# Notes on State Space Models

Xunzhao Yin

February 25, 2019

The state-space model was introduced in Kalman (1960) and Kalman and Bucy (1961). The model was introduced as a method primarily for use in aerospace-related research, where the state equation defines the motion for the position of a spacecraft, and data $y_t$ reflect information that can be observed.

A modern treatment of nonlinear state space models can be found in Douc, Moulines, and Stoffer (2014).

In general, the state space model is characterized by two principles. First, there is a hidden or latent process $x_t$ called the state process. The state process is assumed to be a Markov process. The second condition is that the observations, $y_t$, are independent given the states $x_t$, i.e. the dependence among the observations is generated by states.

# 1 Linear Gaussian Model

An order one, $p$-dimensional vector autoregression *state equation*

$$x_t = \Phi x_{t-1} + w_t \tag{1}$$

$$w_t \sim \text{i.i.d. } \mathrm{N}_p(0, Q)$$

$$x_0 \sim \mathrm{N}_p(\mu_0, \Sigma_0)$$

We observe a $q$-dimensional *observation equation*

$$y_t = A_t x_t + v_t \tag{2}$$

$$v_t \sim \text{i.i.d. } \mathrm{N}_q(0, R)$$

where $A_t$ is a $q \times p$ *measurement or observation matrix*. And, for simplicity, we assume $x_0$, $\{w_t\}$ and $\{v_t\}$ are uncorrelated.

More generally, suppose we have an $r \times 1$ vector of inputs $u_t$,

$$x_t = \Phi x_{t-1} + \Upsilon u_t + w_t \tag{3}$$

$$y_t = A_t x_t + \Gamma u_t + v_t \tag{4}$$

where $\Upsilon$ is $p \times r$ and $\Gamma$ is $q \times r$.

While the model seems simplistic, it is quite general. For example, if the state process is VAR(2), we may write the state equation as a $2p$-dimensional process

$$\begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Phi_2 \\ I & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} w_t \\ 0 \end{pmatrix}$$

And the observation equation as the $q$-dimensional process,

$$y_t = [A_t | 0] \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} + v_t$$

The questions of general interest relate to estimating the unknown parameters contained in $\Phi, \Upsilon, Q, \Gamma, A_t$, and $R$, and estimating or forecasting values of the underlying unobserved process $x_t$.

An equivalence exists between stationary ARMA models and stationary state space models.

# 2 Filtering, Smoothing, and Forecasting

Our aim is to produce estimators for $x_t$, given the data $y_{1:s} = \{y_1, \dots, y_s\}$, to time $s$. State estimation is an essential component of parameter estimation. When $s < t$, the problem is called forecasting or prediction. When $s = t$, filtering, and when $s > t$, smoothing. In addition to these estimates, we would also want to measure their precision. The solution to these problems is accomplished via the Kalman filter and smoother.

Denote

$$x_t^s = \mathrm{E}\,(x_t | y_{1:s})$$
$$P_{t_1,t_2}^s = \mathrm{E}\left\{\left(x_{t_1} - x_{t_1}^s\right)\left(x_{t_2} - x_{t_2}^s\right)' \middle| y_{1:s}\right\}$$

When $t_1 = t_2 = t$, we write $P_t^s$ for convenience.

**Property 2.1** (The Kalman Filter)**.** *For the state-space model specified in (3) and (4), with initial conditions $x_0^0 = \mu_0$ and $P_0^0 = \Sigma_0$, for $t = 1, \dots, n$,*

$$x_t^{t-1} = \Phi x_{t-1}^{t-1} + \Upsilon u_t$$
$$P_t^{t-1} = \Phi P_{t-1}^{t-1} \Phi' + Q$$

*with*

$$x_t^t = x_t^{t-1} + K_t(y_t - A_t x_t^{t-1} - \Gamma u_t)$$
$$P_t^t = [I - K_t A_t] P_t^{t-1}$$

*where*

$$K_t = P_t^{t-1} A_t' [A_t P_t^{t-1} A_t' + R]^{-1}$$

*is called the Kalman gain. Important byproducts of the filter are the innovations (prediction errors)*

$$\epsilon_t = y_t - \mathrm{E}(y_t | y_{1:t-1}) = y_t - A_t x_t^{t-1} - \Gamma u_t \tag{5}$$

*and the corresponding variance-covariance matrices*

$$\Sigma_t = \mathrm{Var}[A_t(x_t - x_t^{t-1}) + v_t] = A_t P_t^{t-1} A_t' + R \tag{6}$$

*for $t = 1, \dots, n$.*

*Proof.* Please refer to Shumway and Stoffer (2017, P296). □

**Corollary 2.2** (Kalman Filter: The Time-Varying Case)**.** *If, in (3) and (4), any or all of the parameters are time dependent, $\Phi = \Phi_t$, $\Upsilon = \Upsilon_t$, $Q = Q_t$ in the state equation or $\Gamma = \Gamma_t$, $R = R_t$ in the observation equation, or the dimension of the observation equation is time dependent, $q = q_t$, Property 2.1 holds with the appropriate substitutions.*

Now, we explore the model from a density point of view. In this part, we drop the inputs $u_t$ from the model. Letting $\mathrm{p}_\Theta(\cdot)$ denote a generic density function with parameters $\Theta$,

**Corollary 2.3** (Kalman Filter: densities)**.** *In terms of densities, the Kalman filter is an updating scheme, where to determine the forecast densities, we have*

$$\mathrm{p}_\Theta(x_t|y_{1:t-1}) = \int_{\mathbb{R}^p} \mathrm{p}_\Theta(x_t|x_{t-1})\mathrm{p}_\Theta(x_{t-1}|y_{1:t-1})dx_{t-1}$$

*Once we have the predictor, the filter density is obtained as*

$$\mathrm{p}_\Theta(x_t|y_{1:t}) \propto \mathrm{p}_\Theta(y_t|x_t)\mathrm{p}_\Theta(x_t|y_{1:t-1}).$$

**Property 2.4** (The Kalman Smoother)**.** *For the state-space model specified in (3) and (4), with initial conditions $x_n^n$ and $P_n^n$ obtained via Property 2.1, for $t = n, n-1, \ldots, 1$,*

$$\begin{aligned} x_{t-1}^n &= x_{t-1}^{t-1} + J_{t-1}\left(x_t^n - x_t^{t-1}\right) \\ P_{t-1}^n &= P_{t-1}^{t-1} + J_{t-1}\left(P_t^n - P_t^{t-1}\right)J_{t-1}' \end{aligned}$$

*where*

$$J_{t-1} = P_{t-1}^{t-1}\Phi'\left(P_t^{t-1}\right)^{-1}.$$

*Proof.* Denote $\eta_t = (v_{t:n}', w_{t+1:n}')'$. Note that $\eta_t$ is independent of $y_{1:t-1}$, as well as $(x_t - x_t^{t-1})$. And note that $(x_{t-1}, y_{1:t-1}, x_t - x_t^{t-1}, \eta_t)$ are jointly normally distributed. So, according to Shumway and Stoffer (2017, p. 497, Eq. (B.9)),

$$\begin{aligned} &\mathrm{E}(x_{t-1}|y_{1:t-1}, x_t - x_t^{t-1}, \eta_t) \\ &= \mathrm{E}(x_{t-1}|y_{1:t-1}, x_t - x_t^{t-1}) \\ &\quad + \mathrm{Cov}(x_{t-1}, \eta_t|y_{1:t-1}, x_t - x_t^{t-1})\mathrm{Var}(\eta|y_{1:t-1}, x_t - x_t^{t-1})^{-1}\left[\eta - \mathrm{E}(\eta|y_{1:t}, x_t - x_t^{t-1})\right] \\ &= \mathrm{E}(x_{t-1}|y_{1:t-1}, x_t - x_t^{t-1}) \end{aligned}$$

because $\mathrm{Cov}(x_{t-1}, \eta_t | y_{1:t-1}, x_t - x_t^{t-1}) = 0$. And, again according to that property of conditional normal distribution,

$$\begin{aligned}
&\mathrm{E}(x_{t-1}|y_{1:t-1}, x_t - x_t^{t-1}) \\
&= \mathrm{E}(x_{t-1}|y_{1:t-1}) \\
&\quad + \mathrm{Cov}(x_{t-1}, x_t - x_t^{t-1}|y_{1:t-1})\mathrm{Var}(x_t - x_t^{t-1}|y_{1:t-1})^{-1}\left[x_t - x_t^{t-1} - \mathrm{E}(x_t - x_t^{t-1}|y_{1:t-1})\right]
\end{aligned}$$

where

$$\mathrm{E}(x_{t-1}|y_{1:t-1}) = x_{t-1}^{t-1},$$

$$\begin{aligned}
\mathrm{Cov}(x_{t-1}, x_t - x_t^{t-1}|y_{1:t-1}) &= \mathrm{Cov}\left(x_{t-1}, \Phi(x_{t-1} - x_{t-1}^{t-1}) + w_t \middle| y_{1:t-1}\right) \\
&= P_{t-1}^{t-1}\Phi',
\end{aligned}$$

$$\mathrm{Var}(x_t - x_t^{t-1}|y_{1:t-1}) = P_t^{t-1},$$

and

$$\mathrm{E}(x_t - x_t^{t-1}|y_{1:t-1}) = 0.$$

Therefore,

$$\mathrm{E}(x_{t-1}|y_{1:t-1}, x_t - x_t^{t-1}, \eta_t) = x_{t-1}^{t-1} + P_{t-1}^{t-1}\Phi'\left(P_t^{t-1}\right)^{-1}(x_t - x_t^{t-1}).$$

Note that, given $y_{1:t-1}$, $x_t^{t-1}$ is given by the Kalman predictor. And given $x_t^{t-1}$ and $x_t - x_t^{t-1}$, we get $x_t$. Combining $x_t$ with $\eta_t$, we get $y_t$. In addition, given $x_t$ and $\eta_t$, we get $x_{t+1}$. Then we get $y_{t+1}$ ... In short, $y_{1:t-1}$, $x_t - x_t^{t-1}$ and $\eta_t$ generate $y_{1:n}$. Therefore,

$$\mathrm{E}\left[\mathrm{E}(x_{t-1}|y_{1:t-1}, x_t - x_t^{t-1}, \eta_t) \middle| y_{1:n}\right] = \mathrm{E}(x_{t-1}|y_{1:n})$$

of which the RHS is $x_{t-1}^n$, and the LHS is equal to $x_{t-1}^{t-1} + P_{t-1}^{t-1}\Phi'\left(P_t^{t-1}\right)^{-1}(x_t^n - x_t^{t-1})$. In summary,

$$x_{t-1}^n = x_{t-1}^{t-1} + P_{t-1}^{t-1}\Phi'\left(P_t^{t-1}\right)^{-1}(x_t^n - x_t^{t-1})$$

□

When we discuss maximum likelihood estimation via the EM algorithm [1] in the next section, we will need a set of recursions for obtaining $P^n_{t,t-1}$, which is given below

**Property 2.5** (The Lag-one Covariance Smoother). *For the state-space model specified in (3) and (4), with $K_t, J_t (t = 1, \ldots, n)$, and $P^n_n$ obtained from Property 2.1 and Property 2.4, and with initial condition*

$$P^n_{n,n-1} = (I - K_n A_n)\Phi P^{n-1}_{n-1}$$

*for $t = n, n-1, \ldots, 2$,*

$$P^n_{t-1,t-2} = P^{t-1}_{t-1} J'_{t-2} + J_{t-1}\left(P^n_{t,t-1} - \Phi P^{t-1}_{t-1}\right) J'_{t-2}.$$

[1] An expectation-maximization algorithm is an iterative method for finding maximum likelihood or maximum a posteriori (MAP) estimates of parameters, where the model depends on unobserved latent variables.

# 3 Maximum Likelihood Estimation

$\Theta$ represents the vector of unknown parameters: $\mu_0, \Sigma_0, \Phi, Q, R, \Upsilon, \Gamma$.

We use maximum likelihood under the assumption that the initial state is normal, $x_0 \sim \mathrm{N}_p(\mu_0, \Sigma_0)$, and the errors are normal, $w_t \sim$ i.i.d. $\mathrm{N}_p(0, Q)$ and $v_t \sim$ i.i.d. $\mathrm{N}_q(0, R)$. For simplicity, we continue to assume $\{w_t\}$ and $\{v_t\}$ are uncorrelated.

## 3.1 Newton-Raphson Procedure

The likelihood is computed using the innovations $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$, defined by (5). Note the innovations are independent Gaussian random vectors with zero means. Their covariance matrices are given by (6).

The innovation form of the Likelihood of the data $y_{1:n}$ was first given by Schweppe (1965). Ignoring a constant,

$$-\ln L_Y(\Theta) = \frac{1}{2}\sum_{t=1}^{n} \ln |\Sigma_t(\Theta)| + \frac{1}{2}\sum_{t=1}^{n} \epsilon_t(\Theta)'\Sigma_t(\Theta)^{-1}\epsilon_t(\Theta)$$

A Newton-Raphson estimation procedure can be used.

1. Select initial values for the parameters, $\Theta^{(0)}$.

2. Run the Kalman filter, Property (2.1), using $\Theta^{(0)}$, to obtain a set of innovations and error covariances, $\{\epsilon_t^{(0)}; t = 1, 2, \ldots, n\}$ and $\{\Sigma_t^{(0)}; t = 1, \ldots, n\}$.

3. Run one iteration of a Newton-Raphson procedure with $-\ln L_Y(\Theta)$ as the criterion function, to obtain a new set of estimates, $\Theta^{(1)}$.

4. At iteration j, repeat step 2 using $\Theta^{(j)}$ to obtain a new set of innovation values $\{\epsilon_t^{(j)}; t = 1, 2, \ldots, n\}$ and $\{\Sigma_t^{(j)}; t = 1, \ldots, n\}$. Then repeat step 3 to obtain a new estimate $\Theta^{(j+1)}$. Stop when the estimates or the likelihood stabilize.

## 3.2 Shumway-Stoffer Procedure

In addition to Newton-Raphson, Shumway and Stoffer (1982) presented a conceptually simpler procedure based on the Expectation-maximization algorithm.

For the sake of brevity, we ignore the inputs $u_t$ in the model.

The basic idea is that if we could observe the states, $x_{0:n} = \{x_0, x_1, \ldots, x_n\}$, in addition to the observations $y_{1:n}$, then we would consider $x_{0:n}, y_{0:n}$ as the complete data, with joint density

$$\mathrm{p}_\Theta(x_{0:t}, y_{0:t}) = \mathrm{p}_{\mu_0,\Sigma_0}(x_0) \prod_{t=1}^{n} \mathrm{p}_{\Phi,Q}(x_t|x_{t-1}) \prod_{t-1}^{n} \mathrm{p}_R(y_t|x_t)$$

Under the Gaussian assumption and ignore constant,

$$\begin{aligned} -2\ln L_{X,Y}(\Theta) = {} & \ln|\Sigma_0| + (x_0-\mu_0)'\Sigma_0^{-1}(x_0-\mu_0) \\ & + n\ln|Q| + \sum_{t=1}^{n}(x_t - \Phi x_{t-1})'Q^{-1}(x_t - \Phi x_{t-1}) \\ & + n\ln|R| + \sum_{t=1}^{n}(y_t - A_t x_t)'R^{-1}(y - A_t x_t) \end{aligned}$$

Thus, if we did have the complete data, we could then use the results from multivariate normal theory to easily obtain the MLEs.

But we do not have the complete data. So we write, at iteration $j$,

$$Q\left(\Theta|\Theta^{(j-1)}\right) = \mathrm{E}\left\{-2\ln_{X,Y}(\Theta)|y_{1:n}, \Theta^{(j-1)}\right\}$$

### 3.2.1 Expectation Step

To calculate the conditional expectation, we can use Property 2.4. This property yields

$$\begin{aligned} Q\left(\Theta|\Theta^{(j-1)}\right) = {} & \ln|\Sigma_0| + \mathrm{tr}\left\{\Sigma_0^{-1}[P_0^n + (x_0^n-\mu_0)(x_0^n-\mu_0)']\right\} \\ & + n\ln|Q| + \mathrm{tr}\left\{Q^{-1}[S_{11} - S_{10}\Phi' - \Phi S_{10}' + \Phi S_{00}\Phi']\right\} \\ & + n\ln|R| + \mathrm{tr}\left\{R^{-1}\sum_{t=1}^{n}[(y_t - A_t x_t^n)(y_t - A_t x_t^n)' + A_t P_t^n A_t']\right\}, \end{aligned} \tag{7}$$

where

$$S_{11} = \sum_{t=1}^{n}(x_t^n x_t^{n\prime} + P_t^n)$$

$$S_{10} = \sum_{t=1}^{n}(x_t^n x_{t-1}^{n\prime} + P_{t,t-1}^n)$$

$$S_{00} = \sum_{t=1}^{n}(x_{t-1}^n x_{t-1}^{n\prime} + P_{t-1}^n)$$

In obtaining $Q(\cdot|\cdot)$, we made repeated use of fact $\mathrm{E}(x_s x_t'|y_{1:n}) = x_s^n x_t^{n\prime} + P_{s,t}^n$; it is important to note that one does not simply replace $x_t$ with $x_t^n$ in the likelihood.

### 3.2.2 Maximization Step

Minimizing (7) with respect to $\Theta$ yields the updated estimates

$$\Phi^{(j)} = S_{10} S_{00}^{-1}$$

$$Q^{(j)} = n^{-1}\left(S_{11} - S_{10} S_{00}^{-1} S_{10}'\right)$$

$$R^{(j)} = n^{-1}\sum_{t=1}^{n}[(y_t - A_t x_t^n)(y_t - A_t x_t^n)' + A_t P_t^n A_t']$$

$$\mu_0^{(j)} = x_0^n$$

$$\sigma_0^{(j)} = P_0^n$$

The iterative procedure is as follows:

1. Choosing initial values $\Theta^{(0)}$
2. On iteration $j$, perform the E-Step: Using $\Theta^{(j-1)}$ and Properties 2.1, 2.4 and 2.5 to obtain the smoothed values $x_t^n, P_t^n$ and $P_{t,t-1}^n$, $t = 1, 2, \ldots, n$, and calculate $S_{11}, S_{10}, S_{00}$.
3. Perform the M-Step: Update the estimates $\Theta^{(j)}$.
4. Repeat Steps (2) - (3) to convergence.

## 3.3 Steady State and Asymptotic Distribution of the MLEs

The consistency and asymptotic normality of the estimators are established under general conditions. But for simplicity, we assume $A_t \equiv A$ for all $t$, and the eigenvalues of $\Phi$ are within the unit circle. And we drop the inputs $u_t$ for all $t$.

### 3.3.1 Stability of the Filter

As $t \to \infty$, the innovation covariance matrix $\Sigma_t \to \Sigma$, the steady-state covariance matrix of the stable innovations.

When the process is in steady-state, as can be seen from Property 2.1, steay-state predictor can be written as

$$x_{t+1}^{t} = \Phi x_{t}^{t-1} + \Phi K \epsilon_t$$
$$\epsilon_t \sim \text{i.i.d. } (0, \Sigma)$$

and the observation can be written as

$$y_t = A x_t^{t-1} + \epsilon_t$$

They two together make up the *steady-state innovations form* of the state space model.

### 3.3.2 Asymptotic Distribution of the Estimators

We denote the true parameters by $\Theta_0$.

**Property 3.1.** *Under general conditions, let $\hat{\Theta}$ be the estimator of $\Theta_0$ obtained by maximizing the innovations likelihood, $L_Y(\Theta)$. Then, as $n \to \infty$,*

$$\sqrt{n}\left(\hat{\Theta} - \Theta_0\right) \xrightarrow{d} \mathrm{N}\left(0, \mathcal{I}(\Theta_0)^{-1}\right),$$

*where $\mathcal{I}(\Theta_0)^{-1}$ is the asymptotic information matrix given by*

$$\mathcal{I}(\Theta) = \lim_{n \to \infty} n^{-1} \mathrm{E}\left[-\frac{\partial^2 \ln L_Y(\Theta)}{\partial \Theta \partial \Theta'}\right]$$

## 3.4 Example: A Dynamic Factor Model

Based on Section 3.5, Kim and Nelson (1999).

In econometrics, a dynamic factor (also known as a diffusion index) is a series which measures the co-movement of many time series.

The index of coincident economic indicators issued by the Department of Commerce (DOC) is developed for the purpose of summarizing the state of macroeconomic activity.

Stock and Watson (1991) developed a dynamic factor model of the coincident economic indicators, based on the idea that the co-movements in many macroeconomic variables have a common element that can be captured by a single underlying, unobserved variable.

let $Y_{1t}, Y_{2t}, Y_{3t}$ and $Y_{4t}$ be the logs of four coincident variables: industrial production, personal income less transfer payments, total manufacturing and trade sales, and employees on nonagricultural payrolls. Unit root tests for each of these U.S. series suggest that one cannot reject the null hypothesis of a unit root. In addition, these four series do not seem to be cointegrated[2]. Thus, Stock and Watson (1991) consider the following dynamic factor model

$$\Delta Y_{it} = D_i + \gamma_i \Delta C_t + e_{it},\ i = 1, 2, 3, 4 \tag{8}$$

$$(\Delta C_t - \delta) = \phi_1(\Delta C_{t-1} - \delta) + \phi_2(\Delta C_{t-2} - \delta) + w_t,\ w_t \sim \text{i.i.d. N}(0, 1) \tag{9}$$

$$e_{it} = \psi_{i1} e_{i,t-1} + \psi_{i2} e_{i,t-2} + \epsilon_{it},\ \epsilon_{it} \sim \text{i.i.d. N}(0, \sigma_i^2),\ i = 1, 2, 3, 4 \tag{10}$$

where $\Delta C_t$ is the common component; roots of $(1 - \phi_1 L - \phi_2 L^2) = 0$ lie outside the unit circle; roots of $(1 - \psi_{i1} L - \psi_{i2} L^2) = 0,\ i = 1, 2, 3, 4$, lie outside the unit circle; and all of the shocks are assumed to be independent.

Note that

$$\text{E}(\Delta Y_{it}) = D_i + \gamma_i \delta$$

Given the corresponding sample first moment, $\Delta\bar{Y}_i$, however, the parameters $D_i$ and $\delta$ are not separately identified. To see this point, suppose you make the mean of $\Delta C$ a little bit larger, i.e. $\delta' = \delta + \epsilon$, and make the mean of $\Delta y_i$ correspondingly smaller, i.e. $D_i' = D_i - \gamma_i \epsilon$. Then, you see that the likelihood will not change. The simplest way to address the identification problem is that you force $\Delta C$ has zero mean. Stock and Watson (1991) adopt a different strategy by suggesting writing the model in deviation from means:

$$\Delta y_{it} = \gamma_i \Delta c_t + e_{it},\ i = 1, 2, 3, 4$$

$$\Delta c_t = \phi_1 \Delta c_{t-1} + \phi_2 \Delta c_{t-2} + w_t,\ w_t \sim \text{i.i.d. N}(0, 1)$$

$$e_{it} = \psi_{i1} e_{i,t-1} + \psi_{i2} e_{i,t-2} + \epsilon_{it},\ \epsilon_{it} \sim \text{i.i.d. N}(0, \sigma_i^2),\ i = 1, 2, 3, 4$$

where $\Delta y_{it} = \Delta Y_{it} - \overline{\Delta Y}_i$ and $\Delta c_t = \Delta C_t - \delta$.

[2]A collection $X_1, X_2, \ldots, X_k$ of time series are cointegrated if, firstly, each of the series is integrated of order 1 and, secondly, a linear combination of this collection is integrated of order zero.

### 3.4.1 State-space Model

A state-space representation of this model, with the observation variable $\Delta y_t = (\Delta y_{1t}, \Delta y_{2t}, \Delta y_{3t}, \Delta y_{4t})'$, is given by

$$\Delta y_t = \begin{pmatrix} \gamma_1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma_2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \gamma_3 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \gamma_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} x_t,$$

$$x_t = \begin{pmatrix} \phi_1 & \phi_2 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \psi_{11} & \psi_{12} & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & \psi_{41} & \psi_{42} \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} x_{t-1} + \varepsilon_t,$$

where the state variable $x_t = (\Delta c_t, \Delta c_{t-1}, e_{1t}, e_{1,t-1}, e_{2t}, e_{2,t-1}, e_{3,t}, e_{3,t-1}, e_{4t}, e_{4,t-1})'$ and the state noise $\varepsilon_t = (w_t, 0, \epsilon_{1t}, 0, \epsilon_{2t}, 0, \epsilon_{3t}, 0, \epsilon_{4t}, 0)$

### 3.4.2 An Estimate of Delta

We need to estimate $\delta$ to construct a new coincident index, $C_t^t,\ t = 1, 2, \ldots, n$.

Note that, from Property 2.1 with $A_t = A$ and no exogenous inputs $u_t$,

$$x_t^t = (I - K_t A)\Phi x_{t-1}^{t-1} + K_t \Delta y_t$$

where $x_t$, $\Delta y_t$, $A$ and $\Phi$ are as defined in Section 3.4.1. So, recursively, the filter $x_t^t$ is a function of current and past observations $\Delta y_{1:t}$ :

$$x_t^t = f(L)\Delta y_t$$

Harvey (1989) show that for a stationary transition equation, the Kalman gain, $K_t$, approaches s steady-state Kalman gain, $K$, as $t \to \infty$. In this model, if one prints $K_t$ for

$t = 1, 2, \ldots, n$, one will notice that $K_t$ converges to a steady-state value reasonably fast. Thus, roughly speaking, for a reasonably large $t$,

$$x_t^t = (I - KA)\Phi x_{t-1}^{t-1} + K\Delta y_t$$

or

$$x_t^t = [I - (I - KA)\Phi L]^{-1} K\Delta y_t$$

If we use $W(L)$ to denote the first row of $[I - (I - KA)\Phi L]^{-1} K$, then

$$\Delta c_t^t = W(L)\Delta y_t \tag{11}$$

So, by definition,

$$\Delta C_t^t - \delta = W(L)\left(\Delta Y_t - \overline{\Delta Y}\right)$$

where $\Delta Y = (\Delta Y_1, \Delta Y_2, \Delta Y_3, \Delta Y_4)$, and $\overline{\Delta Y} = (\overline{\Delta Y}_1, \overline{\Delta Y}_2, \overline{\Delta Y}_3, \overline{\Delta Y}_4)$.

Actually, if the Kalman filter is applied to the model given by (8) - (10), we get the following relationship *(according to P52, Chapter 3, Kim and Nelson (1999))*

$$\Delta C_t^t = W(L)\Delta Y_t \tag{12}$$

Combine the previous two equations, and we get

$$\delta = W(L)\overline{\Delta Y} \tag{13}$$

$$= W(1)\overline{\Delta Y} \tag{14}$$

The lase equation holds because $\overline{\Delta Y}$ is the vector of sample averages and does not depend on time $t$.

### 3.4.3 My Findings

However, I find that Equation (12) is not true. That equation comes from Equation (7), Stock and Watson (1991). Their rationale behind that equation is unclear in that paper.

To show that it is wrong, denote $\Delta Y_t = (\Delta Y_{1t}, \Delta Y_{2t}, \Delta Y_{3t}, \Delta Y_{4t})'$, $D = (D_1, D_2, D_3, D_4)'$, $x_t = (\Delta C_t, \Delta C_{t-1}, e_{1t}, e_{1,t-1}, \ldots, e_{4t}, e_{4,t-1})'$, and $\boldsymbol{d} = (\delta, 0, \ldots, 0)'$. Then a state-space representation of the model given by Equations (8) to (10) is

$$\Delta Y_t = D + Ax_t$$
$$x_t = (I - \Phi)\boldsymbol{d} + \Phi x_{t-1} + \varepsilon_t$$

with $\varepsilon_t$ define above in Section 3.4.1, and

$$A = \begin{pmatrix} \gamma_1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma_2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \gamma_3 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \gamma_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\Phi = \begin{pmatrix} \phi_1 & \phi_2 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \psi_{11} & \psi_{12} & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \psi_{41} & \psi_{42} \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

If we apply Kalman filter to the model, for a reasonably large $t$, the Kalman filter $K_t = K$ and

$$x_t^{t-1} = (I - \Phi)\boldsymbol{d} + \Phi x_{t-1}^{t-1}$$
$$x_t^t = x_t^{t-1} + K(\Delta Y_t - D - Ax_t^{t-1})$$

Or, equivalently,

$$x_t^t = (I - KA)\Phi x_{t-1}^{t-1} + (I - KA)(I - \Phi)\boldsymbol{d} + K(\Delta Y_t - D)$$

That is

$$x_t^t - \boldsymbol{d} = (I - KA)\Phi\left(x_{t-1}^{t-1} - \boldsymbol{d}\right) - KA\boldsymbol{d} + K(\Delta Y_t - D)$$

We get

$$x_t^t - \boldsymbol{d} = [I - (I - KA)\Phi L]^{-1} K(\Delta Y_t - D - A\boldsymbol{d})$$

The first element of $x_t^t$ is $\Delta C_t^t$. So

$$\Delta C_t^t - \delta = W(L)(\Delta Y_t - D - A\boldsymbol{d}) \tag{15}$$

which is different from Equation (12), and, actually,

$$D + A\boldsymbol{d} = (D_1 + \gamma_1\delta, \ldots, D_4 + \gamma_4\delta)' = (\overline{\Delta Y}_{1t}, \ldots, \overline{\Delta Y}_{4t})'.$$

So Equation (15) is exactly the same as Equation (11). Intuitively, you cannot get more information by applying the Kalman filter to another identical model.

Besides that, I think we do not need $\delta$ or $C_{1:n}$. $\Delta C_{1:n}$, instead of $C_{1:n}$, is the economic indicator I care about. $\Delta c_{1:n}$ tells how much $\Delta C_{1:n}$ is stronger or weaker than normal. Taking $\delta = 100$ and construct $\Delta C_t = \Delta c_t + \delta$, we normalize the economic indicator $\Delta C_{1:n}$ to 100. However, this indicator tells us nothing more than whether the economy is growing above or below average speed.

# 4 Missing Data Modifications

For notational simplicity, we assume the model is of the form (1) and (2).

Suppose at time $t$, we partition the $q \times 1$ observation vector into two parts, $y_t^{(1)}$, a $q_{1t} \times 1$ components, and $y_t^{(2)}$, a $q_{2t} \times 1$ component, where $q_{1t} + q_{2t} = q$. Then, write the partitioned observation equation

$$\begin{pmatrix} y_t^{(1)} \\ y_t^{(2)} \end{pmatrix} = \begin{pmatrix} A_t^{(1)} \\ A_t^{(2)} \end{pmatrix} x_t + \begin{pmatrix} v^{(1)} \\ v^{(2)} \end{pmatrix},$$

where $A_t^{(1)}$ and $A_t^{(2)}$ are, respectively, the $q_{1t} \times p$ and $q_{2t} \times p$ partitioned observation matrices, and

$$\mathrm{Cov}\begin{pmatrix} v^{(1)} \\ v^{(2)} \end{pmatrix} = \begin{pmatrix} R_t^{11} & R_t^{12} \\ R_t^{21} & R_t^{22} \end{pmatrix}$$

denotes the covariance matrix of the measurement errors.

In the missing data case where $y_t^{(2)}$ is not observed, we may modify the observation equation so that the model becomes

$$\begin{aligned} x_t &= \Phi x_{t-1} + w_t, \\ y_t^{(1)} &= A_t^{(1)} x_t + v_t. \end{aligned}$$

Essentially, we leave out the missing data such that they are not a player of the game any more.

# 5 Structural Models: Signal Extraction and Forecasting

Structural models are component models in which each component may be thought of as explaining a specific type of behavior, e.g. trend, seasonal, and irregular components. Consequently, each component has a direct interpretation as to the nature of the variation in the data.

**Example 5.1.** Consider the Johnson & Johnson quarterly earnings per share series displayed in Figure 1.1 (Chapter 1 of the text. Not in this Chapter 6.). There are 84 quarters measured from the first quarter of 1960 to the last quarter of 1980. The series is highly unstationary, and there is both an increasing trend signal and a seasonal component that cycles every four quarters. We consider the series to be expressed as

$$y_t = T_t + S_t + v_t$$

where $T_t$ is trend, $S_t$ is the seasonal component, and $v_t$ a white noise. Suppose we allow the trend to increase exponentially

$$T_t = \phi T_{t-1} + w_{t1}$$

where the coefficient $\phi > 1$ characterized the increase. Let the seasonal component be modeled as

$$S_t + S_{t-1} + S_{t-2} + S_{t-3} = w_{t2}$$

which corresponds to assuming the component is expected to sum to zero over a complete period of four quarters.

To express the model in state-space form, let $x_t = (T_t, S_t, S_{t-1}, S_{t-2})'$ be the state vector. So the observation equation

$$y_t = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} T_t \\ S_t \\ S_{t-1} \\ S_{t-2} \end{pmatrix} + v_t$$

with the state equation written as

$$\begin{pmatrix} T_t \\ S_t \\ S_{t-1} \\ S_{t-2} \end{pmatrix} = \begin{pmatrix} \phi & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} T_{t-1} \\ S_{t-1} \\ S_{t-2} \\ S_{t-3} \end{pmatrix} + \begin{pmatrix} w_{t1} \\ w_{t2} \\ 0 \\ 0 \end{pmatrix}.$$

# 6 State-Space Models with Correlated Errors

Here, we write the state-space model as

$$x_{t+1} = \Phi x_t + \Upsilon u_{t+1} + \Theta w_t \quad t = 0, 1, \ldots, n \tag{16}$$

$$y_t = A_t x_t + \Gamma u_t + v_t \quad t = 1, 2, \ldots, n \tag{17}$$

$$x_0 \sim \mathrm{N}_p(\mu_0, \Sigma_0)$$

$$w_t \sim \text{i.i.d. } \mathrm{N}_m(0, Q)$$

$$v_t \sim \text{i.i.d. } \mathrm{N}_q(0, R)$$

$$\mathrm{Cov}(w_t, v_t) = S$$

both $w_t$ and $v_T$ are independent or $x_0$, $\Phi$ is $p \times p$, $\Upsilon$ is $p \times r$, $\Theta$ is $p \times m$, $A_t$ is $q \times p$, $\Gamma$ is $q \times r$, and $S$ is $m \times q$.

Note that this model starts the state noise process at $t = 0$. And the inclusion of the matrix $\Theta$ allows us to avoid using a singular state noise process.

**Property 6.1** (The Kalman Filter with Correlated Noise)**.** *For the state-space model specified above, with initial conditions $x_1^0$ and $P_1^0$, for $t = 1, \ldots, n$,*

$$x_{t+1}^t = \Phi x_t^{t-1} + \Upsilon u_{t+1} + K_t \epsilon_t$$

$$P_{t+1}^t = \Phi P_t^{t-1} \Phi' + \Theta Q \Theta' - K_t \Sigma_t K_t'$$

*where $\epsilon_t = y_t - A_t x_t^{t-1} - \Gamma u_t$ and the gain matrix is given by*

$$K_t = [\Phi P_t^{t-1} A_t' + \Theta S][A_t P_t^{t-1} A_t' + R]^{-1}.$$

*The filter values are given by*

$$x_t^t = x_t^{t-1} + P_t^{t-1} A_t' [A_t P_t^{t-1} A_t' + R]^{-1} \epsilon_t \tag{18}$$

$$P_t^t = P_t^{t-1} - P_t^{t-1} A_t' [A_t P_t^{t-1} A_t' + R]^{-1} A_t P_t^{t-1}. \tag{19}$$

Note that the gain matrix $K$ differs here from Property 2.1. But the filter values are symbolically the same as before.

To initialize the filter, we note that

$$x_1^0 = \mathrm{E}(x_1) = \Phi\mu_0 + \Upsilon u_1$$

$$P_1^0 = \mathrm{Var}(x_1) = \Phi\Sigma_0\Phi' + \Theta Q\Theta'$$

Property 2.4 (the smoother) still holds. Note that, although the smoother stays the same, the Gibbs sampling procedure has changed. Please refer to Section 13.2.2.

## 6.1 ARMAX Models

Consider a $k$-dimensional ARMAX model given by

$$y_t = \Upsilon u_t + \sum_{j=1}^{p} \Phi_j y_{t-j} + \sum_{k=1}^{q} \Theta_k v_{t-k} + v_t. \tag{20}$$

where the $\Phi$s and $\Theta$s are $k \times k$ matrices, $\Upsilon$ is $k \times r$, $u_t$ is the $t \times 1$ input, and $v_t$ is a $k \times 1$ white noise process.

**Property 6.2** (A state-space Form of ARMAX). *For $p \geq q$, let*

$$F = \begin{pmatrix} \Phi_1 & I & 0 & \cdots & 0 \\ \Phi_2 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Phi_{p-1} & 0 & 0 & \cdots & I \\ \Phi_p & 0 & 0 & \cdots & 0 \end{pmatrix} \quad G = \begin{pmatrix} \Theta_1 + \Phi_1 \\ \vdots \\ \Theta_q + \Phi_q \\ \Phi_{q+1} \\ \vdots \\ \Phi_p \end{pmatrix} \quad H = \begin{pmatrix} \Upsilon \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

*where $F$ is $kp \times kp$, $G$ is $kp \times k$, and $H$ is $kp \times r$. Then, the state-space model given by*

$$x_{t+1} = Fx_t + Hu_{t+1} + Gv_t \tag{21}$$

$$y_t = Ax_t + v_t \tag{22}$$

*where $A = [I, 0, \ldots, 0]$ is $k \times pk$ and $I$ is the $k \times k$ identity matrix, implies the ARMAX model (20). If $p < q$, set $\Phi_{p+1} = \cdots = \Phi_1 = 0$, in which case (21) and (22) still apply.*

## 6.2 Multivariate Regression with Autocorrelated Errors

$$y_t = \Gamma u_t + \varepsilon_t$$

where $\varepsilon_t$ is vector ARMA$(p, q)$, $y_t$ is $k \times 1$, $u_t$ is $t$, and $\Gamma$ is $k \times r$.

To put the model in state-space form, notice that $\varepsilon_t = y - \Gamma u_t$ is a $k$-dimensional ARMA(p, q) process with no exogenous inputs. Directly from Property 6.2, the following state-space model implies ARMA process $\varepsilon_T$

$$x_{t+1} = F x_t + G v_t$$

$$y_t - \Gamma u_t = A x_t + v_t$$

# 7 Bootstrapping State Space Models

Although under general conditions the MLEs of the parameters of a state space model are consistent and asymptotically normal, times series data are often of short or moderate length.

In this section, we discuss an algorithm for bootstrapping state space models in order to get a finite sample distribution of the MLEs.

# 8 Smoothing Splines and the Kalman Smoother

# 9 Hidden Markov Models and Switching Autoregression

We have been focusing on linear Gaussian state space models where the state process is a continuous-valued.

But this section will focus on the case where the states are a discrete-valued Markov chain. This type of models are typically called ”hidden Markov models” or ”Markov-switching models”. These models were developed in Goldfeld and Quandt (1973) and Lindgren (1978).

Here, we assume the states, $S_{1:n}$, are a first order Markov chain taking values in a finite state space $\{1, \ldots, m\}$, with steady-state probabilities

$$\pi_j = \Pr(S_t = j) \tag{23}$$

and transition probabilities

$$\pi_{ij} = \Pr(S_{t+1} = j | S_t = i) \tag{24}$$

for $t = 0, 1, 2, \ldots$, and $i, j = 1, \ldots, m$.

## 9.1 Serially Uncorrelated Observations

If the observations are serially uncorrelated given the states (as in Model (49)), we denote

$$\mathrm{p}_j(y_t) = \mathrm{p}(y_t | S_t = j).$$

Note that no particular model is given. We are discussing the procedure in general.

Write

$$\pi_j(t \mid s) = \Pr(S_t = j | y_{1:s})$$

Let $\Theta$ denote the parameters of interest. Given data $y_{1:n}$, the likelihood is given by

$$L_Y(\Theta) = \prod_{t=1}^{n} \mathrm{p}_\Theta(y_t | y_{1:t-1}) \tag{25}$$

By the conditional serial uncorrelation,

$$\begin{aligned}
\mathrm{p}_\Theta(y_t|y_{1:t-1}) &= \sum_{j=1}^{m} \Pr(S_t = j|y_{1:t-1})\mathrm{p}_\Theta(y_t|S_t = j, y_{1:t-1}) \\
&= \sum_{j=1}^{m} \pi_j(t \mid t-1)\mathrm{p}_j(y_t)
\end{aligned}$$

Consequently,

$$\ln L_Y(\Theta) = \sum_{t=1}^{n} \ln\left(\sum_{j=1}^{m} \pi_j(t \mid t-1)\mathrm{p}_j(y_t)\right) \tag{26}$$

As in the linear Gaussian case, we need filters and smoothers of the state.

**Property 9.1** (HMM Filter)**.** *For $t = 1, \ldots, n$, given $\pi_j(t-1|t-1)$ for $j = 1, 2, \ldots, m$,*

$$\begin{aligned}
\pi_j(t \mid t-1) &= \sum_{i=1}^{m} \pi_i(t-1 \mid t-1)\pi_{ij} \\
\pi_j(t \mid t) &= \frac{\pi_j(t \mid t-1)\mathrm{p}_j(y_t)}{\sum_{i=1}^{m} \pi_i(t \mid t-1)\mathrm{p}_i(y_t)}
\end{aligned}$$

*with initial conditions $\pi_i(1|0) = \pi_i$, for $i = 1, 2, \ldots, m$.*

Maximum likelihood can then proceed.

In addition, the EM algorithm applies here as well. Firstly, denote

$$\pi_{ij}(t \mid n) = \Pr(S_t = i, S_{t+1} = j|y_{1:n}).$$

**Property 9.2** (HMM Smoother)**.** *Given $\pi_k(t+1|n)$ for $k = 1, \ldots, m$,*

$$\begin{aligned}
\pi_{jk}(t|n) &= \pi_k(t+1|n)\frac{\pi_{jk}\pi_j(t|t)}{\sum_j \pi_{jk}\pi_j(t|t)} \\
\pi_j(t|n) &= \sum_k \pi_{jk}(t|n)
\end{aligned}$$

*for $j = 1, \ldots, m$.*

*Proof.*

$$\begin{aligned}
\pi_{jk}(t|n) &= \Pr(S_t = j, S_{t+1} = k|y_{1:n}) \\
&= \pi_k(t+1|n)\Pr(S_t = j|S_{t+1} = k, y_{1:n})
\end{aligned}$$

where

$$\begin{aligned}
\Pr(S_t = j|S_{t+1} = k, y_{1:n}) &= \frac{\Pr(S_t = j, y_{t+1:n}|S_{t+1} = k, y_{1:t})}{\mathrm{p}(y_{t+1:n}|S_{t+1} = k, y_{1:t})} \\
&= \frac{\mathrm{p}(y_{t+1:n}|S_t = j, S_{t+1} = k, y_{1:t}) \Pr(S_t = j|S_{t+1} = k, y_{1:t})}{\mathrm{p}(y_{t+1:n}|S_{t+1} = k, y_{1:t})} \\
&= \frac{\mathrm{p}(y_{t+1:n}|S_{t+1} = k) \Pr(S_t = j|S_{t+1} = k, y_{1:t})}{\mathrm{p}(y_{t+1:n}|S_{t+1} = k)} \\
&= \Pr(S_t = j|S_{t+1} = k, y_{1:t})
\end{aligned}$$

and

$$\begin{aligned}
\Pr(S_t = j|S_{t+1} = k, y_{1:t}) &= \frac{\Pr(S_t = j, S_{t+1} = k|y_{1:t})}{\Pr(S_{t+1} = k|y_{1:t})} \\
&= \frac{\pi_{jk}\pi_j(t|t)}{\sum_j \pi_{jk}\pi_j(t|t)}
\end{aligned}$$

$\square$

Secondly, the general complete data likelihood, as before, still has the form of

$$\ln \mathrm{p}_\Theta(S_{0,n}, y_{1:n}) = \ln \mathrm{p}_\Theta(S_0) + \sum_{t=1}^{n} \ln \mathrm{p}_\Theta(S_t|S_{t-1}) + \sum_{t=1}^{n} \ln \mathrm{p}_\Theta(y_t|S_t)$$

It is useful to define $I_j(t) = 1$ if $S_t = j$ and 0 otherwise, and $I_{ij} = 1$ if $(S_{t-1}, S_t) = (i, j)$ and 0 otherwise, for $i, j = 1, \ldots, m$. Then the complete data likelihood can be written as

$$\ln \mathrm{p}_\Theta(S_{0:n}, y_{1:n}) = \sum_{j=1}^{m} I_j(0) \ln \pi_j + \sum_{t=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m} I_{ij} \ln \pi_{ij} + \sum_{t=1}^{n}\sum_{j=1}^{m} I_j(t) \ln \mathrm{p}_j(y_t) \tag{27}$$

Thirdly, as before, we need to maximize $Q(\Theta|\Theta') = \mathrm{E}[\ln \mathrm{p}_\Theta(S_{0:n}, y_{1:n})|y_{1:n}, \Theta']$. Given the current estimates of parameters $\Theta'$, run the filter Property 9.1 and the smoother Property 9.2, and then, as is evident from (27), update the first two estimates as

$$\begin{aligned}
\hat{\pi}_j &= \pi'_j(0|n) \\
\hat{\pi}_{ij} &= \frac{\sum_{t=1}^{n} \pi'_{ij}(t|n)}{\sum_{t=1}^{n}\sum_{k=1}^{m} \pi'_{ik}(t|n)}
\end{aligned}$$

Finally, the third term in (27) will require knowing the distribution of $p_j(y_t)$, and this will depend on the particular model.

## 9.2 Serially Correlated Observations - Switching Autoregressions

Generally, data $y_{1:n}$ can be serially correlated. For example,

$$y_t = \phi_0^{(S_t)} + \sum_{i=1}^{p} \phi_i^{(S_t)} y_{t-i} + \sigma^{(S_t)} v_t$$

$$v_t \sim \text{i.i.d. N}(0, 1)$$

We write

$$\mathrm{p}_j(y_t) = \mathrm{p}(y_t | S_t = j, y_{t-1:t-p})$$

which is Gaussian.

Then, as in (26), the conditional likelihood is given by

$$\ln L_Y(\Theta | y_{1:p}) = \sum_{t=p+1}^{n} \ln \left( \sum_{j=1}^{m} \pi_j(t|t-1) \mathrm{p}_j(y_t) \right)$$

where $\pi_j(t|t-1)$ is still given by Property (9.1), and $\mathrm{p}_j(y_t)$ given above.

## 9.3 Serially Correlated Observations - Switching ARs with Historic Regimes

Based on Chapter 4, Kim and Nelson (1999).

In addition, $y_t$ may not depend only on state $S_t$ but also on $S_{1:t-1}$. For example, an AR($k$) model with first order, $m$-state Markov-switching mean and variance can be written as

$$\phi(L)(y_t - \mu_{S,t}) = e_t$$

$$e_t \sim \mathrm{N}(0, \sigma^2_{S,t})$$

$$\mu_{S,t} = \mu_1 S_{1t} + \mu_2 S_{2t} + \ldots + \mu_m S_{mt}$$

$$\sigma^2_{S,t} = \sigma^2_1 S_{1t} + \sigma^2_2 S_{2t} + \ldots + \sigma^2_m S_{mt}$$

where $S_{it} = 1$ if $S_t = i$, and $S_{it} = 0$ otherwise, $i = 1, 2, \ldots, m$.

### 9.3.1 Autoregression of Order 1

$$y_t - \mu_{S,t} = \phi_1(y_{t-1} - \mu_{S,t-1}) + e_t$$
$$e_t \sim \text{i.i.d. } \mathrm{N}(0, \sigma^2_{S,t})$$

The likelihood function is, again, given by (25), where

$$\mathrm{p}_\Theta(y_t|y_{1:t-1}) = \sum_{S_t}\sum_{S_{t-1}} \mathrm{p}_\Theta(y_t|y_{1:t-1}, S_t, S_{t-1}) \Pr(S_t|S_{t-1}) \Pr(S_{t-1}|y_{1:t-1})$$

with $\Pr(S_{t-1}|y_{1:t-1})$ given by the following filter:

**Property 9.3** (The Hamilton Filter: AR(1))**.** *For* $t = 1, 2, \ldots, n$, *given* $\Pr(S_{t-1}|y_{1:t-1})$ *for* $S_{t-1} = 1, 2, \ldots, m$,

$$\Pr(S_t, S_{t-1}|y_{1:t-1}) = \Pr(S_t|S_{t-1}) \Pr(S_{t-1}|y_{1:t-1})$$
$$\Pr(S_t, S_{t-1}|y_{1:t}) = \frac{\mathrm{p}_\Theta(y_t|S_t, S_{t-1}, y_{1:t-1}) \Pr(S_t, S_{t-1}|y_{1:t-1})}{\sum_{S_t}\sum_{S_{t-1}} \mathrm{p}_\Theta(y_t|S_t, S_{t-1}, y_{1:t-1}) \Pr(S_t, S_{t-1}|y_{1:t-1})}$$
$$\Pr(S_t|y_{1:t}) = \sum_{S_{t-1}} \Pr(S_t, S_{t-1}|y_{1:t-1})$$

*with initial conditions* $\Pr(S_0 = i) = \pi_i$, *for* $i = 1, 2, \ldots, m$.

And the corresponding smoother is given by the following property.

**Property 9.4** (The Hamilton Smoother: AR(1))**.** *Given* $\Pr(S_{t+1}|y_{1:n})$ *and* $\mathrm{p}(y_{t+1:n}|S_t, S_{t+1}, y_{1:t})$ *for* $S_t, S_{t+1} = 1, \ldots, m$,

$$\mathrm{p}(S_t, y_{t+1:n}|S_{t+1}, y_{1:t}) = \mathrm{p}(y_{t+1,n}|S_t, S_{t+1}, y_{1:t}) \frac{\Pr(S_{t+1}|S_t) \Pr(S_t|y_{1:t})}{\sum_{S_t} \Pr(S_{t+1}|S_t) \Pr(S_t|y_{1:t})}$$
$$\Pr(S_t|S_{t+1}, y_{1:n}) = \frac{\mathrm{p}(S_t, y_{t+1:n}|S_{t+1}, y_{1:t})}{\sum_{S_t} \mathrm{p}(S_t, y_{t+1:n}|S_{t+1}, y_{1:t})}$$
$$\Pr(S_t, S_{t+1}|y_{1:n}) = \Pr(S_{t+1}|y_{1:n}) \Pr(S_t|S_{t+1}, y_{1:n})$$
$$\Pr(S_t|y_{1:n}) = \sum_{S_{t+1}} \Pr(S_t, S_{t+1}|y_{1:n})$$

*In addition,*

$$\mathrm{p}(y_{t:n}|S_{t-1}, S_t, y_{1:t-1}) = \mathrm{p}(y_t|S_{t-1}, S_t, y_{1:t-1}) \left[\sum_{S_{t+1}} \mathrm{p}(y_{t+1:n}|S_t, S_{t+1}, y_{1:t}) \Pr(S_{t+1}|S_t)\right]$$

*with initial conditions* $\Pr(S_n|y_{1:n})$ *and* $\mathrm{p}(y_n|S_{n-1}, S_n, y_{1:t-1})$, *for* $S_n, S_{n-1} = 1, \ldots, m$.

*Proof.* To see that last equation,

$$\mathrm{p}(y_{t:n}|S_{t-1}, S_t, y_{1:t-1}) = \mathrm{p}(y_t|S_{t-1}, S_t, y_{1:t-1})\mathrm{p}(y_{t+1:n}|S_{t-1}, S_t, y_{1:t})$$

where

$$\begin{aligned}
\mathrm{p}(y_{t+1:n}|S_{t-1}, S_t, y_{1:t}) &= \sum_{S_{t+1}} \mathrm{p}(y_{t+1:n}, S_{t+1}|S_{t-1}, S_t, y_{1:t}) \\
&= \sum_{S_{t+1}} \mathrm{p}(y_{t+1:n}|S_{t-1}, S_t, S_{t+1}, y_{1:t}) \Pr(S_{t+1}|S_{t-1}, S_t, y_{1:t}) \\
&= \sum_{S_{t+1}} \mathrm{p}(y_{t+1:n}|S_t, S_{t+1}, y_{1:t}) \Pr(S_{t+1}|S_t)
\end{aligned}$$

The second, third and fourth equations are straight forward. To see the first equation,

$$\begin{aligned}
\Pr(S_t, y_{t+1:n}|S_{t+1}, y_{1:t}) &= \mathrm{p}(y_{t+1,n}|S_t, S_{t+1}, y_{1:t}) \Pr(S_t|S_{t+1}, y_{1:t}) \\
&= \mathrm{p}(y_{t+1,n}|S_t, S_{t+1}, y_{1:t}) \frac{\Pr(S_{t+1}|S_t) \Pr(S_t|y_{1:t})}{\sum_{S_t} \Pr(S_{t+1}|S_t) \Pr(S_t|y_{1:t})}
\end{aligned}$$

□

Note that Kim and Nelson (1999) made a mistake on this smoothing part from Page 68 to 69. They mistakenly believe

$$\Pr(S_t|S_{t+1}, y_{1:n}) = \Pr(S_t|S_{t+1}, y_{1:t})$$

in their Equation (4.42). But actually, for an AR(1) model

$$\Pr(S_t|S_{t+1}, y_{1:n}) = \Pr(S_t|S_{t+1}, y_{1:t}) \frac{\mathrm{p}(y_{t+1:n}|S_t, S_{t+1}, y_{1:t})}{\mathrm{p}(y_{t+1:n}|S_{t+1}, y_{1:t})}$$

Therefore, their equations (4.40) and (4.41) do not work for an AR(1) model. They work for the model specified in Section (9.1) and, actually, they are exactly the same as Property 9.2.

### 9.3.2 Autoregression of Order 2

$$y_t - \mu_{S,t} = \phi_1(y_{t-1} - \mu_{S,t-1}) + \phi_2(y_{t-2} - \mu_{S,t-2}) + e_t$$

$$e_t \sim \text{i.i.d. } \mathrm{N}(0, \sigma^2_{S,t})$$

The likelihood function is given by (25), where

$$
\begin{aligned}
&\mathrm{p}(y_t|y_{1:t-1})\\
&= \sum_{S_t}\sum_{S_{t-1}}\sum_{S_{t-2}} \mathrm{p}(y_t|S_t, S_{t-1}, S_{t-2}, y_{1:t-1}) \Pr(S_t|S_{t-1}) \Pr(S_{t-1}, S_{t-2}|y_{1:t-1})
\end{aligned}
$$

with $\Pr(S_{t-1}, S_{t-2}|y_{1:t-1})$ given by the following filter:

**Property 9.5** (The Hamilton Filter: AR(2))**.** *For* $t = 1, 2, \ldots, n$*, given* $\Pr(S_{t-1}, S_{t-2}|y_{1:t-1})$*, for* $S_{t-1} = 1, \ldots, m$ *and* $S_{t-2} = 1, \ldots, m$*,*

$$
\begin{aligned}
\Pr(S_t, S_{t-1}, S_{t-2}|y_{1:t-1}) &= \Pr(S_t|S_{t-1}) \Pr(S_{t-1}, S_{t-2}|y_{1:t-1})\\
\Pr(S_t, S_{t-1}, S_{t-2}|y_{1:t}) &= \frac{\mathrm{p}(y_t|S_t, S_{t-1}, S_{t-2}, y_{1:t-1}) \Pr(S_t, S_{t-1}, S_{t-2}|y_{1:t-1})}{\sum_{S_t}\sum_{S_{t-1}}\sum_{S_{t-2}} \mathrm{p}(y_t|S_t, S_{t-1}, S_{t-2}, y_{1:t-1}) \Pr(S_t, S_{t-1}, S_{t-2}|y_{1:t-1})}\\
\Pr(S_t, S_{t-1}|y_{1:t}) &= \sum_{S_{t-2}} \Pr(S_t, S_{t-1}, S_{t-2}|y_{1:t})
\end{aligned}
$$

*with initial conditions* $\Pr(S_0 = j, S_{-1} = i) = \pi_{ij}\pi_i$*, for* $i, j = 1, \ldots, m$*. And*

$$
\Pr(S_t|y_{1:t}) = \sum_{S_{t-1}} \Pr(S_t, S_{t-1}|y_{1:t})
$$

And the corresponding smoother is given by the following property.

**Property 9.6** (The Hamilton Smoother: AR(2))**.** *Given* $\Pr(S_{t+1}|y_{1:n})$ *and* $\mathrm{p}(y_{t+1:n}|S_{t-1}, S_t, S_{t+1}, y_{1:t})$ *for* $S_{t-1}, S_t, S_{t+1} = 1, \ldots, m$*,*

$$
\mathrm{p}(S_t, y_{t+1:n}|S_{t+1}, y_{1:t}) = \sum_{S_{t-1}} \left[ \mathrm{p}(y_{t+1:n}|S_{t-1}, S_t, S_{t+1}, y_{1:t}) \frac{\Pr(S_{t-1}, S_t, S_{t+1}|y_{1:t})}{\sum_{S_{t-1}}\sum_{S_t} \Pr(S_{t-1}, S_t, S_{t+1}|y_{1:t})} \right]
$$

*with* $\Pr(S_{t-1}, S_t, S_{t+1}|y_{1:t})$ *given by the above Hamilton Filter procedure 9.5.*

$$
\begin{aligned}
\Pr(S_t|S_{t+1}, y_{1:n}) &= \frac{\mathrm{p}(S_t, y_{t+1:n}|S_{t+1}, y_{1:n})}{\sum_{S_t} \mathrm{p}(S_t, y_{t+1:n}|S_{t+1}, y_{1:t})}\\
\Pr(S_t, S_{t+1}|y_{1:n}) &= \Pr(S_{t+1}|y_{1:n}) \Pr(S_t|S_{t+1}, y_{1:n})\\
\Pr(S_t|y_{1:n}) &= \sum_{S_{t+1}} \Pr(S_t, S_{t+1}|y_{1:n}).
\end{aligned}
$$

*In addition,*

$$
\begin{aligned}
&\mathrm{p}(y_{t:n}|S_{t-2}, S_{t-1}, S_t, y_{1:t-1})\\
&= \mathrm{p}(y_t|S_{t-2}, S_{t-1}, S_t, y_{1:t-1}) \left[ \sum_{S_{t+1}} \mathrm{p}(y_{t+1:n}|S_{t-1}, S_t, S_{t+1}, y_{1:t}) \Pr(S_{t+1}|S_t) \right]
\end{aligned}
$$

*with initial conditions* $\Pr(S_n|y_{1:n})$ *and* $\mathrm{p}(y_n|S_{n-2}, S_{n-1}, S_n, y_{1:n-1})$ *for* $S_{n-2}, S_{n-1}, S_n = 1, \ldots, m$.

*Proof.* To see the first equation,

$$\begin{aligned}\mathrm{p}(S_t, y_{t+1:n}|S_{t+1}, y_{1:t}) &= \sum_{S_{t-1}} \mathrm{p}(S_{t-1}, S_t, y_{t+1:n}|S_{t+1}, y_{1:t}) \\ &= \sum_{S_{t-1}} \mathrm{p}(y_{t+1:n}|S_{t-1}, S_t, S_{t+1}, y_{1:t}) \Pr(S_{t-1}, S_t|S_{t+1}, y_{1:t})\end{aligned}$$

where

$$\Pr(S_{t-1}, S_t|S_{t+1}, y_{1:t}) = \frac{\Pr(S_{t-1}, S_t, S_{t+1}|y_{1:t})}{\sum_{S_{t-1}} \sum_{S_t} \Pr(S_{t-1}, S_t, S_{t+1}|y_{1:t})}$$

To derive the last equation,

$$\begin{aligned}&\mathrm{p}(y_{t:n}|S_{t-2}, S_{t-1}, S_t, y_{1:t-1}) \\ &= \mathrm{p}(y_t|S_{t-2}, S_{t-1}, S_t, y_{1:t-1}) \mathrm{p}(y_{t+1:n}|S_{t-2}, S_{t-1}, S_t, y_{1:t}) \\ &= \mathrm{p}(y_t|S_{t-2}, S_{t-1}, S_t, y_{1:t-1}) \sum_{S_{t+1}} \mathrm{p}(y_{t+1:n}, S_{t+1}|S_{t-2}, S_{t-1}, S_t, y_{1:t}) \\ &= \mathrm{p}(y_t|S_{t-2}, S_{t-1}, S_t, y_{1:t-1}) \left[\sum_{S_{t+1}} \mathrm{p}(y_{t+1:n}|S_{t-1}, S_t, S_{t+1}, y_{1:t}) \Pr(S_{t+1}|S_t)\right]\end{aligned}$$

□

Note that Kim and Nelson's (1999) equation (4.43) on their Page 70 is also wrong, because they mistakenly believe, for an AR(k) process

$$\Pr(S_{t-k+1}|S_{t-k+2}, \ldots, S_t, S_{t+1}|y_{1:n}) \neq \Pr(S_{t-k+1}|S_{t-k+2}, \ldots, S_t, S_{t+1}|y_{1:t}).$$

They say their algorithm is vastly more efficient than those in Hamilton 1989 and Lam (1990) in terms of simplicity and computation time. I guess Hamilton (1989) algorithm could be the same as Property 9.6.

## 9.4 Steady-State Probabilities

The transition probabilities for a first order, $m$-state Markov-switching process $S_t$ in equation (24) can be put in the following matrix notation

$$\mathbf{\Pi} = \begin{pmatrix} \pi_{11} & \pi_{21} & \cdots & \pi_{m1} \\ \pi_{12} & \pi_{22} & \cdots & \pi_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{1m} & \pi_{2m} & \cdots & \pi_{mm} \end{pmatrix}$$

where $i_m' \mathbf{\Pi} = i_m'$ with $i_m = (1\ 1 \ldots 1)'$. If we let $\boldsymbol{\pi}$ be a vector of $m \times 1$ steady-state probabilities, we have

$$\boldsymbol{\pi} = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_m \end{pmatrix}.$$

Then $i_m' \boldsymbol{\pi} = 1$, and, according to the definition of steady-state probabilities, $\boldsymbol{\pi} = \mathbf{\Pi}\boldsymbol{\pi}$. Thus

$$(I - \mathbf{\Pi})\boldsymbol{\pi} = \mathbf{0}$$

where $\mathbf{0}$ is $m \times 1$. So

$$\begin{pmatrix} I - \mathbf{\Pi} \\ i_m' \end{pmatrix} \boldsymbol{\pi} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}, \quad \text{or } A\boldsymbol{\pi} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}$$

Therefore,

$$\boldsymbol{\pi} = (A'A)^{-1} A' \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}$$

or, the last column of the matrix $(A'A)^{-1}A'$.

# 10 State-Space Models with Switching

## 10.1 A Simple Model

This section concentrates on the method presented in Shumway and Stoffer (1991). The starting point is the state space model given by

$$x_t = \Phi x_{t-1} + w_t \tag{28}$$

$$y_t = A_{S_t} x_t + v_t \tag{29}$$

$$w_t \sim \text{i.i.d. N}(0, Q)$$

$$v_t \sim \text{i.i.d. N}(0, R)$$

$$\text{Cov}(w_t, v_s) = 0 \quad \forall s,\ \forall t$$

**Example 10.1** (Tracking Multiple Targets)**.** We want to track three moving targets using a vector $y_t = (y_{t1}, y_{t2}, y_{t3})'$ of sensors. And we do not know at any given point in time which target any given sensor has detected. Hence, it is the structure os the measurement matrix $A_t$ in (29) that is changing.

One example of measurement matrix

$$A_t = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

All possible detection configurations will define a set, $\{M_1, M_2, \ldots, M_m\}$, of possible values for $A_t$.

**Example 10.2** (Modeling Economic Change)**.** Lam (1990) has given the following generalization of Hamilton's (1989) model for detecting positive and negative growth periods in the economy. Suppose the data are generated by

$$y_t = z_t + n_t \tag{30}$$

where $z_t$ is an autoregressive series and $n_t$ is a random walk with a drift that switches between two values. For the purpose of illustration,

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + w_t$$

$$n_t = n_{t-1} + \alpha_0 + \alpha_1 S_t$$

where $S_t = 0$ or 1, depending on whether the system is in state 1 or state 2.

(30) can be written as

$$\Delta y_t = z_t - z_{t-1} + \alpha_0 + \alpha_1 S_t$$

which we may take as the observation equation (29)

$$\Delta y_t = A_{S_t} x_t + v_t$$

with state vector

$$x_t = (z_t, z_{t-1}, \alpha_0, \alpha_1)'$$

and

$$A_1 = [1, -1, 1, 0]$$

$$A_2 = [1, -1, 1, 1]$$

determining the two economic conditions.

The state equation (28) is

$$\begin{pmatrix} z_t \\ z_{t-1} \\ \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & & \\ 1 & & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ z_{t-2} \\ \alpha_0 \\ \alpha_1 \end{pmatrix} + \begin{pmatrix} w_t \\ \\ \\ \\ \end{pmatrix}$$

For $j = 1, \ldots, m$ and $t = 1, 2, \ldots, n$, denote

$$\pi_j(t) = \Pr(S_t = j)$$

$$\pi_j(t|t) = \Pr(S_t = j | y_{1:t})$$

We want to obtain estimators of the configuration probabilities $\pi_j(t|t)$, the predicted and filtered state estimators $x_t^{t-1}$ and $x_t^t$, and the corresponding error covariance matrices $P_t^{t-1}$ and $P_t^t$. And we need to estimate the parameters $\Theta$. Our focus will be on maximum likelihood estimation.

### 10.1.1 Predictors and Filters

The predictors and filters and their associated error variance-covariance matrices are given by

$$
\begin{aligned}
x_t^{t-1} &= \Phi x_{t-1}^{t-1} && (31)\\
P_t^{t-1} &= \Phi P_{t-1}^{t-1} \Phi' + Q \\
x_{tj}^t &= x_t^{t-1} + K_{tj}\epsilon_{tj} \\
P_{tj}^t &= (I - K_{tj}A_j)P_t^{t-1} \\
x_t^t &= \sum_{j=1}^m \pi_j(t|t) x_{tj}^t \\
P_t^t &= \sum_{j=1}^m \pi_j(t|t) \left[ P_{tj}^t + (x_{tj}^t - x_t^t)(x_{tj}^t - x_t^t)' \right] && (32)\\
K_{tj} &= P_t^{t-1} A_j' \Sigma_{tj}^{-1}
\end{aligned}
$$

where the innovation values

$$
\begin{aligned}
\epsilon_{tj} &= y_t - A_j x_t^{t-1} \\
\Sigma_{tj} &= A_j P_t^{t-1} A_j' + R && (33)
\end{aligned}
$$

for $j = 1, \ldots, m$.

Note that the counterpart of the above Equation (32), i.e. Equation (6.166) on Page 348 in Shumway and Stoffer (2016), is wrong. It can be derived using the law of total variance, $\mathrm{Var}(X) = \mathrm{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}[\mathrm{E}(X|Y)]$. For more details, refer to Section 10.2.1.

Next, we derive the filters $\pi_j(t|t)$. Let

$$\mathrm{p}_j(t|t-1) = \mathrm{p}(y_t \mid y_{1:t-1}, S_t = j)$$

for $j = 1, \ldots, m$. Then

$$\pi_j(t|t) = \frac{\pi_j(t|t-1)\mathrm{p}_j(t|t-1)}{\sum_{k=1}^m \pi_k(t|t-1)\mathrm{p}_k(t|t-1)}$$

where $\pi_j(t|t-1)$ depends on the model's specifications.

If the distribution $\pi_j(t)$, for $j = 1, \ldots, m$, has been specified before observing $y_{1:t}$, then

$$\pi_j(t|t-1) = \pi_j(t)$$

For example, if the investigator has no reason to prefer one state over another at time $t$, the choice of uniform priors $\pi_j(t) = m^{-1}$, for $j = 1, \ldots, m$, will suffice. In another case, if $\{S_t\}$ is a hidden Markov chain with stationary transition probabilities $\pi_{ij} = \Pr(S_t = j | S_{t-1} = i)$, for $i, \; j = 1, \ldots, m$, we have

$$\pi_j(t|t-1) = \sum_{i=1}^{m} \pi_{ij} \pi_i(t-1|t-1)$$

To evaluate $\mathrm{p}_j(t|t-1)$, notice that

$$\begin{aligned} \mathrm{E}(y_t | y_{1:t-1}, S_t = j) &= \mathrm{E}(A_{S_t} x_t + v_t | y_{1:t-1}, S_t) \\ &= A_j x_t^{t-1} \end{aligned}$$

and

$$\mathrm{Var}(y_t | y_{1:t-1}, S_t = j) = A_j P_t^{t-1} A_j' + R.$$

Therefore,

$$\mathrm{p}_j(t|t-1) = \mathrm{g}\left(y_t; A_j x_t^{t-1}, \Sigma_{tj}\right).$$

where $x_t^{t-1}$ and variance $\Sigma_{tj}$ are given in (31) and (33).

So the filter algorithm for a state-space model with Markov switching is as follows:

**Property 10.3** (Shumway and Stoffer Filter)**.** *Given $x_{t-1}^{t-1}$, $P_{t-1}^{t-1}$ and $\pi_j(t-1|t-1)$ for $j = 1, \ldots, m$ and $t = 1, \ldots, n$,*

$$\begin{gathered} x_t^{t-1} = \Phi x_{t-1}^{t-1}, \quad P_t^{t-1} = \Phi P_{t-1}^{t-1} \Phi' + Q \\ \epsilon_{tj} = y_t - A_j x_t^{t-1}, \quad \Sigma_{tj} = A_j P_t^{t-1} A_j' + R, \quad K_{tj} = P_t^{t-1} A_j' \Sigma_{tj}^{-1} \\ x_{tj}^t = x_t^{t-1} + K_{tj} \epsilon_{tj}, \quad P_{tj}^t = (I - K_{tj} A_j) P_t^{t-1} \\ \pi_j(t|t-1) = \sum_{i=1}^{m} \pi_{ij} \pi_i(t-1|t-1) \\ \mathrm{p}_j(t|t-1) = \mathrm{g}\left(y_t; A_j x_t^{t-1}, \Sigma_{tj}\right) \\ \pi_j(t|t) = \frac{\pi_j(t|t-1) \mathrm{p}_j(t|t-1)}{\sum_{k=1}^{m} \pi_k(t|t-1) \mathrm{p}_k(t|t-1)} \\ x_t^t = \sum_{j=1}^{m} \pi_j(t|t) x_{tj}^t \\ P_t^t = \sum_{j=1}^{m} \pi_j(t|t) \left[ P_{tj}^t + (x_{tj}^t - x_t^t)(x_{tj}^t - x_t^t)' \right] \end{gathered}$$

### 10.1.2 Maximum Likelihood Estimation

The vector of unknown parameters $\Theta = \{\mu_0, \Sigma_0, \Phi, Q, R, \pi_{ij}\}$.

The joint density of data is

$$f(y_1, \ldots, y_n) = \prod_{t=1}^{n} \sum_{j=1}^{m} \Pr(S_t = j \mid y_{1:t-1}) \mathrm{p}(y_t | S_t = j, y_{1:t-1})$$

and hence, the likelihood can be written as

$$\ln L_Y(\Theta) = \sum_{t=1}^{n} \ln \left( \sum_{j=1}^{m} \pi_j(t|t-1) \mathrm{p}_j(t|t-1) \right)$$

We may consider maximizing the likelihood as a function of the parameters $\Theta$, or we may consider applying the EM algorithm to the complete data likelihood.

## 10.2 A More General Case

Based on Chapter 5, Kim and Nelson (1999).

$$x_t = \Phi_{S\,t} x_{t-1} + \mu_{S\,t} + w_t \tag{34}$$

$$y_t = A_{S\,t} x_t + \Gamma_{S\,t} z_t + v_t \tag{35}$$

$$\begin{pmatrix} w_t \\ v_t \end{pmatrix} \sim \mathrm{N} \left( 0 \quad \begin{pmatrix} Q_{S\,t} & 0 \\ 0 & R_{S\,t} \end{pmatrix} \right)$$

Denote

$$x_{t,j}^{t-1,i} = \mathrm{E}(x_t | y_{1:t-1}, S_t = j, S_{t-1} = i)$$

$$P_{t,j}^{t-1,i} = \mathrm{E}\left[ (x_t - x_{t,j}^{t-1,i})(x_t - x_{t,j}^{t-1,i})' \mid y_{1:t-1}, S_t = j, S_{t-1} = i \right]$$

$$x_{t,j}^{t,i} = \mathrm{E}(x_t | y_{1:t}, S_t = j, S_{t-1} = i)$$

$$P_{t,j}^{t,i} = \mathrm{E}\left[ (x_t - x_{t,j}^{t,i})(x_t - x_{t,j}^{t,i})' | y_{1:t}, S_t = j, S_{t-1} = i \right]$$

$$x_{t,j}^{t} = \mathrm{E}(x_t | y_{1:t}, S_t = j)$$

$$P_{t,j}^{t} = \mathrm{E}\left[ (x_t - x_{t,j}^{t})(x_t - x_{t,j}^{t})' \mid y_{1:t-1}, S_t = j \right]$$

### 10.2.1 Kim Filter

Conditional on $S_{t-1} = i$ and $S_t = j$, the Kalman filter algorithm is as follows: for $i = 1, \ldots, m$, given $x_{t-1,i}^{t-1}$ and $P_{t-1,i}^{t-1}$,

$$
\begin{aligned}
x_{t,j}^{t-1,i} &= \mu_j + \Phi_j x_{t-1,i}^{t-1} \\
P_{t,j}^{t-1,i} &= \Phi_j P_{t-1,i}^{t-1} \Phi_j + Q_j \\
\epsilon_{t,j}^{t-1,i} &= y_t - A_j x_{t,j}^{t-1,i} - \Gamma_j z_t \\
\Sigma_{t,j}^{t-1,i} &= A_j P_{t,j}^{t-1,i} A_j' + R_j \\
x_{t,j}^{t,i} &= x_{t,j}^{t-1,i} + P_{t,j}^{t-1,i} A_j' \left(\Sigma_{t,j}^{t-1,i}\right)^{-1} \epsilon_{t,j}^{t-1,i} \\
P_{t,j}^{t,i} &= \left[I - P_{t,j}^{t-1,i} A_j' \left(\Sigma_{t,j}^{t-1,i}\right)^{-1} A_j\right] P_{t,j}^{t-1,i}
\end{aligned}
$$

for $j = 1, \ldots, m$.

In order to suppress $i$,

$$x_{t,j}^t = \sum_i \Pr(S_{t-1} = i | y_{1:t}, S_t = j) x_{t,j}^{t,i}$$

and

$$
\begin{aligned}
P_{t,j}^t &= \mathrm{E}\left[(x_t - x_{t,j}^t)(x_t - x_{t,j}^t)' | y_{1:t}, S_t = j\right] \\
&= \sum_i \Pr(S_{t-1} = i | y_{1:t}, S_t = j) \mathrm{E}\left[(x_t - x_{t,j}^t)(x_t - x_{t,j}^t)' | y_{1:t}, S_t = j, S_{t-1} = i\right] \\
&= \sum_i \Pr(S_{t-1} = i | y_{1:t}, S_t = j) \left\{\mathrm{E}\left[(x_t - x_{t,j}^{t,i})(x_t - x_{t,j}^{t,i})' | y_{1:t}, S_t = j, S_{t-1} = i\right]\right. \\
&\quad \left. + (x_{t,j}^{t,i} - x_{t,j}^t)(x_{t,j}^{t,i} - x_{t,j}^t)'\right\} \\
&= \sum_i \Pr(S_{t-1} = i | y_{1:t}, S_t = j) \left[P_{t,j}^{t,i} + (x_{t,j}^{t,i} - x_{t,j}^t)(x_{t,j}^{t,i} - x_{t,j}^t)'\right]
\end{aligned}
$$

Note that

$$\Pr(S_{t-1} = i | y_{1:t}, S_t = j) = \frac{\Pr(S_t = j, S_{t-1} = i | y_{1:t})}{\Pr(S_t = j | y_{1:t})}$$

where both $\Pr(S_t = j, S_{t-1} = i | y_{1:t})$ and $\Pr(S_t = j | y_{1:t})$ can be calculated using the Hamilton filter specified in Property 9.3.

The Hamilton filter needs $\mathrm{p}(y_t | y_{1:t-1}, S_t = j, S_{t-1} = i)$. Note that

$$
\begin{aligned}
\mathrm{E}(y_t | y_{1:t-1}, S_t = j, S_{t-1} = i) &= \mathrm{E}(A_{S_t} x_t + \Gamma_{S_t} z_t + v_t | y_{1:t-1}, S_t = j, S_{t-1} = i) \\
&= A_j x_{t,j}^{t-1,i} + \Gamma_j z_t
\end{aligned}
$$

and

$$\mathrm{Var}(y_t|y_{1:t-1}, S_t = j, S_{t-1} = i) = \Sigma_{t,j}^{t-1,i}$$

Therefore,

$$\mathrm{p}(y_t|y_{1:t-1}, S_t = j, S_{t-1} = i) = \mathrm{g}\left(y_t; A_j x_{t,j}^{t-1,i} + \Gamma_j z_t, \Sigma_{t,j}^{t-1,i}\right)$$

Kim's (1994) basic filter can be summarized as follows

**Property 10.4** (Kim Filter)**.** *Given* $x_{t-1,j}^{t-1}$*,* $P_{t-1,j}^{t-1}$ *and* $\Pr(S_{t-1} = i|y_{1:t-1})$ *for* $i = 1, \ldots, m$*, we calculate, for* $t = 1, \ldots, n$*,*

$$\begin{aligned}
x_{t,j}^{t-1,i} &= \mu_j + \Phi_j x_{t-1,i}^{t-1} \\
P_{t,j}^{t-1,i} &= \Phi_j P_{t-1,i}^{t-1} \Phi_j + Q_j \\
\epsilon_{t,j}^{t-1,i} &= y_t - A_j x_{t,j}^{t-1,i} - \Gamma_j z_t \\
\Sigma_{t,j}^{t-1,i} &= A_j P_{t,j}^{t-1,i} A_j' + R_j \\
x_{t,j}^{t,i} &= x_{t,j}^{t-1,i} + P_{t,j}^{t-1,i} A_j' \left(\Sigma_{t,j}^{t-1,i}\right)^{-1} \epsilon_{t,j}^{t-1,i} \\
P_{t,j}^{t,i} &= \left[I - P_{t,j}^{t-1,i} A_j' \left(\Sigma_{t,j}^{t-1,i}\right)^{-1} A_j\right] P_{t,j}^{t-1,i} \\
\mathrm{p}(y_t|y_{1:t-1}, S_t = j, S_{t-1} = i) &= \mathrm{g}\left(y_t; A_j x_{t,j}^{t-1,i} + \Gamma_j z_t, \Sigma_{t,j}^{t-1,i}\right) \\
\Pr(S_t = j, S_{t-1} = i|y_{1:t-1}) &= \pi_{ij} \Pr(S_{t-1} = i|y_{1:t-1}) \\
\Pr(S_t = j, S_{t-1} = i|y_{1:t}) &= \frac{\mathrm{p}(y_t|S_t = j, S_{t-1} = i, y_{1:t-1}) \Pr(S_t = j, S_{t-1} = i|y_{1:t-1})}{\sum_j \sum_i \mathrm{p}(y_t|S_t = j, S_{t-1} = i, y_{1:t-1}) \Pr(S_t = j, S_{t-1} = i|y_{1:t-1})} \\
\Pr(S_t = j|y_{1:t}) &= \sum_i \Pr(S_t = j, S_{t-1} = i|y_{1:t}) \\
\Pr(S_{t-1} = i|y_{1:t}, S_t = j) &= \frac{\Pr(S_t = j, S_{t-1} = i|y_{1:t})}{\Pr(S_t = j|y_{1:t})} \\
x_{t,j}^t &= \sum_i \Pr(S_{t-1} = i|y_{1:t}, S_t = j) x_{t,j}^{t,i} \\
P_{t,j}^t &= \sum_i \Pr(S_{t-1} = i|y_{1:t}, S_t = j) \left[P_{t,j}^{t,i} + (x_{t,j}^{t,i} - x_{t,j}^t)(x_{t,j}^{t,i} - x_{t,j}^t)'\right]
\end{aligned}$$

In the filter presented above, we derived the distribution of $x_t$ conditional on $y_{1:t-1}, S_t$ and $S_{t-1}$. Instead, we could have derived the distribution of $x_t$ conditional on $y_{1:t-1}, S_t, S_{t-1}$ and $S_{t-2}$ to obtain more accurate inferences. In general, as we carry more states at each iteration, we can get more efficient inferences, but only at the cost of increased computation time and the model's tractability. As a rule of thumb, when only $S_t$ or $S_t$ and $S_{t-1}$ show up in the state-space representation, carrying $m^2$ (because $S_t, S_{t-1} = 1, \ldots, m$) states is usually enough. When $S_t, S_{t-1}, \ldots, S_{t-r}(r > 1)$ show up in the state-space representation, Kim (1994) recommends carrying at least $m^{r+1}$ states at each iteration.

### 10.2.2 My Filter

Similar to Shumway and Stoffer's filter 10.3, we can derive the distribution of $x_t$ based on $y_{1:t-1}$ and only $S_t$. In contrast to Kim's filter 10.4, $S_{t-1}$ is not necessary.

**Property 10.5.** *Given $x_{t-1}^{t-1}, P_{t-1}^{t-1}$ and $\pi_i(t-1|t-1)$ for $i = 1, 2, \ldots, m$,*

$$x_{t,j}^{t-1} = \Phi_j x_{t-1}^{t-1} + \mu_j, \quad P_{t,j}^{t-1} = \Phi_j P_{t-1}^{t-1} \Phi_j' + Q_j$$

$$\epsilon_{t,j} = y_t - A_j x_{t,j}^{t-1} - \Gamma_j z_t, \quad \Sigma_{t,j} = A_j P_{t,j}^{t-1} A_j' + R_j, \quad K_{t,j} = P_{t,j}^{t-1} A_j' \Sigma_{t,j}^{-1}$$

$$x_{t,j}^{t} = x_{t,j}^{t-1} + K_{t,j}\epsilon_{t,j}, \quad P_{t,j}^{t} = \left(I - K_{t,j}A_j\right) P_{t,j}^{t-1}$$

$$\pi_j(t|t-1) = \sum_i \pi_{ij}\pi_i(t-1|t-1)$$

$$\mathrm{p}_j(t|t-1) = \mathrm{g}\left(y_t; A_j x_{t,j}^{t-1} + \Gamma_j z_t, \Sigma_{t,j}\right)$$

$$\pi_j(t|t) = \frac{\pi_j(t|t-1)\mathrm{p}_j(t|t-1)}{\sum_k \pi_k(t|t-1)\mathrm{p}_j(t|t-1)}$$

$$x_t^t = \sum_j \pi_j(t|t) x_{t,j}^t$$

$$P_t^t = \sum_j \pi_j(t|t)\left[P_{t,j}^t + (x_{t,j}^t - x_t^t)(x_{t,j}^t - x_t^t)'\right]$$

To see how to derive $P_t^t$,

$$\begin{aligned}
P_t^t &= \mathrm{E}\left[(x_t - x_t^t)(x_t - x_t^t)'|y_{1:t}\right] \\
&= \mathrm{E}\left\{\mathrm{E}\left[(x_t - x_t^t)(x_t - x_t^t)'|y_{1:t}, S_t\right] \mid y_{1:t}\right\} \\
&= \sum_j \pi_j(t|t)\mathrm{E}\left[(x_t - x_t^t)(x_t - x_t^t)'|y_{1:t}, S_t = j\right] \\
&= \sum_j \pi_j(t|t)\mathrm{E}\left[(x_t - x_{t,j}^t + x_{t,j}^t - x_t^t)(x_t - x_{t,j}^t + x_{t,j}^t - x_t^t)'|y_{1:t}, S_t = j\right] \\
&= \sum_j \pi_j(t|t)\left[P_{t,j}^t + (x_{t,j}^t - x_t^t)(x_{t,j}^t - x_t^t)'\right]
\end{aligned}$$

### 10.2.3 Smoothing

**Property 10.6.** *[My Smoother Part I] For $S_t, S_{t+1} = 1, \dots, m$, given* $\Pr(S_{t+1}|y_{1:n})$ *and* $p(y_{t+1:n}|S_t, S_{t+1}, y_{1:t})$

$$\begin{aligned}
\mathrm{p}(S_t, y_{t+1:n}|S_{t+1}, y_{1:t}) &= \mathrm{p}(y_{t+1,n}|S_t, S_{t+1}, y_{1:t})\frac{\Pr(S_{t+1}|S_t)\Pr(S_t|y_{1:t})}{\sum_{S_t}\Pr(S_{t+1}|S_t)\Pr(S_t|y_{1:t})} \\
\Pr(S_t|S_{t+1}, y_{1:n}) &= \frac{\mathrm{p}(S_t, y_{t+1:n}|S_{t+1}, y_{1:t})}{\sum_{S_t}\mathrm{p}(S_t, y_{t+1:n}|S_{t+1}, y_{1:t})} \\
\Pr(S_t, S_{t+1}|y_{1:n}) &= \Pr(S_{t+1}|y_{1:n})\Pr(S_t|S_{t+1}, y_{1:n}) \\
\Pr(S_t|y_{1:n}) &= \sum_{S_{t+1}}\Pr(S_t, S_{t+1}|y_{1:n})
\end{aligned}$$

*In addition,*

$$\mathrm{p}(y_{t:n}|S_{t-1}, S_t, y_{1:t-1}) \approx \mathrm{p}(y_t|S_{t-1}, S_t, y_{1:t-1})\left[\sum_{S_{t+1}}\mathrm{p}(y_{t+1:n}|S_t, S_{t+1}, y_{1:t})\Pr(S_{t+1}|S_t)\right]$$

*where*

$$y_t|S_{t-1} = i, S_t = j, y_{1:t-1} \sim \mathbf{N}\left(A_j x_{t,j}^{t-1,i} + \Gamma_j z_t, A_j P_{t,j}^{t-1,i} A_j' + R_j\right)$$

*with initial conditions* $\Pr(S_n|y_{1:n})$ *and* $\mathrm{p}(y_n|S_{n-1}, S_n, y_{1:t-1})$ *for* $S_{n-1}, S_n = 1, \dots, m$.

*Proof.* The derivation of the first four equations are the same as in Property 9.4.

To see the fourth one,

$$\mathrm{p}(y_{t:n}|S_{t-1}, S_t, y_{1:t-1}) = \mathrm{p}(y_t|S_{t-1}, S_t, y_{1:t-1})\mathrm{p}(y_{t+1:n}|S_{t-1}, S_t, y_{1:t})$$

where

$$\begin{aligned}
\mathrm{p}(y_{t+1:n}|S_{t-1}, S_t, y_{1:t}) &= \sum_{S_{t+1}}\mathrm{p}(y_{t+1:n}, S_{t+1}|S_{t-1}, S_t, y_{1:t}) \\
&= \sum_{S_{t+1}}\mathrm{p}(y_{t+1:n}|S_{t-1}, S_t, S_{t+1}, y_{1:t})\Pr(S_{t+1}|S_t)
\end{aligned}$$

and we approximate $\mathrm{p}(y_{t+1:n}|S_{t-1}, S_t, S_{t+1}, y_{1:t}) \approx \mathrm{p}(y_{t+1:n}|S_t, S_{t+1}, y_{1:t})$.

To see the fifth equation, recall that

$$y_t = A_{S_t} x_t + \Gamma_{S_t} z_t + v_t, \quad v_t \sim \mathrm{N}(0, R_{S_t}).$$

Conditional on $S_{t-1} = i, S_t = j$ and $y_{1:t-1}$,

$$x_t | S_{t-1} = i, S_t = j, y_{1:t-1} \sim \mathrm{N}\left(x_{t,j}^{t-1,i}, P_{t,j}^{t-1,i}\right)$$
$$v_t | S_{t-1} = i, S_t = j, y_{1:t-1} \sim \mathrm{N}(0, R_j).$$

where $x_{t,j}^{t-1,i}$ and $P_{t,j}^{t-1,i}$ are given by the Kim filter 10.4. So $y_t | S_{t-1} = i, S_t = j, y_{1:t-1}$ is normally distributed.

To derive the variance of that normal distribution, note that

$$\begin{aligned}
&\mathrm{Cov}(x_t, v_t | S_{t-1} = i, S_t = j, y_{1:t-1}) \\
&= \mathrm{E}\left[(x_t - x_{t,j}^{t-1,i}) v_t \mid S_{t-1} = i, S_t = j, y_{1:t-1}\right] \\
&= \mathrm{E}\left\{\mathrm{E}\left[(x_t - x_{t,j}^{t-1,i}) v_t \mid x_{t-1}, S_{t-1} = i, S_t = j, y_{1:t-1}\right] \,\middle|\, S_{t-1} = i, S_t = j, y_{1:t-1}\right\}
\end{aligned}$$

Recall that

$$x_t = \Phi_{S_t} x_{t-1} + \mu_{S_t} + w_t, \quad w_t \sim (0, Q_{S_t})$$
$$\mathrm{Cov}(w_s, v_t) = 0, \quad s, t = 1, \ldots, n.$$

We see

$$\mathrm{E}\left[(x_t - x_{t,j}^{t-1,i}) v_t \mid x_{t-1}, S_{t-1} = i, S_t = j, y_{1:t-1}\right] = 0$$

Therefore,

$$\mathrm{Cov}(x_t, v_t | S_{t-1} = i, S_t = j, y_{1:t-1}) = 0.$$

□

Now, we derive the $x_t^n$.

**Property 10.7** (Smoother Part II)**.** *For $k = 1, \ldots, m$, given $x_{t+1,k}^n = \mathrm{E}(x_{t+1} | y_{1:n}, S_{t+1} = k)$ and $P_{t+1,k}^n = \mathrm{Var}(x_{t+1} | y_{1:n}, S_{t+1} = k)$,*

$$x_{t,j}^{n,k} \approx x_{t,j}^t + P_{t,j}^t \Phi_k' \left(P_{t+1,k}^{t,j}\right)^{-1} \left[x_{t+1,k}^n - x_{t+1,k}^{t,j}\right],$$
$$P_{t,j}^{n,k} \approx P_{t,j}^t + P_{t,j}^t \Phi_k' \left(P_{t+1,k}^{t,j}\right)^{-1} \left[P_{t+1,k}^n - P_{t+1,k}^{t,j}\right] \left(P_{t+1,k}^{t,j}\right)^{-1} \Phi_k P_{t,j}^t$$

*where* $x^{n,k}_{t,j} \equiv \mathrm{E}(x_t|y_{1:n}, S_t = j, S_{t+1} = k)$. *And*

$$x^n_{t,j} = \pi_{jk} x^{n,k}_{t,j}$$
$$P^n_{t,j} = \sum_k \pi_{jk} \left[ P^{n,k}_{t,j} + (x^{n,k}_{t,j} - x^n_{t,j})(x^{n,k}_{t,j} - x^n_{t,j})' \right]$$
$$x^n_t = \Pr(S_t = j|y_{1:n}) x^n_{t,j}$$
$$P^n_t = \sum_j \Pr(S_t = j|y_{1:n}) \left[ P^n_{t,j} + (x^n_{t,j} - x^n_j)(x^n_{t,j} - x^n_t)' \right]$$

*Proof.* Let $\eta_{t+1} \equiv (v'_{t+1:n}, w'_{t+2:n})'$.

Given $y_{1:t}$, $x^t_{t+1,k}$ can be derived through Kim's predictor. Given $x^t_{t+1,k}$ and $x_{t+1} - x^t_{t+1,k}$, we get $x_{t+1}$. Combining $x_{t+1}$ with $\eta_{t+1}$ and $S_{t+1:n}$, we get $y_{t+1}$. Given $x_{t+1}, \eta_{t+1}$ and $S_{t+1:n}$, we get $x_{t+2}$. Then we get $y_{t+2}$ ... We see $y_{1:t}, x_{t+1} - x^t_{t+1,k}, \eta_{t+1}$ and $S_{t+1:n}$ generate $y_{1:n}$.

$$\begin{aligned}
&\mathrm{E}(x_t|y_{1:t}, x_{t+1} - x^{t,j}_{t+1,k}, \eta_{t+1}, S_{t:n}) \\
&= \mathrm{E}(x_t|y_{1:t}, x_{t+1} - x^{t,j}_{t+1,k}, S_{t:n}) \\
&= \mathrm{E}(x_t|y_{1:t}, x_{t+1} - x^{t,j}_{t+1,k}, S_t, S_{t+1}) \\
&= \mathrm{E}(x_t|y_{1:t}, S_t, S_{t+1}) \\
&\quad + \mathrm{Cov}(x_t, x_{t+1} - x^{t,j}_{t+1,k}|y_{1:t}, S_t, S_{t+1}) \mathrm{Var}(x_{t+1} - x^{t,j}_{t+1,k}|y_{1:t}, S_t, S_{t+1})^{-1} \\
&\quad\ \left[ x_{t+1} - x^{t,j}_{t+1,k} - \mathrm{E}(x_{t+1} - x^{t,j}_{t+1,k}|y_{1:t}, S_t, S_{t+1}) \right]
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{E}(x_t|y_{1:t}, S_t, S_{t+1}) &= \mathrm{E}(x_t|y_{1:t}, S_t) \\
&= x^t_{t,j},
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}(x_t, x_{t+1} - x^{t,j}_{t+1,k}|y_{1:t}, S_t, S_{t+1}) &= \mathrm{Cov}(x_t, x_t - x^t_{t,j}|y_{1:t}, S_t)\Phi'_k \\
&= P^t_{t,j}\Phi'_k,
\end{aligned}$$

and

$$\mathrm{Var}(x_{t+1} - x^{t,j}_{t+1,k}|y_{1:t}, S_t, S_{t+1})^{-1} = P^{t,j}_{t+1,k}$$
$$\mathrm{E}(x_{t+1} - x^{t,j}_{t+1,k}|y_{1:t}, S_t, S_{t+1}) = 0.$$

So

$$\mathrm{E}(x_t|y_{1:t}, x_{t+1} - x^{t,j}_{t+1,k}, \eta_{t+1}, S_{t:n}) = x^t_{t,j} + P^t_{t,j}\Phi'_k\left(P^{t,j}_{t+1,k}\right)^{-1}(x_{t+1} - x^{t,j}_{t+1,k})$$

Therefore,

$$\begin{aligned}
\mathrm{E}(x_t|y_{1:n}, S_t, S_{t+1}) &= \mathrm{E}\left[\mathrm{E}(x_t|y_{1:t}, x_{t+1} - x^{t,j}_{t+1,k}, \eta_{t+1}, S_{t:n})\middle|\, y_{1:n}, S_t, S_{t+1}\right] \\
&= x^t_{t,j} + P^t_{t,j}\Phi'_k\left(P^{t,j}_{t+1,k}\right)^{-1}\left[\mathrm{E}(x_{t+1}|y_{1:n}, S_t, S_{t+1}) - x^{t,j}_{t+1,k}\right] \\
&\approx x^t_{t,j} + P^t_{t,j}\Phi'_k\left(P^{t,j}_{t+1,k}\right)^{-1}\left[\mathrm{E}(x_{t+1}|y_{1:n}, S_{t+1}) - x^{t,j}_{t+1,k}\right]
\end{aligned}$$

Note that I do not know how small the difference is between $\mathrm{E}(x_{t+1}|y_{1:n}, S_t, S_{t+1})$ and $\mathrm{E}(x_{t+1}|y_{1:n}, S_{t+1})$. P107, Kim and Nelson (1999), did not even realize here is an approximation.

$P^n_{t,j}$ and $P^n_t$ can be derived easily using the law of total variance:

$$\mathrm{Var}X = \mathrm{E}\left[\mathrm{Var}(X|Y)\right] + \mathrm{Var}\left[\mathrm{E}(X|Y)\right]$$

□

An alternative algorithm to Property 10.7 is given below.

Denote $x^{n,n}_t = \mathrm{E}(x_t|y_{1:n}, S_{1:n})$.

$$\begin{aligned}
x^n_t &= \mathrm{E}\left(x^{n,n}_t|y_{1:n}\right) \\
&= \sum_{S_{1:n}} x^{n,n}_t \Pr(S_{1:n}|y_{1:n})
\end{aligned}$$

where $\Pr(S_{1:n}|y_{1:n})$ can be derived using Property 10.6:

$$\Pr(S_{1:n}|y_{1:n}) = \Pr(S_n|y_{1:n})\Pr(S_{n-1}|y_{1:n}, S_n)\dots\Pr(S_t|y_{1:n}, S_{t+1})\dots\Pr(S_1|y_{1:n}, S_2)$$

and $x^{n,n}_t$ can be derived using the basic Kalman filter 2.1 and smoother 2.4.

This algorithm is highly computer-intensive in light of the number of possible $S_{1:n}$ paths.

# 11 State-Space Models with Heteroskedasticy

Based on Chapter 6, Kim and Nelson (1999).

As suggested by Hamilton and Susmel (1994), Markov switching heteroscedasticity discussed in Section 10.2 may be more appropriate for low-frequency data over a long period of time, whereas ARCH-type heteroscedasticity to be discussed may be more appropriate for high-frequency data over a short period of time.

Besides, Dieobold (1986), Lastrapes (1989), Lamoureux and Lastrapes (1990), and others suggest that a failure to allow for regimes shifts leads to an overstatement of the persistence of the variance of a series, possibly leading to an integrated ARCH or GARCH.

## 11.1 ARCH($q$) Model

Based on Wikipedia.

Let $\epsilon_t$ denote the error terms. These $\epsilon_{1:n}$ are split into a stochastic piece $z_t$ and a time-dependent standard deviation $\sigma_t$ so that

$$\epsilon_t = \sigma_t z_t.$$

The random variable $z_t$ is a strong white noise process [3]. The series $\sigma_t^2$ is modeled by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_q \epsilon_{t-q}^2$$

where $\alpha_i > 0$, $i = 0, 1, \ldots, q$.

## 11.2 GARCH($p$, $q$) Model

Based on Wikipedia.

If an ARMA model is assume for the error variance, the model is a generalized ARCH model.

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \ldots \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_p \sigma_{t-p}^2$$

[3] A process $\epsilon_t$ is a strong sense white noise if $\epsilon_t$ is i.i.d. with mean 0 and finite variance.

## 11.3 LGARCH and EGARCH

Based on Malmsten (2004).

A logarithmic GARCH$(p, q)$ model

$$\log \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \log \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \log \sigma_{t-j}^2,$$

is a special case of an exponential GARCH$(p, q)$ model:

$$\log \sigma_t^2 = \omega + \sum_{i=1}^{q} g_i(\epsilon_{t-i}) + \sum_{j=1}^{p} \beta_j \log \sigma_{t-j}^2.$$

## 11.4 IGARCH

Based on Wikipedia.

Integrated GARCH is a restricted version of the GARCH model, where the persistence parameters sum up to one, and imports a unit root in the GARCH process.

$$\sum_i \alpha_i + \sum_j \beta_j = 1$$

## 11.5 State-Space Models with ARCH Disturbances

Based on Chapter 6, Kim and Nelson (1999).

Harvey, Ruiz, and Sentana (1992) proposed the following model:

$$x_t = \Phi x_{t-1} + \Upsilon u_t + w_t + \lambda w_t^*$$

$$y_t = A x_t + \Gamma u_t + v_t + \Lambda v_t^*$$

$$w_t \sim \mathrm{N}(0, Q), \quad v_t \sim \mathrm{N}(0, R)$$

$$w_t^* \sim \mathrm{N}(0, h_{1t}), \quad v_2^* \sim \mathrm{N}(0, h_{2t})$$

where $\lambda$ is $p \times 1$, $\Lambda$ is $q \times 1$, $w_t^*$ and $v_t^*$ are $1 \times 1$, and

$$h_{1t} = \alpha_0 + \alpha_1 w_{t-1}^{*2}$$

$$h_{2t} = \gamma_0 + \gamma_1 v_{t-1}^{*2}.$$

Because of the past unobserved shocks $w^*_{t-1}$ and $v^*_{t-1}$, both $h_{1t}$ and $h_{2t}$ are unobserved. Thus, the Kamman filter introduced in Property 2.1 is not operable without approximations.

# 12 Stochastic Volatility

Stochastic volatility (SV) models are an alternative to GARCH-type models.

Let $r_t$ denote the returns of some financial asset. Assume

$$r_t = \sigma_t \varepsilon_t, \tag{36}$$

where $\varepsilon_t$ is an i.i.d. sequence with zero mean and unit variance, and the volatility process, $\sigma_t$, is a non-negative process,

$$\sigma_t = \beta \exp\left\{\frac{x_t}{2}\right\},$$

with

$$x_t = \phi_0 + \phi_1 x_{t-1} + w_t, \quad w_t \sim \text{i.i.d. N}(0, \sigma_w^2). \tag{37}$$

$w_t$ and $\varepsilon_t$ are assumed to be mutually independent. $|\phi_1| < 1$.

Define $y_t = \log r_t^2$, and $v_t = \log \varepsilon_t^2$. Then Equation (36) may be rewritten as

$$y_t = \alpha + x_t + v_t. \tag{38}$$

A constant is needed in either Equation (37) or (38), but not both.

If $\varepsilon_t^2$ has a log-normal distribution, Equations (37) and (38) would form a Gaussian state-space model. Unfortunately, that assumption does not seem to work well.

Instead, one often keeps the ARCH normality assumption, e.g. $\varepsilon_t \sim$ i.i.d. N(0, 1), in which case, $v_t$ is non-Gaussian and distributed as the log of a chi-squared distribution with one degree of freedom. The density is highly skewed and given by

$$f(v) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(e^v - v)\right\}, \quad -\infty < v < \infty.$$

Note that the density is not flexible because there are no free parameters to be estimated. Kim, Shephard and Chib (1998) proposed modeling the log of a chi-squared variable by a mixture of seven normals to approximate the first four moments.

But the assumption that $\varepsilon_t$ is Gaussian is unrealistic for some applications. More generally, and in an effort to to keep matters simple, the observation equation (38) is written as

$$y_t = \alpha + x_t + \eta_t \tag{39}$$

where $\eta_t$ is a white noise, whose distribution is a mixture of two normals, one centered at zero

$$\eta_t = I_t z_{t0} + (1 - I_t) z_{t1},$$

where $I_t$ is an i.i.d. Bernoulli process, $\Pr(I_t = 0) = \pi_0$, $\Pr(I_t = 1) = \pi_1$; $z_{t0} \sim$ i.i.d. $\mathrm{N}(0, \sigma_0^2)$, and $z_{t1} \sim$ i.i.d. $\mathrm{N}(\mu_1, \sigma_1^2)$. So

$$\eta_t \sim \mathrm{N}\left((1 - I_t)\mu_1, I_t\sigma_0^2 + (1 - I_t)\sigma_1^2\right). \tag{40}$$

So Equations (37), (39) and (40) constitute a state-space model with regime switching. Property 10.4, or Property 10.5, applies here. In particular,

**Property 12.1.** *Given $x_{t-1}^{t-1}$ and $P_{t-1}^{t-1}$, then for $j = 0, 1$,*

$$\begin{gathered}
x_t^{t-1} = \phi_0 + \phi_1 x_{t-1}^{t-1}, \quad P_t^{t-1} = \phi_1^2 P_{t-1}^{t-1} + \sigma_w^2 \\
\epsilon_t = y_t - \alpha - x_t^{t-1}, \quad \Sigma_{t,j} = P_t^{t-1} + \sigma_j^2, \quad K_{t,j} = P_t^{t-1} \Big/ \Sigma_{t,j} \\
x_{t,j}^t = x_t^{t-1} + K_{t,j}\epsilon_t, \quad P_{t,j}^t = P_t^{t-1} - K_{t,j} P_t^{t-1} \\
\mathrm{p}_j(t|t-1) = \mathrm{g}\left(y_t; \alpha + x_t^{t-1}, \Sigma_{t,j}\right) \\
\pi_j(t|t) = \frac{\pi_j \mathrm{p}_j(t|t-1)}{\sum_j \pi_j \mathrm{p}_j(t|t-1)} \\
x_t^t = \sum_j x_{t,j}^t \\
P_t^t = \sum_j \pi_j(t|t)\left[P_{t,j}^t + (x_{t,j}^t - x_t^t)^2\right].
\end{gathered}$$

Note that Equation (6.206), Page 359, Shumway and Stoffer (2016) is wrong.

# 13 State Space Models and Gibbs Sampling

We assume that the model is given by (1) and (2) for the sake of brevity, even though inputs are allowed in the model.

Inferences, within the classical approach, about the state variables are conditional on the estimated values of the parameters:

$$x_t^s = \mathrm{E}(x_t | y_{1:s}, \Theta = \hat{\Theta})$$

We now consider some Bayesian approaches to fitting linear Gaussian state space models via Markov chain Monte Carlo (MCMC) methods. Within the Bayesian approach, both the model's parameters and the state variables are treated as random variables. In contrast to the classical approach, inferences on $x_{0:n}$ is based on the joint distribution of $x_{0:n}$ and $\Theta$, not a conditional distribution.

The most common MCMC method is the Gibbs sampler, which is developed by Hastings (1970) in the statistical setting and by Geman and Geman (1984) in the context of image restoration. The basic strategy is to use conditional distributions to set up a Markov chain to obtain samples from a joint distribution.

In this case, Frühwirth-Schnatter (1994) and Carter and Kohn (1994) established the MCMC procedure we will discuss here.

**Example 13.1** (Gibbs Sampling for the Bivariate Normal)**.** Suppose we wish to obtain samples from a bivariate normal distribution,

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right],$$

where $|\rho| < 1$, but we can only generate samples from a univariate normal.

- The univariate conditionals are

$$(X|Y = y) \sim \mathrm{N}(\rho y, 1 - \rho^2)$$
$$(Y|X = x) \sim \mathrm{N}(\rho x, 1 - \rho^2).$$

  We can simulate from these distributions.

- Construct a Markov chain: Pick $X^{(0)} = x_0$, and then iterate the process $X^{(0)} \to Y^{(0)} \to X^{(1)} \to Y^{(1)} \to \cdots X^{(k)} \to Y^{(k)} \to \cdots$, where

$$(Y^{(k)}|X^{(k)} = x_k) \sim \mathrm{N}(\rho x_k, 1-\rho^2)$$
$$(X^{(k)}|Y^{(k-1)} = y_{k-1}) \sim \mathrm{N}(\rho y_{k-1}, 1-\rho^2).$$

- The joint distribution of $(X^{(k)}, Y^{(k)})$ is

$$\begin{pmatrix} X^{(k)} \\ Y^{(k)} \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \rho^{2k} x_0 \\ \rho^{2k+1} x_0 \end{pmatrix}, \begin{pmatrix} 1-\rho^{4k} & \rho(1-\rho^{4k}) \\ \rho(1-\rho^{4k}) & 1-\rho^{4k+2} \end{pmatrix}\right].$$

Thus for any starting value $x_0$, $\left(X^{(k)}, Y^{(k)}\right) \xrightarrow{d} (X, Y)$; the speed depends on $\rho$.

- One would run the chain and throw away the initial $n_0$ sampled values (burnin) and retain the rest.

For state space models, the main objective is to obtain the posterior density of the parameters $\mathrm{p}(\Theta|y_{1:n})$ or the states $\mathrm{p}(x_{0:n}|y_{1:n})$ if the states are meaningful. It is generally easier to get samples from the full posterior $\mathrm{p}(\Theta, x_{0:n}|y_{1:n})$ and then marginalize to obtain $\mathrm{p}(\Theta|y_{1:n})$ and $\mathrm{p}(x_{0:n}|y_{1:n})$. And to get $\mathrm{p}(\Theta, x_{0:n}|y_{1:n})$, the most popular method is to run a full Gibbs sampler.

**Procedure 13.1.** [Gibbs Sampler for State Space Models] Pick $\Theta^{(0)}$. Then run Step 2 below to get $x_{0:n}^{(0)}$. And then

1. draw $\Theta' \sim \mathrm{p}(\Theta|x_{0:n}, y_{1:n})$.

2. draw $x'_{0:n} \sim \mathrm{p}(x_{0:n}|\Theta', y_{1:n})$.

3. Throw away the initial $n_0$ sampled values $\left(\Theta^{(k)}, x_{0:n}^{(k)}\right)$ and retain the rest.

## 13.1 Procedure 1

Procedure 13.1 - 1 is generally much easier. To accomplish Procedure 13.1 - 1, note that

$$\mathrm{p}(\Theta|x_{0:n}, y_{1:n}) \propto \mathrm{p}(\Theta)\mathrm{p}(x_0|\Theta) \prod_{t=1}^{n} \mathrm{p}(x_t|x_{t-1}, \Theta)\mathrm{p}(y_t|x_t, \Theta) \tag{41}$$

where $p(\Theta)$ is the prior on the parameters, and $p(x_0|\Theta)$ is the *pseudo*-prior on $x_0$. The prior often depends on "hyperparameters". For simplicity, these hyperparameters are assumed to be known. The parameters are sometimes conditionally independent (meaning that a parameter can be drawn from a univariate distribution independent of other parameters).

**Example 13.2** (A Univariate Model)**.** In the univariate model

$$x_t = \phi x_{t-1} + w_t$$

$$y_t = x_t + v_t$$

$$w_t \sim \text{i.i.d. } \mathrm{N}(0, \sigma_w^2)$$

$$v_t \sim \text{i.i.d. } \mathrm{N}(0, \sigma_v^2)$$

and $w_t$ is independent of $v_t$. We can chose the priors

$$\phi \sim \mathrm{N}(\mu_\phi, \sigma_\phi^2)$$
$$\sigma_w^2 \sim \mathrm{IG}\left(\frac{a_0}{2}, \frac{b_0}{2}\right)$$
$$\sigma_v^2 \sim \mathrm{IG}\left(\frac{c_0}{2}, \frac{d_0}{2}\right)$$

Then from (41) we know

$$\phi|\sigma_w^2, x_{0:n}, y_{1:n} \sim \mathrm{N}(\cdot, \cdot)$$

$$\sigma_w^2|\phi, x_{0:n}, y_{1:n} \sim \mathrm{IG}(\cdot, \cdot)$$

$$\sigma_v^2|x_{0:n}, y_{1:n} \sim \mathrm{IG}(\cdot, \cdot)$$

So we can draw $\sigma_v^2$ from the univariate inverse gamma distribution, and $(\phi, \sigma_w^2)$ via Gibbs sampling.

**Example 13.3** (Structural Model)**.** Consider the Johnson & John quarterly earnings per quarter series discussed in Example 5.1.

The parameters to be estimated are the transition parameter, $\phi > 1$, the observation noise variance, $\sigma_v^2$, and the state noise variances, $\sigma_{w,11}^2$ and $\sigma_{w,22}^2$.

Write $\phi = 1 + \beta$. As is typical[4], we put a Normal-Inverse Gamma prior on $(\beta, \sigma^2_{w,11})$, i.e.,

$$\beta|\sigma^2_{w,11} \sim \mathrm{N}(b_0, \sigma^2_{w,11} B_0),$$
$$\sigma^2_{w,11} \sim \mathrm{IG}\left(\frac{n_0}{2}, \frac{n_0 s_0^2}{2}\right),$$

with known hyperparameters $b_0, B_0, n_0, s_0^2$.

We also use IG priors for the other two variance parameters, $\sigma^2_v \sim \mathrm{IG}\left(\frac{n_0}{2}, \frac{n_0 s_0^2}{2}\right)$ and $\sigma_{w,22} \sim \mathrm{IG}\left(\frac{n_0}{2}, \frac{n_0 s_0^2}{2}\right)$.

## 13.2 Procedure 2

Procedure 13.1 - 2 is generally difficult. For linear Gaussian models, however, it is easy to perform.

Carter and Kohn (1994) establish the so-called forward-filtering, backward-sampling (FFBS) algorithm.

For Procedure 13.1 - 2, write the *pseudo*-posterior $\mathrm{p}(x_{0:n}|\Theta, y_{1:n})$ as $\mathrm{p}_\Theta(x_{0:n}|y_{1:n})$ to save space. Because of the Markov structure,

$$\mathrm{p}_\Theta(x_{0:n}|y_{1:n}) = \mathrm{p}_\Theta(x_n|y_{1:n})\mathrm{p}_\Theta(x_{n-1}|x_n, y_{1:n-1}) \cdots \mathrm{p}_\Theta(x_t|x_{t+1}, y_{1:t}) \cdots \mathrm{p}_\Theta(x_0|x_1) \tag{42}$$

From (42), we see that we must obtain the densities $\mathrm{p}_\Theta(x_t|x_{t+1}, y_{1:t})$.

### 13.2.1 Uncorrelated $\boldsymbol{w}_t, \boldsymbol{v}_t$

When $\boldsymbol{w}_t$ and $\boldsymbol{v}_t$ are uncorrelated as is the case in (1) and (2) ,

$$\begin{aligned} \mathrm{p}_\Theta(x_t|x_{t+1}, y_{1:t}) &\propto \mathrm{p}_\Theta(x_t|y_{1:t})\mathrm{p}_\Theta(x_{t+1}|x_t, y_{1:t}) \\ &= \mathrm{p}_\Theta(x_t|y_{1:t})\mathrm{p}_\Theta(x_{t+1}|x_t) \end{aligned} \tag{43}$$

where

$$x_t|y_{1:t} \sim \mathrm{N}^\Theta_p(x^t_t, P^t_t)$$
$$x_{t+1}|x_t \sim \mathrm{N}^\Theta_p(\Phi x_t, Q)$$

[4] Different from Example 13.2, the prior for $\beta$ here is conditional on the variance parameter

So $x_t|x_{t+1}, y_{1:t}$ is a normal distribution with mean and variance given by

$$m_t = x_t^t + J_t(x_{t+1} - x_{t+1}^t) \tag{44}$$

$$V_t = P_t^t - J_t P_{t+1}^t J_t' \tag{45}$$

for $t = n-1, n-2, \ldots, 0$, where $P_{t+1}^t$ is given in Property 2.1 and $J_t$ is defined in Property 2.4.

Hence, given $\Theta$, the algorithm is to first sample $x_n$ from a $\mathrm{N}_p^{\Theta}(x_n^n, P_n^n)$, and then sample $x_t$, for $t = n-1, n-2, \ldots, 0$, from a $\mathrm{N}_p^{\Theta}(m_t, V_t)$, where the conditioning value of $x_{t+1}$ is the value previously sampled. Note that the *pseudo*-prior, $\mathrm{p}_{\Theta}(x_0)$, is needed in order to sample $x_0$.

### 13.2.2 Correlated $\boldsymbol{w}_t, \boldsymbol{v}_t$

But when $\boldsymbol{w}_t$ and $\boldsymbol{v}_t$ are correlated as in Section 6, things are a little different:

$$\begin{aligned}
\mathrm{p}_{\Theta}(x_t|x_{t+1}, y_{1:t}) &\propto \mathrm{p}_{\Theta}(x_t|y_{1:t})\mathrm{p}_{\Theta}(x_{t+1}|x_t, y_{1:t}) \\
&= \mathrm{p}_{\Theta}(x_t|y_{1:t})\mathrm{p}_{\Theta}(x_{t+1}|x_t, y_t).
\end{aligned}$$

To derive the conditional mean and conditional variance, according to Shumway and Stoffer (2017, p. 497, Eq. (B.9)),

$$\mathrm{E}(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \boldsymbol{y}_{1:t}) = \boldsymbol{x}_t^t + \mathrm{Cov}(\boldsymbol{x}_t, \boldsymbol{x}_{t+1}|\boldsymbol{y}_{1:t})\left(P_{t+1}^t\right)^{-1}\left(\boldsymbol{x}_{t+1} - \boldsymbol{x}_{t+1}^t\right), \tag{46}$$

$$\mathrm{Var}(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \boldsymbol{y}_{1:t}) = \boldsymbol{P}_t^t - \mathrm{Cov}(\boldsymbol{x}_t, \boldsymbol{x}_{t+1}|\boldsymbol{y}_{1:t})\left(P_{t+1}^t\right)^{-1}\mathrm{Cov}'(\boldsymbol{x}_t, \boldsymbol{x}_{t+1}|\boldsymbol{y}_{1:t}) \tag{47}$$

where

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{x}_t, \boldsymbol{x}_{t+1}|\boldsymbol{y}_{1:t}) &= \mathrm{Cov}(\boldsymbol{x}_t, \boldsymbol{\Phi}\boldsymbol{x}_t + \boldsymbol{w}_t|\boldsymbol{y}_{1:t}) \\
&= \boldsymbol{P}_t^t\boldsymbol{\Phi}' + \mathrm{Cov}(\boldsymbol{x}_t, \boldsymbol{w}_t|\boldsymbol{y}_{1:t}),
\end{aligned}$$

and $\mathrm{Cov}(\boldsymbol{x}_t, \boldsymbol{w}_t|\boldsymbol{y}_{1:t})$ is the top-right block of $\mathrm{Var}\left((\boldsymbol{x}_t', \boldsymbol{w}_t')'|\boldsymbol{y}_{1:t}\right)$. Then, again, according to Shumway and Stoffer (2017, p. 497, Eq. (B.9)),

$$\begin{aligned}
&\mathrm{Var}\left((\boldsymbol{x}_t', \boldsymbol{w}_t')'|\boldsymbol{y}_{1:t}\right) \\
&= \mathrm{Var}\left((\boldsymbol{x}_t', \boldsymbol{w}_t')'|\boldsymbol{y}_{1:t-1}\right) \\
&\quad - \mathrm{Cov}\left((\boldsymbol{x}_t', \boldsymbol{w}_t')', \boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}\right)\mathrm{Var}^{-1}(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1})\mathrm{Cov}'\left((\boldsymbol{x}_t', \boldsymbol{w}_t')', \boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}\right),
\end{aligned}$$

where

$$\mathrm{Var}\left((\boldsymbol{x}_t', \boldsymbol{w}_t')' | \boldsymbol{y}_{1:t-1}\right) = \begin{pmatrix} \boldsymbol{P}_t^t & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix},$$

and

$$\mathrm{Cov}\left((\boldsymbol{x}_t', \boldsymbol{w}_t')', \boldsymbol{y}_t | \boldsymbol{y}_{1:t-1}\right) = \begin{pmatrix} \boldsymbol{P}_t^{t-1}\boldsymbol{A}' \\ \boldsymbol{S} \end{pmatrix}.$$

Therefore,

$$\begin{aligned} &\mathrm{Cov}(\boldsymbol{x}_t, \boldsymbol{x}_{t+1} | \boldsymbol{y}_{1:t}) \\ &= \boldsymbol{P}_t^t \boldsymbol{\Phi}' - \left[ \begin{pmatrix} \boldsymbol{P}_t^{t-1}\boldsymbol{A}' \\ \boldsymbol{S} \end{pmatrix} (\boldsymbol{A}\boldsymbol{P}\boldsymbol{A}' + \boldsymbol{R})^{-1} \begin{pmatrix} \boldsymbol{A}\boldsymbol{P}_t^{t-1} & \boldsymbol{S}' \end{pmatrix} \right]_{(1:m,\ (m+1):2m)}. \end{aligned}$$

Plug it into Equations (46) and (47), and we get the conditional distribution of $\boldsymbol{x}_t | \boldsymbol{x}_{t+1}, \boldsymbol{y}_{1:t}$.

## 13.3 Procedure 2 When Q is Singular

Consider the following dynamic factor model with AR(1) common and individual components

**Example 13.4** (A Dynamic Factor Model - AR(1))**.**

$$\begin{aligned} y_{1t} &= \gamma_1 C_t + e_{1t} \\ y_{2t} &= \gamma_2 C_t + e_{2t} \\ C_t &= \phi C_{t-1} + w_t, \quad w_t \sim \mathrm{N}(0, 1) \\ e_{it} &= \psi_i e_{i,t-1} + \epsilon_{it}, \quad \epsilon_{it} \sim \text{i.i.d. } \mathrm{N}(0, \sigma_i^2), \quad i = 1, 2 \end{aligned}$$

A state-space representation of the model is given by

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \gamma_1 & 1 & 0 \\ \gamma_2 & 0 & 1 \end{pmatrix} \begin{pmatrix} C_t \\ e_{1t} \\ e_{2t} \end{pmatrix}$$

$$(y_t = Ax_t)\,,$$

$$\begin{pmatrix} C_t \\ e_{1t} \\ e_{2t} \end{pmatrix} = \begin{pmatrix} \phi & & \\ & \psi_1 & \\ & & \psi_2 \end{pmatrix} \begin{pmatrix} C_{t-1} \\ e_{1,t-1} \\ e_{2,t-2} \end{pmatrix} + \begin{pmatrix} w_t \\ \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

$$(x_t = \Phi x_{t-1} + v_t)\,.$$

Note that

$$Q \equiv \mathrm{E}(v_t v_t') = \begin{pmatrix} 1 & & \\ & \sigma_1^2 & \\ & & \sigma_2^2 \end{pmatrix}$$

is positive definite.

However, in many state-space models, the covariance matrix of the shocks to the transition equation, $Q$, may not be positive definite. Then Procedure 13.1 - 2, elaborated in the preceding subsection, needs to be modified. Consider the following extension of Example 13.4:

**Example 13.5** (A Dynamic Factor Model - AR(1))**.** Suppose the common component

$$C_t = \phi_1 C_{t-1} + \phi_2 C_{t-2} + w_t, \quad w_t \sim \text{i.i.d. N}(0, 1)$$

where roots of $(1 - \phi_1 L - \phi_2 L^2) = 0$ lie outside the complex unit circle.

A state-space representation of the model may be given by

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \gamma_1 & 1 & 0 & 0 \\ \gamma_2 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} C_t \\ e_{1t} \\ e_{2t} \\ C_{t-1} \end{pmatrix}$$

$$(y_t = Ax_t)\,,$$

$$\begin{pmatrix} C_t \\ e_{1t} \\ e_{2t} \\ C_{t-1} \end{pmatrix} = \begin{pmatrix} \phi & & & \phi_2 \\ & \psi_1 & & \\ & & \psi_2 & \\ 1 & & & \end{pmatrix} \begin{pmatrix} C_{t-1} \\ e_{1,t-1} \\ e_{2,t-2} \\ C_{t-2} \end{pmatrix} + \begin{pmatrix} w_t \\ \epsilon_{1t} \\ \epsilon_{2t} \\ 0 \end{pmatrix}$$

$$(x_t = \Phi x_{t-1} + v_t)\,.$$

Note that

$$Q \equiv \mathrm{E}(v_t v_t') = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sigma_1^2 & 0 & 0 \\ 0 & 0 & \sigma_2^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Here, the fourth row of the transition equation describes an identity, and $Q$ is singular. When we draw $x_t$ conditioning on $x_{t+1}$ from (43), the variance of the distribution, $V_t$ given by (45), would be singular.

#### 13.3.1 Wrong Way

Based on Section 8.2, Chapter 8, Kim and Nelson (1999). The same approach is advocated by Shumway and Stoffer (2017), as can be seen in the code of Example 6.27, Page 374. However, according to my experience this does not work as good as my approach as introduced in Section 13.3.2. Actually, this approach here in this section is not correct. But anyway, we introduce it here because of its popularity.

According to Kim and Nelson (1999)'s approach, in Example A Dynamic Factor Model - AR(2), to generate $x_t$, only the first three rows of $x_{t+1}$ should be the conditioning factors.

In general, suppose that the first $J \times J$ block of the $Q$ matrix, denoted by $Q^*$, is positive-definite and that all the other elements of the $Q$ matrix are 0s. Then in generating $x_t$, only the first $J$ rows of $x_{t+1}$, denoted by $x^*_{t+1}$, can be conditioning factors in the distribution of $x_t$ in (43).

In the present case, our purpose is to generate $x_{0:n}$ from the following joint distribution:

$$\mathrm{p}_\Theta(x_{0:n}|y_{0:n}) = \mathrm{p}_\Theta(x_n|y_{1:n}) \prod_{t=1}^{n-1} \mathrm{p}_\Theta(x_t|x^*_{t+1}, y_{1:t}) \tag{48}$$

where, if we denote the first $J$ rows of $\Phi$ as $\Phi^*$,

$$\mathrm{p}_\Theta(x_t|x^*_{t+1}, y_{1:t}) \sim \mathrm{N}^\Theta_p(m^*_t, V^*_t)$$

$$m^*_t = x^t_t + P^t_t\Phi^{*\prime}\left(\Phi^* P^t_t\Phi^{*\prime} + Q^*\right)^{-1}\left(x^*_{t+1} - x^{*t}_{t+1}\right)$$

$$V^*_t = P^t_t - P^t_t\Phi^{*\prime}\left(\Phi^* P^t_t\Phi^{*\prime} + Q^*\right)^{-1}\Phi^* P^t_t$$

So the general procedure for sampling the state vector is to first draw $x_n \sim N^\Theta_p(x^n_n, P^n_n)$, only keep the first $J$ elements, and then generate $x_t$, for $t = n-1, n-2, \ldots, 0$, from a $\mathrm{N}^\Theta_p(m^*_t, V^*_t)$. *(In Kim and Nelson (1999), the state vector starts from Time 1, instead of Time 0.)*

### 13.3.2 Correct Way

Notations are as in Example 13.5. To see how to draw a sample $\boldsymbol{x}_{1:n}$, note that

$$\begin{aligned}
\mathrm{p}(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \boldsymbol{y}_{1:t}) &\propto \mathrm{p}(\boldsymbol{x}_t, \boldsymbol{x}_{t+1}|\boldsymbol{y}_{1:t}) \\
&= \mathrm{p}(C_{t-1}, C_t, e_{1t}, e_{2t}, C_{t+1}, e_{1t}, e_{2t}|\boldsymbol{y}_{1:t}) \\
&= \mathrm{p}(C_{t+1}, e_{1t}, e_{2t}|\boldsymbol{x}_t, \boldsymbol{y}_{1:t})\mathrm{p}(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t}),
\end{aligned}$$

from which you get the distribution of $\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \boldsymbol{y}_{1:t}$, and it is singular. This is intuitive: because $\boldsymbol{x}_{t+1}$ contains $C_t$, given $\boldsymbol{x}_{t+1}$, $C_t$ is known with no uncertainty. Therefore, when generating $\boldsymbol{x}_t$, we borrow $C_t$ from $\boldsymbol{x}_{t+1}$ and only worry about the sampling of $(e_{1t}, e_{2t}, C_{t-1})$.

# 14 Markov-Switching Models and Gibbs-Sampling

Based on Chapter 9, Kim and Nelson (1999).

In Markov-switching models (hidden Markov models), some parameters are dependent on an unobserved state variable ($S_t$) that is an outcome of an unobserved, discrete-time Markov process.

In the classical framework, inference on Markov-switching models consists of first estimating the model's unknown parameters, then making inferences on the unobserved Markov-switching variables, $S_{1:n}$, conditional on the parameter estimates.

This chapter provides a Bayesian approach to infer on Markov switching models. In the Bayesian analysis, both the parameters and the Markov switching variable, $S_t$, are treated as random variables. Thus, inference on $S_{1:n}$ is based on a joint distribution, not a conditional distribution. Albert and Chib (1993) have made the Bayesian analysis of Markov-switching models easy to implement.

## 14.1 A Basic Model

Consider the following simple model with Markov-switching mean and variance:

$$\begin{aligned} y_t &= \mu_{S_t} + e_t \qquad (49) \\ e_t &\sim \mathrm{N}(0, \sigma^2_{S_t}) \\ \mu_{S_t} &= \mu_0 + \mu_1 S_t \\ \sigma^2_{S_t} &= \sigma^2_0(1 - S_t) + \sigma^2_1 S_t \\ \mu_1 &> 0 \end{aligned}$$

where $S_t$ evolves according to a two-state, first-order Markov-switching process with the following transition probabilities

$$\Pr[S_t = 0|S_{t-1} = 0] = q, \Pr[S_t = 1|S_{t-1} = 1] = p$$

In the Bayesian approach, along with $S_t, t = 1, 2, \ldots, n$, the model's parameters, $\mu_0, \mu_1$, $\sigma^2_0, \sigma^2_1$, $p$ and $q$, are treated as random variables. The joint posterior density

$$\mathrm{p}(S_{1:n}, \mu_0, \mu_1, \sigma^2_0, \sigma^2_1, p, q|y_{1:n}) = \mathrm{p}(\mu_0, \mu_1, \sigma^2_0, \sigma^2_1|y_{1:n}, S_{1:n}) \cdot \mathrm{p}(p, q|S_{1:n}) \cdot \mathrm{p}(S_{1:n}|y_{1:n})$$

assumes that, conditional on $S_{1:n}$, the transition probabilities, $p$ and $q$, are independent of both the other four parameters of the model and the observations, $y_{1:n}$.

Thus, using arbitrary starting values for the parameters $\Theta^{(0)}$,

1. generate $S_{1:n} \sim \Pr_\Theta(S_{1:n}|y_{1:n})$;
2. generate $p, q \sim \mathrm{p}(p, q|S_{1:n})$;
3. generate $\mu_0, \mu_1, \sigma_0^2, \sigma_1^2 \sim \mathrm{p}(\mu_0, \mu_1, \sigma_0^2, \sigma_1^2|S_{1:n}, y_{1:n})$;
4. repeat the above three steps until convergence occurs.

### 14.1.1 Step 1: Generating the Hidden Markov States

From (42), conditional on the parameters $\Theta$, the unobserved Markov state

$$\Pr(S_{1:n}|y_{1:n}) = \Pr(S_n|y_{1:n}) \prod_{t=1}^{n-1} \mathrm{p}(S_t|S_{t+1}, y_{1:t})$$

where $\mathrm{p}_\Theta(S_n|y_{1:n})$ is known from Property 9.1, and

$$\Pr(S_t|S_{t+1}, y_{1:t}) = \frac{\Pr(S_{t+1}|S_t)\Pr(S_t|y_{1:t})}{\sum_{S_t}\Pr(S_{t+1}|S_t)\Pr(S_t|y_{1:t})} \tag{50}$$

with $\Pr(S_t|y_{1:t})$ given by Property (9.1).

Therefore, we draw first $S_n$ from $\Pr(S_n|y_{1:n})$, and then draw $S_t$ from $\Pr_\Theta(S_t|S_{t+1}, y_{1:t})$ for $t = n-1, n-2, \ldots, 1$.

Note that both $S_n|y_{1:n}$ and $S_t|S_{t+1}, y_{1:t}$, $t = n-1, n-2, \ldots, 1$, are subject to binary distributions. We can generate them using a uniform distribution. For example, we generate a random number from a uniform distribution between 0 and 1. If the generated number is less than $\Pr(S_t = 1|y_{1:t})$, we set $S_t = 1$. Otherwise, $S_t$ is set equal to 0.

### 14.1.2 Step 2: Generating the Transition Probabilities

From (41),

$$\mathrm{p}(p, q|S_{1:n}) \propto \mathrm{p}(p, q)\mathrm{p}(S_{1:n}|p, q) \tag{51}$$

where $\mathrm{p}(p, q)$ is the prior, and we are going to use beta distributions as conjugate priors for $p$ and $q$.

**Definition 14.1** (Beta Distribution)**.** A beta distributed variable, $z \sim \beta(\alpha_0, \alpha_1)$, has the density function

$$p(z) \propto z^{\alpha_0 - 1}(1-z)^{\alpha_1 - 1}, \quad \text{for } 0 < z < 1$$
$$p(z) = 0, \quad \text{for } z \geq 1 \text{or } z \leq 0$$

The mean and variance is given by

$$E(z) = \frac{\alpha_0}{\alpha_0 + \alpha_1}$$
$$\mathrm{Var}(z) = \frac{\alpha_0 \alpha_1}{(\alpha_0 + \alpha_1)^2(\alpha_0 + \alpha_1 + 1)}.$$

Assume independent beta distributions for the priors of $p$ and $q$:

$$p \sim \beta(u_{11}, u_{10})$$
$$q \sim \beta(u_{00}, u_{01}),$$

where $u_{ij}, i, j = 0, 1$, are known hyperparameters. Then

$$p(p, q) \propto p^{u_{11}-1}(1-p)^{u_{10}-1} q^{u_{00}-1}(1-q)^{u_{01}-1}$$

Then the second component of (51),

$$p(S_{1:n}|p, q) = p^{n_{11}}(1-p)^{n_{10}} q^{n_{00}}(1-q)^{n_{01}}.$$

where $n_{ij}$ refers to the counted transitions from state $i$ to $j$, and $\sum_i \sum_j n_{ij} = n$.

Therefore,

$$p(p, q|S_{1:n}) \propto p^{u_{11}+n_{11}-1}(1-p)^{u_{10}+n_{10}-1} q^{u_{00}+n_{00}-1}(1-q)^{u_{01}+n_{01}-1}$$

which suggests that the pseudo posterior distribution is given by the two independent beta distributions:

$$p|S_{1:n} \sim \beta(u_{11} + n_{11}, u_{10} + n_{10}) \tag{52}$$
$$q|s_{1:n} \sim \beta(u_{00} + n_{00}, u_{01} + n_{01}) \tag{53}$$

### 14.1.3 Step 3 Part 1

Write $\mu = (\mu_0, \mu_1)'$. In this part, we generate $\mu$ conditional on $\sigma_0^2, \sigma_1^2, S_{1:n}, y_{1:n}$.

Note that the model (49) can be written as

$$y_t = (1, S_t)\mu + e_t$$
$$e_t \sim \mathrm{N}(0, \sigma_{S_t}^2)$$

Write it in matrix notation

$$Y = X\mu + V, \quad V \sim \mathrm{N}(0, \Sigma)$$

where

$$X = \begin{pmatrix} 1 & S_1 \\ \vdots & \vdots \\ 1 & S_n \end{pmatrix}$$
$$\Sigma = \begin{pmatrix} \sigma_{S_1}^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{S_n}^2 \end{pmatrix}$$

We assume a normal (conditional) prior for $\mu$:

$$\mu | \sigma_0^2, \sigma_1^2 \sim \mathrm{N}(b_0, B_0)$$

Then, the pseudo (conditional) posterior distribution is given by

$$\begin{aligned} \mathrm{p}(\mu|\sigma_0^2, \sigma_1^2, S_{1:n}, y_{1:n}) &\propto \mathrm{p}(\mu|\sigma_0^2, \sigma_1^2)\mathrm{p}(S_{1:n}, y_{1:n}|\mu, \sigma_0^2, \sigma_1^2) \\ &= \mathrm{p}(\mu|\sigma_0^2, \sigma_1^2)\mathrm{p}(y_{1:n}|S_{1:n}, \mu, \sigma_0^2, \sigma_1^2)\mathrm{p}(S_{1:n}) \\ &\propto \mathrm{p}(\mu|\sigma_0^2, \sigma_1^2)\mathrm{p}(y_{1:n}|S_{1:n}, \mu, \sigma_0^2, \sigma_1^2) \end{aligned}$$

where

$$\mathrm{p}(\mu|\sigma_0^2, \sigma_1^2) \propto \exp\left\{-\frac{1}{2}(\mu - b_0)' B_0^{-1}(\mu - b_0)\right\}$$

and

$$\mathrm{p}(y_{1:n}|S_{1:n}, \mu, \sigma_0^2, \sigma_1^2) \propto \exp\left\{-\frac{1}{2}(Y - X\mu)'\Sigma^{-1}(Y - X\mu)\right\}.$$

Therefore, the pseudo (conditional) posterior distribution of $\mu$ is also of normal form:

$$\mu|\sigma_0^2, \sigma_1^2, S_{1:n}, y_{1:n} \sim \mathrm{N}(b_1, B_1)$$

where

$$b_1 = (B_0^{-1} + X'\Sigma^{-1}X)^{-1}(B_0^{-1}b_0 + X'\Sigma^{-1}Y) \tag{54}$$

$$B_1 = (B_0^{-1} + X'\Sigma^{-1}X)^{-1} \tag{55}$$

Observe that the $b_1$ above has an alternative representation

$$b_1 = (B_0^{-1} + X'\Sigma^{-1}X)^{-1}[B_0^{-1}b_0 + (X'\Sigma^{-1}X)(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y]$$

where $(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$ is the sample GLS estimate of $\mu$, with $(X'\Sigma^{-1}X)^{-1}$ being the variance of the GLS estimate. Thus the posterior mean, $b_1$, if an weighted average of $b_0$ and the GLS estimate, the weights being the relative precision of the two measures, $(B_0^{-1} + X'\Sigma^{-1}X)^{-1}B_0^{-1}$ and $(B_0^{-1} + X'\Sigma^{-1}X)^{-1}(X'\Sigma X)^{-1}$.

To constrain $\mu_1 > 0$, if the generated value of $\mu_1$ is less or equal to 0, we discard the draws.

#### 14.1.4 Step 3 Part 2

In this part, we generate $\sigma_0^2$ and $\sigma_1^2$ conditional on $\mu_0, \mu_1, S_{1:n}, y_{1:n}$. We are going to use inverse gamma distributions as the conjugate priors for $\sigma_0^2$ and $\sigma_1^2$.

**Definition 14.2** (Gamma Distribution and Inverse Gamma Distribution)**.** For $t = 1, 2, \ldots, v$, let $z_t$ be i.i.d. normal with mean 0 and variance $1/\delta$. Let $W = \sum_{t=1}^{v} z_t^2$. Then, we have

$$W \sim \Gamma\left(\frac{v}{2}, \frac{\delta}{2}\right),$$

where the density function is given by

$$\mathrm{p}(w) \propto w^{\frac{v}{2}-1} \exp\left\{-\frac{w\delta}{2}\right\},$$

with

$$\mathrm{E}(W) = \frac{v}{\delta}$$
$$\mathrm{Var}(W) = 2\frac{v}{\delta^2}.$$

And we say $1/W$ is subject to an inverse gamma distribution

$$\frac{1}{W} \sim \mathrm{IG}\left(\frac{v}{2}, \frac{\delta}{2}\right)$$

We can first generate $\sigma_0^2$, conditional on $\sigma_1^2$, and then we can generate $\sigma_1^2$ conditional on $\sigma_0^2$. To generate $\sigma_0^2$ conditional on $\sigma_1^2$, we assume an inverse gamma (conditional) prior for $\sigma_0^2$, i.e.

$$\frac{1}{\sigma_0^2}|\sigma_1^2, \mu_0, \mu_1 \sim \Gamma\left(\frac{v_0}{2}, \frac{\delta_0}{2}\right)$$

with

$$\mathrm{p}\left(\frac{1}{\sigma_0^2}|\sigma_1^2, \mu_0, \mu_1\right) \propto \left(\frac{1}{\sigma_0^2}\right)^{\frac{v_0}{2}-1} \exp\left\{-\frac{\delta_0}{2\sigma_0^2}\right\}.$$

We know that

$$\mathrm{p}\left(\frac{1}{\sigma_0^2}|\sigma_1^2, \mu_0, \mu_1, S_{1:n}, y_{1:n}\right) \propto \mathrm{p}\left(\frac{1}{\sigma_0^2}|\sigma_1^2, \mu_0, \mu_1\right) \mathrm{p}\left(y_{1:n}|S_{1:n}, \sigma_0^2, \sigma_1^2, \mu_0, \mu_1\right)$$

where

$$\begin{aligned}
\mathrm{p}\left(y_{1:n}|S_{1:n}, \sigma_0^2, \sigma_1^2, \mu_0, \mu_1\right) &= (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(Y - X\mu)'\Sigma^{-1}(Y - X\mu)\right\} \\
&= (2\pi)^{-\frac{n}{2}} (\sigma_0^2)^{-\frac{n_0}{2}} (\sigma_1^2)^{-\frac{n_1}{2}} \exp\left\{-\frac{1}{2}(Y - X\mu)'\Sigma^{-1}(Y - X\mu)\right\} \\
&\propto \left(\frac{1}{\sigma_0^2}\right)^{\frac{n_0}{2}} \exp\left\{-\frac{1}{2}(Y - X\mu)'\Sigma^{-1}(Y - X\mu)\right\} \\
&\propto \left(\frac{1}{\sigma_0^2}\right)^{\frac{n_0}{2}} \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{t \in N_0}(y_t - \mu_0)^2\right\}
\end{aligned}$$

with $N_0 = \{t : S_t = 0\}$. So the likelihood function of $1/\delta_0^2$ depends only on the values of $y_t$ for which $S_t = 0$.

So the conditional pseudo-posterior

$$\mathrm{p}(\frac{1}{\sigma_0^2}|\sigma_1^2, \mu_0, \mu_1, S_{1:n}, y_{1:n}) \sim \Gamma\left(\frac{v_1}{2}, \frac{\delta_1}{2}\right)$$

with

$$v_1 = v_0 + n_0$$

$$\delta_1 = \delta_0 + \sum_{t \in N_0}(y_t - \mu_0)^2$$

where $n_0$ is the cardinality of $N_0$.

Similarly, the likelihood of $1/\sigma_1^2$ depends only on the $y_t$ for which $S_t = 1$. If we use an inverse gamma prior for $\sigma_1^2$,

$$\frac{1}{\sigma_1^2}|\sigma_0^2, \mu_0, \mu_1 \sim \Gamma\left(\frac{v_2}{2}, \frac{\delta_2}{2}\right)$$

the conditional pseudo posterior

$$\frac{1}{\sigma_1^2}|\sigma_0^2, \mu_0, \mu_1, S_{1:n}, y_{1:n} \sim \Gamma\left(\frac{v_3}{2}, \frac{\delta_3}{2}\right)$$

with

$$v_3 = v_2 + n_1$$
$$\delta_3 = \delta_2 + \sum_{t \in N_1} (y_t - \mu_1)^2.$$

where $N_1 = \{t : S_t = 1\}$ and $n_1$ is the cardinality of $N_1$.

## 14.2 A Three-State Markov Switching Mean-Variance Model of the Real Interest Rate

Assessing the existence of a unit root in the real interest rate is potentially interesting in many aspects. Literature on the tests have reported mixed results. Perron (1990) rejects the unit root hypothesis by incorporating regime shifts in the ex post real interest rate. He argues that unit roots test are biased toward nonrejection of the unit root hypothesis when the series contains a sudden change in the mean.

Garcia and Perron (1996) employ Hamilton's (1989) Markov-Switching model to emplicitly account for regime shifts in an auto regressive model of the ex post real interest rate. If $y_t$ denotes the ex post real interest rate calculated by subtracting the CPI inflation rate from the three-month Treasury bill rate,

$$y_t - \mu_{S,t} = \phi_1(y_{t-1} - \mu_{S,t-1}) + \phi_2(y_{t-2} - \mu_{S,t-2}) + e_t \tag{56}$$
$$e_t \sim \mathrm{N}(0, \sigma_{S,t}^2)$$
$$\mu_{S,t} = \mu_1 S_{1t} + \mu_2 S_{2t} + \mu_3 S_{3t}$$
$$\sigma_{S,t}^2 = \sigma_1^2 S_{1t} + \sigma_2^2 S_{2t} + \sigma_3^2 S_{3t}$$

with $S_{jt} = 1$ if $S_t = j$, and $S_{jt} = 0$ otherwise, $j = 1, 2, 3$. Denote

$$\pi_{ij} = \Pr[S_t = j | S_{t-1} = i],$$
$$\sum_{j=1}^{3} \pi_{ij} = 1.$$

Of particular interest is the sum of the AR coefficients $(\phi_1 + \phi_2)$. Perron's (1990) conjecture is that, if the true data generating process is given by the above model, yet one forces the mean of the series, $\mu_{S,t}$, to be constant in estimation, then the sum of the AR coefficients will be biased toward unity even though the true sum of close to 0. The MLE estimation results can be found on Page 84 of Kim and Nelson (1999).

This section provides a Bayesian alternative to the analysis of the above model. To guarantee the identification within the Gibbs-sampling framework, the constraint $\mu_1 < \mu_2 < \mu_3$ is needed.

Defining $\boldsymbol{\sigma^2} = [\sigma_1^2, \sigma_2^2, \sigma_3^2]'$, $\boldsymbol{\pi} = [\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}, \pi_{31}, \pi_{32}]'$, $\boldsymbol{\mu} = [\mu_1, \mu_2, \mu_3]'$, and $\boldsymbol{\phi} = [\phi_1, \phi_2]'$, the Gibbs-sampling procedure is given by successively iteration of the following five steps:

1. Generate $S_{1:n}$, conditional on $\boldsymbol{\sigma^2}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\phi}$ and $y_{1:n}$.

2. Generate $\boldsymbol{\pi}$, conditional on $S_{1:n}$.

3. Generate $\boldsymbol{\sigma^2}$, conditional on $\boldsymbol{\mu}, \boldsymbol{\phi}, S_{1:n}$ and $y_{1:n}$.

4. Generate $\boldsymbol{\mu}$, conditional on $\boldsymbol{\sigma^2}, \boldsymbol{\phi}, S_{1:n}$ and $y_{1:n}$.

5. Generate $\boldsymbol{\phi}$, conditional on $\boldsymbol{\sigma^2}, \boldsymbol{\mu}, S_{1:n}, y_{1:n}$.

### 14.2.1 Generating the Hidden Markov States

From (42), conditional on the parameters $\Theta$,

$$\Pr(S_{1:n} | y_{1:n}) = \Pr(S_n | y_{1:n}) \prod_{t=1}^{n-1} \Pr(S_t | S_{t+1}, y_{1:t}) \tag{57}$$

where

$$\Pr(S_t | S_{t+1}, y_{1:t}) = \frac{\Pr(S_{t+1} | S_t) \Pr(S_t | y_{1:t})}{\sum_{S_t} \Pr(S_{t+1} | S_t) \Pr(S_t | y_{1:t})}$$

with $\Pr(S_t|y_{1:t})$, $t = 1, 2, \ldots, n$ given by Property 9.5.

Because $S_t$ is a three-state Markov-switching variable, we need to pay special attention to its generation based on the uniform distribution. First, we calculate

$$\Pr(S_t = 1|S_{t+1}, y_{1:t}) = \frac{\Pr(S_{t+1}|S_t = 1)\Pr(S_t = 1|y_{1:n})}{\sum_{j=1}^{3}\Pr(S_{t+1}|S_t = j)\Pr(S_t = j|y_{1:n})}$$

Then, we generate a random number from the uniform distribution. If the generated number is less than or equal to $\Pr(S_t = 1|S_{t+1}, y_{1:t})$, we set $S_t = 1$; if it is greater than $\Pr(S_t = 1|S_{t+1}, y_{1:t})$, we calculate

$$\Pr(S_t = 2|S_{t+1}, y_{1:t}, S_t \neq 1) = \frac{\Pr(S_{t+1}|S_t = 2)\Pr(S_t = 2|y_{1:t})}{\sum_{j=2}^{3}\Pr(S_{t+1}|S_t = j)\Pr(S_t = j|y_{1:t})}$$

and we generate another random number from the uniform distribution. If the generated number is less than or equal to $\Pr(S_t = 2|S_{t+1}, y_{1:t}, S_t \neq 1)$, we set $S_t = 2$; otherwise, we set $S_t = 3$.

### 14.2.2 Generating the Transition Probabilities

Conditional on $S_{1:n}$, the transition probabilities are assumed to be independent of $y_{1:n}$ and the other parameters of the model. Denote $\boldsymbol{\pi} = [\boldsymbol{\pi}_{ii}, \boldsymbol{\pi}_{ij}]$ where $\boldsymbol{\pi}_{ii} = [\pi_{11}, \pi_{22}, \pi_{33}]$ and $\boldsymbol{\pi}_{ij} = [\pi_{12}, \pi_{21}, \pi_{31}]$. Then we can write

$$\mathrm{p}(\boldsymbol{\pi}|S_{1:n}) = \mathrm{p}(\boldsymbol{\pi}_{ii}|S_{1:n})\mathrm{p}(\boldsymbol{\pi}_{ij}|\boldsymbol{\pi}_{ii}, S_{1:n})$$

Therefore, we generate $\boldsymbol{\pi}_{ii}$ conditional on $S_{1:n}$ first, and then generate $\boldsymbol{\pi}_{ij}$ conditional on $\boldsymbol{\pi}_{ii}$ and $S_{1:n}$.

Similar to the previous example, by taking the beta family of distributions as conjugate priors, it can be shown that, for $i = 1, 2, 3$,

$$\pi_{ii}|S_{1:n} \sim \beta(u_{ii} + n_{ii}, \bar{u}_{ii} + \bar{n}_{ii})$$

where $u_{ii}$ and $\bar{u}_{ii}$ are known hyperparameters of the priors, $n_{ii}$ the number of transitions from regime/state $i$ to $i$, and $\bar{n}_{ii}$ the number of transitions from regime $i$ to a different regime.

For $j = 1, 2, 3$ and $j \neq i$, denote $\bar{\pi}_{ij} = \Pr(S_t = j|S_{t-1} = i, S_t \neq i)$. Given that $\boldsymbol{\pi}_{ii}$ is generated, $\boldsymbol{\pi}_{ij}$ can be calculated from

$$\pi_{ij} = \bar{\pi}_{ij}(1 - \pi_{ii})$$

where $\bar{\pi}_{ij}$ can be generated by taking the following beta priors:

$$\bar{\pi}_{12} \sim \beta(u_{12}, u_{13})$$
$$\bar{\pi}_{21} \sim \beta(u_{21}, u_{23})$$
$$\bar{\pi}_{31} \sim \beta(u_{31}, u_{32})$$

which is followed by

$$\mathrm{p}(\bar{\pi}_{12}, \bar{\pi}_{21}, \bar{\pi}_{31}) \propto \bar{\pi}_{12}^{u_{12}-1}(1-\bar{\pi}_{12})^{u_{13}-1}\bar{\pi}_{21}^{u_{21}-1}(1-\bar{\pi}_{21})^{u_{23}-1}\bar{\pi}_{31}^{u_{31}-1}(1-\bar{\pi}_{31})^{u_{32}-1}$$

So the pseudo-posterior

$$\begin{aligned}
&\mathrm{p}(\bar{\pi}_{12}, \bar{\pi}_{21}, \bar{\pi}_{31}|\boldsymbol{\pi}_{ii}, S_{1:n}) \\
&\propto \Pr(S_{1:n}|\bar{\pi}_{12}, \bar{\pi}_{21}, \bar{\pi}_{31}, \boldsymbol{\pi}_{ii})\mathrm{p}(\bar{\pi}_{12}, \bar{\pi}_{21}, \bar{\pi}_{31}|\boldsymbol{\pi}) \\
&= \Pr(S_{1:n}|\boldsymbol{\pi})\mathrm{p}(\bar{\pi}_{12}, \bar{\pi}_{21}, \bar{\pi}_{31}) \\
&\propto \bar{\pi}_{12}^{u_{12}+n_{12}-1}(1-\bar{\pi}_{12})^{u_{13}+n_{13}-1}\bar{\pi}_{21}^{u_{21}+n_{21}-1}(1-\bar{\pi}_{21})^{u_{23}+n_{23}-1}\bar{\pi}_{31}^{u_{31}+n_{31}-1}(1-\bar{\pi}_{31})^{u_{32}+n_{32}-1}
\end{aligned}$$

which suggests that $\bar{\pi}_{12}, \bar{\pi}_{21}, \bar{\pi}_{31}$ can be generated by the following beta distribution

$$\bar{\pi}_{12}|S_{1:n} \sim \beta(u_{12}+n_{12}, u_{13}+n_{13})$$
$$\bar{\pi}_{21}|S_{1:n} \sim \beta(u_{21}+n_{21}, u_{23}+n_{23})$$
$$\bar{\pi}_{31}|S_{1:n} \sim \beta(u_{31}+n_{31}, u_{32}+n_{32})$$

### 14.2.3 Generating $\mu_1, \mu_2, \mu_3$

Rearranging Equation (56), we have

$$y_t^* = \mu_1 S_{1t}^* + \mu_2 S_{2t}^* + \mu_3 S_{3t}^* + e_t$$

where

$$y_t^* = y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2}$$
$$S_{it}^* = S_{it} - \phi_1 S_{i,t-1} - \phi_2 S_{i,t-2}, \quad i = 1, 2, 3$$

In matrix notation,

$$Y = X\boldsymbol{\mu} + V, \quad V \sim \mathrm{N}(0, \Sigma) \tag{58}$$

where $\Sigma$ is an $(n-2) \times (n-2)$ diagonal matrix.

We assume a normal (conditional) prior for $\boldsymbol{\mu}$

$$\boldsymbol{\mu}|\boldsymbol{\sigma^2}, \boldsymbol{\phi} \sim \mathrm{N}(b_0, B_0)_{I(\mu_1<\mu_2<\mu_3)}$$

where $I(\cdot)$ refers to an indicator function.

Then, in a similar way to the previous example and with the fact that

$$\mathrm{p}(y_{1:n}|\boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{\phi}, S_{1:n}) = \mathrm{p}(y^*_{1:n}|\boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{\phi}, S_{1:n}),$$

we can get

$$\boldsymbol{\mu}|\boldsymbol{\sigma^2}, \boldsymbol{\phi}, S_{1:n}, y_{1:n} \sim \mathrm{N}(b_1, B_1)_{I(\mu_1<\mu_2<\mu_3)}$$

where $b_1$ and $B_1$ are given by Equations (54) and (55).

To meet the constraint $\mu_1 < \mu_2 < \mu_3$, we employ rejection sampling.

#### 14.2.4 Generating $\sigma_1^2, \sigma_2^2, \sigma_3^2$

We first generate $\sigma_1^2$ conditional on $\sigma_2^2$ and $\sigma_3^2$, then generate $\sigma_2^2$ conditional on $\sigma_1^2$ and $\sigma_3^2$, and at last generate $\sigma_3^2$ conditional on $\sigma_1^2$ and $\sigma_2^2$.

We assume an inverse Gamma prior for $\sigma_1^2$

$$\frac{1}{\sigma_1^2}|\sigma_2^2, \sigma_3^2, \boldsymbol{\mu}, \boldsymbol{\phi} \sim \Gamma\left(\frac{v_1}{2}, \frac{\delta_1}{2}\right)$$

i.e.

$$\mathrm{p}\left(\frac{1}{\sigma_1^2}|\sigma_2^2, \sigma_3^2, \boldsymbol{\mu}, \boldsymbol{\phi}\right) \propto \left(\frac{1}{\sigma_1^2}\right)^{\frac{v_1}{2}-1} \exp\left\{-\frac{\delta_1}{2\sigma_1^2}\right\}$$

We know

$$\begin{aligned}\mathrm{p}\left(\frac{1}{\sigma_1^2}|\sigma_2^2, \sigma_3^2, \boldsymbol{\mu}, \boldsymbol{\phi}, y_{1:n}, S_{1:n}\right) &\propto \mathrm{p}\left(\frac{1}{\sigma_1^2}|\sigma_2^2, \sigma_3^2, \boldsymbol{\mu}, \boldsymbol{\phi}\right) \mathrm{p}(y_{1:n}|\boldsymbol{\sigma^2}, \boldsymbol{\mu}, \boldsymbol{\phi}, S_{1:n}) \Pr(S_{1:n}) \\ &\propto \mathrm{p}\left(\frac{1}{\sigma_1^2}|\sigma_2^2, \sigma_3^2, \boldsymbol{\mu}, \boldsymbol{\phi}\right) \mathrm{p}(y_{1:n}|\boldsymbol{\sigma^2}, \boldsymbol{\mu}, \boldsymbol{\phi}, S_{1:n})\end{aligned}$$

with

$$\begin{aligned}
\mathrm{p}(y_{1:n}|\boldsymbol{\sigma^2},\boldsymbol{\mu},\boldsymbol{\phi},S_{1:n}) &\propto |\Sigma|^{\frac{1}{2}}\exp\left\{-\frac{1}{2}(Y-X\boldsymbol{\mu})'\Sigma^{-1}(Y-X\boldsymbol{\mu})\right\}\\
&\propto \left(\frac{1}{\sigma_1^2}\right)^{\frac{n_1}{2}}\exp\left\{-\frac{1}{2\sigma_1^2}\sum_{t\in N_1}(y_t^*-\mu_1S_{1t}^*-\mu_2S_{2t}^*-\mu_3S_{3t}^*)^2\right\}
\end{aligned}$$

where $Y, X$ and $\Sigma$ are as in Equation (58), $N_1 = \{t : S_t = 1\}$, and $n_1$ is the cardinality of $N_1$.

So

$$\begin{aligned}
&\frac{1}{\sigma_1^2}|\sigma_2^2,\sigma_3^2,\boldsymbol{\mu},\boldsymbol{\phi},y_{1:n},S_{1:n}\\
&\sim \Gamma\left(\frac{v_1+n_1}{2},\frac{\delta_1+\sum_{t\in N_1}\left(y_t^*-\mu_1S_{1t}^*-\mu_2S_{2t}^*-\mu_3S_{3t}^*\right)^2}{2}\right)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\frac{1}{\sigma_2^2}|\sigma_1^2,\sigma_3^2,\boldsymbol{\mu},\boldsymbol{\phi},y_{1:n},S_{1:n}\\
&\sim \Gamma\left(\frac{v_2+n_2}{2},\frac{\delta_1+\sum_{t\in N_2}\left(y_t^*-\mu_1S_{1t}^*-\mu_2S_{2t}^*-\mu_3S_{3t}^*\right)^2}{2}\right)\\
&\frac{1}{\sigma_3^2}|\sigma_1^2,\sigma_2^2,\boldsymbol{\mu},\boldsymbol{\phi},y_{1:n},S_{1:n}\\
&\sim \Gamma\left(\frac{v_3+n_3}{2},\frac{\delta_1+\sum_{t\in N_3}\left(y_t^*-\mu_1S_{1t}^*-\mu_2S_{2t}^*-\mu_3S_{3t}^*\right)^2}{2}\right)
\end{aligned}$$

### 14.2.5 Generating $\phi_1, \phi_2$

If we let $y_t^{**} = y_t - \mu_{S,t}$, Equation (56) is given by

$$y_t^{**} = \phi_1 y_{t-1}^{**} + \phi_2 y_{t-2}^{**} + e_t$$

In matrix notation, we have

$$Y^* = X^*\boldsymbol{\phi} + V, \quad V \sim \mathrm{N}(0,\Sigma)$$

Choose a normal prior

$$\boldsymbol{\phi}|\boldsymbol{\mu},\boldsymbol{\sigma^2} \sim \mathrm{N}(c_0, C_0)_{I(\phi(L)\in\mathbb{S})}$$

where $c_0$ and $C_0$ are known hyperparameters of the prior distribution, and $I(\cdot)$ refers to an indicator function used to denote that the roots of $(1-\phi L-\phi L^2)=0$ lie outside the unit circle, or, the process is stable $\phi(L)\in\mathbb{S}$. Then, given the fact that

$$\mathrm{p}(y_{1:n}|\boldsymbol{\mu},\boldsymbol{\sigma^2},\boldsymbol{\phi},S_{1:n})=\mathrm{p}(y^{**}_{1:n}|\boldsymbol{\mu},\boldsymbol{\sigma^2},\boldsymbol{\phi},S_{1:n}),$$

the posterior, similar to (54) and (55),

$$\boldsymbol{\phi}|\boldsymbol{\mu},\boldsymbol{\sigma^2},y_{1:n},S_{1:n}\sim\mathrm{N}(c_1,C_1)_{I(\phi(L)\in\mathbb{S})}$$

where

$$c_1=(C_0^{-1}+X^{*\prime}\Sigma^{-1}X^{*})^{-1}(C_0^{-1}c_0+X^{*\prime}\Sigma^{-1}Y^{*})$$
$$C_1=(C_0^{-1}+X^{*\prime}\Sigma^{-1}X^{*\prime})^{-1}$$

We, again, adopt rejection sampling to achieve the constraint that $\phi(L)\in\mathbb{S}$.

## 14.3 Hamilton's Markov Switching Model of Business Fluctuations

Hamilton (1989) allows the mean of the growth in real GNP to be evolving according to a two-state Markov-switching process. Growth in real GNP is modeled as an AR(4) process:

$$(\Delta y_t-\mu_{S,t})=\sum_{i=1}^{4}\phi_i(\Delta y_{t-i}-\mu_{S,t-i})+e_t \tag{59}$$

$$e_t\sim\text{i.i.d. }N(0,\sigma^2)$$

$$\mu_{S,t}=\mu_0(1-S_t)+\mu_1S_t \tag{60}$$

$$\Pr(S_t=1|S_{t-1}=1)=p,\quad \Pr(S_t=0|S_{t-1}=0)=q$$

### 14.3.1 Dummy: 1983:I - 1995:III

When applied to the sample period employed by Hamilton (1989): 1952:II - 1984:IV, the model can be estimated with MLE well. When the sample is extended to include 11 more years of more recent data (1985:I - 1995:III), the model fails to provide reasonable parameter estimates (via MLE).

One potential reason is that the model lacks a mechanism to account for a productivity growth slowdown in the more recent sample. Another potential reason is that the gap between average growth rates of output during boom and recession could be smaller than in earlier sample, because of stabilizing monetary policy. To account for these possibilities, Equation (60) is replaced by

$$\mu_{S,t} = (\mu_0 + \mu_0^* D_t)(1 - S_t) + (\mu_1 + \mu_1^* D_t) S_t$$

where $D_t$ is a dummy variable set equal to 1 for the subsample 1983:I - 1995:III.

We estimate and infer via the Bayesian approach. The Gibbs-sampling procedure for this example is the same as the previous one except for the $\boldsymbol{\mu}$ part.

Define $\boldsymbol{\mu} = (\mu_0, \mu_1, \mu_0^*, \mu_1^*)$. Then $\boldsymbol{\mu}$ can be generated in a similar way to the preceding example.

Take a normal prior

$$\boldsymbol{\mu}|\sigma^2, \boldsymbol{\phi} \sim \mathrm{N}(b_0, B_0)_{I(\mu_0<\mu_1)}$$

Then

$$\mathrm{p}(\boldsymbol{\mu}|\sigma^2, \boldsymbol{\phi}, \Delta y_{1:n}, D_{1:n}, S_{1:n}) \propto \mathrm{p}(\boldsymbol{\mu}|\sigma^2, \boldsymbol{\phi})\mathrm{p}(\Delta y_{1:n}|\boldsymbol{\mu}, \sigma^2, \boldsymbol{\phi}, D_{1:n}, S_{1:n})$$

Define

$$\Delta y_t^* = \Delta y_t - \sum_{i=1}^{4} \phi_i \Delta y_{1-i}$$

$$S_{0t}^* = (1 - S_t) - \sum_{i=1}^{4} \phi_i (1 - S_{t-i})$$

$$S_{1t}^* = S_t - \sum_{i=1}^{4} \phi_i S_{t-i}$$

Then

$$\mathrm{p}(\Delta y_{1:n}|\boldsymbol{\mu}, \sigma^2, \boldsymbol{\phi}, D_{1:n}, S_{1:n}) = \mathrm{p}(\Delta y_{1:n}^*|\boldsymbol{\mu}, \sigma^2, \boldsymbol{\phi}, D_{1:n}, S_{1:n})$$

and Equation (59) could be written as

$$\begin{aligned}\Delta y_t^* &= (\mu_0 + \mu_0^* D_t) S_{0t}^* + (\mu_1 + \mu_1^* D_t) S_{1t}^* + e_t \\ &= \mu_0 S_{0t}^* + \mu_1 S_{1t}^* + \mu_0^* D_t S_{0t}^* + \mu_1^* D_t S_{1t}^* + e_t\end{aligned}$$

In matrix form,

$$Y = X\boldsymbol{\mu} + V, \quad V \sim \mathrm{N}(0, \sigma^2 I)$$

Therefore, the posterior distribution

$$\boldsymbol{\mu}|\sigma^2, \boldsymbol{\phi}, \Delta y_{1:n}, D_{1:n}, S_{1:n} \sim \mathrm{N}(b_1, B_1)_{I(\mu_0<\mu_1)}$$

where

$$b_1 = (B_0^{-1} + \sigma^{-2}X'X)^{-1}(B_0^{-1}b_0 + \sigma^{-2}X'Y)$$
$$B_1 = (B_0^{-1} + \sigma^{-2}X'X)^{-1}$$

## 14.4 A Three-State Markov-Switching Variance Model of Stock Returns

Stock returns tend to exhibit nonnormal distributions in the form of skewness and excess kurtosis, a fact known at least since Fama (1963) and Mandlebrot (1963). The pronounced peak and heavy tails in the distribution of stock returns, as mentioned in Turner, Startz and Nelson (1989), are typical of densities of normal observations subject to heteroskedasticity.

One specification used in the study of stock heteroskedasticity is the GARCH model.

Later, Hamilton and Susmel (1994) proposed a SWARTCH (Switching ARCH) model in which parameters of an ARCH process come from one of several different regimes. One simplified version of the model is

$$y_t = \sigma_{S,t} u_t$$
$$u_t = h_t v_t, \quad v_t \sim \text{i.i.d. Student's } t$$
$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \beta d_{t-1} u_{t-1}^2$$

where $d_{t-1}$ is a dummy variable introduced to capture leverage effects [5]. The estimates using weekly stock returns show that $\hat{\alpha}_1 + \hat{\alpha}_2 = 0.48$. Note that $(\hat{\alpha}_1 + \hat{\alpha}_2)^4 \approx 0.05$, which

[5] The correlation between an asset return and its changes of volatility.

means that the ARCH effects die out almost completely after a month, suggesting that no ARCH term may be necessary in modeling monthly stock returns.

Kim, Nelson, and Startz (1998) models monthly stock returns as a pure Markov switching variance model

$$
\begin{aligned}
y_t &\sim \mathrm{N}(0, \sigma^2_{S,t}) \\
\sigma^2_{S,t} &= \sigma^2_1 S_{1t} + \sigma^2_2 S_{2t} + \sigma^2_3 S_{3t} \\
\sigma^2_1 &< \sigma^2_2 < \sigma^2_3
\end{aligned}
$$

After estimating the parameters, they calculate the variance of the stock returns in the following way

$$
\mathrm{E}(\sigma^2_{S,t}|y_{1:n}) = \hat{\sigma}^2_1 \Pr(S_t = 1|y_{1:n}) + \hat{\sigma}^2_2 \Pr(S_t = 2|y_{1:n}) + \hat{\sigma}^2_3 \Pr(S_t = 3|y_{1:n}).
$$

Then they standardize the stock returns with the variance calculated above, applied ARCH tests and find no ARCH effects. In addition, the standardized returns show little excess kurtosis. After a Jarque-Bera joint test of normality, they conclude that they cannot reject the hypothesis that the standardized returns are normally distributed at a 5% significance level

# 15 State-Space Models with Markov Switching and Gibbs Sampling

Based on Chapter 10, Kim and Nelson (1999).

Consider the model given by Equations (34) and (35). Inferences via Bayesian Gibbs-sampling is a straightforward application of the methods presented in the above two sections.

1. (an application of Section 13) Conditional on parameters of the model, $S_{1:n}$, and the observed data $y_{1:n}$, generate $x_{1:n}$ from
$$\mathrm{p}(x_{1:n}|y_{1:n}, S_{1:n}) = \mathrm{p}(x_n|y_{1:n}, S_{1:n}) \prod_{t=1}^{n-1} \mathrm{p}(x_t|x_{t+1}, y_{1:t}, S_{1:t}),$$
where $\mathrm{p}(x_t|x_{t+1}, y_{1:t}, S_{1:t})$ is given by (43).

2. (an application of Section 14) Conditional on parameters of the model, $x_{1:n}$ and the observed data $y_{1:n}$, generate $S_{1:n}$ from
$$\Pr(S_{1:n}|y_{1:n}, x_{1:n}) = \Pr(S_n|y_{1:n}, x_{1:n}) \prod_{t=1}^{n-1} \Pr(S_t|S_{t+1}, y_{1:t}, x_{1:t}),$$
where $\Pr(S_t|S_{t+1}, y_{1:n}, x_{1:n})$ is given by (50).

3. Conditional on $x_{1:n}, S_{1:n}$ and $y_{1:n}$, generate parameters of the model from appropriate distributions.

The model can easily be extended to include more than one Markov-switching variables.

$$x_t = \Phi_{S_{1t}} x_{t-1} + \mu_{S_{1t}} + w_t$$
$$y_t = A_{S_{2t}} x_t + \Gamma_{S_{2t}} z_t + v_t$$
$$\begin{pmatrix} w_t \\ v_t \end{pmatrix} \sim \mathrm{N}\left( 0 \quad \begin{pmatrix} Q_{S_{1t}} & 0 \\ 0 & R_{S_{2t}} \end{pmatrix} \right)$$

where $S_{1t}$ and $S_{2t}$ are independent of each other.

**Example 15.1.** One might consider the following unobserved component model of the ex ante real interest rate:

$$y_t = r_t + e_t, \quad e_t \sim \mathrm{N}(0, \sigma^2_{e,S_{1t}})$$

where $y_t$ is the ex post real interest rate, $r_t$ is the unobserved ex ante real interest rate, and $e_t$ is the negative of the inflation forecast error. The ex ante real interest rate is modeled as follows *(I can not see why $r_t$ is ex ante real interest rate. Actually, it is just the AR component of the ex post real interest rate and nothing more. If $r_t$ is called the ex ante rate, then $e_t$ is the inflation forecast error. But, as I said, I do not think they have anything to do with expected inflation.)*

$$\phi(L)(r_t - \mu_{S_{2t}}) = w_t, \quad w_t \sim \mathrm{N}(0, \sigma^2_{w,S_{2t}})$$

where $S_{2t}$ is independent of $S_{1t}$, because there is no reason to believe that variance of the inflation forecast errors switches its regime whenever the ex ante real interest rate switches its regimes.

## 15.1 A Dynamic Factor Model with Markov Switching: Business Cycle Turning Points and a New Coincident Index

Burns and Mitchell (1946) established two defining characteristics of the business cycle: comovement among economic variables through the cycle and nonlinearity in the evolution of the business cycle, that is, regime switching at the turning points of the business cycle.

As noted by Diebold and Rudebusch (1996), these two aspects of the business cycle have generally been considered in isolation from each other. In Stock and Watson's dynamic factor model (1989, 1991, 1993), introduced in section 3.4, comovement among economic variables is captured by a composite index. And Hamilton's regime switching model introduced in section 14.3 features nonlinearity in an individual economic variable.

After an extensive survey of related literature, and based on both theoretical and empirical foundations, Diebold and Rudebusch proposes a multivariate dynamic factor model with regime switching that encompasses the two key features of the business cycle.

In this section, we deal with inferences on the model based on Kim and Nelson's (1998) approach via Bayesian Gibbs-sampling.

Let $Y_{it}, i = 1, 2, 3, 4$ be the logs of four variables: industrial production, total personal income less transfer payments, total manufacturing and trade sales, and employees on nonagricultural payrolls.

$$\begin{gathered}
\Delta Y_{it} = \gamma_i(L)\Delta C_t + D_i + e_{it}, \quad i = 1, 2, 3, 4 \\
\psi_i(L)e_{it} = \epsilon_{it}, \quad \epsilon_{it} \sim \text{i.i.d. } \mathrm{N}(0, \sigma_i^2) \\
\phi(L)(\Delta C_t - \mu_{St} - \delta) = w_t, \quad w_t \sim \text{i.i.d. } \mathrm{N}(0, 1) \\
\mu_{St} = \mu_0 + \mu_1 S_t, \quad \mu_1 > 0, \quad S_t = \{0, 1\} \\
\pi_{11} = p, \quad \pi_{00} = q
\end{gathered} \tag{61}$$

where $C_t$ is the new composite index; $w_t$ and $\epsilon_{it}$ are independent of each other for all $t$ and $i$; and the variance of $w_t$ is taken to be unity for identification of the model.

For expositional purposes, we assume $\gamma_i(L) = \gamma_i$ *(A brief digression, $\gamma_i$ is the reason why the variance of $w_t$ needs to be normalized)*, $\psi_i(L) = 1 - \psi_{i1}L - \psi_{i2}L^2$, for $i = 1, 2, 3, 4$, and $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$.

The linear dynamic factor has a mean, $\mu_{St} + \delta$, switching between the two regimes of the business cycle. Imposing a mean of 0 on the $\mu_{St}$ process, $\delta$ determines the long-run growth rate of the index and $\mu_{St}$ produces deviations from that long-run growth rate (Page 129, Chapter 5, Kim and Nelson (1999)).

Similar to section 3.4, the above model is not identified, as the mean of each series $\mathrm{E}[\Delta Y_{it}] = D_i + \gamma_i\delta$ is overparameterized (refer to 3.4 for more details). Thus, we consider the model in deviation from mean form

$$\Delta y_{it} = \gamma_i \Delta c_t + e_{it}, \quad i = 1, 2, 3, 4 \tag{62}$$

$$\phi(L)(\Delta c_t - \mu_{St}) = w_t, \quad w_t \sim \text{i.i.d. } \mathrm{N}(0, 1) \tag{63}$$

where $\Delta y_{it} = \Delta Y_{it} - \overline{\Delta Y}_i$ and $\Delta c_t = \Delta C_t - \delta$.

My comment: by subtracting $D_i + \gamma_i\delta$, a zero mean is implicitly imposed on $\mu_{St}$. To see this point, Equation (62) requires that $\Delta c_t$ has a zero mean, while Equation (63) stipulates that $\Delta c_t$ and $\mu_{St}$ share the same mean.

### 15.1.1 Generating $\Delta c_{1:n}$, Conditional on $\Delta y_{i,1:n}$, $i = 1, 2, 3, 4$, $S_{1:n}$ and $\Theta$

We have five state variables: $\Delta c_t$ and $e_{it}, i = 1, 2, 3, 4$. If we multiply both sides of (62) by $\psi_i(L)$, we have

$$\Delta y_{it}^{*} = \gamma_i \psi_i(L) \Delta c_t + \epsilon_{it}, \quad \epsilon_{it} \sim \text{i.i.d. N}(0, \sigma_i^2), \quad i = 1, 2, 3, 4 \tag{64}$$

Then equations (64) and (63) can be written in the following state-space representation form:

$$\begin{pmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \end{pmatrix} = \begin{pmatrix} \phi(L)\mu_{S\,t} \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \phi_1 & \phi_2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \Delta c_{t-1} \\ \Delta c_{t-2} \\ \Delta c_{t-3} \end{pmatrix} + \begin{pmatrix} w_t \\ 0 \\ 0 \end{pmatrix}$$

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} \Delta y_{1t}^{*} \\ \Delta y_{2t}^{*} \\ \Delta y_{3t}^{*} \\ \Delta y_{4t}^{*} \end{pmatrix} = \begin{pmatrix} \gamma_1 & -\gamma_1\psi_{11} & -\gamma_1\psi_{12} \\ \gamma_2 & -\gamma_2\psi_{21} & -\gamma_2\psi_{22} \\ \gamma_3 & -\gamma_3\psi_{31} & -\gamma_3\psi_{32} \\ \gamma_4 & -\gamma_4\psi_{41} & -\gamma_4\psi_{42} \end{pmatrix} \begin{pmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \epsilon_{4t} \end{pmatrix}$$

$$R = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{pmatrix}$$

Denote $x_t = (\Delta c_t, \Delta c_{t-1}, \Delta c_{t-2})'$, and $y_{1:n} = (\Delta y_{1,1:n}^{*\prime}, \ldots, \Delta y_{4,1:n}^{*\prime})'$. The fact that the model is linear and Gaussian conditional on $S_{1:n}$ allows us to employ the procedures in Property 2.1 and is the reason why no procedure of the kind in Property 10.4 need be employed.

Then the FFBS algorithm introduced in section 13.2 can be used to generate $x_{1:n}$. More specifically, for the reason that $Q$ is singular in this model, Equation (48) is used to generate $\Delta c_{1:n}$.

### 15.1.2 Generating $\{\gamma_i, \psi_{i1}, \psi_{i2}, \sigma_i^2\}$, Conditional on $\Delta c_{1:n}$ and $\Delta y_{i,1:n}$, $i = 1, 2, 3, 4$

Conditional on $\Delta c_{1:n}$, Equation (63) and the switching regime become irrelevant. The model (62) collapses to four independent regression equations with autocorrelated disturbances.

For $i = 1, 2, 3, 4$, denote $\boldsymbol{\psi_i} = (\psi_{i1}, \psi_{i2})'$. We employ normal priors for $\gamma_i$ and $\boldsymbol{\psi_i}$, and an inverted Gamma distribution for $\sigma_i^2$ in the following way:

$$\gamma_i|\boldsymbol{\psi_i}, \sigma_i^2 \sim \mathrm{N}(\underline{a}_i, \underline{A}_i)$$
$$\boldsymbol{\psi_i}|\sigma_i^2 \sim \mathbf{N}(\underline{b}_i, \underline{B}_i)_{I(\psi(L)\in\mathbb{S})}$$
$$\sigma_i^2|\boldsymbol{\psi_i} \sim \mathrm{IG}(\frac{f_i}{2}, \frac{F_i}{2})$$

Then the pseudo-posterior of $\gamma_i$

$$\begin{aligned}
\mathrm{p}(\gamma_i|\boldsymbol{\psi_i}, \sigma_i^2, \Delta c_{1:n}, \Delta y_{i,1:n}) &\propto \mathrm{p}(\gamma_i|\boldsymbol{\psi_i}, \sigma_i^2)\mathrm{p}(\Delta y_{i,1:n}, \Delta c_{1:n}|\gamma_i, \boldsymbol{\psi_i}, \sigma_i^2) \\
&= \mathrm{p}(\gamma_i|\boldsymbol{\psi_i}, \sigma_i^2)\mathrm{p}(\Delta y_{i,1:n}|\gamma_i, \boldsymbol{\psi_i}, \sigma_i^2, \Delta c_{1:n})\mathrm{p}(\Delta c_{1:n}) \\
&\propto \mathrm{p}(\gamma_i|\boldsymbol{\psi_i}, \sigma_i^2)\mathrm{p}(\Delta y_{i,1:n}|\gamma_i, \boldsymbol{\psi_i}, \sigma_i^2, \Delta c_{1:n})
\end{aligned}$$

where

$$\mathrm{p}(\Delta y_{i,1:n}|\gamma_i, \boldsymbol{\psi_i}, \sigma_i^2, \Delta c_{1:n}) = \mathrm{p}(\Delta y^*_{i,1:n}|\gamma_i, \boldsymbol{\psi_i}, \sigma_i^2, \Delta c_{1:n})$$

with $\Delta y^*_{i,1:n}$ defined in Equation (64). Let's denote $\Delta c^*_t = \psi_i(L)\Delta c_t$, $\Delta c^*_{1:n} = (\Delta c^*_1, \ldots, \Delta c^*_n)'$ and $\Delta y^*_{i,1:n} = (\Delta y^*_{i,1:n}, \ldots, \Delta y^*_{i,1:n})'$. Then in matrix notation, Equation (64) can be written as

$$\Delta y^*_{i,1:n} = \gamma_i \Delta c^*_{1:n} + \epsilon_{i,1:n}, \quad \epsilon_{i,1:n} \sim \mathbf{N}(0, \sigma_i^2 I)$$

So

$$\begin{aligned}
\mathrm{p}(\Delta y_{i,1:n}|\gamma_i, \boldsymbol{\psi_i}, \sigma_i^2, \Delta c_{1:n}) &= \mathrm{p}(\Delta y^*_{i,1:n}|\gamma_i, \boldsymbol{\psi_i}, \sigma_i^2, \Delta c_{1:n}) \\
&= \mathrm{p}(\Delta y^*_{i,1:n}|\gamma_i, \sigma_i^2, \Delta c^*_{1:n}) \\
&\propto \exp\left\{-\frac{(\Delta y^*_{i,1:n} - \gamma_i\Delta c^*_t)'(\Delta y^*_{i,1:n} - \gamma_i\Delta c^*_t)}{2\sigma_i^2}\right\}
\end{aligned}$$

Therefore, the pseudo-posterior of $\gamma_i$

$$\gamma_i|\boldsymbol{\psi_i}, \sigma_i^2, \Delta c_{1:n}, \Delta y_{i,1:n} \sim \mathrm{N}(\overline{a}_i, \overline{A}_i)$$

with

$$\overline{a}_i = \left(\underline{A}_i^{-1} + \sigma_i^{-2}\Delta c_{1:n}^{*\prime}\Delta c_{1:n}^{*}\right)^{-1}\left(\underline{A}_i^{-1}\underline{a}_i + \sigma_i^{-2}\Delta c_{1:n}^{*\prime}\Delta y_{i,1:n}^{*}\right)$$
$$\overline{A}_i = \left(\underline{A}_i^{-1} + \sigma_i^{-2}\Delta c_{1:n}^{*\prime}\Delta c_{1:n}^{*}\right)^{-1}$$

The pseudo-posterior of $\boldsymbol{\psi_i}$, according to Equation (61),

$$\begin{aligned}\mathrm{p}(\boldsymbol{\psi_i}|\gamma_i, \sigma_i^2, \Delta c_{1:n}, \Delta y_{i,1:n}) &= \mathrm{p}(\boldsymbol{\psi_i}|\sigma_i^2, e_{i,1:n}) \\ &\propto \mathrm{p}(\boldsymbol{\psi_i}|\sigma_i^2)\mathrm{p}(e_{i,1:n}|\boldsymbol{\psi_i}, \sigma_i^2)\end{aligned}$$

where $e_{i,1:n} = (e_{i,1}, \ldots, e_{i,n})'$; $\mathrm{p}(e_{i,1:n}|\boldsymbol{\psi_i}, \sigma_i^2)$ can be get from Equation (61), which can be written in the following matrix notation:

$$e_{i,1:n} = E_i\boldsymbol{\psi_i} + \epsilon_{i,1:n}, \quad \epsilon_{i,1:n} \sim \mathbf{N}(0, \sigma_i^2 I) \tag{65}$$

with $E_i = (e_{i,0:n-1}, e_{i,-1:n-2})$ being an $n \times 2$ matrix. So

$$\mathrm{p}(e_{i,1:n}|\boldsymbol{\psi_i}, \sigma_i^2) \propto \exp\left\{-\frac{(e_{i,1:n} - E_i\boldsymbol{\psi_i})'(e_{i,1:n} - E_i\boldsymbol{\psi_i})}{2\sigma_i^2}\right\}.$$

Therefore, the pseudo-posterior of $\boldsymbol{\psi_i}$

$$\boldsymbol{\psi_i}|\gamma_i, \sigma_i^2, \Delta c_{1:n}, \Delta y_{i,1:n}^{*} \sim \mathbf{N}(\overline{b}_i, \overline{B}_i)_{I(\psi(L)\in\mathbb{S})}$$

with

$$\overline{b}_i = \left(\underline{B}_i^{-1} + \sigma_i^{-2}E_i'E_i\right)^{-1}\left(\underline{B}_i^{-1}\underline{b}_i + \sigma_i^{-2}E_i'e_{i,1:n}\right)$$
$$\overline{B}_i = \left(\underline{B}_i^{-1} + \sigma_i^{-2}E_i'E_i\right)^{-1}$$

The pseudo-posterior of $\sigma_i^2$

$$\begin{aligned}\mathrm{p}(\sigma_i^2|\gamma_i, \boldsymbol{\psi_i}, \Delta c_{1:n}, \Delta y_{i,1:n}) &= \mathrm{p}(\sigma_i^2|\boldsymbol{\psi_i}, e_{i,1:n}) \\ &\propto \mathrm{p}(\sigma_i^2|\boldsymbol{\psi_i})\mathrm{p}(e_{i,1:n}|\boldsymbol{\psi_i}, \sigma_i^2)\end{aligned}$$

where

$$\mathrm{p}(\sigma_i^2|\boldsymbol{\psi_i}, e_{i,1:n}) = \left(\frac{1}{\sigma_i^2}\right)^{\frac{f_i}{2}-1}\exp\left\{-\frac{F_i}{2\sigma_i^2}\right\}$$

and, according to Equation (65),

$$\mathrm{p}(e_{i,1:n}|\boldsymbol{\psi_i}, \sigma_i^2) \propto \left(\frac{1}{\sigma_i^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{(e_{i,1:n} - E_i\boldsymbol{\psi_i})'(e_{i,1:n} - E_i\boldsymbol{\psi_i})}{2\sigma_i^2}\right\}.$$

Therefore, the pseudo-posterior of $\sigma_i^2$

$$\sigma_i^2|\gamma_i, \boldsymbol{\psi_i}, \Delta c_{1:n}, \Delta y_{i,1:n} \sim \mathrm{IG}\left(\frac{f_i + n}{2}, \frac{F_i + (e_{i,1:n} - E_i\boldsymbol{\psi_i})'(e_{i,1:n} - E_i\boldsymbol{\psi_i})}{2}\right).$$

### 15.1.3 Generating $S_{1:n}, \phi_1, \phi_2, \mu_0, \mu_1$, $p$ and $q$, Conditional on $\Delta c_{1:n}$

Conditional on $\Delta c_{1:n}$, the observed data $\Delta y_{1:n}$ carry no information about $S_{1:n}$ or the parameters $\{\phi_1, \phi_2, \mu_0, \mu_1, p, q\}$ beyond that contained in $\Delta c_{1:n}$. Thus we focus on Equation (63), which is an autoregression with Markov-switching mean with $\Delta c_{1:n}$ treated as data for the model. The procedure is the same as in Section 14.2.

$S_{1:n}$ can be generated via the FFBS procedure from Equation (57).

$p$ and $q$ can be generated according to the two beta posteriors given by (52) and (53).

And similar to Section 14.2.3, to generate $\mu_0$ and $\mu_1$, we rearrange Equation (63)

$$\Delta c_t^{**} = (1 - \phi_1 - \phi_2)\mu_0 + S_t^{**}\mu_1 + w_t, \quad w_t \sim \mathrm{N}(0, 1)$$

with

$$\Delta c_t^{**} = \Delta c_t - \phi_1 \Delta c_{t-1} - \phi_2 \Delta c_{t-2}$$
$$S_t^{**} = S_t - \phi_1 S_{t-1} - \phi_2 S_{t-2}$$

In matrix notation,

$$\Delta c_{1:n}^{**} = X\boldsymbol{\mu} + w_{1:n}, \quad w_{1:n} \sim \mathbf{N}(0, I)$$

where

$$X = \begin{pmatrix} 1 - \phi_1 - \phi_2 & S_1^{**} \\ \vdots & \vdots \\ 1 - \phi_1 - \phi_2 & S_n^{**} \end{pmatrix}$$

If we assume a normal prior for $\boldsymbol{\mu} \equiv (\mu_0, \mu_1)'$:

$$\boldsymbol{\mu}|\boldsymbol{\phi}, \sim \mathbf{N}(\underline{g}, \underline{G}),$$

with $\boldsymbol{\phi} \equiv (\phi_1, \phi_2)'$, then the posterior

$$\boldsymbol{\mu}|\boldsymbol{\phi}, \Delta c_{1:n}, S_{1:n} \sim \mathbf{N}(\overline{g}, \overline{G})$$

with

$$\overline{g} = \left(\underline{G}^{-1} + X'X\right)^{-1} \left(\underline{G}^{-1}\underline{g} + X'\Delta c_{1:n}^{**}\right)$$
$$\overline{G} = \left(\underline{G}^{-1} + X'X\right)^{-1}$$

$\boldsymbol{\phi}$ can be generated, similar to Section 14.2.5, from a normal posterior:

$$\boldsymbol{\phi}|\boldsymbol{\mu}, \Delta c_{1:n}, S_{1:n} \sim \mathbf{N}(\overline{h}, \overline{H})_{I(\phi(L)\in\mathbb{S})}$$

with

$$\overline{h} = \left(\underline{H}^{-1} + X^{*\prime}X^{*}\right)^{-1} \left(\underline{H}^{-1}\underline{h} + X^{*\prime}Y^{*}\right)$$
$$\overline{H} = \left(\underline{H}^{-1} + X^{*\prime}X^{*}\right)^{-1}$$

where

$$X^{*} = \begin{pmatrix} \Delta c_0 - \mu_{S,0} & \Delta c_{-1} - \mu_{S,-1} \\ \vdots & \vdots \\ \Delta c_{n-1} - \mu_{S,n-1} & \Delta c_{n-2} - \mu_{S,n-2} \end{pmatrix}$$
$$Y^{*} = (\Delta c_1 - \mu_{S,1}, \ldots, \Delta c_n - \mu_{S,n})'$$

### 15.1.4 Calculation of the Composite Coincident Indicator Index, $C_t$

$\Delta C_t = \Delta c_t + \delta$ for $t = 1, \ldots, n$. Same as Equation (14), $\delta$ is calculated as

$$\delta = W(1)\overline{\Delta Y}$$

with $\overline{\Delta Y} \equiv (\overline{\Delta Y_1}, \ldots, \overline{\Delta Y_4})'$, and $W(1) \equiv (1, 0, \ldots, 0)\,[I - (I - KA)\Phi]^{-1}\,K$, where $A, \Phi$ and $K$ are as defined in Section 3.4.3.

As I pointed out in Section 3.4.3, I see no theoretical basis for doing that.

# References


Dieobold, F. X. (1986). Modeling the persistence of conditional variances: A comment. *Econometric Reviews*, *5*(1), 51–56.

Douc, R., Moulines, E., & Stoffer, D. (2014). *Nonlinear time series: theory, methods and applications with r examples*. CRC Press.

Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of econometrics*, *64*(1), 307–333.

Harvey, A., Ruiz, E., & Sentana, E. (1992). Unobserved component time series models with arch disturbances. *Journal of Econometrics*, *52*(1-2), 129–157.

Kim, C.-J., & Nelson, C. R. (1999). State-space models with regime switching: classical and gibbs-sampling approaches with applications. *MIT Press Books*, *1*.

Kim, C.-J., Nelson, C. R., & Startz, R. (1998). Testing for mean reversion in heteroskedastic data based on gibbs-sampling-augmented randomization. *Journal of Empirical finance*, *5*(2), 131–154.

Lamoureux, C. G., & Lastrapes, W. D. (1990). Persistence in variance, structural change, and the garch model. *Journal of Business & Economic Statistics*, *8*(2), 225–234.

Lastrapes, W. D. (1989). Exchange rate volatility and us monetary policy: an arch application. *journal of Money, Credit and Banking*, *21*(1), 66–77.

Malmsten, H. (2004). *Evaluating exponential garch models* (No. 564). SSE/EFI Working paper Series in Economics and Finance. (Working Paper)

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: with r examples*. Springer Science & Business Media.