Generalizations of Linear Time Series Models 18

Autoregressive moving-average models have become the predominant approach in the analysis of economic, especially macroeconomic time series. The success of these parametric models is due to a mature and by now well-understood statistical theory which has been the subject of this book. The main assumption behind this theory is its linear structure. Although convenient, the assumption of a constant linear structure turned out to be unrealistic in many empirical applications. The evolution of economies and the economic dynamics are often not fully captured by constant coefficient linear models. Many time series are subject to structural breaks which manifest themselves as a sudden change in the model coefficients by going from one period to another. The detection and dating of such structural breaks is the subject of Sect. 18.1. Alternatively, one may think of the model coefficients as varying over time. Such models have proven to be very flexible and able to generate a variety of non-linear features. We present in Sects. 18.2 and 18.3 two variants of such models. In the first one, the model parameters vary in a systematic way with time. They are, for example, following an autoregressive process. In the second one, the parameters switch between a finite number of states according to a hidden Markov chain. These states are often identified as regimes which have a particular economic meaning, for example as booms and recessions. Further parametric and nonparametric methods for modeling and analyzing nonlinear time series can be found in Fan and Yao (2003).

18.1 Structural Breaks

There is a extensive literature dealing with the detection and dating of structural breaks in the context of time series. This literature is comprehensibly summarized in Perron (2006), among others. A compact account can also be found in Aue and Horváth (2011) where additional testing procedures, like the CUSUM test, are

K. Neusser, *Time Series Econometrics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-319-32862-1_18

presented. In this short exposition we follow Bai et al. (1998) and focus on Chow type test procedures. For the technical details the interested reader is referred to these papers.

18.1.1 Methodology

Consider, for the ease of exposition, a VAR(1) process which allows for a structural break at some known date t_b:

$$X_t = d_t(t_b)\left(c^{(1)} + \Phi^{(1)}X_{t-1}\right) + (1 - d_t(t_b))\left(c^{(2)} + \Phi^{(2)}X_{t-1}\right) + Z_t \tag{18.1}$$

where

$$d_t(t_b) = \begin{cases} 1, & t \le t_b; \\ 0, & t > t_b. \end{cases}$$

Thus, before time t_b the coefficients of the VAR process are given by $c^{(1)}$ and $\Phi^{(1)}$ whereas after t_b they are given by $c^{(2)}$ and $\Phi^{(2)}$. The error process $\{Z_t\}$ is assumed to be IID$(0, \Sigma)$ with Σ positive definite.[1] Suppose further that the roots of $\Phi^{(1)}(z)$ as well as those of $\Phi^{(2)}(z)$ are outside the unit circle. The process therefore is stationary and admits a causal representation with respect to $\{Z_t\}$ before and after date t_b.

The assumption of a structural break at some known date t_b can then be investigated by testing the hypothesis

$$\mathbf{H}_0 : c^{(1)} = c^{(2)} \text{ and } \Phi^{(1)} = \Phi^{(2)} \quad \text{against} \quad \mathbf{H}_1 : c^{(1)} \neq c^{(2)} \text{ or } \Phi^{(1)} \neq \Phi^{(2)}.$$

The standard way to test such a hypothesis is via the F-statistic. Given a sample ranging from period 0 to period T, the strategy is to partition all variables and matrices along the break date t_b. Following the notation and the spirit of Sect. 13.2, define $Y = \text{vec}(Y^{(1)}, Y^{(2)})$ where $Y^{(1)} = (X_1, X_2, \ldots, X_{t_b})$ and $Y^{(2)} = (X_{t_b+1}, X_{t_b+2}, \ldots, X_T)$, $Z = (Z_1, Z_2, \ldots, Z_T)$, and

$$\mathbf{X}^{(1)} = \begin{pmatrix} 1 & X_{1,0} & \ldots & X_{n,0} \\ 1 & X_{1,1} & \ldots & X_{n,1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1,t_b-1} & \ldots & X_{n,t_b-1} \end{pmatrix} \qquad \mathbf{X}^{(2)} = \begin{pmatrix} 1 & X_{1,t_b} & \ldots & X_{n,t_b} \\ 1 & X_{1,t_b+1} & \ldots & X_{n,t_b+1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1,T-1} & \ldots & X_{n,T-1} \end{pmatrix},$$

[1] Generalization to higher order VAR models is straightforward. For changes in the covariance matrix Σ see Bai (2000). For the technical details the reader is referred to the relevant literature.

then the model (18.1) can be written as

$$Y = \text{vec}(Y^{(1)}, Y^{(2)}) = \underbrace{\begin{pmatrix} \mathbf{X}^{(1)} \otimes I_n & 0 \\ 0 & \mathbf{X}^{(2)} \otimes I_n \end{pmatrix}}_{\mathbf{X}} \underbrace{\text{vec}(c^{(1)}, \Phi^{(1)}, c^{(2)}, \Phi^{(2)})}_{\beta} + \text{vec}\, Z.$$

The least-squares estimator becomes

$$\hat{\beta} = \text{vec}(\hat{c}^{(1)}, \widehat{\Phi}^{(1)}, \hat{c}^{(2)}, \widehat{\Phi}^{(2)}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y$$
$$= \begin{pmatrix} ((\mathbf{X}^{(1)'}\mathbf{X}^{(1)})^{-1}\mathbf{X}^{(1)'}) \otimes I_n & 0 \\ 0 & ((\mathbf{X}^{(2)'}\mathbf{X}^{(2)})^{-1}\mathbf{X}^{(2)'}) \otimes I_n \end{pmatrix} \text{vec}(Y^{(1)}, Y^{(2)}).$$

This amounts to estimate the model separately over the two sample periods. Note that as in Sect. 13.2 the GLS estimator is numerically identical to the OLS estimator because the same regressors are used for each equation. The corresponding Wald-test can be implemented by defining $R = (I_{n^2+n}, -I_{n^2+n})$ and computing the F-statistic

$$\text{F}(t_b) = (R\hat{\beta})'[R(\mathbf{X}'(I_T \otimes \widehat{\Sigma}_{t_b}^{-1})\mathbf{X})R']^{-1}(R\hat{\beta}) \tag{18.2}$$

where $\hat{\beta} = \text{vec}(\hat{c}^{(1)}, \widehat{\Phi}^{(1)}, \hat{c}^{(2)}, \widehat{\Phi}^{(2)})$ and where $\widehat{\Sigma}_{t_b}$ is computed from the least-squares residuals $\widehat{Z}_t$ as $\widehat{\Sigma}_{t_b} = \frac{1}{T}\sum_{t=1}^{T} \widehat{Z}_t \widehat{Z}_t'$ given break date t_b. Under the standard assumptions made in Sect. 13.2, the test statistic $\text{F}(t_b)/(n^2 + n)$ converges for $T \to \infty$ to a chi-square distribution with $n^2 + n$ degrees of freedom.[2] This test is known in the literature as the Chow test.

The previous analysis assumed that the potential break date t_b is known. This assumption often turns out to be unrealistic in practice. The question then arises how to determine a potential break date. Quandt (1960) proposed a simple procedure: compute the Chow-test for all possible break dates and take as a candidate break date the date where the F-statistic reaches its maximal value. Despite its simplicity, Quandt's procedure could not be implemented coherently because it was not clear which distribution to use for the construction of the critical values. This problem remained open for more than thirty years until the contribution of Andrews (1993).[3] Denote by $\lfloor x \rfloor$ the value of x rounded to the nearest integer towards minus infinity, then the maximum Wald statistic and the logarithm of the Andrews and Ploberger (1994) exponential Wald statistic can be written as follows:

[2] As the asymptotic theory requires that t_b/T does not go to zero, one has to assume that both the number of periods before and after the break go to infinity.

[3] A textbook version of the test can be found in Stock and Watson (2011).

$$\sup \mathrm{F}: \quad \sup_{\tau \in (\tau^*, 1-\tau^*)} \mathrm{F}(\lfloor T\tau \rfloor)$$

$$\exp \mathrm{F}: \quad \log \int_{\tau^*}^{1-\tau^*} \exp(\tfrac{1}{2} F(\lfloor T\tau \rfloor)) \mathrm{d}\tau$$

where τ^* denotes the percentage of the sample which is trimmed. Usually, τ^* takes the value of 0.15 or 0.10. Critical values for low degrees of freedom are tabulated in Andrews (1993, 2003) and Stock and Watson (2011). It is possible to construct an asymptotic confidence interval for the break date. The corresponding formulas can be found in Bai et al. (1998, p. 401–402).

18.1.2 An Example

The use of the structural break test is demonstrated using historical data for the United Kingdom. The data consist of logged per capita real GDP, logged per capita real government expenditures, logged per capita real government revenues, the inflation based on the consumer price index, and a long-term interest rate over a sample period from 1830 to 2003. The basis for the analysis consists of a five variable VAR(2) model including a constant term and a linear trend. Three alternative structural break modes are investigated: break in the intercept, break in the intercept and the time trend, and break in all coefficients, including the VAR coefficients. The corresponding F-statistics are plotted in Fig. 18.1 against all possible break dates allowing for a trimming value of 10 %. The horizontal lines show for all three alternative break modes the corresponding critical values for the supF test given 5 % significance levels. These critical values have been obtained from Monte Carlo simulations as in Andrews (1993, 2003) and are given as 18.87, 28.09, and 97.39.[4]

Figure 18.1 shows that for all three modes a significant structural break occurs. The corresponding values of the supF statistics are 78.06, 104.75, and 285.22. If only the deterministic parts are allowed to change, the break date is located in 1913. If all coefficients are allowed to change, the break is dated in 1968. However, all three F-statistics show a steep increase in 1913. Thus, if only one break is allowed 1913 seems to be the most likely one.[5] The breaks are quite precisely dated. The corresponding standard errors are estimated to be two years for the break in the intercept only and one year for the other two break modes.

[4] Assuming a trimming value of 0.10 Andrews (2003, table I) reports critical values of 18.86 for $p = 5$ which corresponds to changes in the intercept only and 27.27 for $p = 10$ which corresponds to changes in intercept and time trend.

[5] See Perron (2006) for a discussion of multiple breaks.

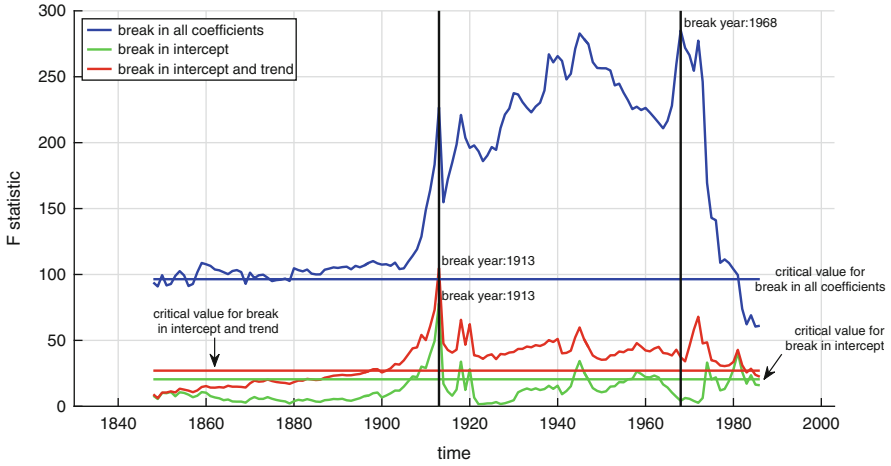

Fig. 18.1 Analysis of breaks dates with the sup F test statistic for historical UK time series

18.2 Time-Varying Parameters

This section discusses time-varying coefficient vector autoregressive models (TVC-VAR models). This model class retains the flavor of VAR models but assumes that they are only valid locally. Consider for this purpose a VAR(1) model with time-varying autoregressive coefficients Φ_t:

$$X_{t+1} = \Phi_t X_t + Z_{t+1}, \qquad Z_t \sim \text{IID}(0, \Sigma) \text{ with } \Sigma > 0, \quad t \in \mathbb{Z}. \tag{18.3}$$

This model can be easily generalized to higher order VAR's (see below) or, alternatively, one may think of Eq. (18.3) as a higher order VAR in companion form. The autoregressive coefficient matrix is assumed to be stochastic. Thus, Φ_t is a random $n \times n$ matrix. Models of this type have been widely discussed in the probabilistic literature because they arise in many diverse contexts. In economics, Eq. (18.3) can be interpreted as the probabilistic version describing the value of a perpetuity, i.e. the present discounted value of a permanent commitment to pay a certain sum each period. Thereby Z_t denotes the random periodic payments and Φ_t the random cumulative discount factors. The model also plays an important role in the characterization of the properties of volatility models as we have seen in Sect. 8.1 (see in particular the proofs of Theorems 8.1 and 8.3). In this presentation, the above model is interpreted as a locally valid VAR process.

A natural question to ask is under which conditions Eq. (18.3) admits a stationary solution. An answer to this question can be found by iterating the equation backwards in time:

$$
\begin{aligned}
X_t &= \Phi_{t-1}X_{t-1} + Z_t = \Phi_{t-1}(\Phi_{t-2}X_{t-2} + Z_{t-1}) + Z_t \\
&= Z_t + \Phi_{t-1}Z_{t-1} + \Phi_{t-1}\Phi_{t-2}X_{t-2} \\
&= Z_t + \Phi_{t-1}Z_{t-1} + \Phi_{t-1}\Phi_{t-2}Z_{t-2} + \Phi_{t-1}\Phi_{t-2}\Phi_{t-3}X_{t-3} \\
&\quad \dots \\
&= \sum_{j=0}^{k} \left(\prod_{i=1}^{j} \Phi_{t-i} \right) Z_{t-j} + \left(\prod_{i=1}^{k+1} \Phi_{t-i} \right) X_{t-k-1}, \quad k = 0, 1, 2, \dots
\end{aligned}
$$

where it is understood that $\prod_{i=1}^{0} = I_n$. This suggests as a solution candidate

$$
\begin{aligned}
X_t &= \lim_{k\to\infty} \sum_{j=0}^{k} \left(\prod_{i=1}^{j} \Phi_{t-i} \right) Z_{t-j} \\
&= Z_t + \Phi_{t-1}Z_{t-1} + \Phi_{t-1}\Phi_{t-2}Z_{t-2} + \Phi_{t-1}\Phi_{t-2}\Phi_{t-3}Z_{t-3} + \dots
\end{aligned} \tag{18.4}
$$

Based on results obtained by Brandt (1986) and extended by Bougerol and Picard (1992b), we can cite the following theorem.

Theorem 18.1 (Solution TVC-VAR(1))**.** *Let* $\{(\Phi_t, Z_t)\}$ *be a strictly stationary ergodic process such that*

(i) $\mathbb{E}(\log^+ \|\Phi_t\|) < \infty$ *and* $\mathbb{E}(\log^+ \|Z_t\|) < \infty$ *where* x^+ *denotes* $\max\{x, 0\}$*;*
(ii) the top Lyapounov exponent γ *defined as*

$$
\gamma = \inf_{n\in\mathbb{N}} \left\{ \mathbb{E}\left(\frac{1}{n+1} \log \|\Phi_0\Phi_{-1}\dots\Phi_{-n}\| \right) \right\}
$$

is strictly negative.

Then X_t *as defined in Eq.* (18.4) *converges a.s. and* $\{X_t\}$ *is the unique strictly stationary solution of equation* (18.3)*.*

Remark 18.1. The Lyapunov exponent measures the rate of separation of nearly trajectories in a dynamic system. The top Lyapunov exponent gives the largest of these rates. It is used to characterize the stability of a dynamic system (see Colonius and Kliemann (2014)).

Remark 18.2. Although Theorem 18.1 states only sufficient conditions, these assumptions can hardly be relaxed.

The solution (18.4), if it exists, is similar to a causal representation. The matrix sequence $\{\prod_{i=1}^{h} \Phi_{t+h-i}\}_{h=0,1,2,\dots} = \{I_n, \Phi_t, \Phi_{t+1}\Phi_t, \Phi_{t+2}\Phi_{t+1}\Phi_t, \dots\}$ represents the effect of an impulse in period t to X_{t+h}, $h = 0, 1, 2, \dots$ and can therefore be interpreted as impulse response functions. In contrast to the impulse response

functions studied so far, they are clearly random and time-dependent because the effect of Z_t depends on future coefficients. In particular, the effect of Z_t on X_{t+h}, $h \geq 1$, is not the same as the effect of Z_{t-h} on X_t. Nevertheless it is possible to construct meaningful impulse response functions by Monte Carlo simulations. One may then report the mean of the impulse responses or some quantiles for different time periods.[6] Alternatively, one may ignore the randomness and time-dependency and define "local" impulse responses as Φ_t^h, $h = 0, 1, 2, \ldots$. Note, however, that the impulse responses so defined still vary with time. Irrespectively how the impulse responses are constructed, they can be interpreted in the same way as in the case of constant coefficients. In particular, we may use some of the identification schemes discussed in Chap. 15 and compute the impulse responses with respect to structural shocks. Similar arguments apply to the forecast error variance decomposition (FEVD).

The model is closed by fixing the law of motion for Φ_t. As already mentioned in Sect. 17.1.1 there are several possibilities. In this presentation we adopt the following flexible autoregressive specification:

$$\beta_{t+1} - \bar{\beta} = F(\beta_t - \bar{\beta}) + V_{t+1} \qquad V_t \sim \mathrm{WN}(0, Q) \tag{18.5}$$

where $\beta_t = \operatorname{vec} \Phi_t$ denotes the n^2 vector of stacked coefficients. Q is assumed to be fixed and is, usually, specified as a diagonal matrix. If the eigenvalues of F are inside the unit circle, the autoregressive model is mean-reverting and $\bar{\beta}$ can be interpreted as the average coefficient vector. The formulation in Eq. (18.5) is, however, not restricted to this case and allows explicitly the possibility that $\{\beta_t\}$ follows a random walk. This specification has become very popular in the empirical macroeconomic literature and was initially adopted by Cogley and Sargent (2001) to analyze the dynamics of inflation across different policy regimes.[7]

The model consisting of Eqs. (18.3) and (18.5) can be easily reformulated as a state space model by defining $\xi_t = \beta_t - \bar{\beta}$ as the state vector. The state and the measurement equation can then be written as:

$$\text{state equation:} \qquad \xi_{t+1} = F\xi_t + V_{t+1} \tag{18.6}$$

$$\text{measurement equation:} \qquad X_t = (X_{t-1}' \otimes I_n)\bar{\beta} + (X_{t-1}' \otimes I_n)\xi_t + Z_t. \tag{18.7}$$

Conditional on initial values for the coefficients and their covariances, the state space model can be estimated by maximum likelihood by applying the Kalman filter (see Sect. 17.3 and Kim and Nelson (1999)). One possibility to initialize the Kalman filter is to estimate the model for some initial sample period assuming fixed coefficients and extract from these estimates the corresponding starting values.

[6] Potter (2000) discusses the primal problems of defining impulse responses in a nonlinear context.

[7] They allow for a correlation between V_t and Z_t.

As it turns out, allowing time-variation only in the coefficients of the VAR model overstates the role attributed to structural changes. We therefore generalize the model to allow for time-varying volatility. More specifically, we also allow Σ in Eq. (18.3) to vary with time. The modeling of the time-variation in Σ is, however, not a straightforward task because we must ensure that in each period Σ_t is a symmetric positive definite matrix. One approach is to specify a process especially designed for modeling the dynamics of covariance matrices. This so-called Wishart autoregressive process was first introduced to economics by Gouriéroux et al. (2009) and successfully applied by Burren and Neusser (2013). It leads to a nonlinear state space system which can be estimated with the particle filter, a generalization of the Kalman filter.

Another more popular approach was initiated by Cogley and Sargent (2005) and Primiceri (2005). It is based on the Cholesky factorization of the time-varying covariance matrix Σ_t. Using the same notation as in Sect. 15.3 Σ_t is decomposed as

$$\Sigma_t = B_t \Omega_t B_t' \tag{18.8}$$

where B_t is a time-varying lower triangular matrix with ones on the diagonal and Ω_t a time-varying diagonal matrix with strictly positive diagonal elements.[8] The logged diagonal elements of Ω_t are then assumed to evolve as independent univariate random walks. This specification can be written in matrix terms as

$$\Omega_t = \Omega_{t-1} \exp(D_t) \tag{18.9}$$

where D_t is a diagonal matrix with $\operatorname{diag}(D_t) \sim \mathrm{WN}(0, \Omega_D)$. In the above formulation exp denotes the matrix exponential.[9] Taking the matrix logarithm, we get exactly the formulation of Cogley and Sargent (2005) and Primiceri (2005). For the time evolution of B_t we propose a similar specification:

$$B_t = B_{t-1} \exp(C_t) \tag{18.10}$$

where C_t is a strictly lower triangular matrix, i.e. C_t is a lower triangular matrix with zeros on the diagonal. The non-zero entries of C_t, denoted by $[C_t]_{i>j}$, are assumed to follow a multivariate white noise process with diagonal covariance matrix Σ_B, i.e. $[C_t]_{i>j} \sim \mathrm{WN}(0, \Sigma_B)$. It can be shown that the matrix exponential of strictly lower triangular matrices are triangular matrices with ones on the diagonal. As the set of triangular matrices with ones on the diagonal form a group, called the *unipotent group* and denoted by SLT_n, the above specification is well-defined. Moreover, this formulation is a very natural one as the set of strictly lower triangular matrices

[8] It is possible to consider other short-run type identification schemes (see Sect. 15.3) than the Cholesky factorization.

[9] The matrix exponential of a matrix A is defined as $\exp(A) = \sum_{i=0}^{\infty} \frac{1}{i!} A^i$ where A is any matrix. Its inverse $\log(A)$ is defined only for $\|A\| < 1$ and is given by $\log(A) = \sum_{i=1}^{\infty} \frac{(-1)^{i-1}}{i} A^i$.

is the tangent space of SLT_n at the identity (see Baker 2002, for details). Thus, Eq. (18.10) can be interpreted as a log-linearized version of B_t. The technique proposed for the evolution of B_t in Eq. (18.10) departs from Primiceri (2005) who models each element of the inverse of B_t and therefore misses a coherent system theoretic approach. See Neusser (2016) for details.

Although this TVC-VAR model with time-varying volatility can in principle also be estimated by maximum likelihood, this technique can hardly be implemented successfully in practice. The main reason is that the likelihood function of such a model, even when the dimension and the order of the VAR is low, is a very high dimensional nonlinear object with probably many local maxima. Moreover, as the variances governing the time-variation are small, at least for some of the coefficients, the likelihood function is flat in some regions of the parameter space. These features make maximization of the likelihood function a very difficult, if not impossible, task in practice. For these reasons, Bayesian techniques have been used almost exclusively. There is, however, also a conceptional issue involved. As the Bayesian approach does not strictly distinguish between fixed "true" parameters and random samples, it is better suited to handle TVC-VAR models which treat the parameters as random. In this monograph, we will not tackle the Bayesian approach but refer to the relevant literature. See for example Primiceri (2005), Negro and Primiceri (2015), Cogley and Sargent (2005), Canova (2007) and Koop and Korobilis (2009) among others.

The Minnesota Prior

Although, we will not discuss the Bayesian approach to VAR modeling, it is nevertheless instructive to portray the so-called Minnesota prior applied by Doan et al. (1984) to TVC-VAR models. This prior has gained some reputation in connection to forecasting with VAR models and as a way to specify the initial distribution for the Kalman filter in time-varying models. The combination of the prior distribution with the likelihood function delivers via Bayes' rule a posterior distribution of the parameters which can then be analyzed using simulation methods.

The Minnesota prior is based on the a priori belief that each variable follows a random walk with no interaction among the variables nor among the coefficients of the VAR equations. We expose one version of the Minnesota prior in the general context of a TVC-VAR model of order p with time-varying constant term c_t:

$$X_t = c_t + \Phi_{t-1}^{(1)} X_{t-1} + \ldots + \Phi_{t-1}^{(p)} X_{t-p} + Z_t. \tag{18.11}$$

This model can be written compactly as

$$X_t = (\mathbf{X}_{t-1}' \otimes I_n) \operatorname{vec} \Phi_{t-1} + Z_t = (\mathbf{X}_{t-1}' \otimes I_n)\beta_{t-1} + Z_t \tag{18.12}$$

where $\mathbf{X}_{t-1} = (1, X'_{t-1}, \ldots, X'_{t-p})'$, $\Phi_{t-1} = (c_t, \Phi^{(1)}_{t-1}, \ldots, \Phi^{(p)}_{t-1})$, and $\beta_t = \text{vec}\,\Phi_t$. Assuming for β_t the same autoregressive form as in Eq. (18.5), the state space representation (18.6) and (18.7) also applies to the TVC-VAR(p) model with X_{t-1} replaced by $\mathbf{X}_{t-1}$. Note that the dimension of the state equation can become very high because β_t is a $n + n^2p$ vector.

Taking date 0 as the initial date, the prior distribution of the autoregressive parameters is supposed to be normal:

$$\beta_0 = \text{vec}\,\Phi_0 \sim \text{N}(\bar{\beta}, P_{0|0})$$

where $\bar{\beta} = \text{vec}(0, I_n, 0, \ldots, 0)$. This implies that the mean for all coefficients, including the constant term, is assumed to be zero except for the own lag coefficients of order one $[\Phi^{(1)}_0]_{ii}$, $i = 1, \ldots, n$, which are assumed to be one. The covariance matrix $P_{0|0}$ is taken as being diagonal so that there is no correlation across coefficients. Thus, the prior specification amounts to assuming that each variable follows a random walk with no interaction with other variables.

The strength of this belief is governed by a number of so-called hyperparameters which regulate the diagonal elements of $P_{0|0}$. The first one, γ^2, controls the confidence placed on the assumption that $[\Phi^{(1)}_0]_{ii} = 1$:

$$\left[\Phi^{(1)}_0\right]_{ii} \sim \text{N}(1, \gamma^2), \qquad i = 1, 2, \ldots, n.$$

A small (large) value of γ^2 thus means more (less) confidence. As the lag order increases more confidence is placed on the assumption $[\Phi^{(h)}_0]_{ii} = 0$:

$$\left[\Phi^{(h)}_0\right]_{ii} \sim \text{N}\left(0, \frac{\gamma^2}{h}\right), \qquad h = 2, \ldots, p \text{ and } i = 1, \ldots, n$$

Instead of the harmonic decline other schemes have been proposed. For $h = 1, \ldots, p$ the off-diagonal elements of $\Phi^{(h)}_0$ are assumed to have prior distribution

$$\left[\Phi^{(h)}_0\right]_{ij} \sim \text{N}\left(0, \frac{w^2\gamma^2\hat{\tau}_i^2}{h\hat{\tau}_j^2}\right), \qquad i, j = 1, \ldots, n, i \neq j, h = 1, 2, \ldots, p.$$

Thereby $\hat{\tau}_i^2/\hat{\tau}_j^2$ represents a correction factor which accounts for the magnitudes of X_{it} relative to X_{jt}. Specifically, $\hat{\tau}_i^2$ is the residual variance of a univariate AR(1) model. The hyperparameter w^2 is assumed to be strictly smaller than one. This represents the belief that $X_{j,t-h}$ is less likely to be important as an explanation for $X_{i,t}$, $i \neq j$, than the own lag $X_{i,t-h}$. Finally, the strength of the belief that the constant terms are zero is

$$c_{i0} = N(0, g\hat{\tau}_i).$$

This completes the specification for the prior belief on β_0. Combining all elements we can write $P_{0|0}$ as a block diagonal matrix with diagonal blocks:

$$P_{0|0} = \begin{pmatrix} P_{0|0}^{(c)} & 0 \\ 0 & P_{0|0}^{(\phi)} \end{pmatrix}$$

where $P_{0|0}^{(c)} = g \times \text{diag}(\hat{\tau}_1, \ldots, \hat{\tau}_n)$ and $P_{0|0}^{(\phi)} = \text{diag}(\text{vec}(G \otimes \Upsilon))$. Thereby, G and Υ are defined as

$$G = (\gamma^2, \gamma^2/2, \ldots, \gamma^2/p)$$

$$[\Upsilon]_{ij} = \begin{cases} 1, & i = j; \\ w^2(\hat{\tau}_i^2/\hat{\tau}_j^2), & i \neq j. \end{cases}$$

According to Doan et al. (1984) the preferred values for the three hyperparameters are $g = 700$, $\gamma^2 = 0.07$, and $w^2 = 0.01$.

Thus, for a bivariate TVC-VAR(2) model the mean vector is given by $\bar{\beta} = (0, 0, 1, 0, 0, 1, 0, 0, 0, 0)'$ with diagonal covariance matrix $P_{0|0}$:

$$P_{0|0} = \begin{pmatrix} P_{0|0}^{(c)} & 0 & 0 \\ 0 & P_{0|0}^{(1)} & 0 \\ 0 & 0 & P_{0|0}^{(2)} \end{pmatrix}$$

with

$$P_{0|0}^{(c)} = g \begin{pmatrix} \hat{\tau}_1^2 & 0 \\ 0 & \hat{\tau}_2^2 \end{pmatrix},$$

$$P_{0|0}^{(1)} = \gamma^2 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & w^2\hat{\tau}_2^2/\hat{\tau}_1^2 & 0 & 0 \\ 0 & 0 & w^2\hat{\tau}_1^2/\hat{\tau}_2^2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$P_{0|0}^{(2)} = \frac{\gamma^2}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & w^2\hat{\tau}_2^2/\hat{\tau}_1^2 & 0 & 0 \\ 0 & 0 & w^2\hat{\tau}_1^2/\hat{\tau}_2^2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Next we specify the parameters of the state transition equation (18.5). Following Doan et al. (1984), $F = \pi_F I_{n+pn^2}$ with $\pi_F = 0.999$ and $Q = \pi_Q P_{0|0}$ with $\pi_Q = 10^{-7}$. The proportionality factor does, however, not apply to the constant terms. For these terms, the corresponding diagonal elements of Q, $[Q]_{ii}$, $i = 1, \ldots, n$, are set to $\pi_Q [P_{0|0}]_{i(n+1),i(n+1)}$, $i = 1, \ldots, n$. The reason for this correction is that the prior put on the constants is rather loose as expressed by the high value of g. The final component is a specification for Σ, the variance of Z_t. This matrix is believed to be diagonal with $\Sigma = \pi_\Sigma \text{diag}(\hat{\tau}_1^2, \ldots, \hat{\tau}_n^2)$ and $\pi_\Sigma = 0.9$.

With these ingredients the state space model is completely specified. Given observations $\mathbf{X}_1, \ldots, \mathbf{X}_t$, the Kalman filter produces a sequence of $\beta_{t+1|t}$, $t = 1, 2, \ldots$ and one-period ahead forecasts $X_{t+1|t}$ computed as

$$X_{t+1|t} = (\mathbf{X}_t' \otimes I_n)\beta_{t+1|t}.$$

Doan et al. (1984) suggest to compute an *approximate* h period ahead forecast by treating the forecast from the previous periods as if they were actual observations.

18.3 Regime Switching Models

The regime switching model is similar to the time-varying model discussed in the previous section. The difference is that the time-varying parameters are governed by a hidden Markov chain with a finite state space $\mathbb{S} = \{1, 2, \ldots, k\}$. Usually, the number of states k is small and is equal in practice to two or maximal three. The states have usually an economic connotation. For example, if k equals two, state 1 might correspond to a boom phase whereas state 2 to a recession. Such models have a long tradition in economics and have therefore been used extensively. Seminal references include Goldfeld and Quandt (1973, 1976), Hamilton (1994b), Kim and Nelson (1999), Krolzig (1997), and Maddala (1986). Frühwirt-Schnatter (2006) presents a detailed statistical analysis of regime switching models.

The starting point of our presentation of the regime switching model is again the TVC-VAR(1) as given in Eq. (18.3). We associate to each state $j \in \mathbb{S}$ a coefficient matrix $\Phi^{(j)}$. Thus, in the regime switching model the coefficients Φ_t can only assume a finite number values $\Phi^{(1)}, \ldots, \Phi^{(k)}$ depending on the state of the Markov chain. The actual value assigned to Φ_t is governed by a Markov chain defined through a fixed but unknown transition probability matrix P where

$$[P]_{ij} = \mathbf{P}\left(\Phi_t = \Phi^{(j)} | \Phi_{t-1} = \Phi^{(i)}\right) \qquad i, j = 1, \ldots, k. \tag{18.13}$$

Thus, $[P]_{ij}$ is the probability that Φ_t assumes value $\Phi^{(j)}$ given that it assumed in the previous period the value $\Phi^{(i)}$. The probability that Φ_{t+h} is in state j given that Φ_t was in state i is therefore $[P^h]_{ij}$. The definition of the transition matrix in Eq. (18.13) implies that P is a stochastic matrix, i.e. that $[P]_{ij} \geq 0$ and $\sum_{j=1}^{k} [P]_{ij} = 1$. Moreover, we assume that the chain is regular meaning that it is ergodic (irreducible)

and aperiodic.[10] This is equivalent to the existence of a fixed integer $m > 0$ such that P^m has only strictly positive entries (see Berman and Plemmons 1994, Chapter 8). Regular Markov chains have a unique ergodic (stationary) distribution vector π with strictly positive entries and determined by $\pi' P = \pi'$. This distribution is approached from any initial distribution vector π_0, i.e. $\lim_{t\to\infty} \pi_0' P^t = \pi'$. Moreover, $\lim_{t\to\infty} P^t = P^\infty$ where P^∞ is a transition matrix with all rows equal to π'.

Given this setup we can again invoke Theorem 18.1 and claim that a (strictly) stationary solution of the form of Eq. (18.4) exists if all the autoregressive matrices $\Phi^{(j)}, j = 1, 2, \ldots, k$, have eigenvalues strictly smaller than one.

Given observations $x_T, x_{T-1}, \ldots, x_1, x_0$ for $X_T, X_{T-1}, \ldots, X_1, X_0$, a maximum likelihood approach can be set up to estimate the unknown parameters $\Phi^{(1)}, \ldots, \Phi^{(k)}, \Sigma, P$.[11] Collect these parameters into a vector θ and denote by $s_t \in \mathbb{S}$ the state of the Markov chain in period t and by $\mathcal{X}_t = (x_t, x_{t-1}, \ldots, x_1, x_0)$ the information available up to period t. Write the conditional density of x_t given $s_t = j$ and observations $\mathcal{X}_{t-1}$ as

$$f(x_t | s_t = j, \mathcal{X}_{t-1}; \theta).$$

The joint density of $(x_t, s_t = j)$ is

$$f(x_t, s_t = j | \mathcal{X}_{t-1}; \theta) = f(x_t | s_t = j, \mathcal{X}_{t-1}; \theta) \times P(s_t = j | \mathcal{X}_{t-1}; \theta)$$

where in analogy to the Kalman filter the expressions $P(s_t = j | \mathcal{X}_{t-1}; \theta)$, $j = 1, \ldots, k$, are called the predicted transition probabilities. The conditional marginal density of x_t then becomes

$$f(x_t | \mathcal{X}_{t-1}; \theta) = \sum_{j=1}^{k} f(x_t | s_t = j, \mathcal{X}_{t-1}; \theta) \times P(s_t = j | \mathcal{X}_{t-1}; \theta).$$

In the case of $Z_t \sim IID\, N(0, \Sigma)$ the above density is a finite mixture of Gaussian distributions (see Frühwirt-Schnatter 2006, for details). The (conditional) log likelihood function, finally, is therefore given by

$$\ell(\theta) = \sum_{t=1}^{T} \log f(x_t | \mathcal{X}_{t-1}; \theta).$$

[10] A chain is called ergodic or irreducible if for every states i and j there is a strictly positive probability that the chain moves from state i to state j in finitely many steps. A chain is called aperiodic if it can return to any state i at irregular times. See, among others, Norris (1998) and Berman and Plemmons (1994) for an introduction to Markov chains and its terminology.

[11] The presentation of the maximum likelihood approach follows closely the exposition by Hamilton (1994b, chapter 22) where more details can be found.

In order to evaluate the likelihood function note that the joint density of $(x_t, s_t = j)$ may also be factored as

$$f(x_t, s_t = j|\mathcal{X}_{t-1}; \theta) = P(s_t = j|\mathcal{X}_t; \theta) \times f(x_t|\mathcal{X}_{t-1}; \theta).$$

Combining these expressions one obtains an expression for the filtered transition probabilities $P(s_t = j|\mathcal{X}_t; \theta)$:

$$\begin{aligned} P(s_t = j|\mathcal{X}_t; \theta) &= \frac{f(x_t|s_t = j, \mathcal{X}_{t-1}; \theta) \times P(s_t = j|\mathcal{X}_{t-1}; \theta)}{f(x_t|\mathcal{X}_{t-1}; \theta)} \\ &= \frac{f(x_t|s_t = j, \mathcal{X}_{t-1}; \theta) \times P(s_t = j|\mathcal{X}_{t-1}; \theta)}{\sum_{j=1}^{k} f(x_t|s_t = j, \mathcal{X}_{t-1}; \theta) \times P(s_t = j|\mathcal{X}_{t-1}; \theta)} \end{aligned} \tag{18.14}$$

Next period's predicted transition probabilities are then obtained by multiplication with the transition matrix:

$$\begin{pmatrix} P(s_{t+1} = 1|\mathcal{X}_t; \theta) \\ \vdots \\ P(s_{t+1} = k|\mathcal{X}_t; \theta) \end{pmatrix} = P' \times \begin{pmatrix} P(s_t = 1|\mathcal{X}_t; \theta) \\ \vdots \\ P(s_t = k|\mathcal{X}_t; \theta) \end{pmatrix} \tag{18.15}$$

Given initial probabilities $P(s_1 = j|\mathcal{X}_0; \theta)$, $j = 1, \ldots, k$, and a fixed value for θ, Eqs. (18.14) and (18.15) can be iterated forward to produce a sequence of predicted transition probabilities $(P(s_t = 1|\mathcal{X}_{t-1}; \theta), \ldots, P(s_t = k|\mathcal{X}_{t-1}; \theta))'$, $t = 1, 2, \ldots, T$ which can be used to evaluate the Gaussian likelihood function. Numerical procedures must then be used for the maximization of the likelihood function. This task is not without challenge because the likelihood function of Gaussian mixture models typically has singularities and many local maxima. Kiefer (1978) showed that there exists a bounded local maximum which yields a consistent and asymptotically normal estimate for θ for which standard errors can be constructed in the usual way. In practice, problems encountered during the maximization can be alleviated by experimentation with alternative starting values. Thereby the initial probability $(P(s_1 = 1|\mathcal{X}_0; \theta), \ldots, P(s_1 = k|\mathcal{X}_0; \theta)$ could either be treated as additional parameters as in Goldfeld and Quandt (1973) or set to the uniform distribution. For technical details and alternative estimation strategies, like the EM algorithm, see Hamilton (1994b, chapter 22) and in particular Frühwirt-Schnatter (2006).

By reversing the above recursion it is possible to compute smoothed transition probabilities $P(s_t = j|\mathcal{X}_T; \theta)$ (see Kim 1994):

$$P(s_t = j|\mathcal{X}_T; \theta) = P(s_t = j|\mathcal{X}_t; \theta) \sum_{i=1}^{k} [P]_{ij} \frac{P(s_{t+1} = i|\mathcal{X}_T; \theta)}{P(s_{t+1} = i|\mathcal{X}_t; \theta)}$$

The iteration is initialized with $P(s_T = j|\mathcal{X}_T; \theta)$ which has been computed in the forward recursion.

The basic model can and has been generalized in several dimensions. The most obvious one is the inclusion of additional lags beyond the first one. The second one concerns the possibility of a regime switching covariance matrix Σ. These modifications can be accommodated using the methods outlined above. Thirdly, one may envision time-varying transition probabilities to account for duration dependence. In business cycle analysis, for example, the probability of moving out of a recession may depend on how long the economy has been in the recession regime. This idea can be implemented by modeling the transition probabilities via a logit specification:

$$[P]_{ij} = \frac{\exp(z_t'\alpha_i)}{1 + \exp(z_t'\alpha_i)} \qquad i \neq j$$

where z_t includes a constant and a set of additional variables. These additional variables can be some exogenous variables, but more interestingly may include some lagged variables x_{t-d} (Krolzig 1997). Note that the transition probabilities do not only depend on z_t, but also on the state. The resulting model has some features shared with the smooth transition autoregressive model of Granger and Teräsvirta (1993). Early economic applications of regime switching models with time-varying transition probabilities can be found in Diebold et al. (1994), Filardo (1994), and Filardo and F.Gordon (1998).

An important aspect in practice is the determination of the number of regimes. Unfortunately, there is no direct test available for the null hypothesis $k = m$ against the alternative $k = m + 1$. The reason is that the likelihood contains parameters which are only present under the alternative. The parameters describing the $m + 1$-th state are unidentified under the null hypothesis. The problem has been analyzed by Andrews and Ploberger (1994) in a general theoretical context. Alternatively, one may estimate the model under the null hypothesis and conduct a series of specification tests as proposed by Hamilton (1996). It has also been suggested to use the information criteria like AIC and BIC to determine the number of regimes (Frühwirt-Schnatter 2006, p. 346–347):

$$\text{AIC} = -2\ell(\theta) + 2k(k-1)$$
$$\text{BIC} = -2\ell(\theta) + \log(T)k(k-1)$$

where $k(k-1)$ are the free parameters in the transition matrix P.