this model (5), we can consider the ARMA(p,q) with a constant term, where the above transition equation is replaced as the following two equations:

$$\begin{aligned} &\text{(Measurement equation)} && y_t = z\alpha_t, \\ &\text{(Transition equation)} && \alpha_t - \overline{\alpha} = C\beta_t, \\ & && \beta_t = D\beta_{t-1} + E\eta_t. \end{aligned}$$

C represents a selection matrix consisting of zeros and ones with the characteristic: $CC' = I_m$, where I_m is the unity matrix. Or equivalently, the two equations related to the transition equation are compactly rewritten as:

$$\alpha_t - \overline{\alpha} = CDC'(\alpha_{t-1} - \overline{\alpha}) + CE\eta_t,$$

where $A = CDC'$ and $B = CE$ in the transition equation of the system (5). Kirchen (1988) concluded that this extension yields better results than the transition equation in the system (5).

2.3 Seasonal Component Model

As another application, we can take an estimation of seasonal components. A time-series consists of seasonal, cyclical, and irregular components. Each component is unobservable. The Kalman filter is applied to estimate each component separately.

The following suggestion by Pagan (1975) is essentially a combination of an econometric model for the cyclical components with the filtering and estimation of the seasonal components formulated in a state-space form. See Chow (1983).

Assume, first, that an endogenous variable y_t is the sum of cyclical, seasonal, and irregular components, as given by:

$$y_t = y_t^c + y_t^s + v_t,$$

and, second, that the cyclical component y_t^c is represented as the following model:

$$y_t^c = Ay_{t-1}^c + Cx_t + u_t,$$

where x_t is a $k \times 1$ vector of exogenous variables and u_t denotes a random disturbance. Finally, an autoregressive seasonal model is assumed for the seasonal component, i.e.,

$$y_t^s = By_{t-m}^s + w_t,$$

where w_t represents a random residual and m can be 4 for a quarterly model and 12 for a monthly model. Combining these three equations above, we can construct the following state-space form:

$$\begin{array}{lll} \text{(Measurement equation)} & & y_t = z\alpha_t + v_t, \\ \text{(Transition equation)} & & \alpha_t = M\alpha_{t-1} + Nx_t + \eta_t, \end{array} \tag{6}$$

where z, α_t, M, N and η_t are given by:

$$\begin{array}{ccccc} z = (1, 1, 0, \cdots, 0), & \alpha_t = \begin{pmatrix} y_t^c \\ y_t^s \\ \alpha_{2t} \\ \vdots \\ \alpha_{mt} \end{pmatrix}, & M = \left(\begin{array}{c|cc} A & & 0 \\ \hline & 0 & I_{m-1} \\ 0 & B & 0 \end{array}\right), & N = \begin{pmatrix} C \\ 0 \end{pmatrix}, & \eta_t = \begin{pmatrix} u_t \\ w_t \\ 0 \end{pmatrix}. \\ 1 \times (m+1) & (m+1) \times 1 & (m+1) \times (m+1) & (m+1) \times k & (m+1) \times 1 \end{array}$$

In this example, α_t is the unobservable variable to be estimated by the Kalman filter. In the system (6), we can take each variable as $Z_t = z$, $d_t = 0$, $S_t = 1$, $\epsilon_t = v_t$, $T_t = M$, $c_t = Nx_t$, and $R_t = I$.

2.4 Prediction of Final Data Based on Preliminary Data

It is well known that economic indicators are usually reported according to the following two steps: (i) first of all, the preliminary data are reported, and (ii) then we can obtain the final or revised data after a while (see Table 1). The problem is how to estimate the final data (or the revised data) while only the preliminary data are available.

In the case of annual data on the U.S. national accounts, the preliminary data at the present time are reported at the beginning of the next year. The revision process is performed over a few years and every decade, which is shown in Table 1, taking an example of the nominal gross domestic product data (GDP, billion dollars). In Table 1, the preliminary data of 1988 and 1992 are taken from *Survey of Current Business* (January, 1989 and January, 1993). The rest of the data in Table 1 are from *Economic Report of the President* (ERP), published from 1984 to 1995. Each column indicates the year when ERP is published, while each row represents the data of the corresponding year. The superscript p denotes the preliminary data, the superscript r implies the data revised in the year corresponding to each column, and NA indicates that the data are not available (i.e., the data have not yet been published). For instance, take the GDP data of 1984 (see the corresponding row in Table 1). The preliminary GDP data of 1984 was reported in 1985 (i.e., 3616.3), and it was revised in 1986 for the first time (i,e., 3726.7). In 1987 and 1988, the second and third revised data were published, respectively (i.e., 3717.5 and 3724.8). Since it was not revised in 1989, the GDP data of 1984 published in 1989 is given by 3724.8. Moreover, the GDP data of 1984 was revised as 3777.2 in 1992. Also, as another example, the GDP data of 1975 was revised from 1531.9 to 1580.9 in 1986 and to 1585.9 in 1992.

Thus, each data series is revised every year for the first few years and thereafter less frequently. This implies that we cannot really know the true final data, because the data are revised forever while the preliminary data are reported only once. Therefore, it might be possible to consider that the final data are unobservable, which leads to estimation of the final data given the preliminary data.

There is a wide literature dealing with the data revision process. Conrad and Corrado (1979) applied the Kalman filter to improve upon published preliminary estimates of monthly retail sales, using an ARIMA model. Howrey (1978, 1984) used the preliminary data in econometric forecasting and obtained the substantial improvements in forecast accuracy if the preliminary and revised data are used optimally.

In the context of the revision process, the Kalman filter is used as follows. There is some relationship between the final and preliminary data, because they are originally same data (see, for example, Conrad and Corrado (1979)). This

Table 1: Revision Process of U.S. National Accounts (Nominal GDP)

	1984	1985	1986	1987	1988	1989
1975	1531.9	1531.9	1580.9^r	1580.9	1580.9	1580.9
1976	1697.5	1697.5	1761.7^r	1761.7	1761.7	1761.7
1977	1894.9	1894.9	1965.1^r	1965.1	1965.1	1965.1
1978	2134.3	2134.3	2219.1^r	2219.1	2219.1	2219.1
1979	2375.2	2375.2	2464.4^r	2464.4	2464.4	2464.4
1980	2586.4^r	2586.4	2684.4^r	2684.4	2684.4	2684.4
1981	2904.5^r	2907.5^r	3000.5^r	3000.5	3000.5	3000.5
1982	3025.7^r	3021.3^r	3114.8^r	3114.8	3114.8	3114.8
1983	3263.4^p	3256.5^r	3350.9^r	3355.9^r	3355.9	3355.9
1984	NA	3616.3^p	3726.7^r	3717.5^r	3724.8^r	3724.8
1985	NA	NA	3951.8^p	3957.0^r	3970.5^r	3974.1^r
1986	NA	NA	NA	4171.2^p	4201.3^r	4205.4^r
1987	NA	NA	NA	NA	4460.2^p	4497.2^r
1988	NA	NA	NA	NA	NA	4837.8^p

	1990	1991	1992	1993	1994	1995
1975	1580.9	1580.9	1585.9^r	1585.9	1585.9	1585.9
1976	1761.7	1761.7	1768.4^r	1768.4	1768.4	1768.4
1977	1965.1	1965.1	1974.1^r	1974.1	1974.1	1974.1
1978	2219.1	2219.1	2232.7^r	2232.7	2232.7	2232.7
1979	2464.4	2464.4	2488.6^r	2488.6	2488.6	2488.6
1980	2684.4	2684.4	2708.0^r	2708.0	2708.0	2708.0
1981	3000.5	3000.5	3030.6^r	3030.6	3030.6	3030.6
1982	3114.8	3114.8	3149.6^r	3149.6	3149.6	3149.6
1983	3355.9	3355.9	3405.0^r	3405.0	3405.0	3405.0
1984	3724.8	3724.8	3777.2^r	3777.2	3777.2	3777.2
1985	3974.1	3974.1	4038.7^r	4038.7	4038.7	4038.7
1986	4197.2^r	4197.2	4268.6^r	4268.6	4268.6	4268.6
1987	4493.8^r	4486.7^r	4539.9^r	4539.9	4539.9	4539.9
1988	4847.3^r	4840.2^r	4900.4^r	4900.4	4900.4	4900.4
1989	5199.6^p	5163.2^r	5244.0^r	5250.8^r	5250.8	5250.8
1990	NA	5424.4^p	5513.8^r	5522.2^r	5546.1^r	5546.1
1991	NA	NA	5671.8^p	5677.5^r	5722.8^r	5724.8^r
1992	NA	NA	NA	5945.7^p	6038.5^r	6020.2^r
1993	NA	NA	NA	NA	6374.0^p	6343.3^r
1994	NA	NA	NA	NA	NA	6736.9^p

relationship is referred to as the measurement equation, where the final data is unobservable but the preliminary data is observed. The equation obtained by some economic theory is related to the final data, rather than the preliminary data. This equation is taken as the transition equation. We can represent this problem with the state-space form:

$$\begin{aligned} &\text{(Measurement equation)} \qquad && y_t^p = \gamma y_t^f + u_t, \\ &\text{(Transition equation)} \qquad && y_t^f = \theta_1 y_{t-1}^f + \theta_2 x_t + v_t, \end{aligned} \tag{7}$$

where y_t^p and y_t^f denote the preliminary data and the final data, respectively. y_t^f represents the state-variable. For $t = 1, \cdots, T$, the preliminary data y_t^p are observed, while we do not have the final data y_t^f. x_t is assumed to be exogenous, nonstochastic and available for $t = 1, \cdots, T$. θ_1, θ_2 and γ are the parameters to be estimated. u_t and v_t are error terms, which are mutually and independently distributed. Suppose that the relationship between y_t and x_t:

$$y_t = \theta_1 y_{t-1} + \theta_2 x_t + v_t,$$

which is derived from an economic theory. The equation obtained by some economic theory is related to the final data, rather than the preliminary data. Therefore, y_t should be replaced by the final data, not the preliminary data, because the preliminary data include the measurement errors. If we estimate with such data, the appropriate results cannot be obtained. Therefore, as in the system (7), the transition equation in this example is given by

$$y_t^f = \theta_1 y_{t-1}^f + \theta_2 x_t + v_t.$$

Mariano and Tanizaki (1995) also examined the above problem, where several predictors are introduced and compared using the U.S. national accounts.

2.5 Estimation of Permanent Consumption

The next application is concerned with an estimation of permanent consumption. Total consumption consists of permanent and transitory consumption. This relationship is represented by an identity equation, which is taken as the measurement equation where total consumption is observable but both permanent and transitory consumption are unobservable. Permanent consumption depends on life-time income expected in the future, which is defined as permanent income. Solving a dynamic optimization problem, maximizing a quadratic utility function of the representative agent with respect to permanent consumption for each time and assuming that the discount rate is equal to the reciprocal of the gross rate of return on savings, it is shown that permanent consumption follows a random walk, i.e., this implies solving the following utility maximization problem:

$$\max_{\{c_t^p\}} \ \mathrm{E}_0 \left(\sum_t \beta^t u(c_t^p) \right), \ \text{subject to} \ A_{t+1} = R_t (A_t + y_t - c_t),$$

where $0 < \beta < 1$, $u(c_t^p) = -\frac{1}{2}(\bar{c} - c_t^p)^2$, $c_t = c_t^p + c_t^T$ and $\beta R_t = 1$. c_t, c_t^p, c_t^f, R_t, A_t, y_t, β, $u(\cdot)$ and $\mathrm{E}_t(\cdot)$ denote per capita total consumption, per capita permanent consumption, per capita transitory consumption, the real gross rate of return on savings between periods t and $t+1$, the stock of assets at the beginning of period t, per capita labor income, the discount rate, the representative utility function and the mathematical expectation given information up to t, respectively. See Hall (1978).

Next, the assumption on transitory consumption is as follows. In a cross sectional framework,

$$\sum_i C_{it}^T = 0,$$

which is assumed by Friedman (1957), where C_{it}^T denotes transitory consumption of the i-th household at time t. Moreover, Friedman (1957) assumed that C_{it}^T is independent of permanent income, transitory income and permanent consumption. Also, see Branson (1979). Note that the relationship between c_t^T and C_{it}^T is:

$$c_t^T = \frac{1}{L_t} \sum_i C_{it}^T,$$

where L_t is the number of population at time t. We can approximate as

$$\frac{1}{L_t}\sum_i C_{it}^T \approx \mathrm{E}(C_{it}^T),$$

which is equal to zero. Here, assuming that $C_{it}^T \equiv \epsilon_{it}$ is identically distributed with mean zero and variance σ_ϵ^2 , transitory consumption of the representative agent (i.e., c_t^T) is given by a random shock with mean zero and variance σ_ϵ^2/L_t. Thus, the model to this problem is given by:

$$\begin{aligned} &\text{(Measurement equation)} \qquad && c_t = c_t^p + c_t^T, \\ &\text{(Transition equation)} && c_t^p = c_{t-1}^p + \eta_t, \end{aligned} \tag{8}$$

where the error terms $c_t^T = \epsilon_t$ and η_t are mutually and independently distributed. ϵ_t has a distribution with mean zero and variance σ_ϵ^2/L_t, and η_t is a random variable with mean zero and variance σ_η^2, i.e.,

$$\begin{pmatrix}\epsilon_t\\ \eta_t\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_\epsilon^2/L_t & 0\\ 0 & \sigma_\eta^2\end{pmatrix}\right)$$

In the state-space model (8), c_t^p is regarded as the state-variable which is estimated by the Kalman filtering technique. Thus, we can estimate permanent and transitory consumption separately. We can consider this example with the nonlinear transition equation, where the utility function of the representative agent is assumed to be a constant relative risk aversion type of utility function.

For the other applications, we can find estimation of the rational expectation models (for example, see Burmeister and Wall (1982), Engle and Watson (1987) and McNelis and Neftci (1983)). See Harvey (1987) for a survey of applications of the Kalman filter model.

3 Derivations of Kalman Filter Algorithm

In this paper, we derive the filtering algorithm of the state-space form for the linear case given by equations (1) and (2). One may consider the problem of estimating α_t using information up to time s, i.e., $Y_s = \{y_1, y_2, \cdots, y_s\}$. Denote by $\mathrm{E}(\alpha_t|Y_s) \equiv a_{t|s}$ the conditional expectation of α_t given Y_s. The evaluation of $a_{t|s}$ is known as filtering if $t = s$, smoothing if $t < s$, and prediction if $t > s$, respectively. In this paper we focus only on the filtering problem. Finally, $\mathrm{Cov}(\alpha_t|Y_s) \equiv \Sigma_{t|s}$ is defined as the conditional covariance matrix of α_t given Y_s.

The filtering algorithm in this case is known as the Kalman filter, which is given by the following equations:

$$a_{t|t-1} = T_t a_{t-1|t-1} + c_t, \tag{9}$$

$$\Sigma_{t|t-1} = T_t \Sigma_{t-1|t-1} T_t' + R_t Q_t R_t', \tag{10}$$

$$y_{t|t-1} = Z_t a_{t|t-1} + d_t, \tag{11}$$

$$F_{t|t-1} = Z_t \Sigma_{t|t-1} Z_t' + S_t H_t S_t', \tag{12}$$

$$K_t = \Sigma_{t|t-1} Z_t' F_{t|t-1}^{-1}, \tag{13}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_t F_{t|t-1} K_t', \tag{14}$$

$$a_{t|t} = a_{t|t-1} + K_t(y_t - y_{t|t-1}), \tag{15}$$

for $t = 1, \cdots, T$, where Z_t, d_t, S_t, T_t, c_t, R_t, Q_t, H_t, $a_{0|0}$ and $\Sigma_{0|0}$ are known for now. See Anderson and Moore (1979), Gelb (1974) and Jazwinski (1970) for the Kalman filter algorithm.

Since equations (9) and (10) essentially predict the parameter α_t using the information up to time $t-1$, they are called the prediction equations. On the other hand, equations (14) and (15) play the role of combining the new observation obtained at time t (i.e., y_t) and the past information up to time $t-1$ (i.e., Y_{t-1}). They are known as the updating equations. Moreover, equation (13) is called the Kalman gain. An interpretation on the Kalman gain K_t depends on the derivation process. When the filtering algorithm is derived from the minimum mean square linear estimator, K_t is chosen such that the filtering estimate $a_{t|t}$ has minimum variance (see Section 3.3). In this case, K_t

is referred to as the Kalman gain. Equations (11) and (12) represent the prediction estimate and its variance of the observed data y_t.

According to the Kalman filter algorithm shown above, $a_{1|0}$ and $\Sigma_{1|0}$ are readily calculated from prediction equations (9) and (10) since we assume that $a_{0|0}$ and $\Sigma_{0|0}$ are known. $a_{1|1}$ and $\Sigma_{1|1}$ are obtained by substituting $a_{1|0}$ and $\Sigma_{1|0}$ with K_1 into equations (14) and (15). Thus, $a_{t|t}$ and $\Sigma_{t|t}$, $t = 1, \cdots, T$, are calculated recursively from (9) – (15), once Q_t, H_t, $a_{0|0}$ and $\Sigma_{0|0}$ are assumed to be known.

There are several ways to interpret the Kalman filter algorithm. Their interpretations are related to the derivations of the algorithm, i.e., the derivation under the assumption of normality (Harvey (1989)), by orthogonal projection (Brockwell and Davis (1987) and Chow (1983)), from the mixed estimation (Cooley (1977), Harvey (1981) and Diderrich (1985)), from the generalized least squares method (Sant (1977)), and by the minimum mean square linear estimator (Burridge and Wallis (1988) and Harvey (1989)). Here, we discuss three of these derivations, i.e., derivation under normality assumption for the error terms (Section 3.1), use of mixed estimation procedure (Section 3.2) and interpretation of minimum mean square linear estimator (Section 3.3).

3.1 Derivation under Normality Assumption

A derivation of the Kalman filter algorithm under normality assumption comes from a recursive Bayesian estimation. If ϵ_t and η_t are mutually, independently and normally distributed and both measurement and transition equations are linear in the state-vector and the error term as in equations (1) and (2), then the conventional linear recursive algorithm represented by equations (9) – (15) can be obtained. The derivation is shown as follows.

Let $P(\cdot|\cdot)$ be a conditional density function. The prediction and updating equations, based on the probability densities, are given as follows (see, for example, Kitagawa (1987), Kramer and Sorenson (1988), and Harvey (1989)):

(Prediction equation)

$$\begin{aligned} P(\alpha_t|Y_{t-1}) &= \int P(\alpha_t, \alpha_{t-1}|Y_{t-1}) \mathrm{d}\alpha_{t-1} \\ &= \int P(\alpha_t|\alpha_{t-1}, Y_{t-1}) P(\alpha_{t-1}|Y_{t-1}) \mathrm{d}\alpha_{t-1} \\ &= \int P(\alpha_t|\alpha_{t-1}) P(\alpha_{t-1}|Y_{t-1}) \mathrm{d}\alpha_{t-1}, \end{aligned} \tag{16}$$

(Updating equation)

$$\begin{aligned} P(\alpha_t|Y_t) &= P(\alpha_t|y_t, Y_{t-1}) \\ &= \frac{P(\alpha_t, y_t|Y_{t-1})}{P(y_t|Y_{t-1})} \\ &= \frac{P(\alpha_t, y_t|Y_{t-1})}{\displaystyle\int P(\alpha_t, y_t|Y_{t-1}) \mathrm{d}\alpha_t} \\ &= \frac{P(y_t|\alpha_t, Y_{t-1}) P(\alpha_t|Y_{t-1})}{\displaystyle\int P(y_t|\alpha_t, Y_{t-1}) P(\alpha_t|Y_{t-1}) \mathrm{d}\alpha_t} \\ &= \frac{P(y_t|\alpha_t) P(\alpha_t|Y_{t-1})}{\displaystyle\int P(y_t|\alpha_t) P(\alpha_t|Y_{t-1}) \mathrm{d}\alpha_t}, \end{aligned} \tag{17}$$

for $t = 1, \cdots, T$. See Appendix A.1 for the derivation in detail. Equations (16) and (17) are already a recursive algorithm of the density functions, which is called the density-based filtering algorithm in this paper. $P(\alpha_t|\alpha_{t-1})$

is obtained from the transition equation (2) if the distribution of the error term η_t is specified, and also $P(y_t|\alpha_t)$ is derived from the measurement equation (1) given the specific distribution of the error term ϵ_t. Thus, the filtering density $P(\alpha_t|Y_t)$ is obtained recursively, given the distributions $P(\alpha_t|\alpha_{t-1})$ and $P(y_t|\alpha_t)$.

Now, we derive the filtering algorithm represented by equations (9) – (15), i.e., density-based filtering algorithm.

If α_0, ϵ_t and η_t are normally distributed, it is known that $P(\alpha_t|Y_s)$, $s = t, t-1$, is expressed by the Gaussian distribution:

$$P(\alpha_t|Y_s) = \Phi(\alpha_t - a_{t|s}, \Sigma_{t|s}),$$

where $\Phi(\alpha_t - a_{t|s}, \Sigma_{t|s})$ denotes the normal density with mean $a_{t|s}$ and variance $\Sigma_{t|s}$, i.e.,

$$\Phi(\alpha_t - a_{t|s}, \Sigma_{t|s}) = (2\pi)^{-k/2}|\Sigma_{t|s}|^{-1/2} \exp\left(-\frac{1}{2}(\alpha_t - a_{t|s})'\Sigma_{t|s}^{-1}(\alpha_t - a_{t|s}) \right).$$

Note as follows:

$$\Phi(\alpha_t - a_{t|s}, \Sigma_{t|s}) \equiv N(a_{t|s}, \Sigma_{t|s}).$$

Then, from equation (16), one-step ahead prediction density $P(\alpha_t|Y_{t-1})$ is given by:

$$\begin{aligned}
P(\alpha_t|Y_{t-1}) &= \int P(\alpha_t|\alpha_{t-1})P(\alpha_{t-1}|Y_{t-1})\mathrm{d}\alpha_{t-1} \\
&= \int \Phi(\alpha_t - T_t\alpha_{t-1} - c_t, R_tQ_tR_t')\ \Phi(\alpha_{t-1} - a_{t-1|t-1}, \Sigma_{t-1|t-1})\mathrm{d}\alpha_{t-1} \\
&= \Phi(\alpha_t - T_ta_{t-1|t-1} - c_t, T_t\Sigma_{t-1|t-1}T_t' + R_tQ_tR_t') \\
&\equiv \Phi(\alpha_t - a_{t|t-1}, \Sigma_{t|t-1}),
\end{aligned} \tag{18}$$

where $P(\alpha_t|\alpha_{t-1})$ and $P(\alpha_{t-1}|Y_{t-1})$ are rewritten as follows:

$$\begin{aligned}
&P(\alpha_t|\alpha_{t-1}) = \Phi(\alpha_t - T_t\alpha_{t-1} - c_t, R_tQ_tR_t'), \\
&P(\alpha_{t-1}|Y_{t-1}) = \Phi(\alpha_{t-1} - a_{t-1|t-1}, \Sigma_{t-1|t-1}).
\end{aligned}$$

The third equality in equation (18) is proved in Proof I of Appendix A.2.

Thus, comparing each argument (moment) in the two normal densities given by the fourth line and the fifth line in equation (18), we can derive the prediction equations (9) and (10).

Next, from equation (17), to derive the updating equations, equation (17) is calculated as follows:

$$\begin{aligned}
P(\alpha_t|Y_t) &= \frac{P(y_t|\alpha_t)P(\alpha_t|Y_{t-1})}{\displaystyle\int P(y_t|\alpha_t)P(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t} \\
&= \frac{\Phi(y_t - Z_t\alpha_t - d_t, S_tH_tS_t')\ \Phi(\alpha_t - a_{t|t-1}, \Sigma_{t|t-1})}{\displaystyle\int \Phi(y_t - Z_t\alpha_t - d_t, S_tH_tS_t')\ \Phi(\alpha_t - a_{t|t-1}, \Sigma_{t|t-1})\mathrm{d}\alpha_t} \\
&= \Phi\big(\alpha_t - a_{t|t-1} - K_t(y_t - y_{t|t-1}), \Sigma_{t|t-1} - K_tF_{t|t-1}K_t'\big) \\
&\equiv \Phi(\alpha_t - a_{t|t}, \Sigma_{t|t}),
\end{aligned} \tag{19}$$

where

$$\begin{aligned}
&y_{t|t-1} \equiv \mathrm{E}(y_t|Y_{t-1}) = Z_ta_{t|t-1} + d_t, \\
&F_{t|t-1} \equiv \mathrm{Var}(y_t|Y_{t-1}) = Z_t\Sigma_{t|t-1}Z_t' + S_tH_tS_t', \\
&K_t = \Sigma_{t|t-1}Z_t'F_{t|t-1}^{-1}.
\end{aligned}$$

Note that $P(y_t|\alpha_t)$ and $P(\alpha_t|Y_{t-1})$ are rewritten as:

$$P(y_t|\alpha_t) = \Phi(y_t - Z_t\alpha_t - d_t, S_tH_tS_t'),$$
$$P(\alpha_t|Y_{t-1}) = \Phi(\alpha_t - a_{t|t-1}, \Sigma_{t|t-1}).$$

The third equality in equation (19) is proved in Proof II of Appendix A.2.

Therefore, the updating equations (13), (14) and (15) are obtained, comparing each argument in equation (19). Thus, the filtering algorithm is represented by (9) – (15).

There are two notable points:

(1) Even if the lagged dependent variables are contained in the measurement and transition equations, i.e., even if Z_t, d_t, S_t, T_t, c_t and R_t in equations (1) and (2) depend on the lagged dependent variables $y_{t-1}, \cdots, y_1$, we can derive the same algorithm as equation (9) – (15) under normality assumption for the error terms ϵ_t and η_t (see Harvey, 1989, p.156). In this case, the probability densities $P(y_t|\alpha_t)$ and $P(\alpha_t|\alpha_{t-1})$ in (16) and (17) are replaced by:

$$P(y_t|\alpha_t, y_{t-1}, \cdots, y_1) \qquad \text{(i.e., } P(y_t|\alpha_t, Y_{t-1}) \text{ in general)},$$
$$P(\alpha_t|\alpha_{t-1}, y_{t-1}, \cdots, y_1) \qquad \text{(i.e., } P(\alpha_t|\alpha_{t-1}, Y_{t-1}) \text{ in general)}.$$

Also, see Appendix A.1.

(2) Even if the system is nonlinear, i.e., even if the measurement equation (1) and the transition equation (2) are given by:

$$\text{(Measurement equation)} \qquad q_t(y_t, \alpha_t) = \epsilon_t, \tag{20}$$
$$\text{(Transition equation)} \qquad f_t(\alpha_t, \alpha_{t-1}) = \eta_t, \tag{21}$$

respectively, the prediction and updating equations represented by the densities (16) and (17) are unchanged (see Appendix A.1). We can deal with a nonlinear system as the general case of the state-space model (20) and (21).

Thus, under the normality and linearity assumptions, from equations (16) and (17), we can derive the linear recursive Kalman filter algorithm given by equations (9) – (15).

The following is an alternative derivation based on normality assumption, which is very simple and easy. The prediction equations (9) and (10) are easily obtained by taking the conditional expectation and variance given information up to time $t-1$ on both sides of equation (2).

Since the transition equation (2) is linear and α_{t-1} and η_t are normal, α_t (which is a sum of α_{t-1} and η_t) is also normal. Therefore, when the initial value α_0 is normal, α_t for $t = 1, \cdots, T$ are normal.

In order to obtain the updating equations (13) – (15), consider the joint probability density function of α_t and y_t, i.e., $P(\alpha_t, y_t|Y_{t-1})$, which are normal. For the measurement equation, when α_t and ϵ_t are normal, y_t (which is a sum of α_t and ϵ_t) is also normal. Accordingly, given information Y_{t-1}, the joint density function $P(\alpha_t, y_t|Y_{t-1})$ is as follows:

$$\begin{pmatrix} \alpha_t \\ y_t \end{pmatrix} \sim N\left(\begin{pmatrix} a_{t|t-1} \\ y_{t|t-1} \end{pmatrix}, \begin{pmatrix} \Sigma_{t|t-1} & \Sigma_{t|t-1}Z_t' \\ Z_t\Sigma_{t|t-1} & F_{t|t-1} \end{pmatrix} \right). \tag{22}$$

Note that each component of conditional mean and variance is computed as:

$$\mathrm{E}(\alpha_t|Y_{t-1}) \equiv a_{t|t-1} = T_t a_{t|t-1} + c_t,$$
$$\mathrm{E}(y_t|Y_{t-1}) \equiv y_{t|t-1} = Z_t a_{t|t-1} + d_t,$$
$$\mathrm{E}\big((\alpha_t - a_{t|t-1})(\alpha_t - a_{t|t-1})'|Y_{t-1}\big) \equiv \Sigma_{t|t-1},$$
$$\mathrm{E}\big((y_t - y_{t|t-1})(\alpha_t - a_{t|t-1})'|Y_{t-1}\big) = Z_t\mathrm{E}\big((\alpha_t - a_{t|t-1})(\alpha_t - a_{t|t-1})'|Y_{t-1}\big) = Z_t\Sigma_{t|t-1},$$
$$\mathrm{E}\big((y_t - y_{t|t-1})(y_t - y_{t|t-1})'|Y_{t-1}\big) \equiv F_{t|t-1}.$$

Here, applying Lemma 3 to equation (22), the conditional distribution $P(\alpha_t|y_t, Y_{t-1}) \equiv P(\alpha_t|Y_t)$ is obtained as:

$$\alpha_t \sim N(a_{t|t-1} - K_t(y_t - y_{t|t-1}),\ \Sigma_{t|t-1} - K_t F_{t|t-1} K_t'),$$

where K_t is given by equation (13). Since $P(\alpha_t|Y_t)$ implies

$$\alpha_t \sim N(a_{t|t},\ \Sigma_{t|t}).$$

Comparing each argument, the following updating equations can be obtained.

$$a_{t|t} = a_{t|t-1} - K_t(y_t - y_{t|t-1}),$$
$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_t F_{t|t-1} K_t',$$

which are given by equations (14) and (15).

Thus, based on normality assumption, two derivations were introduced in this section. When the measurement and transition equations are linear and the error terms in the two equations are normal, the Kalman filter algorithm can be obtained from equations (16) and (17) to equations (9) – (15).

In the following two sections, the Kalman filter algorithm is derived without normality assumption.

3.2 Derivation by Mixed Estimation

Next, we derive the Kalman filter algorithm by utilizing the mixed estimation method (so-called Goldberger-Theil estimation), which is introduced by Harvey (1981), Diderrich (1985) and Tanizaki (1993a). According to this estimation method, we do not have to give a specific distribution to the error terms ϵ_t and η_t because it is based on the generalized least squares estimation.

The prediction equations are simply derived by taking the conditional expectation and variance given information up to time $t-1$ on both sides of equation (2). They are given by (9) and (10). α_t is rewritten based on prior information as follows. α_t consists of one-step ahead prediction $a_{t|t-1}$ and error term ξ_t, i.e.,

$$\alpha_t = a_{t|t-1} + \xi_t, \tag{23}$$

where mean and variance of ξ_t are given by:

$$\mathrm{E}(\xi_t|Y_{t-1}) = 0$$
$$\mathrm{Var}(\xi_t|Y_{t-1}) = \Sigma_{t|t-1}.$$

Again, we write the measurement equation (1) together with equation (23) as follows:

$$y_t = Z_t\alpha_t + d_t + S_t\epsilon_t, \tag{1}$$

$$a_{t|t-1} = \alpha_t - \xi_t. \tag{23}$$

Equation (23) represents the prior information and the present sample is described by equation (1). Accordingly, equations (1) and (23) hold all the information available at time t. By the assumption of no correlation between ϵ_t and η_t, the error terms ϵ_t and ξ_t are independently distributed as follows:

$$\mathrm{E}\begin{pmatrix}\epsilon_t\\ \xi_t\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}, \qquad \mathrm{Var}\begin{pmatrix}\epsilon_t\\ \xi_t\end{pmatrix} = \begin{pmatrix}H_t & 0\\ 0 & \Sigma_{t|t-1}\end{pmatrix},$$

where we do not have to assume a functional form of the distribution of the error terms ϵ_t and ξ_t.

Equations (1) and (23) are rewritten as:

$$\begin{pmatrix}y_t - d_t\\ a_{t|t-1}\end{pmatrix} = \begin{pmatrix}Z_t\\ I_k\end{pmatrix}\alpha_t + \begin{pmatrix}S_t & 0\\ 0 & -I_k\end{pmatrix}\begin{pmatrix}\epsilon_t\\ \xi_t\end{pmatrix}. \tag{24}$$

Let $a_{t|t}$ be the generalized least squares estimate of the state-variable α_t in equation (24). Apply the generalized least squares method (GLS) to (24), and we can obtain $a_{t|t}$ which is the filtering estimate of α_t.

$$\begin{aligned}a_{t|t} &= \left((Z_t' \ \ I_k)\begin{pmatrix}S_tH_tS_t' & 0\\ 0 & \Sigma_{t|t-1}\end{pmatrix}^{-1}\begin{pmatrix}Z_t\\ I_k\end{pmatrix}\right)^{-1}(Z_t' \ \ I_k)\begin{pmatrix}S_tH_tS_t' & 0\\ 0 & \Sigma_{t|t-1}\end{pmatrix}^{-1}\begin{pmatrix}y_t - d_t\\ a_{t|t-1}\end{pmatrix}\\ &= \big(Z_t'(S_tH_tS_t')^{-1}Z_t + \Sigma_{t|t-1}^{-1}\big)^{-1}\big(Z_t'(S_tH_tS_t')^{-1}(y_t - d_t) + \Sigma_{t|t-1}^{-1}a_{t|t-1}\big).\end{aligned} \tag{25}$$

Note that we have:

$$\mathrm{Var}\left(\begin{pmatrix} S_t & 0 \\ 0 & -I_k \end{pmatrix} \begin{pmatrix} \epsilon_t \\ \xi_t \end{pmatrix} \right) = \begin{pmatrix} S_t H_t S_t' & 0 \\ 0 & \Sigma_{t|t-1} \end{pmatrix}.$$

The above GLS estimate of α_t obtained from equation (24), i.e., $a_{t|t}$ in equation (25), implies $\mathrm{E}(\alpha_t|Y_t)$, because the present sample y_t and the past information Y_{t-1} are included in equation (24).

And it is easily shown that:

$$\begin{aligned} \Sigma_{t|t} &= (Z_t'(S_t H_t S_t')^{-1} Z_t + \Sigma_{t|t-1}^{-1})^{-1} \\ &= \Sigma_{t|t-1} - \Sigma_{t|t-1} Z_t' F_{t|t-1}^{-1} Z_t \Sigma_{t|t-1} \\ &= \Sigma_{t|t-1} - K_t F_{t|t-1} K_t', \end{aligned} \tag{26}$$

where

$$K_t = \Sigma_{t|t-1} Z_t' F_{t|t-1}^{-1}.$$

Note that Lemma 1 in Appendix A.2 is used for the second equality in equation (26). Substituting equation (26) into equation (25), we can rewrite equation (25) as:

$$\begin{aligned} a_{t|t} &= \big(Z_t'(S_t H_t S_t')^{-1} Z_t + \Sigma_{t|t-1}^{-1}\big)^{-1} \big(Z_t'(S_t H_t S_t')^{-1}(y_t - d_t) + \Sigma_{t|t-1}^{-1} a_{t|t-1}\big) \\ &= \big(\Sigma_{t|t-1} - \Sigma_{t|t-1} Z_t' F_{t|t-1}^{-1} Z_t \Sigma_{t|t-1}\big)\big(Z_t'(S_t H_t S_t')^{-1}(y_t - d_t) + \Sigma_{t|t-1}^{-1} a_{t|t-1}\big) \\ &= a_{t|t-1} + \Sigma_{t|t-1} Z_t' F_{t|t-1}^{-1}(y_t - Z_t a_{t|t-1} - d_t) \\ &= a_{t|t-1} + K_t(y_t - y_{t|t-1}). \end{aligned} \tag{27}$$

Thus, the algorithm of Kalman filter is represented by equations (9) – (15).

(26) and (27) represent the updating equations. The updating problem is to utilize the added sample with the prior information to improve the parameter estimates.

The advantage of this approach is that we do not have to assume any distribution function for the error terms ϵ_t and η_t in equations (1) and (2).

3.3 Minimum Mean Square Linear Estimator

The Kalman filter as minimum mean square linear estimator do not require normality assumption for the error terms. Prediction equations (9) and (10) are similarly obtained by taking the conditional mean and variance given information up to time $t-1$ on both sides of the transition equation (2).

In order to derive the updating equations (13) – (15), first, note that $a_{t|t}$ is written as:

$$a_{t|t} = A_t a_{t-1|t-1} + B_t c_t + D_t d_t + K_t y_t, \tag{28}$$

because $a_{t|t}$ denotes the conditional mean of α_t given information Y_t and the measurement and transition equations are linear (also, here we consider a linear estimator only). Therefore, As shown above, $a_{t|t}$ is represented by a linear function of the past information (i.e., $a_{t-1|t-1}$ which includes information up to time $t-1$, Y_{t-1}) and the present sample (i.e., c_t, d_t and y_t). We obtain A_t, B_t, D_t and K_t such that $a_{t|t}$ is the minimum mean square linear estimator.

Define $e_t = \alpha_t - a_{t|t}$. Then, e_t is transformed into:

$$\begin{aligned} e_t &= \alpha_t - A_t a_{t-1|t-1} - B_t c_t - D_t d_t - K_t y_t \\ &= (T_t \alpha_{t-1} + c_t + R_t \eta_t) - A_t(\alpha_{t-1} - e_{t-1}) - B_t c_t - D_t d_t - K_t\Big(Z_t(T_t \alpha_{t-1} + c_t + R_t \eta_t) + d_t + S_t \epsilon_t\Big) \\ &= A_t e_{t-1} + (T_t - A_t - K_t Z_t T_t)\alpha_{t-1} + (I_k - B_t - K_t Z_t) c_t - (D_t + K_t) d_t + (I_k - K_t Z_t) R_t \eta_t - K_t S_t \epsilon_t. \end{aligned}$$

In calculating, the followings are used.

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t,$$

$$a_{t-1|t-1} = \alpha_{t-1} - e_{t-1},$$

$$\begin{aligned} y_t &= Z_t \alpha_t + d_t + S_t \epsilon_t \\ &= Z_t (T_t \alpha_{t-1} + c_t + R_t \eta_t) + d_t + S_t \epsilon_t. \end{aligned}$$

In order for $a_{t|t}$ to be unbiased, the expectation of e_t yields zero. Accordingly, we have:

$$\begin{aligned} &T_t - A_t - K_t Z_t T_t = 0, \\ &I_k - B_t - K_t Z_t = 0, \\ &D_t + K_t = 0. \end{aligned}$$

Therefore, eliminating A_t, B_t and D_t, e_t is rewritten as follows:

$$e_t = (I_k - K_t Z_t)(T_t e_{t-1} + R_t \eta_t) - K_t S_t \epsilon_t.$$

In the above equation, we still have K_t to be computed. Next, we have to obtain K_t such that e_t has minimum variance. Since e_{t-1}, η_t and ϵ_t are mutually independent, variance of e_t, i.e., $\Sigma_{t|t}$, is defined as:

$$\begin{aligned} \Sigma_{t|t} &= (I_k - K_t Z_t)(T_t \Sigma_{t-1|t-1} T_t' + R_t Q_t R_t')(I_k - K_t Z_t)' + K_t S_t H_t S_t' K_t' \\ &= (I_k - K_t Z_t) \Sigma_{t|t-1} (I_k - K_t Z_t)' + K_t S_t H_t S_t' K_t' \end{aligned} \tag{29}$$

Note that variance of e_t can be defined as $\Sigma_{t|t}$ because e_t includes information up to time t, i.e., Y_t.

It is easily shown that e_t has minimum variance when K_t takes the following:

$$K_t = \Sigma_{t|t-1} Z_t' F_{t|t-1}^{-1},$$

where

$$F_{t|t-1} = Z_t \Sigma_{t|t-1} Z_t' + S_t H_t S_t',$$

which denotes the conditional variance of y_t given information at time $t-1$, Y_{t-1}. Taking into account equations (9) and (10), equations (28) and (29) are represented as:

$$a_{t|t} = a_{t|t-1} + K_t (y_t - Z_t a_{t|t-1} - d_t),$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_t F_{t|t-1} K_t'.$$

Thus, the Kalman filter algorithm is derived. K_t is called the Kalman gain, because it is obtained such that $a_{t|t}$ has minimum variance.

In this section, we focused on derivation of the Kalman filter algorithm. The same algorithm was derived from the three different ways; derivation based on normal density in Section 3.1, and distribution-free derivations in Sections 3.2 and 3.3. In Section 3.3, the Kalman filter is derived as the minimum mean square estimator. Using this derivation, we can easily obtain the filtering algorithm even when the error terms ϵ_t and η_t in the state-space model represented by equations (1) and (10) are correlated with each other. See Burridge and Wallis (1988) and Harvey (1989).

4 Estimation of Unknown Parameters

Finally, we discuss estimation of unknown parameters, using the maximum likelihood function procedure. In this section, the likelihood function to be maximized is constructed by assuming normality for the error terms ϵ_t and η_t.

In the case where Z_t, d_t, H_t, T_t, c_t, R_t and Q_t depend on the unknown parameter vector to be estimated, say θ, the following innovation form of the likelihood function is maximized with respect to θ (see Chow (1983) and Watson and Engle (1983)).

$$
\begin{aligned}
P(y_T, y_{T-1}, \cdots, y_1) &= P(y_T|Y_{T-1})P(y_{T-1}|Y_{T-2}) \cdots P(y_2|y_1)P(y_1) \\
&= \prod_{t=1}^{T} P(y_t|Y_{t-1}) \\
&= \prod_{t=1}^{T} (2\pi)^{-g/2}|F_{t|t-1}|^{-1/2} \exp\left(-\frac{1}{2}(y_t - y_{t|t-1})' F_{t|t-1}^{-1}(y_t - y_{t|t-1})\right).
\end{aligned}
\tag{30}
$$

Note that, in equation (30), the initial density function $P(y_1)$ is approximated as $P(y_1|Y_0)$. According to this maximization procedure, we cannot obtain a solution in closed form. The simplicity of using the innovation form is that $y_{t|t-1}$ and $F_{t|t-1}$ are computed in the filtering algorithm (9) – (15), where no extra algorithm for the maximization procedure is required. Therefore, the innovation form (30) is broadly used for the maximum likelihood procedure.

For another estimation method, the expectation of the log-likelihood function is maximized. This estimation is called the EM algorithm, where unobservable variables are replaced by the conditionally expected values given all the observed data $\{y_1, y_2, \cdots, y_T\}$, i.e., Y_T. Since the state-space model is originally developed for estimation of unobservable variables, the Kalman filter model is consistent with the EM algorithm in the concept and therefore the EM algorithm is often used for estimation of unknown parameters in the state-space form. See Shumway and Stoffer (1982), Tanizaki (1989, 1993a) and Watson and Engle (1983) for an application of the EM algorithm to the Kalman filter model and Dempster, Laird and Rubin (1977) and Rund (1991) for the EM algorithm. It is known that according to the EM algorithm the convergence is very slow, but it quickly searches the neighborhood of the true parameter. In the case where Z_t, d_t, H_t, T_t and c_t are linear in the unknown parameter vector θ, a solution of unknown parameters can be obtained explicitly by a function of the smoothed estimates. See Watson and Engle (1983). Thus, the expected log-likelihood function is maximized with respect to the unknown parameter vector (i.e., θ), when θ is included in the measurement and transition equations. The problem of using the EM algorithm is that we need the smoothed estimates and therefore the extra algorithm is necessary, which implies a greater computational burden.

5 Summary

In this paper, by the way of introduction to the Kalman filter, we have discussed some applications of the state-space model, the conventional linear recursive algorithm of Kalman filter, its derivations and the estimation problem of unknown parameters.

For the derivations, there are several approaches. Here we took three derivations; the first is based on normality assumption, the second is based on the mixed estimation and the third is taken as the minimum mean square linear estimator. The advantage of the mixed estimation approach and the minimum mean square linear estimator is that we do not have to assume normality for the error terms. The mixed estimation approach is useful for deriving the single-stage iteration filtering algorithm, which is one of the nonlinear filters. The approach under the assumption of normality is used for deriving the nonlinear and/or nonnormal filtering algorithms based on the density functions.

Finally, in the case where the unknown parameters are included in the measurement and transition equations, the maximum likelihood method is used. Here we consider maximizing the innovation form of the likelihood function.

A Appendix

A.1 Density-Based Filtering Algorithm

We prove the third equality in (16) and the fifth equality in (17), where $P(\alpha_t|\alpha_{t-1}, Y_{t-1})$ is replaced by $P(\alpha_t|\alpha_{t-1})$ in equation (16) and $P(y_t|\alpha_t, Y_{t-1})$ is by $P(y_t|\alpha_t)$ in equation (17).

We give the proof more generally by using the nonlinear forms given by equations (20) and (21), which are the following state-space model:

$$\text{(Measurement equation)} \qquad q_t(y_t, \alpha_t) = \epsilon_t, \tag{20}$$

$$\text{(Transition equation)} \qquad f_t(\alpha_t, \alpha_{t-1}) = \eta_t. \tag{21}$$

Let $P_\epsilon(\epsilon_t)$ and $P_\eta(\eta_t)$ be the probability density functions of ϵ_t and η_t. Then, the joint density function of $\eta_t, \cdots, \eta_1$, $\epsilon_{t-1}, \cdots, \epsilon_1$ is given by:

$$P(\eta_t, \cdots, \eta_1, \epsilon_{t-1}, \cdots, \epsilon_1) = P_\eta(\eta_t) \cdots P_\eta(\eta_1) P_\epsilon(\epsilon_{t-1}) \cdots P_\epsilon(\epsilon_1),$$

because $\eta_t, \cdots, \eta_1$, $\epsilon_{t-1}, \cdots, \epsilon_1$ are mutually independent. By transforming the variables to obtain the joint density of $\alpha_t, \cdots, \alpha_1$, $y_{t-1}, \cdots, y_1$, we have

$$\begin{aligned} P(\alpha_t, \cdots, \alpha_1, y_{t-1}, \cdots, y_1) &= P(\alpha_t, \cdots, \alpha_1, Y_{t-1}) \\ &= \prod_{s=1}^{t-1} \left| \frac{\partial q_s(y_s, \alpha_s)}{\partial y_s'} \right| P_\epsilon\big(q_s(y_s, \alpha_s)\big) \prod_{s=1}^{t} \left| \frac{\partial f_s(\alpha_s, \alpha_{s-1})}{\partial \alpha_s'} \right| P_\eta\big(f_s(\alpha_s, \alpha_{s-1})\big). \end{aligned}$$

Here, we can take as:

$$P(y_s|\alpha_s) = \left| \frac{\partial q_s(y_s, \alpha_s)}{\partial y_s'} \right| P_\epsilon\big(q_s(y_s, \alpha_s)\big),$$

which is a conditional density of y_s, given α_s. Similarly, $P(\alpha_s|\alpha_{s-1})$ is given as follows:

$$P(\alpha_s|\alpha_{s-1}) = \left| \frac{\partial f_s(\alpha_s, \alpha_{s-1})}{\partial \alpha_s'} \right| P_\eta\big(f_s(\alpha_s, \alpha_{s-1})\big),$$

which is derived from the density $P_\eta(\eta_s)$.

Therefore, the joint density of $\alpha_t, \cdots, \alpha_1$, $y_{t-1}, \cdots, y_1$ is:

$$P(\alpha_t, \cdots, \alpha_1, Y_{t-1}) = P(\alpha_t|\alpha_{t-1}) \cdots P(\alpha_1|\alpha_0) P(y_{t-1}|\alpha_{t-1}) \cdots P(y_1|\alpha_1),$$

where α_0 is given. Next, to obtain $P(\alpha_t, \alpha_{t-1}, Y_{t-1})$, the above joint-density function is integrated with respect to $\alpha_{t-2}, \cdots, \alpha_1$.

$$\begin{aligned} & P(\alpha_t, \alpha_{t-1}, Y_{t-1}) \\ &= \int \cdots \int P(\alpha_t, \cdots, \alpha_1, Y_{t-1}) \mathrm{d}\alpha_{t-2} \ \cdots \ \mathrm{d}\alpha_1 \\ &= \int \cdots \int P(\alpha_t|\alpha_{t-1}) \cdots P(\alpha_1|\alpha_0) \ P(y_{t-1}|\alpha_{t-1}) \cdots P(y_1|\alpha_1) \mathrm{d}\alpha_{t-2} \cdots \mathrm{d}\alpha_1 \\ &= P(\alpha_t|\alpha_{t-1}) P(y_{t-1}|\alpha_{t-1}) \ \int \cdots \int P(\alpha_{t-1}|\alpha_{t-2}) \cdots P(\alpha_1|\alpha_0) \ P(y_{t-2}|\alpha_{t-2}) \cdots P(y_1|\alpha_1) \mathrm{d}\alpha_{t-2} \cdots \mathrm{d}\alpha_1 \\ &= P(\alpha_t|\alpha_{t-1}) P(y_{t-1}|\alpha_{t-1}) P(\alpha_{t-1}, Y_{t-2}), \end{aligned}$$

where the joint density $P(\alpha_{t-1}, Y_{t-2})$ is given by:

$$P(\alpha_{t-1}, Y_{t-2}) = \int \cdots \int P(\alpha_t|\alpha_{t-1}) \cdots P(\alpha_1|\alpha_0) \ P(y_{t-1}|\alpha_{t-1}) \cdots P(y_1|\alpha_1) \mathrm{d}\alpha_{t-2} \cdots \mathrm{d}\alpha_1.$$

To obtain the joint density of Y_{t-1}, $P(\alpha_t, \alpha_{t-1}, Y_{t-1})$ is integrated with respect to α_t and α_{t-1}, i.e.,

$$\begin{aligned} P(Y_{t-1}) &= \int\!\!\int P(\alpha_t|\alpha_{t-1}) P(y_{t-1}|\alpha_{t-1}) P(\alpha_{t-1}, Y_{t-2}) \mathrm{d}\alpha_t \mathrm{d}\alpha_{t-1} \\ &= \int P(y_{t-1}|\alpha_{t-1}) P(\alpha_{t-1}, Y_{t-2}) \mathrm{d}\alpha_{t-1}. \end{aligned}$$

Therefore, the conditional density function of α_t and α_{t-1} is as follows:

$$\begin{aligned}
P(\alpha_t, \alpha_{t-1}|Y_{t-1}) &= \frac{P(\alpha_t, \alpha_{t-1}, Y_{t-1})}{P(Y_{t-1})} \\
&= \frac{P(\alpha_t|\alpha_{t-1})P(y_{t-1}|\alpha_{t-1})P(\alpha_{t-1}, Y_{t-2})}{\displaystyle\int P(y_{t-1}|\alpha_{t-1})P(\alpha_{t-1}, Y_{t-2})\mathrm{d}\alpha_{t-1}} \\
&= P(\alpha_t|\alpha_{t-1})\frac{P(y_{t-1}|\alpha_{t-1})P(\alpha_{t-1}, Y_{t-2})}{\displaystyle\int P(y_{t-1}|\alpha_{t-1})P(\alpha_{t-1}, Y_{t-2})\mathrm{d}\alpha_{t-1}} \\
&= P(\alpha_t|\alpha_{t-1})\frac{P(y_{t-1}|\alpha_{t-1})P(\alpha_{t-1}|Y_{t-2})}{\displaystyle\int P(y_{t-1}|\alpha_{t-1})P(\alpha_{t-1}|Y_{t-2})\mathrm{d}\alpha_{t-1}} \\
&= P(\alpha_t|\alpha_{t-1})P(\alpha_{t-1}|Y_{t-1}),
\end{aligned}$$

where we can consider $P(\alpha_{t-1}|Y_{t-1})$ as follows:

$$P(\alpha_{t-1}|Y_{t-1}) = \frac{P(y_{t-1}|\alpha_{t-1})P(\alpha_{t-1}|Y_{t-2})}{\displaystyle\int P(y_{t-1}|\alpha_{t-1})P(\alpha_{t-1}|Y_{t-2})\mathrm{d}\alpha_{t-1}},$$

which is equivalent to equation (17). Accordingly, in equation (16), we have:

$$P(\alpha_t|Y_{t-1}) = P(\alpha_t|\alpha_{t-1})P(\alpha_{t-1}|Y_{t-1}).$$

Thus, the third equality in (16) and the fifth equality in (17) are proved, i.e., $P(\alpha_t|\alpha_{t-1}, Y_{t-1})$ is replaced by $P(\alpha_t|\alpha_{t-1})$ in equation (16) and $P(y_t|\alpha_t, Y_{t-1})$ is by $P(y_t|\alpha_t)$ in equation (17).

A.2 Conditional Normal Distribution

A.2.1 Preliminaries: Useful Lemmas

Lemma 1: Let A and C be $n \times n$ and $m \times m$ nonsingular matrices, and B be a $n \times m$ matrix. Then, we have the following formula.

$$(A + BCB')^{-1} = A^{-1} - A^{-1}B(C^{-1} + B'A^{-1}B)^{-1}B'A^{-1}.$$

Lemma 2: Let A, B and D be $k \times k$, $k \times g$ and $g \times g$ matrices, where A and D are nonsingular. Then, we have the following formula.

$$\begin{pmatrix} A & B \\ B' & D \end{pmatrix}^{-1} = \begin{pmatrix} X & Y \\ Y' & Z \end{pmatrix},$$

where X, Y and Z are as follows:

$$\begin{aligned}
X &= (A - BD^{-1}B')^{-1} = A^{-1} + A^{-1}B(D - B'A^{-1}B)^{-1}B'A^{-1}, \\
Y &= -(A - BD^{-1}B')^{-1}BD^{-1} = -A^{-1}B(D - B'A^{-1}B)^{-1}, \\
Z &= (D - B'A^{-1}B)^{-1} = D^{-1} + D^{-1}B'(A - BD^{-1}B')^{-1}BD^{-1}.
\end{aligned}$$

Note that Lemma 1 is used to obtain the second equality in X and Z.

Lemma 3: Let y and x be $k \times 1$ and $g \times 1$ vectors of random variables, which are distributed as the following bivariate normal density:

$$\begin{pmatrix} y \\ x \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}, \ \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix} \right),$$

which implies a joint distribution of x and y, i.e., $P(x, y)$.

Then, the conditional density of y given x (i.e., $P(y|x)$) follows a normal distribution with mean

$$\mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(x - \mu_x)$$

and variance

$$\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}.$$

Moreover, the marginal density of x (i.e., $P(x)$) is also a normal distribution with mean μ_x and variance Σ_{xx}. Accordingly, we have the following:

$$P(y|x) = \Phi\big(y - \mu_y - \Sigma_{yx}\Sigma_{xx}^{-1}(x - \mu_x), \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\big),$$

$$P(x) = \Phi(x - \mu_x, \Sigma_{xx}).$$

Note that $P(x, y) = P(y|x)P(x)$.

Lemma 4: Let A, B, C and D be $k \times k$, $k \times g$, $g \times k$ and $g \times g$ matrices, where A and D are nonsingular. Then, the following formula on the determinant is obtained.

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |D||A - BD^{-1}C|.$$

Applying the above formula to variance in Lemma 3, the following relationship holds:

$$\begin{vmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{vmatrix} = |\Sigma_{xx}||\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}|,$$

where A, B, C and D correspond to Σ_{yy}, Σ_{yx}, Σ_{xy} and Σ_{xx}, respectively.

A.2.2 Proof I

In this appendix, we prove the following equality:

$$\begin{aligned} &\int \Phi(\alpha_t - T_t\alpha_{t-1} - c_t, R_tQ_tR_t')\Phi(\alpha_{t-1} - a_{t-1|t-1}, \Sigma_{t-1|t-1})\mathrm{d}\alpha_{t-1} \\ &= \Phi(\alpha_t - T_ta_{t-1|t-1} - c_t, T_t\Sigma_{t-1|1-1}T_t' + R_tQ_tR_t'), \end{aligned}$$

which is in equation (18).

For simplicity of discussion, each variable is re-defined as follows:

$$\begin{aligned} &x = \alpha_{t-1} - a_{t-1|t-1}, \\ &\Sigma_{xx} = \Sigma_{t-1|t-1}, \\ &y = \alpha_t - T_ta_{t-1|t-1} - c_t, \\ &\Sigma_{yy} = R_tQ_tR_t', \\ &A = T_t. \end{aligned}$$

Substituting each variable, the equality to be proved is given by:

$$\int \Phi(y - Ax, \Sigma_{yy})\Phi(x, \Sigma_{xx})\mathrm{d}x = \Phi(y, A\Sigma_{xx}A' + \Sigma_{yy}).$$

The two normal distributions $\Phi(y - Ax, \Sigma_{yy})$ and $\Phi(x, \Sigma_{xx})$ are written as:

$$\Phi(y - Ax, \Sigma_{yy}) = (2\pi)^{-g/2}|\Sigma_{yy}|^{-1/2}\exp\left(-\frac{1}{2}(y - Ax)'\Sigma_{yy}^{-1}(y - Ax)\right),$$

$$\Phi(x, \Sigma_{xx}) = (2\pi)^{-k/2}|\Sigma_{xx}|^{-1/2}\exp\left(-\frac{1}{2}x'\Sigma_{xx}^{-1}x\right).$$

The dimensions of y and x are given by $g \times 1$ and $k \times 1$, respectively. A denotes a $g \times k$ matrix. Note that $g = k$ in Proof I but $g \neq k$ in Proof II.

A product of these two normal densities is transformed into:

$$\begin{aligned}
&\Phi(y - Ax, \Sigma_{yy})\Phi(x, \Sigma_{xx}) \\
&= (2\pi)^{-g/2}|\Sigma_{yy}|^{-1/2}\exp\left(-\frac{1}{2}(y - Ax)'\Sigma_{yy}^{-1}(y - Ax)\right)(2\pi)^{-k/2}|\Sigma_{xx}|^{-1/2}\exp\left(-\frac{1}{2}x'\Sigma_{xx}^{-1}x\right) \\
&= (2\pi)^{-(g+k)/2}|\Sigma_{yy}|^{-1/2}\ |\Sigma_{xx}|^{-1/2}\exp\left(-\frac{1}{2}\begin{pmatrix} y - Ax \\ x \end{pmatrix}'\begin{pmatrix} \Sigma_{yy} & 0 \\ 0 & \Sigma_{xx} \end{pmatrix}^{-1}\begin{pmatrix} y - Ax \\ x \end{pmatrix}\right) \\
&= (2\pi)^{-(g+k)/2}|\Sigma_{yy}|^{-1/2}\ |\Sigma_{xx}|^{-1/2}\exp\left(-\frac{1}{2}\begin{pmatrix} y \\ x \end{pmatrix}'\begin{pmatrix} I_g & -A \\ 0 & I_k \end{pmatrix}'\begin{pmatrix} \Sigma_{yy} & 0 \\ 0 & \Sigma_{xx} \end{pmatrix}^{-1}\begin{pmatrix} I_g & -A \\ 0 & I_k \end{pmatrix}\begin{pmatrix} y \\ x \end{pmatrix}\right) \\
&= (2\pi)^{-(g+k)/2}|\Sigma_{yy}|^{-1/2}\ |\Sigma_{xx}|^{-1/2}\exp\left(-\frac{1}{2}\begin{pmatrix} y \\ x \end{pmatrix}'\begin{pmatrix} I_g & A \\ 0 & I_k \end{pmatrix}'^{-1}\begin{pmatrix} \Sigma_{yy} & 0 \\ 0 & \Sigma_{xx} \end{pmatrix}^{-1}\begin{pmatrix} I_g & A \\ 0 & I_k \end{pmatrix}^{-1}\begin{pmatrix} y \\ x \end{pmatrix}\right) \\
&= (2\pi)^{-(g+k)/2}|\Sigma_{yy}|^{-1/2}\ |\Sigma_{xx}|^{-1/2}\exp\left(-\frac{1}{2}\begin{pmatrix} y \\ x \end{pmatrix}'\left(\begin{pmatrix} I_g & A \\ 0 & I_k \end{pmatrix}\begin{pmatrix} \Sigma_{yy} & 0 \\ 0 & \Sigma_{xx} \end{pmatrix}\begin{pmatrix} I_g & A \\ 0 & I_k \end{pmatrix}'\right)^{-1}\begin{pmatrix} y \\ x \end{pmatrix}\right) \\
&= (2\pi)^{-(g+k)/2}|\Sigma_{yy}|^{-1/2}\ |\Sigma_{xx}|^{-1/2}\exp\left(-\frac{1}{2}\begin{pmatrix} y \\ x \end{pmatrix}'\begin{pmatrix} A\Sigma_{xx}A' + \Sigma_{yy} & A\Sigma_{xx} \\ \Sigma_{xx}A' & \Sigma_{xx} \end{pmatrix}^{-1}\begin{pmatrix} y \\ x \end{pmatrix}\right).
\end{aligned}$$

Note that, in deriving the above equation, the inverse of the following matrix is used.

$$\begin{pmatrix} I_g & -A \\ 0 & I_k \end{pmatrix}^{-1} = \begin{pmatrix} I_g & A \\ 0 & I_k \end{pmatrix}.$$

Furthermore, using Lemma 4, we have:

$$\begin{vmatrix} A\Sigma_{xx}A' + \Sigma_{yy} & A\Sigma_{xx} \\ \Sigma_{xx}A' & \Sigma_{xx} \end{vmatrix} = |\Sigma_{xx}||\Sigma_{yy}|.$$

Accordingly, the joint distribution of x and y is represented by the following bivariate normal distribution.

$$\begin{pmatrix} y \\ x \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} A\Sigma_{xx}A' + \Sigma_{yy} & A\Sigma_{xx} \\ \Sigma_{xx}A' & \Sigma_{xx} \end{pmatrix}\right), \qquad (31)$$

From Lemma 3, the marginal density of y is given by:

$$P(y) = \Phi(y, A\Sigma_{xx}A' + \Sigma_{yy}),$$

which implies that

$$P(\alpha_t|Y_{t-1}) = \Phi(\alpha_t - T_t a_{t-1|t-1} - c_t, T_t\Sigma_{t-1|1-1}T_t' + R_tQ_tR_t').$$

Thus, the third equality in equation (18) is derived.

A.2.3 Proof II

We prove that the following equation holds.

$$
\begin{aligned}
&\Phi(y_t - Z_t\alpha_t - d_t, S_tH_tS_t')\Phi(\alpha_t - a_{t|t-1}, \Sigma_{t|t-1}) \\
&= \Phi\big(\alpha_t - a_{t|t-1} - K_t(y_t - y_{t|t-1}), \Sigma_{t|t-1} - K_tF_{t|t-1}K_t'\big)\ \Phi(y_t - y_{t|t-1}, F_{t|t-1}),
\end{aligned}
$$

which is utilized in equation (19).

Similarly, we re-define each variable as follows:

$$
\begin{aligned}
&x = \alpha_t - a_{t|t-1}, \\
&\Sigma_{xx} = \Sigma_{t|t-1}, \\
&y = y_t - Z_ta_{t|t-1} - d_t, \\
&\Sigma_{yy} = S_tH_tS_t', \\
&A = Z_t,
\end{aligned}
$$

where x is a $k \times 1$ vector, y is a $g \times 1$ vector, and A is a $g \times k$ matrix.

By taking into account the following three equations,

$$
\begin{aligned}
&y_{t|t-1} = Z_ta_{t|t-1} + d_t, \\
&F_{t|t-1} = Z_t\Sigma_{t|t-1}Z_t' + S_tH_tS_t', \\
&K_t = \Sigma_{t|t-1}Z_t'F_{t|t-1}^{-1},
\end{aligned}
$$

the equation to be proved is represented by:

$$
\begin{aligned}
&\Phi(y - Ax, \Sigma_{yy})\Phi(x, \Sigma_{xx}) \\
&= \Phi\big(x - \Sigma_{xx}A'(A\Sigma_{xx}A' + \Sigma_{yy})^{-1}y,\ \Sigma_{xx} - \Sigma_{xx}A'(A\Sigma_{xx}A' + \Sigma_{yy})^{-1}A\Sigma_{xx}\big)\Phi(y, A\Sigma_{xx}A' + \Sigma_{yy}).
\end{aligned}
$$

We can prove the above equation in the exactly same fashion as Proof I, and accordingly the joint density of x and y (i.e., $P(x, y)$) follows the same distribution as equation (31).

Using Lemma 3, the conditional density of x given y (i.e., $P(x|y)$) and the marginal density of y (i.e., $P(y)$) are derived as:

$$
\begin{aligned}
P(x, y) &= P(x|y)P(y) \\
&= \Phi\big(x - \Sigma_{xx}A'(A\Sigma_{xx}A' + \Sigma_{yy})^{-1}y,\ \Sigma_{xx} - \Sigma_{xx}A'(A\Sigma_{xx}A' + \Sigma_{yy})^{-1}A\Sigma_{xx}\big)\Phi(y, A\Sigma_{xx}A' + \Sigma_{yy}),
\end{aligned}
$$

where

$$
\begin{aligned}
&P(x|y) = \Phi\big(x - \Sigma_{xx}A'(A\Sigma_{xx}A' + \Sigma_{yy})^{-1}y,\ \Sigma_{xx} - \Sigma_{xx}A'(A\Sigma_{xx}A' + \Sigma_{yy})^{-1}A\Sigma_{xx}\big), \\
&P(y) = \Phi(y, A\Sigma_{xx}A' + \Sigma_{yy}).
\end{aligned}
$$

Therefore, the following equation is obtained.

$$
\begin{aligned}
&\Phi(y_t - Z_t\alpha_t - d_t, S_tH_tS_t')\Phi(\alpha_t - a_{t|t-1}, \Sigma_{t|t-1}) \\
&= \Phi\big(\alpha_t - a_{t|t-1} - K_t(y_t - y_{t|t-1}), \Sigma_{t|t-1} - K_tF_{t|t-1}K_t'\big)\ \Phi(y_t - y_{t|t-1}, F_{t|t-1}).
\end{aligned}
$$

Note that we have the following:

$$
\Phi(y_t - y_{t|t-1}, F_{t|t-1}) = \int P(y_t|\alpha_t)P(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t.
$$

Thus, we can obtain the third equality of equation (19).

References

Anderson, B.D.O. and Moore, J.B., 1979, *Optimal Filtering*, Prentice-Hall, New York.

Aoki, M., 1987, *State Space Modeling of Time Series*, Springer-Verlag.

Belsley, D.A., 1973, " On the determination of Systematic Parameter Variation in the Linear Regression Model, " *Annals of Economic and Social Measurement*, Vol.2, pp.487 – 494.

Belsley, D.A. and Kuh, E., 1973, " Time-Varying Parameter Structures: An Overview, " *Annals of Economic and Social Measurement*, Vol.2, No.4, pp.375 – 379.

Branson, W.H., 1979, *Macroeconomic Theory and Policy* (second edition), Haper & Row, Publishers, Inc.

Brockwell, P.A. and Davis, R.A., 1987, *Time Series Theory and Models*, Springer-Verlag.

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

Burridge, P. and Wallis, K.F., 1988, " Prediction Theory for Autoregressive Moving Average Processes, " *Econometric Reviews*, Vol.7, No.1, pp.65 – 95.

Conrad, W. and Corrado, C., 1979, " Application of the Kalman Filter to Revisions in Monthly Retail Sales Estimates, " *Journal of Economic Dynamic and Control*, Vol.1, pp.177 – 198.

Cooley, T.F., 1977, " Generalized Least Squares Applied to Time Varying Parameter Models: A Comment, " *Annals of Economic and Social Measurement*, Vol.6, No.3, pp.313 – 314.

Cooley, T.F. and Prescott, E.C., 1976, " Estimation in the presence of stochastic parameter variation, " *Econometrica*, Vol.44, pp.167 – 183.

Cooley, T.F., Rosenberg, B. and Wall, K.D., 1977, " A Note on Optimal Smoothing for Time Varying Coefficient Problems, " *Annals of Economic and Social Measurement*, Vol.6, No.4, pp.453 – 456.

Cooper, J.P., 1973, " Time-Varying Regression Coefficients: A Mixed Estimation Approach and Operational Limitations of the General Markov Structure, " *Annals of Economic and Social Measurement*, Vol.2, No.4, pp.525 – 530.

Chow, G.C., 1983, *Econometrics*, McGraw-Hill Book Company.

Dempster, A.P., Laird, N.M. and Rubin, D.B., 1977, " Maximum Likelihood from Incomplete Data via the EM Algorithm, " *Journal of the Royal Statistical Society*, Ser.B, Vol.39, pp.1 – 38 (with discussion).

Diderrich, G.T., 1985, " The Kalman Filter from the Perspective of Goldberger-Theil Estimators, " *The American Statistician*, Vol.39, pp.193 – 198.

Dziechciarz, J., 1989, " Changing and Random Coefficient Models. A Survey, " in *Statistical Analysis and Forecasting of Economic Structural Change*, edited by Hackl, P., Springer-Verlag.

Engle, R.F. and Watson, M.W., 1987, " The Kalman Filter: Applications to Forecasting and Rational expectations Models, " in *Advances in Econometrics, Fifth World Congress, Vol.I*, Cambridge University Press.

Friedman, M., 1957, *A Theory of the Consumption Function*, Princeton University Press.

Garbade, K., 1976, " Two Methods of Examining the Stability of Regression Coefficients, " *Journal of the American Statistical Association*, Vol.72, pp.54 – 63.

Gardner, G., Harvey, A.C. and Phillips, G.D.A., 1980, " An Algorithm for Maximum Likelihood Estimation Autoregressive-Moving Average Models by means of Kalman Filtering, " *Applied Statistics*, Vol.29, No.3, pp.311 – 322.

Gelb, A., 1974, *Applied Optimal Estimation*, MIT Press.

Hall, R.E., 1978, “ Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence, ” *Journal of Political Economy*, Vol.86, No.6, pp.971 – 987.

Hannan, E.J. and Deistler, M., 1988, *The Statistical Theory of Linear System*, John Wiley & Sons.

Harvey, A.C., 1981, *Time Series Models*, Philip Allen Publishers Limited, Oxford.

Harvey, A.C., 1987, “ Applications of the Kalman Filter in Econometrics, ” in *Advances in Econometrics, Fifth World Congress, Vol.I*, Cambridge University Press.

Harvey, A.C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Howrey, E.P., 1978, “ The Use of Preliminary Data in Econometric Forecasting, ” *The Review of Economics and Statistics*, Vol.60, pp.193 – 200.

Howrey, E.P., 1984, “ Data Revision, Reconstruction, and Prediction: An Application to Inventory Investment, ” *The Review of Economics and Statistics*, Vol.66, pp.386 – 393.

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

Johnston, J., 1984, *Econometric Methods* (third edition), McGraw-Hill Book Company.

Kalman, R.E., 1960, “ A New Approach to Linear Filtering and Prediction Problems, ” *Journal of Basic Engineering, Transactions ASME*, Ser.D, Vol.82, pp.35 – 45.

Kalman, R.E. and Bucy, R.S., 1961, “ New Results in Linear Filtering and Prediction Theory, ” *Journal of Basic Engineering, Transactions ASME*, Ser.D, Vol.83, pp.95 – 108.

Kitagawa, G., 1987, “ Non-Gaussian State-Space Modeling of Nonstationary Time Series, ” *Journal of the American Statistical Association*, Vol.82, pp.1032 – 1063 (with discussion).

Kirchen, A., 1988, *Schäzung zeitveränderlicher Strukturparameter in ökonometrischen Prognosemodellen*, Frankfurt/ Main: Athenäum.

Kramer, S.C. and Sorenson, H.W., 1988, “ Recursive Bayesian Estimation Using Piece-wise Constant Approximations, ” *Automatica*, Vol.24, No.6, pp.789 – 801.

Laumas, G.S. and Mehra, Y.P., 1976, “ The Stability of the Demand for Money Function: The Evidence from Quarterly Data, ” *The Review of Economics and Statistics*, Vol.58, pp.464 – 468.

Mariano, R.S. and Tanizaki, H., 1995, “ Prediction of Final Data with Use of Preliminary and/or Revised Data, ” *Journal of Forecasting*, Vol.14, No.4, pp.351 – 380.

McNelis, P.D. and Neftci, S.N., 1983, “ Policy-dependent Parameters in the Presence of Optimal Learning: An Application of Kalman Filtering to the Fair and Sargent Supply-side Equations, ” *The Review of Economics and Statistics*, Vol.65, pp.296 – 306.

Nicholls, D.F. and Pagan, A.R., 1985, “ Varying Coefficient Regression, ” in *Handbook of Statistics*, Vol.5, edited by Hannan, E.J., Krishnaiah, P.R. and Rao, M.M., Elsevier Science Publishers B.V., pp.413 – 449.

Pagan, A.R., 1975, “ A Note on the Extraction of Components from Time Series, ” *Econometrica*, Vol.43, pp.163 – 168.

Pagan, A.R., 1980, “ Some Identification and Estimation Results for Regression Models with Stochastically Varying, ” *Journal of Econometrics*, Vol.13, pp.343 – 363.

Rosenberg, B., 1977, “ Estimation Error Covariance in Regression with Sequentially Varying Parameters, ” *Annals of Economic and Social Measurement*, Vol.6, No.4, pp.457 – 462.

Rund, P.A., 1991, “ Extensions of Estimation Methods Using the EM Algorithm, ” *Journal of Econometrics*, Vol.49, pp.305 – 341.

Sarris, A.H., 1973, “ A Bayesian Approach to Estimation of Time Varying Regression Coefficients, ” *Annals of Economic and Social Measurement*, Vol.2, No.4, pp.501 – 523.

Sant, D.T., 1977, “ Generalized Least Squares Applied to Time Varying Parameter Models, ” *Annals of Economic and Measurement*, Vol.6, No.3, pp.301 – 311.

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

Tanizaki, H., 1987, “ Theory and Application of Kalman Filter Model, ” Master Thesis, Kobe University (in Japanese).

Tanizaki, H., 1989, “ The Kalman Filter Model under the Assumption of the First-Order Autoregressive Process in the Disturbance Terms, ” *Economics Letters*, Vol.31, No.2, pp.145 – 149.

Tanizaki, H., 1993a, “ Kalman Filter Model with Qualitative Dependent Variable, ” *The Review of Economics and Statistics*, Vol.75, No.4, pp.747 – 752.

Tanizaki, H., 1993b, *Application of State-Space Model to Economics: Estimation of U.S. and Japanese Macro-Econometric Model by Time-Varying Parameter Model* (Research Monograph in Department of Economics, Kobe-Gakuin University, No.9), Nippon-Hyoron-Sya (in Japanese).

Tanizaki, H., 1996, *Nonlinear Filters: Estimation and Applications* (Second, Revised and Enlarged Edition), Springer-Verlag.

Watson, M.W. and Engle, R.F., 1983, “ Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression Models, ” *Journal of Econometrics*, Vol.23, pp.385 – 400.

Watanabe, N., 1985, “ Notes on the Kalman Filter with Estimated Parameters, ” *Journal of Time Series*, Vol.6, No.4, pp.269 – 278.