17 State-Space Models and the Kalman Filter

The state space representation is a flexible technique originally developed in automatic control engineering to represent, model, and control dynamic systems. Thereby we summarize the unobserved or partially observed state of the system in period t by an m-dimensional vector X_t. The evolution of the state is then described by a VAR of order one usually called the state equation. A second equation describes the connection between the state and the observations given by a n-dimensional vector Y_t. Despite its simple structure, state space models encompass a large variety of model classes: VARMA, respectively VARIMA models,[1] unobserved-component models, factor models, structural time series models which decompose a given time series into a trend, a seasonal, and a cyclical component, models with measurement errors, VAR models with time-varying parameters, etc.

From a technical point of view, the main advantage of state space modeling is the unified treatment of estimation, forecasting, and smoothing. At the center of the analysis stands the Kalman-filter named after its inventor Rudolf Emil Kálmán (Kalman 1960, 1963). He developed a projection based algorithm which recursively produces a statistically optimal estimate of the state. The versatility and the ease of implementation have made the Kalman filter an increasingly popular tool also in the economically oriented times series literature. Here we present just an introduction to the subject and refer to Anderson and Moore (1979), Brockwell and Davis (1991, Chapter 12), Brockwell and Davis (1996, Chapter 8), Hamilton (1994b, Chapter 13), Hamilton (1994a), Hannan and Deistler (1988), or Harvey (1989), and in particular to Durbin and Koopman (2011) and Kim and Nelson (1999) for extensive reviews and further details.

[1] VARIMA models stand for vector autoregressive integrated moving-average models.

K. Neusser, *Time Series Econometrics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-319-32862-1_17

17.1 The State Space Model

We consider a dynamical system whose state at each point in time t is determined by a vector X_t. The evolution of the system over time is then described by a *state equation*. The state is, however, unobserved or only partly observed to the outside observer. Thus, a second equation, called the *observation equation*, is needed to describe the connection of the state to the observations. This relation may be subject to measurement errors. The equation system consisting of state and observation equation is called a state space model which is visualized in Fig. 17.1. The state equation typically consists of a VAR model of order one whereas the observation equation has the structure of multivariate linear regression model.[2] Despite the simplicity of each of these two components, their combination is very versatile and able to represent a great variety of models.

In the case of time invariant coefficients[3] we can set up these two equations as follows:

$$\text{state equation:} \qquad X_{t+1} = FX_t + V_{t+1}, \qquad t = 1, 2, \ldots \tag{17.1}$$

$$\text{observation equation:} \qquad Y_t = A + GX_t + W_t, \qquad t = 1, 2, \ldots \tag{17.2}$$

Thereby X_t denotes an m-dimensional vector which describes the state of the system in period t. The evolution of the state is represented as a vector autoregressive model

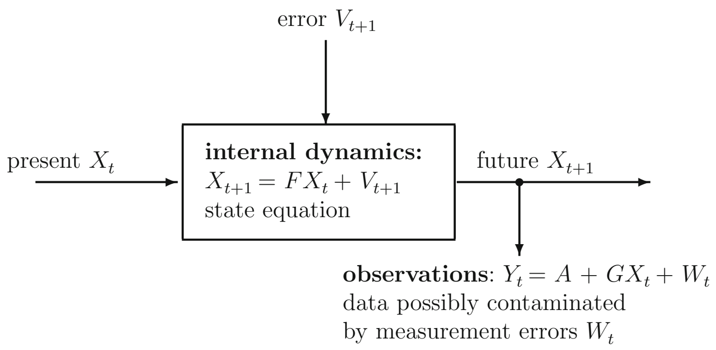

Fig. 17.1 State space model

[2]We will focus on linear dynamic models only. With the availability of fast and cheap computing facilities, non-linear approaches have gained some popularity. See Durbin and Koopman (2011) for an exposition.

[3]For the ease of exposition, we will present first the time-invariant case and analyze the case of time-varying coefficients later.

of order one with coefficient matrix F and disturbances V_{t+1}.[4] As we assume that the state X_t is unobservable or at least partly unobservable, we need a second equation which relates the state to the observations. In particular, we assume that there is a linear time-invariant relation given by A and G of the n-dimensional vector of observations, Y_t, to the state X_t. This relation may be contaminated by measurement errors W_t. The system is initialized in period $t = 1$.

We make the following simplifying assumption of the state space model represented by Eqs. (17.1) and (17.2).

(i) $\{V_t\} \sim \mathrm{WN}(0, Q)$ where Q is a constant nonnegative definite $m \times m$ matrix.
(ii) $\{W_t\} \sim \mathrm{WN}(0, R)$ where R is a constant nonnegative definite $n \times n$ matrix.
(iii) The two disturbances are uncorrelated with each other at all leads and lags, i.e.:

$$\mathbb{E}(W_s V_t') = 0, \qquad \text{for all } t \text{ and } s.$$

(iv) V_t and W_t are multivariate normally distributed.
(v) X_1 is uncorrelated with V_t as well as with W_t, $t = 1, 2, \ldots$

Remark 17.1. In a more general context, we can make both covariance matrices Q and R time-varying and allow for contemporaneous correlations between V_t and W_t (see example Sect. 17.4.1).

Remark 17.2. As both the state and the observation equation may include identities, the covariance matrices need not be positive definite. They can be non-negative definite.

Remark 17.3. Neither $\{X_t\}$ nor $\{Y_t\}$ are assumed to be stationary.

Remark 17.4. The specification of the state equation and the normality assumption imply that the sequence $\{X_1, V_1, V_2, \ldots\}$ is independent so that the conditional distribution X_{t+1} given $X_t, X_{t-1}, \ldots, X_1$ equals the conditional distribution of X_{t+1} given X_t. Thus, the process $\{X_t\}$ satisfies the *Markov property*. As the dimension of the state vector X_t is arbitrary, it can be expanded in such a way as to encompass every component X_{t-1} for any t (see, for example, the state space representation of a VAR(p) model with $p > 1$). However, there remains the problem of the smallest dimension of the state vector (see Sect. 17.3.2).

Remark 17.5. The state space representation is not unique. Defining, for example, a new state vector $\tilde{X}_t$ by multiplying X_t with an invertible matrix P, i.e. $\tilde{X}_t = PX_t$, all properties of the system remain unchanged. Naturally, we must redefine all the system matrices accordingly: $\tilde{F} = PFP^{-1}$, $\tilde{Q} = PQP'$, $\tilde{G} = GP^{-1}$.

[4] In control theory the state equation (17.1) is amended by an additional term HU_t which represents the effect of control variables U_t. These exogenous controls are used to regulate the system.

Given X_1, we can iterate the state equation forward to arrive at:

$$X_t = F^{t-1}X_1 + \sum_{j=1}^{t-1} F^{j-1}V_{t+1-j}, \qquad t = 1, 2, \ldots$$

$$Y_t = A + GF^{t-1}X_1 + \sum_{j=1}^{t-1} GF^{j-1}V_{t+1-j} + W_t, \qquad t = 1, 2, \ldots$$

The state equation is called stable or causal if all eigenvalues of F are inside the unit circle which is equivalent that all roots of $\det(\mathrm{I}_m - Fz) = 0$ are outside the unit circle (see Sect. 12.3). In this case the state equation has a unique stationary solution:

$$X_t = \sum_{j=0}^{\infty} F^{j-1}V_{t+1-j}. \tag{17.3}$$

The process $\{Y_t\}$ is therefore also stationary and we have:

$$Y_t = A + \sum_{j=0}^{\infty} GF^{j-1}V_{t+1-j} + W_t. \tag{17.4}$$

In the case of a stationary state space model, we may do without an initialization period and take $t \in \mathbb{Z}$.

In the case of a stable state equation, we can easily deduce the covariance function for $\{X_t\}$, $\Gamma_X(h)$, $h = 0, 1, 2, \ldots$ According to Sect. 12.4 it holds that:

$$\Gamma_X(0) = F\Gamma_X(0)F' + Q,$$

$$\Gamma_X(h) = F^h\Gamma_X(0), \qquad h = 1, 2, \ldots$$

where $\Gamma_X(0)$ is uniquely determined given the stability assumption. Similarly, we can derive the covariance function for the observation vector, $\Gamma_Y(h)$, $h = 0, 1, 2, \ldots$:

$$\Gamma_Y(0) = G\Gamma_X(0)G' + R,$$

$$\Gamma_Y(h) = GF^h\Gamma_X(0)G', \qquad h = 1, 2, \ldots$$

17.1.1 Examples

The following examples should illustrate the versatility of the state space model and demonstrate how many economically relevant models can be represented in this form.

VAR(p) Process

Suppose that $\{Y_t\}$ follows a n-dimensional VAR(p) process given by $\Phi(L)Y_t = Z_t$, respectively by $Y_t = \Phi_1 Y_{t-1} + \ldots + \Phi_p Y_{t-p} + Z_t$, with $Z_t \sim \text{WN}(0, \Sigma)$. Then the companion form of the VAR(p) process (see Sect. 12.2) just represents the state equation (17.1):

$$X_{t+1} = \begin{pmatrix} Y_{t+1} \\ Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Phi_2 & \ldots & \Phi_{p-1} & \Phi_p \\ I_n & 0 & \ldots & 0 & 0 \\ 0 & I_n & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & I_n & 0 \end{pmatrix} \begin{pmatrix} Y_t \\ Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{pmatrix} + \begin{pmatrix} Z_{t+1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= FX_t + V_{t+1},$$

with $V_{t+1} = (Z'_{t+1}, 0, 0, \ldots, 0)'$ and $Q = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}$. The observation equation is just an identity because all components of X_t are observable:

$$Y_t = (\mathrm{I}_n, 0, 0, \ldots, 0)X_t = GX_t.$$

Thus, $G = (I_n, 0, 0, \ldots, 0)$ and $R = 0$. Assuming that X_t is already mean adjusted, $A = 0$.

ARMA(1,1) Process

The representation of ARMA processes as a state space model is more involved when moving-average terms are involved. Let $\{Y_t\}$ be an ARMA(1,1) process defined by the stochastic difference equation $Y_t = \phi Y_{t-1} + Z_t + \theta Z_{t-1}$ with $Z_t \sim \text{WN}(0, \sigma^2)$ and $\phi\theta \neq 0$.

Define $\{X_t\}$ as the AR(1) process defined by the stochastic difference equation $X_t - \phi X_{t-1} = Z_t$ and $\mathbf{X}_t = (X_t, X_{t-1})'$ as the state vector, then we can write the observation equation as:

$$Y_t = (1, \theta)\mathbf{X}_t = G\mathbf{X}_t$$

with $R = 0$. The state equation is then

$$\mathbf{X}_{t+1} = \begin{pmatrix} X_{t+1} \\ X_t \end{pmatrix} = \begin{pmatrix} \phi & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} Z_{t+1} \\ 0 \end{pmatrix} = F\mathbf{X}_t + V_{t+1},$$

where $Q = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix}$. It is easy to verify that the so defined process $\{Y_t\}$ satisfies the stochastic difference equation $Y_t = \phi Y_{t-1} + Z_t + \theta Z_{t-1}$. Indeed $Y_t - \phi Y_{t-1} = (1, \theta)\mathbf{X}_t - \phi(1, \theta)\mathbf{X}_{t-1} = X_t + \theta X_{t-1} - \phi X_{t-1} - \theta\phi X_{t-2} = (X_t - \phi X_{t-1}) + \theta(X_{t-1} - \phi X_{t-2}) = Z_t + \theta Z_{t-1}$.

If $|\phi| < 1$, the state equation defines a causal process $\{X_t\}$ so that the unique stationary solution is given by Eq. (17.3). This implies a stationary solution for $\{Y_t\}$ too. It is thus easy to verify if this solution equals the unique solution of the ARMA stochastic difference equation.

The state space representation of an ARMA model is not unique. An alternative representation in the case of a causal system is given by:

$$X_{t+1} = \phi X_t + (\phi + \theta)Z_t = FX_t + V_{t+1}$$
$$Y_t = X_t + Z_t = X_t + W_t.$$

Note that in this representation the dimension of the state vector is reduced from two to one. Moreover, the two disturbances $V_{t+1} = (\phi + \theta)Z_t$ and $W_t = Z_t$ are perfectly correlated.

ARMA(p,q) Process

It is straightforward to extend the above representation to ARMA(p,q) models.[5] Let $\{Y_t\}$ be defined by the following stochastic difference equation:

$$\Phi(\mathrm{L})Y_t = \Theta(\mathrm{L})Z_t \qquad \text{with } Z_t \sim \mathrm{WN}(0, \sigma^2) \text{ and } \phi_p\theta_q \neq 0.$$

Define r as $r = \max\{p, q + 1\}$ and set $\phi_j = 0$ for $j > p$ and $\theta_j = 0$ for $j > q$. Then, we can set up the following state space representation with state vector $\mathbf{X}_t$ and observation equation

$$Y_t = (1, \theta_1, \ldots, \theta_{r-1})\mathbf{X}_t$$

where the state vector equals $\mathbf{X}_t = (X_t, \ldots, X_{t-r+2}, X_{t-r+1})'$ and where $\{X_t\}$ follows an AR(p) process $\Phi(\mathrm{L})X_t = Z_t$. The AR(p) process can be transformed into companion form to arrive at the state equation:

$$\mathbf{X}_{t+1} = \begin{pmatrix} \phi_1 & \phi_2 & \ldots & \phi_{r-1} & \phi_r \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \mathbf{X}_t + \begin{pmatrix} Z_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Missing Observations

The state space approach is best suited to deal with missing observations. However, in this situation the coefficient matrices are no longer constant, but time-varying. Consider the following simple example of an AR(1) process for which we have

[5] See also Exercise 17.2.

only observations for the periods $t = 1, \ldots, 100$ and $t = 102, \ldots, 200$, but not for period $t = 101$ which is missing. This situation can be represented in state space form as follows:

$$X_{t+1} = \phi X_t + Z_t$$
$$Y_t = G_t X_t + W_t$$
$$G_t = \begin{cases} 1, \; t = 1, \ldots, 100, 102, \ldots, 200; \\ 0, \; t = 101. \end{cases}$$
$$R_t = \begin{cases} 0, & t = 1, \ldots, 100, 102, \ldots, 200; \\ c > 0, & t = 101. \end{cases}$$

This means that $W_t = 0$ and that $Y_t = X_t$ for all t except for $t = 101$. For the missing observation, we have $G_{101} = Y_{101} = 0$. The variance for this observation is set to $R_{101} = c > 0$.

The same idea can be used to obtain quarterly data when only yearly data are available. This problem typically arises in statistical offices which have to produce, for example, quarterly GDP data from yearly observations incorporating quarterly information from indicator variables (see Sect. 17.4.1). More detailed analysis for the case of missing data can be found in Harvey and Pierce (1984) and Brockwell and Davis (1991, Chapter 12.3).

Time-Varying Coefficients

Consider the regression model with time-varying parameter vector β_t:

$$Y_t = x_t' \beta_t + W_t \tag{17.5}$$

where Y_t is an observed dependent variable, x_t is a K-vector of exogenous regressors, and W_t is a white noise error term. Depending on the specification of the evolution of β_t several models have been proposed in the literature:

Hildreth-Houck :	$\beta_t = \bar{\beta} + v_t$
Harvey-Phillips:	$\beta_t - \bar{\beta} = F(\beta_t - \bar{\beta}) + v_t$
Cooley-Prescott:	$\beta_t = \beta_t^p + v_{1t}$
	$\beta_t^p = \beta_{t-1}^p + v_{2t}$

where v_t, v_{1t}, and v_{2t} are white noise error terms. In the first specification, proposed originally proposed by Hildreth and Houck (1968), the parameter vector is in each period just a random from a distribution with mean $\bar{\beta}$ and variance given by the variance of v_t. Departures from the mean are seen as being only of a transitory nature. In the specification by Harvey and Phillips (1982), assuming that all eigenvalues of F are strictly smaller than one in absolute value, the parameter vector is a mean reverting VAR of order one. In this case, the departures from the mean can have a longer duration depending on the eigenvalues of F. The last specification due to Cooley and Prescott (1973, 1976) views the parameter vector as being subject to transitory and permanent shifts. Whereas shifts in v_{1t} have only a transitory effect on β_t, movements in v_{2t} result in permanent effects.

In the Cooley-Prescott specification, for example, the state is given by $X_t = (\beta_t', \beta_t^{p\prime})'$ and the state equation can be written as:

$$X_{t+1} = \begin{pmatrix} \beta_{t+1} \\ \beta_{t+1}^p \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}}_{=F} \begin{pmatrix} \beta_t \\ \beta_t^p \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}$$

The observation equation then becomes:

$$Y_t = y_t = \begin{pmatrix} x_t \\ 0 \end{pmatrix}' X_t + W_t$$

Thus, $A = 0$ and $G_t = (x_t', 0)$. Note that this an example of a state space model with time-varying coefficients. In Sect. 18.2, we will discuss time-varying coefficient models in the context of VAR models.

Structural Time Series Analysis

An important application of the state space representation in economics is the decomposition of a given time series into several components: trend, cycle, season and irregular component. This type of analysis is usually coined structural time series analysis (See Harvey 1989; Mills 2003). Consider, for example, the additive decomposition of a time series $\{Y_t\}$ into a trend T_t, a seasonal component S_t, a cyclical component $\{C_t\}$, and an irregular or cyclical component W_t:

$$Y_t = T_t + S_t + C_t + W_t.$$

The above equation relates the observed time series to its unobserved components and is called the *basic structural model* (BSM) (Harvey 1989).

The state space representation is derived in several steps. Consider first the case with no seasonal and no cyclical component. The trend is typically viewed as a random walk with time-varying drift δ_{t-1}:

$$T_t = \delta_{t-1} + T_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$$
$$\delta_t = \delta_{t-1} + \xi_t, \qquad \xi_t \sim \mathrm{WN}(0, \sigma_\xi^2).$$

The second equation models the drift as a random walk. The two disturbances $\{\varepsilon_t\}$ and $\{\xi_t\}$ are assumed to be uncorrelated with each other and with $\{W_t\}$. Defining the state vector $X_t^{(T)}$ as $X_t^{(T)} = (T_t, \delta_t)'$, the state and the observation equations become:

$$X_{t+1}^{(T)} = \begin{pmatrix} T_{t+1} \\ \delta_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} T_t \\ \delta_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{t+1} \\ \xi_{t+1} \end{pmatrix} = F^{(T)} X_t^{(T)} + V_{t+1}^{(T)}$$
$$Y_t = (1, 0) X_t^{(T)} + W_t$$

with $W_t \sim \text{WN}(0, \sigma_W^2)$. This representation is called the local linear trend (LLT) model and implies that $\{Y_t\}$ follows an ARIMA(0,2,2) process (see Exercise 17.5.1).

In the special case of a constant drift equal to δ, $\sigma_\xi^2 = 0$ and we have that $\Delta Y_t = \delta + \varepsilon_t + W_t - W_{t-1}$. $\{\Delta Y_t\}$ therefore follows a MA(1) process with $\rho(1) = -\sigma_W^2/(\sigma_\varepsilon^2 + 2\sigma_W^2) = -(2+\kappa)^{-1}$ where $\kappa = \sigma_\varepsilon^2/\sigma_W^2$ is called the signal-to-noise ratio. Note that the first order autocorrelation is necessarily negative. Thus, this model is not suited for time series with positive first order autocorrelation in its first differences.

The seasonal component is characterized by two conditions $S_t = S_{t-d}$ and $\sum_{t=1}^{d} S_t = 0$ where d denotes the frequency of the data.[6] Given starting values $S_1, S_0, S_{-1}, \ldots, S_{-d+3}$, the subsequent values can be computed recursively as:

$$S_{t+1} = -S_t - \ldots - S_{t-d+2} + \eta_{t+1}, \qquad t = 1, 2, \ldots$$

where a noise $\eta_t \sim \text{WN}(0, \sigma_\eta^2)$ is taken into account.[7] The state vector related to the seasonal component, $X_t^{(S)}$, is defined as $X_t^{(S)} = (S_t, S_{t-1}, \ldots, S_{t-d+2})'$ which gives the state equation

$$X_{t+1}^{(S)} = \begin{pmatrix} -1 & -1 & \ldots & -1 & -1 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix} X_t^{(S)} + \begin{pmatrix} \eta_{t+1} \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} = F^{(S)} X_t^{(S)} + V_{t+1}^{(S)}$$

with $Q^{(S)} = \text{diag}(\sigma_\eta^2, 0, \ldots, 0)$.

Combining the trend and the seasonal model to an overall model with state vector X_t given by $X_t = \left(X_t^{(T)\prime}, X_t^{(S)\prime}\right)'$, we arrive at the state equation:

$$X_{t+1} = \begin{pmatrix} F^{(T)} & 0 \\ 0 & F^{(S)} \end{pmatrix} X_t + \begin{pmatrix} V_{t+1}^{(T)} \\ V_{t+1}^{(S)} \end{pmatrix} = F X_t + V_{t+1}$$

with $Q = \text{diag}(\sigma_\varepsilon^2, \sigma_\delta^2, \sigma_\eta^2, 0, \ldots, 0)$. The observation equation then is:

$$Y_t = \begin{pmatrix} 1 & 0 & 1 & 0 & \ldots & 0 \end{pmatrix} X_t + W_t$$

with $R = \sigma_W^2$.

Finally, we can add a cyclical component $\{C_t\}$ which is modeled as a harmonic process (see Sect. 6.2) with frequency λ_C, respectively periodicity $2\pi/\lambda_C$:

$$C_t = A\cos(\lambda_C t) + B\sin(\lambda_C t)$$

[6] Four in the case of quarterly and twelve in the case of monthly observations.

[7] Alternative seasonal models can be found in Harvey (1989) and Hylleberg (1986).

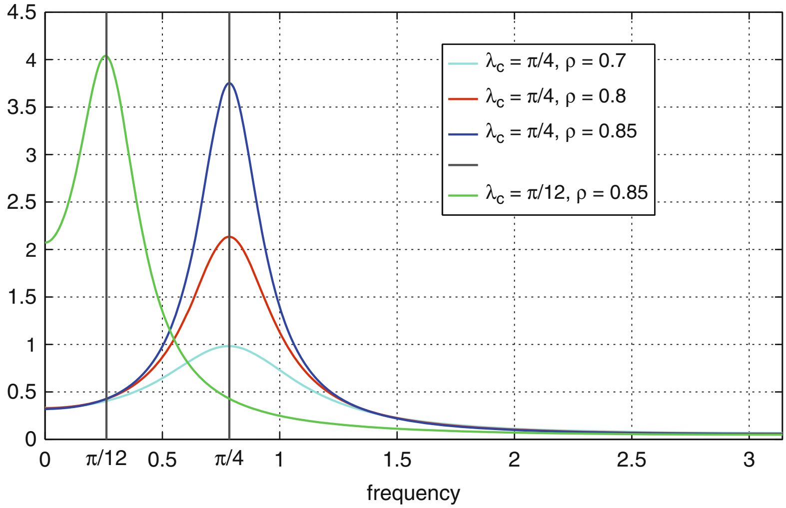

Fig. 17.2 Spectral density of the cyclical component for different values of λ_C and ρ

Following Harvey (1989, p.39), we let the parameters A and B evolve over time by introducing the recursion

$$\begin{pmatrix} C_{t+1} \\ C^*_{t+1} \end{pmatrix} = \rho \begin{pmatrix} \cos\lambda_C & \sin\lambda_C \\ -\sin\lambda_C & \cos\lambda_C \end{pmatrix} \begin{pmatrix} C_t \\ C^*_t \end{pmatrix} + \begin{pmatrix} V^{(C)}_{1,t+1} \\ V^{(C)}_{2,t+1} \end{pmatrix}$$

where $C_0 = A$ and $C^*_0 = B$ and where $\{C^*_t\}$ is an auxiliary process. The dampening factor ρ allows for additional flexibility in the specification. The processes $\{V^{(C)}_{1,t}\}$ and $\{V^{(C)}_{2,t}\}$ are two mutually uncorrelated white noise processes. It is instructive to examine the spectral density (see Sect. 6.1) of the cyclical component in Fig. 17.2. It can be shown (see Exercise 17.5.2) that $\{C_t\}$ follows an ARMA(2,1) process.

The cyclical component can be incorporated into the state space model above by augmenting the state vector X_{t+1} by the cyclical components C_{t+1} and C^*_{t+1} and the error term V_{t+1} by $\{V^{(C)}_{1,t+1}\}$ and $\{V^{(C)}_{2,t+1}\}$. The observations equation has to be amended accordingly. Section 17.4.2 presents an empirical application of this approach.

Dynamic Factor Models

Dynamic factor models are an interesting approach when it comes to modeling simultaneously a large cross-section of times series. The concept was introduced into macroeconomics by Sargent and Sims (1977) and was then developed further and popularized by Quah and Sargent (1993), Reichlin (2003) and Breitung and Eickmeier (2006), among others. The idea is to view each time series Y_{it}, $i = 1, \ldots, n$, as the sum of a linear combination of some joint unobserved factors

$f_t = (f_{1t}, \ldots, f_{rt})'$ and an idiosyncratic component $\{W_{it}\}$, $i = 1, \ldots n$. Dynamic factor models are particularly effective when the number of factors r is small compared to the number of time series n. In practice, several hundred time series are related to a handful factors. In matrix notation we can write the observation equation for the dynamic factor model as follows:

$$Y_t = \Lambda_0 f_t + \Lambda_1 f_{t-1} + \ldots + \Lambda_q f_{t-q} + W_t$$

where Λ_i, $i = 0, 1, \ldots, q$, are $n \times r$ matrices. The state vector X_t equals $(f_t', \ldots, f_{t-q}')'$ if we assume that the idiosyncratic component is white noise, i.e. $W_t = (W_{1t}, \ldots, W_{nt})' \sim \text{WN}(0, R)$. The observation equation can then be written compactly as:

$$Y_t = GX_t + W_t$$

where $G = (\Lambda_0, \Lambda_1, \ldots, \Lambda_q)$. Usually, we assume that R is a diagonal matrix. The correlation between the different time series is captured exclusively by the joint factors.

The state equation depends on the assumed dynamics of the factors. One possibility is to model $\{f_t\}$ as a VAR(p) process with $\Phi(\text{L})f_t = e_t$, $e_t \sim \text{WN}(0, \Sigma)$, and $p \leq q + 1$, so we can use the state space representation of the VAR(p) process from above. For the case $p = 2$ and $q = 2$ we get:

$$X_{t+1} = \begin{pmatrix} f_{t+1} \\ f_t \\ f_{t-1} \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Phi_2 & 0 \\ I_r & 0 & 0 \\ 0 & I_r & 0 \end{pmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \end{pmatrix} + \begin{pmatrix} e_{t+1} \\ 0 \\ 0 \end{pmatrix} = FX_t + V_{t+1}$$

and $Q = \text{diag}(\Sigma, 0, 0)$. This scheme can be easily generalized to the case $p > q + 1$ or to allow for autocorrelated idiosyncratic components, assuming for example that they follow autoregressive processes.

The dimension of these models can be considerably reduced by an appropriate re–parametrization or by collapsing the state space adequately (Bräuning and Koopman 2014). Such a reduction can considerably increase the efficiency of the estimation.

Real Business Cycle Model (RBC Model)

State space models are becoming increasingly popular in macroeconomics, especially in the context of dynamic stochastic general equilibrium (DSGE) models. These models can be seen as generalizations of the real business cycle (RBC) models.[8] In these models a representative consumer is supposed to maximize the utility of his consumption stream over his infinite life time. Thereby, the consumer has the choice to consume part of his income or to invest his savings (part of his

[8]Prototypical models can be found in King et al. (1988) or Woodford (2003). Canova (2007) and Dejong and Dave (2007) present a good introduction to the analysis of DSGE models.

income which is not consumed) at the market rate of interest. These savings can be used as a mean to finance investment projects which increase the economy wide capital stock. The increased capital stock then allows for increased production in the future. The production process itself is subject to a random shocks called technology shocks.

The solution of this optimization problem is a nonlinear dynamic system which determines the capital stock and consumption in every period. Its local behavior can be investigated by linearizing the system around its steady state. This equation can then be interpreted as the state equation of the system. The parameters of this equation F and Q are related, typically in a nonlinear way, to the parameters describing the utility and the production function as well as the process of technology shocks. Thus, the state equation summarizes the behavior of the theoretical model.

The parameters of the state equation can then be estimated by relating the state vector, given by the capital stock and the state of the technology, via the observation equation to some observable variables, like real GDP, consumption, investment, or the interest rate. This then completes the state space representation of the model which can be analyzed and estimated using the tools presented in Sect. 17.3.[9]

17.2 Filtering and Smoothing

As we have seen, the state space model provides a very flexible framework for a wide array of applications. We therefore want to develop a set of tools to handle this kind of models in terms of interpretation and estimation. In this section we will analyze the problem of inferring the unobserved state from the data given the parameters of the model. In Sect. 17.3 we will then investigate the estimation of the parameters by maximum likelihood.

In many cases the state of the system is not or only partially observable. It is therefore of interest to infer from the data $Y_1, Y_2, \ldots, Y_T$ the state vector X_t. We can distinguish three types of problems depending on the information used:

(i) estimation of X_t from $Y_1, \ldots, Y_{t-1}$, known as the *prediction* problem;
(ii) estimation of X_t from $Y_1, \ldots, Y_t$, known as the *filtering* problem;
(iii) estimation of X_t from $Y_1, \ldots, Y_T$, known as the *smoothing* problem.

For the ease of exposition, we will assume that the disturbances V_t and W_t are normally distributed. The recursive nature of the state equation implies that $X_t = F^{t-1}X_1 + \sum_{j=0}^{t-2} F^j V_{t-j}$. Therefore, X_t is also normally distributed for all t, if X_1 is normally distributed. From the observation equation we can infer also

[9]See Sargent (2004) or Fernandez-Villaverde et al. (2007) for systematic treatment of state space models in the context of macroeconomic models. In this literature the use of Bayesian methods is widespread (see An and Schorfheide 2007; Dejong and Dave 2007).

that Y_t is normally distributed, because it is the sum of two normally distributed random variables $A + GX_t$ and W_t. Thus, under these assumptions, the vector $(X_1', \ldots, X_T', Y_1', \ldots, Y_T')'$ is jointly normally distributed:

$$\begin{pmatrix} X_1 \\ \vdots \\ X_T \\ Y_1 \\ \vdots \\ Y_T \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Gamma_X & \Gamma_{YX} \\ \Gamma_{XY} & \Gamma_Y \end{pmatrix} \right)$$

where the covariance matrices Γ_X, Γ_{YX}, Γ_{XY} and Γ_Y can be retrieved from the model given the parameters.

For the understanding of the rest of this section, the following theorem is essential (see standard textbooks, like Amemiya 1994; Greene 2008).

Theorem 17.1. *Let Z be a n-dimensional normally distributed random variable with $Z \sim \mathrm{N}(\mu, \Sigma)$. Consider the partitioned vector $Z = (Z_1', Z_2')'$ where Z_1 and Z_2 are of dimensions $n_1 \geq 1$ and $n_2 \geq 1$, $n = n_1 + n_2$, respectively. The corresponding partitioning of the covariance matrix Σ is*

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where $\Sigma_{11} = \mathbb{V}Z_1$, $\Sigma_{22} = \mathbb{V}Z_2$, and $\Sigma_{12} = \Sigma_{21}' = \mathrm{cov}(Z_1, Z_2) = \mathbb{E}(Z_1 - \mathbb{E}Z_1)'(Z_2 - \mathbb{E}Z_2)$. Then the partitioned vectors Z_1 and Z_2 are normally distributed. Moreover, the conditional distribution of Z_1 given Z_2 is also normal with mean and variance

$$\mathbb{E}(Z_1|Z_2) = \mathbb{E}Z_1 + \Sigma_{12}\Sigma_{22}^{-1}(Z_2 - \mathbb{E}Z_2),$$

$$\mathbb{V}(Z_1|Z_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

This formula can be directly applied to figure out the mean and the variance of the state vector given the observations. Thus, setting $Z_1 = (X_1', \ldots, X_t')'$ and $Z_2 = (Y_1', \ldots, Y_{t-1}')'$, we get the predicted values; setting $Z_1 = (X_1', \ldots, X_t')'$ and $Z_2 = (Y_1', \ldots, Y_t')'$, we get the filtered values; setting $Z_1 = (X_1', \ldots, X_t')'$ and $Z_2 = (Y_1', \ldots, Y_T')'$, we get the smoothed values.

AR(1) Process with Measurement Errors

We illustrate the above ideas by analyzing a univariate AR(1) process with measurement errors[10]:

[10] Sargent (1989) provides an interesting application showing the implications of measurement errors in macroeconomic models.

$$X_{t+1} = \phi X_t + v_{t+1}, \qquad v_t \sim \text{IIDN}(0, \sigma_v^2)$$
$$Y_t = X_t + w_t, \qquad w_t \sim \text{IIDN}(0, \sigma_w^2).$$

For simplicity, we assume $|\phi| < 1$. Suppose that we only have observations Y_1 and Y_2 at our disposal. The joint distribution of $(X_1, X_2, Y_1, Y_2)'$ is normal. The covariances can be computed by applying the methods discussed in Chap. 2:

$$\begin{pmatrix} X_1 \\ X_2 \\ Y_1 \\ Y_2 \end{pmatrix} \sim \text{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \frac{\sigma_v^2}{1-\phi^2}\begin{pmatrix} 1 & \phi & 1 & \phi \\ \phi & 1 & \phi & 1 \\ 1 & \phi & 1 + \frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2} & \phi \\ \phi & 1 & \phi & 1 + \frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2} \end{pmatrix}\right)$$

The smoothed values are obtained by applying the formula from Theorem 17.1:

$$\mathbb{E}\left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \middle| Y_1, Y_2\right) =$$
$$\frac{1}{(1 + \frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2})^2 - \phi^2}\begin{pmatrix} 1 & \phi \\ \phi & 1 \end{pmatrix}\begin{pmatrix} 1 + \frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2} & -\phi \\ -\phi & 1 + \frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2} \end{pmatrix}\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

Note that for the last observation, Y_2 in our case, the filtered and the smoothed values are the same. For X_1 the filtered value is

$$\mathbb{E}(X_1|Y_1) = \frac{1}{1 + \frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2}} Y_1.$$

An intuition for this result can be obtained by considering some special cases. For $\phi = 0$, the observations are not correlated over time. The filtered value for X_1 therefore corresponds to the smoothed one. This value lies between zero, the unconditional mean of X_1, and Y_1 with the variance ratio σ_w^2/σ_v^2 delivering the weights: the smaller the variance of the measurement error the closer the filtered value is to Y_1. This conclusion holds also in general. If the variance of the measurement error is relatively large, the observations do not deliver much information so that the filtered and the smoothed values are close to the unconditional mean.

For large systems the method suggested by Theorem 17.1 may run into numerical problems due to the inversion of the covariance matrix of Y, Σ_{22}. This matrix can become rather large as it is of dimension $nT \times nT$. Fortunately, there exist recursive solutions to this problem known as the Kalman filter, and also the Kalman smoother.

17.2.1 The Kalman Filter

The Kalman filter circumvents the problem of inverting a large $nT \times nT$ matrix by making use of the Markov property of the system (see Remark 17.4). The distribution of X_t given the observations up to period t can thereby be computed recursively from the distribution of the state in period $t-1$ given the information available up to period $t-1$. Starting from some initial distribution in period 0, we can in this way obtain in T steps the distribution of all states. In each step only an $n \times n$ matrix must be inverted. To describe the procedure in detail, we introduce the following notation:

$$\mathbb{E}(X_t | Y_1, \ldots, Y_h) = X_{t|h}$$
$$\mathbb{V}(X_t | Y_1, \ldots, Y_h) = P_{t|h}.$$

Suppose, we have already determined the distribution of X_t conditional on the observations $Y_1, \ldots, Y_t$. Because we are operating in a framework of normally distributed random variables, the distribution is completely characterized by its conditional mean $X_{t|t}$ and variance $P_{t|t}$. The goal is to carry forward these entities to obtain $X_{t+1|t+1}$ and $P_{t+1|t+1}$ having observed an additional data point Y_{t+1}. This problem can be decomposed into a *forecasting* and an *updating* step.

Step 1: Forecasting Step The state equation and the assumption about the disturbance term V_{t+1} imply:

$$X_{t+1|t} = FX_{t|t} \tag{17.6}$$
$$P_{t+1|t} = FP_{t|t}F' + Q$$

The observation equation then allows to compute a forecast of Y_{t+1} where we assume for simplicity that $A = 0$:

$$Y_{t+1|t} = GX_{t+1|t} \tag{17.7}$$

Step 2: Updating Step In this step the additional information coming from the additional observation Y_{t+1} is processed to update the conditional distribution of the state vector. The joint conditional distribution of $(X'_{t+1}, Y'_{t+1})'$ given $Y_1, \ldots, Y_t$ is

$$\begin{pmatrix} X_{t+1} \\ Y_{t+1} \end{pmatrix} \Bigg| Y_1, \ldots, Y_t \sim \mathrm{N}\left(\begin{pmatrix} X_{t+1|t} \\ Y_{t+1|t} \end{pmatrix}, \begin{pmatrix} P_{t+1|t} & P_{t+1|t}G' \\ GP_{t+1|t} & GP_{t+1|t}G' + R \end{pmatrix} \right)$$

As all elements of the distribution are available from the forecasting step, we can again apply Theorem 17.1 to get the distribution of the filtered state vector at time $t+1$:

$$X_{t+1|t+1} = X_{t+1|t} + P_{t+1|t}G'(GP_{t+1|t}G' + R)^{-1}(Y_{t+1} - Y_{t+1|t}) \tag{17.8}$$
$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t}G'(GP_{t+1|t}G' + R)^{-1}GP_{t+1|t} \tag{17.9}$$

where we replace $X_{t+1|t}$, $P_{t+1|t}$, and $Y_{t+1|t}$ by $FX_{t|t}$, $FP_{t|t}F' + Q$, and $GFX_{t|t}$, respectively, which have been obtained from the forecasting step.

Starting from given values for $X_{0|0}$ and $P_{0|0}$, we can therefore iteratively compute $X_{t|t}$ and $P_{t|t}$ for all $t = 1, 2, \ldots, T$. Only the information from the last period is necessary at each step. Inserting Eq. (17.8) into Eq. (17.6) we obtain as a forecasting equation:

$$X_{t+1|t} = FX_{t|t-1} + FP_{t|t-1}G'(GP_{t|t-1}G' + R)^{-1}(Y_t - GX_{t|t-1})$$

where the matrix

$$K_t = FP_{t|t-1}G'(GP_{t|t-1}G' + R)^{-1}$$

is know as the *(Kalman) gain matrix*. It prescribes how the innovation $Y_t - Y_{t|t-1} = Y_t - GX_{t|t-1}$ leads to an update of the predicted state.

Initializing the Algorithm It remains to determine how to initialize the recursion. In particular, how to set the starting values for $X_{0|0}$ and $P_{0|0}$. If X_t is stationary and causal with respect to V_t, the state equation has the solution $X_0 = \sum_{j=0}^{\infty} F^j V_{t-j}$. Thus,

$$X_{0|0} = \mathbb{E}(X_0) = 0$$
$$P_{0|0} = \mathbb{V}(X_0)$$

where $P_{0|0}$ solves the equation (see Sect. 12.4)

$$P_{0|0} = FP_{0|0}F' + Q.$$

According to Eq. (12.4), the solution of the above matrix equation is:

$$\text{vec}(P_{0|0}) = [I - F \otimes F]^{-1}\text{vec}(Q).$$

If the process is not stationary, we can set $X_{0|0}$ to zero and $P_{0|0}$ to infinity. In practice, a very large number is sufficient.

17.2.2 The Kalman Smoother

The Kalman filter determines the distribution of the state at time t given the information available up to this time. In many instances, we want, however, make an optimal forecast of the state given all the information available, i.e. the whole sample. Thus, we want to determine $X_{t|T}$ and $P_{t|T}$. The Kalman filter determines the smoothed distribution for $t = T$, i.e. $X_{T|T}$ and $P_{T|T}$. The idea of the Kalman smoother is again to determine the smoothed distribution in a recursive manner. For this purpose, we let the recursion run backwards. Starting with the last observation in period $t = T$, we proceed back in time by letting t take successively the values $t = T-1, T-2, \ldots$ until the first observation in period $t = 1$.

Using again the linearity of the equations and the normality assumption, we get:

$$\begin{pmatrix} X_t \\ X_{t+1} \end{pmatrix} \Bigg| Y_1, Y_2, \ldots, Y_t \sim \mathrm{N}\left(\begin{pmatrix} X_{t|t} \\ FX_{t|t} \end{pmatrix}, \begin{pmatrix} P_{t|t} & P_{t|t}F' \\ FP_{t|t} & P_{t+1|t} \end{pmatrix} \right)$$

This implies that

$$\mathbb{E}(X_t|Y_1, \ldots, Y_t, X_{t+1}) = X_{t|t} + P_{t|t}F'P_{t+1|t}^{-1}(X_{t+1} - X_{t+1|t}).$$

The above mean is only conditional on all information available up to time t and on the information at time $t+1$. The Markov property implies that this mean also incorporates the information from the observations $Y_{t+1}, \ldots, Y_T$. Thus, we have:

$$\begin{aligned} \mathbb{E}(X_t|Y_1, \ldots, Y_T, X_{t+1}) &= \mathbb{E}(X_t|Y_1, \ldots, Y_t, X_{t+1}) \\ &= X_{t|t} + P_{t|t}F'P_{t+1|t}^{-1}(X_{t+1} - X_{t+1|t}) \end{aligned}$$

Applying the law of iterated expectations or means (see, f.e. Amemiya 1994, p. 78), we can derive $X_{t|T}$:

$$\begin{aligned} X_{t|T} &= \mathbb{E}(X_t|Y_1, \ldots, Y_T) = \mathbb{E}(\mathbb{E}(X_t|Y_1, \ldots, Y_T, X_{t+1})|Y_1, \ldots, Y_T) \\ &= \mathbb{E}(X_{t|t} + P_{t|t}F'P_{t+1|t}^{-1}(X_{t+1} - X_{t+1|t})|Y_1, \ldots, Y_T) \\ &= X_{t|t} + P_{t|t}F'P_{t+1|t}^{-1}(X_{t+1|T} - X_{t+1|t}). \end{aligned} \tag{17.10}$$

The algorithm can now be implemented as follows. In the first step compute $X_{T-1|T}$ according to Eq. (17.10) as

$$X_{T-1|T} = X_{T-1|T-1} + P_{T-1|T-1}F'P_{T|T-1}^{-1}(X_{T|T} - X_{T|T-1}).$$

All entities on the right hand side can readily be computed by applying the Kalman filter. Having found $X_{T-1|T}$, we can again use Eq. (17.10) for $t = T-2$ to evaluate $X_{T-2|T}$:

$$X_{T-2|T} = X_{T-2|T-2} + P_{T-2|T-2}F'P_{T-1|T-2}^{-1}(X_{T-1|T} - X_{T-1|T-2}).$$

Proceeding backward through the sample we can derive a complete sequence of smoothed states $X_{T|T}, X_{T-1|T}, X_{T-2|T}, \ldots, X_{1|T}$. These calculations are based on the computations of $X_{t|t}$, $X_{t+1|t}$, $P_{t|t}$, and $P_{t+1|t}$ which have already been obtained from the Kalman filter. The smoothed covariance matrix $P_{t|T}$ is given as (see Hamilton 1994b, Section 13.6):

$$P_{t|T} = P_{t|t} + P_{t|t}FP_{t+1|t}^{-1}(P_{t+1|T} - P_{t+1|t})P_{t+1|t}^{-1}F'P_{t|t}.$$

Thus, we can compute also the smoothed variance with the aid of the values already determined by the Kalman filter.

AR(1) Process with Measurement Errors (Continued)

We continue our illustrative example of an AR(1) process with measurement errors and just two observations. First, we determine the filtered values for the state vector with the aid of the Kalman filter. To initialize the process, we have to assign a distribution to X_0. For simplicity, we assume that $|\phi| < 1$ so that it makes sense to assign the stationary distribution of the process as the distribution for X_0:

$$X_0 \sim \mathrm{N}\left(0, \frac{\sigma_V^2}{1-\phi^2}\right)$$

Then we compute the forecasting step as the first step of the filter (see Eq. (17.6)):

$$\begin{aligned} X_{1|0} &= \phi X_{0|0} = 0 \\ P_{1|0} &= \phi^2 \frac{\sigma_v^2}{1-\phi^2} + \sigma_v^2 = \frac{\sigma_v^2}{1-\phi^2} \\ Y_{1|0} &= 0. \end{aligned}$$

$P_{1|0}$ was computed by the recursive formula from the previous section, but is, of course, equal to the unconditional variance. For the updating step, we get from Eqs. (17.8) and (17.9):

$$\begin{aligned} X_{1|1} &= \left(\frac{\sigma_v^2}{1-\phi^2}\right)\left(\frac{\sigma_v^2}{1-\phi^2} + \sigma_w^2\right)^{-1} Y_1 = \frac{1}{1+\frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2}} Y_1 \\ P_{1|1} &= \left(\frac{\sigma_v^2}{1-\phi^2}\right) - \left(\frac{\sigma_v^2}{1-\phi^2}\right)^2 \left(\frac{\sigma_v^2}{1-\phi^2} + \sigma_w^2\right)^{-1} \\ &= \frac{\sigma_v^2}{1-\phi^2}\left(1 - \frac{1}{1+\frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2}}\right) \end{aligned}$$

These two results are then used to calculate the next iteration of the algorithm. This will give the filtered values for $t = 2$ which would correspond to the smoothed values because we just have two observations. The forecasting step is:

$$X_{2|1} = \frac{\phi}{1 + \frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2}} Y_1$$

$$P_{2|1} = \frac{\phi^2\sigma_v^2}{1-\phi^2}\left(1 - \frac{1}{1 + \frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2}}\right) + \sigma_v^2$$

Next we perform the updating step to calculate $X_{2|2}$ and $P_{2|2}$. It is easy to verify that this leads to the same results as in the first part of this example.

An interesting special case is obtained when we assume that $\phi = 1$ so that the state variable is a simple random walk. In this case the unconditional variance of X_t and consequently also of Y_t are no longer finite. As mentioned previously, we can initialize the Kalman filter by $X_{0|0} = 0$ and $P_{0|0} = \infty$. This implies:

$$Y_{1|0} = X_{1|0} = X_{0|0} = 0$$

$$P_{1|0} = P_{0|0} + \sigma_V^2 = \infty.$$

Inserting this result in the updating Eqs. (17.8) and (17.9), we arrive at:

$$X_{1|1} = \frac{P_{1|0}}{P_{1|0} + \sigma_W^2}(Y_1 - Y_{1|0}) = \frac{P_{0|0} + \sigma_V^2}{P_{0|0} + \sigma_V^2 + \sigma_W^2} Y_1$$

$$P_{1|1} = P_{1|0} - \frac{P_{1|0}^2}{P_{1|0} + \sigma_W^2} = (P_{0|0} + \sigma_V^2)\left(1 - \frac{P_{1|0}}{P_{1|0} + \sigma_W^2}\right) = \frac{(P_{0|0} + \sigma_V^2)\sigma_W^2}{P_{0|0} + \sigma_V^2 + \sigma_W^2}.$$

Letting $P_{0|0}$ go to infinity, leads to:

$$X_{1|1} = Y_1$$

$$P_{1|1} = \sigma_W^2.$$

This shows that the filtered variance is finite for $t = 1$ although $P_{1|0}$ was infinite.

17.3 Estimation of State Space Models

Up to now we have assumed that the parameters of the system are known and that only the state is unknown. In most economic applications, however, also the parameters are unknown and have therefore to be estimated from the data. One big advantage of the state space models is that they provide an integrated approach to forecasting, smoothing and estimation. In particular, the Kalman filter turns out to be an efficient and quick way to compute the likelihood function. Thus, it seems natural to estimate the parameters of state space models by the method of maximum likelihood. Kim and Nelson (1999) and Durbin and Koopman (2011) provide excellent and extensive reviews of the estimation of state space models using the Kalman filter.

More recently, due to advances in computational methods, in particular with respect to sparse matrix programming, other approaches can be implemented. For example, by giving the states a matrix representation Chan and Jeliazkov (2009) derive a viable and efficient method for the estimation of state space models.

17.3.1 The Likelihood Function

The joint unconditional density of the observations $(Y_1', \ldots, Y_T')'$ can be factorized into the product of conditional densities as follows:

$$
\begin{aligned}
f(Y_1, \ldots, Y_T) &= f(Y_T|Y_1, \ldots, Y_{T-1}) f(Y_1, \ldots, Y_{T-1}) \\
&= \vdots \\
&= f(Y_T|Y_1, \ldots, Y_{T-1}) f(Y_{T-1}|Y_1, \ldots, Y_{T-2}) \ldots f(Y_2|Y_1) f(Y_1)
\end{aligned}
$$

Each conditional density is the density of a normal distribution and is therefore given by:

$$
f(Y_t|Y_1, \ldots, Y_{t-1}) = (2\pi)^{-n/2} (\det \Delta_t)^{-1/2} \exp\left[-\frac{1}{2}(Y_t - Y_{t|t-1})' \Delta_t^{-1} (Y_t - Y_{t|t-1})\right]
$$

where $\Delta_t = G P_{t|t-1} G' + R$. The Gaussian likelihood function L is therefore equal to:

$$
L = (2\pi)^{-(Tn)/2} \left(\prod_{t=1}^{T} \det(\Delta_t)\right)^{-1/2} \exp\left[-\frac{1}{2}\sum_{t=1}^{T}(Y_t - Y_{t|t-1})' \Delta_t^{-1} (Y_t - Y_{t|t-1})\right].
$$

Note that all the entities necessary to evaluate the likelihood function are provided by the Kalman filter. Thus, the evaluation of the likelihood function is a byproduct of the Kalman filter. The maximum likelihood estimator (MLE) is then given by the maximizer of the likelihood function, or more conveniently the log-likelihood function. Usually, there is no analytic solution available so that one must resort to numerical methods. An estimation of the asymptotic covariance matrix can be obtained by evaluating the Hessian matrix at the optimum. Under the usual assumptions, the MLE is consistent and delivers asymptotically normally distributed estimates (Greene 2008; Amemiya 1994).

The direct maximization of the likelihood function is often not easy in practice, especially for large systems involving many parameters. The expectation-maximization algorithm, EM algorithm for short, represents a valid, though slower alternative. As the name indicates, it consists of two steps which have to be carried out iteratively. Based on some starting values for the parameters, the first step (expectation step) computes estimates, $X_{t|T}$, of the unobserved state vector X_t using the Kalman smoother. In the second step (maximization step), the likelihood function is maximized taking the estimates of X_t, $X_{t|T}$, as additional observations. The treatment of $X_{t|T}$ as additional observations, allows to reduce the maximization step to a simple multivariate regression. Indeed, by treating $X_{t|T}$ as if they were known, the state equation becomes a simple VAR(1) which can be readily estimated by linear least-squares to obtain the parameters F and Q. The parameters A, G and R are also easily retrieved from a regression of Y_t on $X_{t|T}$. Based on these new parameter estimates, we go back to step one and derive new estimates for $X_{t|T}$ which are then used in the maximization step. One can show that this procedure maximizes the original likelihood function (see Dempster et al. 1977; Wu 1983). A more detailed analysis of the EM algorithm in the time series context is provided by Brockwell and Davis (1996).[11]

Sometimes it is of interest not only to compute parameter estimates and to derive from them estimates for the state vector via the Kalman filter or smoother, but also to find confidence intervals for the estimated state vector to take the uncertainty into account. If the parameters are known, the methods outlined previously showed how to obtain these confidence intervals. If, however, the parameters have to be estimated, there is a double uncertainty: the uncertainty from the filter and the uncertainty arising from the parameter estimates. One way to account for this additional uncertainty is by the use of simulations. Thereby, we draw a given number of parameter vectors from the asymptotic distribution and compute for each of these draws the corresponding estimates for the state vector. The variation in these estimates is then a measure of the uncertainty arising from the estimation of the parameters (see Hamilton 1994b, Section 13.7).

[11]The analogue to the EM algorithm in the Bayesian context is given by the Gibbs sampler. In contrast to the EM algorithm, we compute in the first step not the expected value of the states, but we draw a state vector from the distribution of state vectors given the parameters. In the second step, we do not maximize the likelihood function, but draw a parameter from the distribution of parameters given the state vector drawn previously. Going back and forth between these two steps, we get a Markov chain in the parameters and the states whose stationary distribution is exactly the distribution of parameters and states given the data. A detailed description of Bayesian methods and the Gibbs sampler can be found in Geweke (2005). Kim and Nelson (1999) discuss this method in the context of state space models.

17.3.2 Identification

As emphasized in Remark 17.5 of Sect. 17.1, the state space representations are not unique. See, for example, the two alternative representations of the ARMA(1,1) model in Sect. 17.1. This non-uniqueness of state space models poses an identification problem because different specifications may give rise to observationally equivalent models.[12] This problem is especially serious if all states are unobservable. In practice, the identification problem gives rise to difficulties in the numerical maximization of the likelihood function. For example, one may obtain large differences for small variations in the starting values; or one may encounter difficulties in the inversion of the matrix of second derivatives.

The identification of state space models can be checked by transforming them into VARMA models and by investigating the issue in this reparameterized setting (Hannan and Deistler 1988). Exercise 17.5.6 invites the reader to apply this method to the AR(1) model with measurement errors. System identification is a special field in systems theory and will not be pursued further here. A systematic treatment can be found in the textbook by Ljung (1999).

17.4 Examples

17.4.1 Disaggregating Yearly Data into Quarterly Ones

The official data for quarterly GDP are released in Switzerland by the State Secretariat for Economic Affairs (SECO). They estimate these data taking the yearly values provided by the Federal Statistical Office (FSO) as given. This division of tasks is not uncommon in many countries. One of the most popular methods for disaggregation of yearly data into quarterly ones was proposed by Chow and Lin (1971).[13] It is a regression based method which can take additional information in the form of indicator variables (i.e. variables which are measured at the higher frequency and correlated at the lower frequency with the variable of interest) into account. This procedure is, however, rather rigid. The state space framework is much more flexible and ideally suited to deal with missing observations. Applications of this framework to the problem of disaggregation were provided by Bernanke et al. (1997:1) and Cuche and Hess (2000), among others. We will illustrate this approach below.

Starting point of the analysis are the yearly growth rates of GDP and indicator variables which are recorded at the quarterly frequency and which are correlated with GDP growth. In our application, we will consider the growth of industrial production (*IP*) and the index on consumer sentiment (*C*) as indicators. Both variables

[12] Remember that, in our context, two representations are equivalent if they generate the same mean and covariance function for $\{Y_t\}$.

[13] Similarly, one may envisage the disaggregation of yearly data into monthly ones or other forms of disaggregation.

are available on a quarterly basis from 1990 onward. For simplicity, we assume that the annualized quarterly growth rate of GDP, $\{Q_t\}$, follows an AR(1) process with mean μ:

$$Q_t - \mu = \phi(Q_{t-1} - \mu) + w_t, \qquad w_t \sim \text{WN}(0, \sigma_w^2)$$

In addition, we assume that GDP is related to industrial production and consumer sentiment by the following two equations:

$$\begin{aligned} IP_t &= \alpha_{IP} + \beta_{IP} Q_t + v_{IP,t} \\ C_t &= \alpha_C + \beta_C Q_t + v_{C,t} \end{aligned}$$

where the residuals $v_{IP,t}$ and $v_{C,t}$ are uncorrelated. Finally, we define the relation between quarterly and yearly GDP growth as:

$$J_t = \frac{1}{4}Q_t + \frac{1}{4}Q_{t-1} + \frac{1}{4}Q_{t-2} + \frac{1}{4}Q_{t-3}, \quad t = 4, 8, 12 \ldots$$

We can now bring these equations into state space form. Thereby the observation equation is given by

$$Y_t = A_t + G_t X_t + W_t$$

with observation and state vectors

$$Y_t = \begin{cases} \begin{pmatrix} J_t \\ IP_t \\ C_t \end{pmatrix}, & t = 4, 8, 12, \ldots; \\ \begin{pmatrix} 0 \\ IP_t \\ C_t \end{pmatrix}, & t \neq 4, 8, 12, \ldots \end{cases}$$

$$X_t = \begin{pmatrix} Q_t - \mu \\ Q_{t-1} - \mu \\ Q_{t-2} - \mu \\ Q_{t-3} - \mu \end{pmatrix}$$

and time-varying coefficient matrices

$$A_t = \begin{cases} \begin{pmatrix} \mu \\ \alpha_{IP} \\ \alpha_C \end{pmatrix}, t = 4, 8, 12, \ldots; \\ \begin{pmatrix} 0 \\ \alpha_{IP} \\ \alpha_C \end{pmatrix}, t \neq 4, 8, 12, \ldots \end{cases}$$

$$G_t = \begin{cases} \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \beta_{IP} & 0 & 0 & 0 \\ \beta_C & 0 & 0 & 0 \end{pmatrix}, t = 4, 8, 12, \ldots; \\ \begin{pmatrix} 0 & 0 & 0 & 0 \\ \beta_{IP} & 0 & 0 & 0 \\ \beta_C & 0 & 0 & 0 \end{pmatrix}, \quad t \neq 4, 8, 12, \ldots \end{cases}$$

$$R_t = \begin{cases} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sigma_{IP}^2 & 0 \\ 0 & 0 & \sigma_C^2 \end{pmatrix}, t = 4, 8, 12, \ldots; \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sigma_{IP}^2 & 0 \\ 0 & 0 & \sigma_C^2 \end{pmatrix}, t \neq 4, 8, 12, \ldots \end{cases}$$

The state equation becomes:

$$X_{t+1} = FX_t + V_{t+1}$$

where

$$F = \begin{pmatrix} \phi & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$Q = \begin{pmatrix} \sigma_w^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

On my homepage http://www.neusser.ch/ you will find a MATLAB code which maximizes the corresponding likelihood function numerically. Figure 17.3 plots the different estimates of GDP growth and compares them with the data released by State Secretariat for Economic Affairs (SECO).

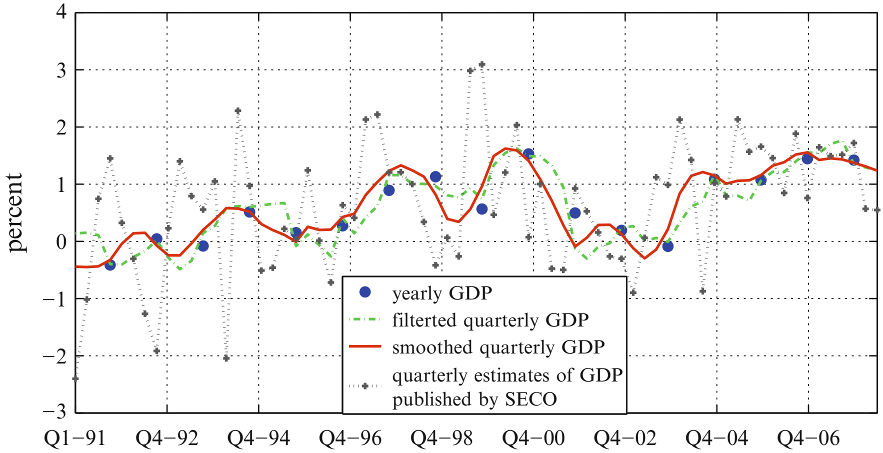

Fig. 17.3 Estimates of quarterly GDP growth rates for Switzerland

17.4.2 Structural Time Series Analysis

A customary practice in business cycle analysis is to decompose a time series into several components. As an example, we estimate a structural time series model which decomposes a times series additively into a local linear trend, a business cycle component, a seasonal component, and an irregular component. This is the specification studied as the basic structural model (BSM) in Sect. 17.1.1. We carry over the specification explained there to apply it to quarterly real GDP of Switzerland. Figure 17.4 shows the smoothed estimates of the various components. In the left upper panel the demeaned logged original series (see Fig. 17.4a) is plotted. One clearly discern the trend and the seasonal variations. The right upper panel shows the local linear trend (LLT). As one can see the trend is not a straight line, but exhibits pronounced waves of low frequency. The business cycle component showed in Fig. 17.4c is much more volatile. The large drop of about 2.5 % in 2008/09 corresponds to the financial markets. The lower right panel plots the seasonal component (see Fig. 17.4d). From a visual inspections, one can infer that the volatility of the seasonal component is much larger than the cyclical component (compare the scale of the two components) so that movements in GDP are dominated by seasonal fluctuations.[14] Moreover, the seasonal component changes its character over time.

[14]The irregular component which is not shown has only very small variance.

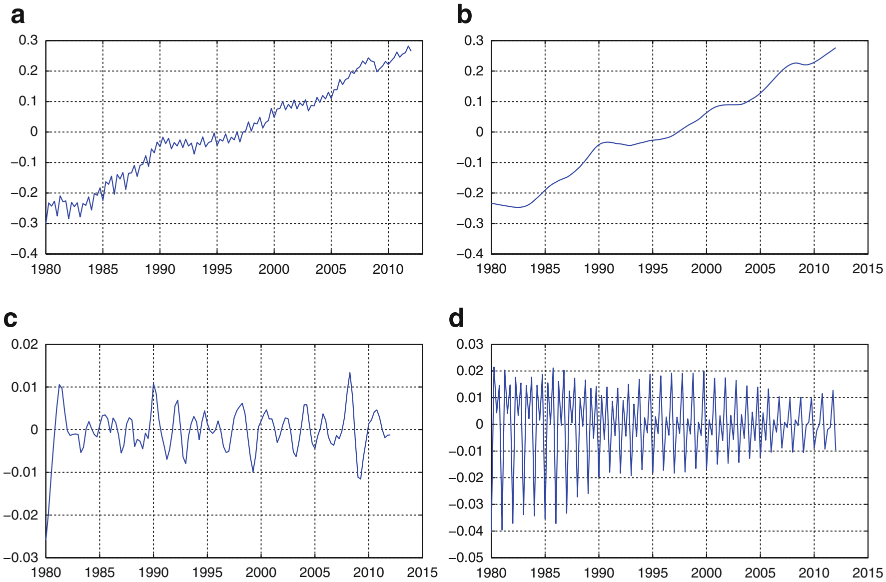

Fig. 17.4 Components of the basic structural model (BSM) for real GDP of Switzerland. (**a**) Logged Swiss GDP (demeaned). (**b**) Local linear trend (LLT). (**c**) Business cycle component. (**d**) Seasonal component

17.5 Exercises

Exercise 17.5.1. *Consider the basic structural time series model for $\{Y_t\}$:*

$$\begin{aligned} Y_t &= T_t + W_t, & W_t &\sim \text{WN}(0, \sigma_W^2) \\ T_t &= \delta_{t-1} + T_{t-1} + \varepsilon_t, & \varepsilon_t &\sim \text{WN}(0, \sigma_\varepsilon^2) \\ \delta_t &= \delta_{t-1} + \xi_t, & \xi_t &\sim \text{WN}(0, \sigma_\xi^2) \end{aligned}$$

where the error terms W_t, ε_t and ξ_t are all uncorrelated with other at all leads and lags.

(i) *Show that $\{Y_t\}$ follows an ARIMA(0,2,2) process.*
(ii) *Compute the autocorrelation function of $\{\Delta^2 Y_t\}$.*

Exercise 17.5.2. *If the cyclical component of the basic structural model for $\{Y_t\}$ is:*

$$\begin{pmatrix} C_t \\ C_t^* \end{pmatrix} = \rho \begin{pmatrix} \cos\lambda_C & \sin\lambda_C \\ -\sin\lambda_C & \cos\lambda_C \end{pmatrix} \begin{pmatrix} C_{t-1} \\ C_{t-1}^* \end{pmatrix} + \begin{pmatrix} V_{1,t}^{(C)} \\ V_{2,t}^{(C)} \end{pmatrix}$$

where $\{V_{1,t}^{(C)}\}$ *and* $\{V_{2,t}^{(C)}\}$ *are mutually uncorrelated white-noise processes.*

(i) *Show that* $\{C_t\}$ *follows an ARMA(2,1) process with ACF given by* $\gamma_h(h) = \rho^h \cos \lambda_C h$.

Exercise 17.5.3. *Write the ARMA(p,q) process* $Y_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$ *as a state space model such that the state vector* X_t *is given by:*

$$X_t = \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p} \\ Z_{t-1} \\ Z_{t-2} \\ \vdots \\ Z_{t-q} \end{pmatrix}.$$

Exercise 17.5.4. *Show that* X_t *and* Y_t *have a unique stationary and causal solution if all eigenvalues of* F *are absolutely strictly smaller than one. Use the results from Sect. 12.3.*

Exercise 17.5.5. *Find the Kalman filter equations for the following system:*

$$X_t = \phi X_{t-1} + w_t$$
$$Y_t = \lambda X_t + v_t$$

where λ *and* ϕ *are scalars and where*

$$\begin{pmatrix} v_t \\ w_t \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_v^2 & \sigma_{vw} \\ \sigma_{vw} & \sigma_w^2 \end{pmatrix} \right).$$

Exercise 17.5.6. *Consider the state space model of an AR(1) process with measurement error analyzed in Sect. 17.2:*

$$X_{t+1} = \phi X_t + v_{t+1}, \qquad v_t \sim \mathrm{IIDN}(0, \sigma_v^2)$$
$$Y_t = X_t + w_t, \qquad w_t \sim \mathrm{IIDN}(0, \sigma_w^2).$$

For simplicity assume that $|\phi| < 1$.

(i) *Show that* $\{Y_t\}$ *is an ARMA(1,1) process given by* $Y_t - \phi Y_{t-1} = Z_t + \theta Z_{t-1}$ *with* $Z_t \sim \mathrm{WN}(0, \sigma_Z^2)$.

(ii) *Show that the parameters of the state space,* $\phi, \sigma_v^2, \sigma_w^2$ *and those of the ARMA(1,1) model are related by the equation*

$$\theta \sigma_Z^2 = -\phi \sigma_w^2$$

$$\frac{1}{1 + \theta^2} = \frac{-\phi \sigma_w^2}{\sigma_v^2 + (1 + \phi^2)\sigma_w^2}$$

(iii) *Why is there an identification problem?*