14 Identification by Heteroskedasticity or Non-Gaussianity

14.1 Introduction

As we have seen in Chapters 8 and 10, the identification of structural VAR models typically relies on economically motivated identifying restrictions. Another strand of the literature exploits certain statistical properties of the data for identification. In particular, changes in the conditional or unconditional volatility of the VAR errors (and hence of the observed variables) can be used to assist in the identification of structural shocks. For example, Rigobon (2003), Rigobon and Sack (2003), and Lanne and Lütkepohl (2008) rely on unconditional heteroskedasticity, whereas Normandin and Phaneuf (2004), Bouakez and Normandin (2010), and Lanne, Lütkepohl, and Maciejowska (2010) exploit conditional heteroskedasticity.

In this chapter, we explain the principle of identification by heteroskedasticity.[1] In Section 14.2, the general modeling strategy is presented and its advantages and limitations are discussed. The central idea is that in a conventional structural VAR analysis the structural shocks are recovered by transforming the reduced-form residuals. As we have seen in previous chapters, this is typically done through exclusion restrictions. The current chapter considers the question of how changes in the volatility of the model errors can be used for this purpose. We show that the assumption that the structural impulse responses are time invariant, as the volatility of the reduced-form shocks changes, provides additional restrictions that can be used to uniquely pin down mutually uncorrelated shocks. There is nothing in this purely statistical identification procedure, however, that ensures that these shocks are also economically meaningful, making it difficult to interpret them as structural VAR shocks.

One way to assess whether all or some of the shocks identified by heteroskedasticity correspond to economic shocks and, hence, can be interpreted

[1] This chapter is partly adapted from Lütkepohl (2013b). Other papers from which we draw are Lütkepohl and Velinov (2016) and Lütkepohl and Netšunajev (2017).

as proper structural shocks is to treat conventional identifying restrictions as overidentifying restrictions within the heteroskedastic model, facilitating formal tests of these restrictions.

In some cases it may also be possible to infer the economic interpretation of these shocks informally from comparisons of the implied impulse responses with impulse response estimates based on conventional structural VAR models, as illustrated in Lütkepohl and Netšunajev (2014). A necessary condition for such comparisons is that the structural impulse responses in the VAR model based on conventional identifying restrictions can be estimated consistently under the assumptions maintained in the explicitly heteroskedastic VAR model.

In Section 14.3, we consider some specific models of changes in volatility. First, models of heteroskedasticity with extraneously generated shifts in the variance are discussed. We then introduce models in which the timing of the volatility change is determined by the data. One example is models in which volatility changes are governed by a Markov regime-switching mechanism. Another example is the smooth-transition model for capturing changes in the error volatility. Third, we examine a model with multivariate generalized autoregressive conditionally heteroskedastic (GARCH) errors. Detailed examples are provided for each of these models.

Throughout Section 14.3, it is assumed that the impact effects of the structural shocks remain constant, as the volatility of the structural shocks changes. In contrast, in Section 14.4 we allow for time-varying impact effects and consider models that permit the primitive shocks of the model to be mutually correlated. Finally, in Section 14.5 we examine an alternative identification strategy that exploits the non-Gaussianity of the VAR model errors in many applications. Section 14.6 summarizes the pros and cons of each approach.

14.2 The Model Setup

14.2.1 The Baseline Model

The point of departure is a K-dimensional reduced-form VAR(p) process,

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t. \tag{14.2.1}$$

As usual, A_j, $j = 1, \ldots, p$, are $K \times K$ VAR coefficient matrices and u_t is a serially uncorrelated zero-mean error term subject to conditional or unconditional heteroskedasticity. The model may contain cointegrated variables and may in fact be set up as a vector error correction model.[2] In addition, the VAR

[2] In the latter case, one may estimate the cointegration relations in a first step without accounting for heteroskedasticity and fix the cointegration parameters at those estimates in the subsequent analysis.

model may include additional unmodeled right-hand side variables. As long as time-invariance of all parameters apart from those in the error covariance structure can be justified, this extension is straightforward.

In earlier chapters, the structural shocks, w_t, were obtained from the reduced-form errors by a linear transformation, $w_t = B_0 u_t$ or, equivalently, $B_0^{-1} w_t = u_t$. The matrix B_0 was chosen such that the components of w_t are instantaneously uncorrelated. Thus, depending on the normalization assumptions,

$$\Sigma_u = B_0^{-1} B_0^{-1\prime} \quad \text{or} \quad \Sigma_u = B_0^{-1} \Sigma_w B_0^{-1\prime},$$

where $\mathbb{E}(u_t u_t') = \Sigma_u$ was assumed to be time-invariant. The identification of the structural shocks relied on normalizing assumptions and additional economically motivated identifying restrictions on the elements of B_0 or B_0^{-1}.

An alternative approach to identification becomes feasible if the unconditional variances of the reduced-form shocks (and, hence, also the unconditional variances of the structural shocks) change during the sample period. To see this suppose that $\mathbb{E}(u_t u_t') = \Sigma_1$ for $t = 1, \dots, T_1$, and $\mathbb{E}(u_t u_t') = \Sigma_2$ for $t > T_1$, where $\Sigma_1 \neq \Sigma_2$. Moreover, suppose that all other VAR parameters in (14.2.1) remain time-invariant. Then there is a matrix decomposition result that ensures the existence of a matrix G and a diagonal matrix $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_K)$ such that

$$\Sigma_1 = GG' \quad \text{and} \quad \Sigma_2 = G\Lambda G' \tag{14.2.2}$$

(see Lütkepohl 1996, chapter 6). This G matrix may be viewed as a solution for the structural impact multiplier matrix B_0^{-1}, which implies that the structural shocks $w_t = G^{-1} u_t$ have variance

$$\mathbb{E}(w_t w_t') = \begin{cases} I_K, & t = 1, \dots, T_1, \\ \Lambda, & t > T_1. \end{cases}$$

Since Λ is diagonal, the structural errors obtained in this way satisfy the basic requirement of being instantaneously uncorrelated across the whole sample. Lanne, Lütkepohl, and Maciejowska (2010) show that the matrix B_0^{-1} obtained from this decomposition is unique apart from changes in the signs and permutations of the columns, if the diagonal elements of Λ are all distinct. If the latter condition is satisfied, we may obtain unique shocks by just imposing the basic requirement that the structural shocks are instantaneously uncorrelated. The possible changes in sign just mean that we consider negative instead of positive shocks and vice versa. The fact that the columns of B_0^{-1} can be permuted means that the ordering of the shocks can be chosen freely. Of course, using the same transformation matrix B_0^{-1} for the whole sample period implies that the impact effects of the shocks are time-invariant as well and only the variances change.

14.2.2 An Illustrative Example

It is useful to review an illustrative example in some detail. Consider the bivariate system

$$u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \begin{bmatrix} b_0^{11} & b_0^{12} \\ b_0^{21} & b_0^{22} \end{bmatrix} \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix},$$

where b_0^{ij} denotes the ij^{th} element of B_0^{-1}. Substituting $G = B_0^{-1}$ into the relations (14.2.2) yields

$$\begin{bmatrix} \sigma_{1,1}^2 & \sigma_{12,1} \\ \sigma_{12,1} & \sigma_{2,1}^2 \end{bmatrix} = \begin{bmatrix} (b_0^{11})^2 + (b_0^{12})^2 & b_0^{11}b_0^{21} + b_0^{12}b_0^{22} \\ b_0^{11}b_0^{21} + b_0^{12}b_0^{22} & (b_0^{21})^2 + (b_0^{22})^2 \end{bmatrix}$$

and

$$\begin{bmatrix} \sigma_{1,2}^2 & \sigma_{12,2} \\ \sigma_{12,2} & \sigma_{2,2}^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \left(b_0^{11}\right)^2 + \lambda_2 \left(b_0^{12}\right)^2 & \lambda_1 b_0^{11} b_0^{21} + \lambda_2 b_0^{12} b_0^{22} \\ \lambda_1 b_0^{11} b_0^{21} + \lambda_2 b_0^{12} b_0^{22} & \lambda_1 \left(b_0^{21}\right)^2 + \lambda_2 \left(b_0^{22}\right)^2 \end{bmatrix}.$$

Thus, we have six distinct equations

$$\begin{aligned}
\sigma_{1,1}^2 &= \left(b_0^{11}\right)^2 + \left(b_0^{12}\right)^2, \\
\sigma_{12,1} &= b_0^{11} b_0^{21} + b_0^{12} b_0^{22}, \\
\sigma_{2,1}^2 &= \left(b_0^{21}\right)^2 + \left(b_0^{22}\right)^2, \\
\sigma_{1,2}^2 &= \lambda_1 \left(b_0^{11}\right)^2 + \lambda_2 \left(b_0^{12}\right)^2, \\
\sigma_{12,2} &= \lambda_1 b_0^{11} b_0^{21} + \lambda_2 b_0^{12} b_0^{22}, \\
\sigma_{2,2}^2 &= \lambda_1 \left(b_0^{21}\right)^2 + \lambda_2 \left(b_0^{22}\right)^2,
\end{aligned}$$

from which we can solve for the six structural parameters $b_0^{11}, b_0^{12}, b_0^{21}, b_0^{22}, \lambda_1, \lambda_2$. The last two parameters are strictly greater than zero. They are the diagonal elements of Λ. The solution is unique (up to sign) if λ_1 and λ_2 are distinct and are ordered such as $\lambda_1 < \lambda_2$. Note that the variances of the structural shocks are normalized to one in the first part of the sample. Hence, the λ_i indicate the changes in the variances from the first to the second volatility regime. In other words, they can be interpreted as the relative variances in the second part of the sample. The condition $\lambda_1 \neq \lambda_2$ which ensures uniqueness just means that the change in variance is not the same for both variables. In fact, for a unique solution it suffices that the variance of one of the variables changes.

It should be clear, however, that this is a purely statistical approach to uniquely identifying mutually uncorrelated shocks that does not necessarily result in economically meaningful shocks, much like orthogonalizing the reduced-form errors using a Cholesky decomposition does not automatically lead to economically meaningful shocks. A case in point is the bivariate demand and supply model considered in Rigobon (2003). Rigobon exploits heteroskedasticity to identify a unique set of mutually uncorrelated shocks. Without further economic identifying assumptions, however, we cannot infer which shock, if any, corresponds to the demand shock and which to the supply shock. Thus, the estimates are not structural in the sense discussed in Chapters 8 and 10.

If the statistical identification conditions for the shocks in the heteroskedastic model are satisfied, all we can say is that there are two mutually uncorrelated shocks that induce time-invariant impact responses (and hence more generally time-invariant impulse response functions) throughout the sample. The assumption of time-invariant impact responses is, of course, also maintained in conventional structural VAR analysis with unmodeled heteroskedasticity. While this assumption may be unrealistic, as discussed in Chapter 18, it is required for the methods discussed in the current chapter.

Although this illustrative example involved only one volatility shift, the idea of identification by heteroskedasticity generalizes easily to more than two volatility regimes. In fact, the volatility may change in every period.

14.2.3 The General Model

Identification by heteroskedasticity was originally proposed in the context of shifts in the unconditional variance, but may also be extended to shifts in the conditional variance as illustrated in Sections 14.3.2 and 14.3.4. In general,

$$\mathbb{E}(u_t u_t') = \Sigma_t = B_0^{-1}\Lambda_t B_0^{-1\prime} \quad \text{or} \quad \mathbb{E}_t(u_t u_t') = \Sigma_t = B_0^{-1}\Lambda_t B_0^{-1\prime} \tag{14.2.3}$$

with diagonal matrices $\Lambda_t = \operatorname{diag}(\lambda_{1t}, \dots, \lambda_{Kt})$, where the Σ_t may alternatively denote the unconditional covariance matrix or the conditional covariance matrix. Of course, a decomposition of Σ_t as in (14.2.3) may not exist. In particular, there may not exist a time-invariant matrix B_0^{-1} such that $\Sigma_t = B_0^{-1}\Lambda_t B_0^{-1\prime}$. In fact, the existence of such a decomposition imposes testable restrictions on Σ_t, allowing us to test whether the data are compatible with this decomposition. In the latter case, we may use B_0^{-1} to transform the reduced-form errors into structural errors with time-invariant impact effects.

In order to estimate and analyze the volatility structure we first need to parameterize and estimate the time-varying diagonal covariance matrices. In Section 14.3 we discuss a number of volatility models that have been used for this purpose in the literature. Some of these models assume a finite number of volatility states, whereas others allow for infinitely many volatility states. In each case we discuss the identification conditions required for the uniqueness of the transformation matrix B_0^{-1} and, hence, of the structural shocks.

If unique structural shocks are obtained via heteroskedasticity, then any further restrictions on B_0^{-1} become overidentifying. In particular, restrictions on the impact or the long-run effects imposed in conventional SVAR analyses are overidentifying and, hence, can be tested against the data. This implication of the present framework is convenient if there are additional competing identifying restrictions such as the exclusion restrictions commonly used in the homoskedastic model. In this case the data can potentially speak against such restrictions or confirm that they are compatible with the data. Of course, it can happen that even this framework does not allow us to discriminate between competing economic theories. In other words, two competing structural forms may both be in line with the data in the present framework. Moreover, rejecting a particular model in the present framework could be the result of the underlying assumptions of the statistical model not being compatible with the data. For example, the assumption of regime-invariant impact effects may be problematic. In that case a particular set of exclusion restrictions may be rejected, although it is actually the modeling framework that is inappropriate.

14.3 Alternative Volatility Models

This section discusses a range of different proposals for modeling changing volatility in this setup, where $\Sigma_t = B_0^{-1}\Lambda_t B_0^{-1\prime}$ with B_0^{-1} being the matrix of impact effects of the structural shocks and Λ_t being a diagonal matrix.

14.3.1 Structural VAR Models with Extraneously Specified Volatility Changes

General Setup. In the original article on identifying mutually uncorrelated VAR shocks by heteroskedasticity, Rigobon (2003) considers changes in the unconditional variances of the shocks at discrete points in time. In the baseline model the date of the change in volatility is deterministic and known. Suppose that the reduced-form error covariance matrix in time period t is

$$\mathbb{E}(u_t u_t') = \begin{cases} \Sigma_1 & \text{for} \quad t \in \mathcal{T}_1, \\ \vdots & \\ \Sigma_M & \text{for} \quad t \in \mathcal{T}_M, \end{cases} \tag{14.3.1}$$

where $\mathcal{T}_m = \{T_{m-1}+1, \dots, T_m\}$, $m = 1, \dots, M$, and there are M mutually exclusive volatility regimes. T_m, for $m > 0$, denotes the period of a change in the volatility regime. It is assumed that $T_0 = 0$ and $T_M = T$. At least two of the error covariance matrices, Σ_m, are assumed to be distinct.

Further suppose that we can decompose the error covariance matrices according to

$$\Sigma_1 = B_0^{-1}B_0^{-1\prime}, \quad \Sigma_m = B_0^{-1}\Lambda_m B_0^{-1\prime}, \quad m = 2, \dots, M, \tag{14.3.2}$$

with diagonal matrices $\Lambda_m = \text{diag}(\lambda_{1m}, \dots, \lambda_{Km})$. Clearly, such a decomposition may not exist if $M > 2$ and arbitrary covariance matrices Σ_m, $m = 1, \dots, M$, are allowed for. Since the decomposition imposes restrictions on the reduced form, however, its validity can be tested against the data as we will see shortly.

Identification. In the context of this model, the shocks w_t are statistically identified if B_0^{-1} is statistically identified. Lanne, Lütkepohl, and Maciejowska (2010, Proposition 1) show that uniqueness of B_0^{-1} (apart from column ordering and column sign changes) is ensured if for any two subscripts $k, l \in \{1, \dots, K\}$, $k \neq l$, there exists an $m \in \{2, \dots, M\}$ such that $\lambda_{km} \neq \lambda_{lm}$. In other words, there must be at least one regime, in which the change in volatility in variable k is different from that in variable l, relative to what it is in regime 1. Although this condition for exact (local) identification is more complicated than in the 2-state case, it is easy to check. Moreover, the identification condition can in principle be examined with statistical tests because, if there are M distinct volatility regimes, the diagonal elements of the Λ_m matrices are identified and, hence, can be estimated consistently with a proper asymptotic distribution under common assumptions. We discuss this issue in more detail later in the chapter.

So far we have assumed that only the restrictions from the covariance decomposition are imposed on B_0^{-1}. In practice, sometimes the elements on the main diagonal of B_0 are also restricted to be one. In that case the variances of the structural shocks in the first volatility regime cannot be normalized to one, and we consider a decomposition of the reduced-form covariance matrices of the form

$$\Sigma_m = B_0^{-1} \Lambda_m^* B_0^{-1\prime}, \quad m = 1, \dots, M, \tag{14.3.3}$$

where the $\Lambda_m^* = \text{diag}(\lambda_{1m}^*, \dots, \lambda_{Km}^*)$ are diagonal matrices. It can be shown that for $M \geq 2$ the matrix B_0 is locally unique, if the following condition is satisfied:

$$\begin{aligned} &\forall k, l \in \{1, \dots, K\} \text{ with } k \neq l, \exists\, m \in \{2, \dots, M\} \\ &\text{such that } \lambda_{km}/\lambda_{k1} \neq \lambda_{lm}/\lambda_{l1}. \end{aligned} \tag{14.3.4}$$

This condition ensures local identification of B_0. If this matrix has no further unit elements apart from those on its main diagonal, column or row permutations are not possible. In that case column or row permutations are not possible, while maintaining the unit diagonal, ensuring the uniqueness of the solution. Enforcing a unit diagonal of B_0 also rules out sign changes of the shocks. Note, however, that a unit diagonal of B_0 does not ensure a unit diagonal of the matrix of impact effects B_0^{-1}. Only the signs of the impact effects are fixed by requiring B_0 to have a unit main diagonal. Moreover, the fact that $w_t = B_0 u_t$ implies that the structural-form covariance matrices

$$\mathbb{E}(w_t w_t') = \Lambda_m^* \quad \text{for} \quad t \in \mathcal{T}_m,$$

are diagonal, as required by the assumption of instantaneously uncorrelated structural shocks.

One requirement for condition (14.3.4) is that the variances of the model variables do not change proportionately. For example, if there are just two volatility states and all variances change proportionally such that for some scalar c, $\Lambda_1 = c\Lambda_2$, then $\lambda_{k2}/\lambda_{k1} = \lambda_{l2}/\lambda_{l1}$, so that condition (14.3.4) is not satisfied.

It is important to reiterate that the identification conditions for B_0 given so far are purely statistical conditions that allow consistent estimation of all the elements of B_0 without imposing any further restrictions on the model. Hence, the implied shocks w_t may not have an economic interpretation. Still, the statistical uniqueness is useful because it may enable us to perform statistical tests of additional economic identifying restrictions conditional on the heteroskedastic model being correctly specified.

Estimation and Inference. Assuming that $u_t \sim \mathcal{N}(0, \Sigma_t)$ with a covariance structure that satisfies (14.3.1) and (14.3.2), the model may be estimated by ML. The log-likelihood function is

$$\log l(\boldsymbol{\alpha}, \boldsymbol{\sigma}) = -\frac{KT}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log(\det(\Sigma_t)) - \frac{1}{2}\sum_{t=1}^{T} u_t'\Sigma_t^{-1}u_t, \tag{14.3.5}$$

where $\boldsymbol{\sigma}$ contains all unknown covariance parameters and the time-invariant VAR parameters including deterministic terms are collected in $\boldsymbol{\alpha} = \text{vec}[\nu, A_1, \ldots, A_p]$ (see Lütkepohl (2005, chapter 17) for further details). The log-likelihood (14.3.5) can also be used for quasi-ML estimation if u_t (and hence y_t) is not normally distributed. Alternatively one may estimate the VAR model by equation-by-equation LS and use the residuals, $\hat{u}_t$, for estimating the covariance matrices as

$$\widehat{\Sigma}_m = \frac{1}{\#\mathcal{T}_m}\sum_{t \in \mathcal{T}_m} \hat{u}_t \hat{u}_t',$$

where $\#\mathcal{T}_m$ denotes the number of observations in $\mathcal{T}_m$. In the next step these estimates can then be used to form the GLS estimator

$$\widehat{\boldsymbol{\alpha}} = \left(\sum_{t=1}^{T} Z_{t-1}Z_{t-1}' \otimes \widehat{\Sigma}_t^{-1}\right)^{-1}\left(\sum_{t=1}^{T}(Z_{t-1} \otimes \widehat{\Sigma}_t^{-1})y_t\right), \tag{14.3.6}$$

where $Z_{t-1} = (1, y'_{t-1}, \dots, y'_{t-p})'$ and $\widehat{\Sigma}_t = \widehat{\Sigma}_m$ for $t \in \mathcal{T}_m$. If the VAR process is stable, these estimators have standard asymptotic properties. Although this type of model typically requires nonlinear estimation, these estimation methods are simple enough to allow the use of bootstrap methods for inference on the structural impulse responses.

The structural parameters can be estimated by substituting $B_0^{-1}B_0^{-1\prime}$ for Σ_1, $B_0^{-1}\Lambda_m B_0^{-1\prime}$ for Σ_m, $m = 2, \dots, M$, and $y_t - (Z_t' \otimes I_K)\boldsymbol{\alpha}$ for u_t in (14.3.5) and maximizing the log-likelihood. Alternatively, the $\boldsymbol{\alpha}$ parameters can be estimated in a first step, and in the second step the concentrated log-likelihood is optimized only with respect to the structural parameters B_0^{-1} and Λ_m, $m = 2, \dots, M$.

GMM is an alternative estimation method for the structural parameters. In that case, the structural parameters are estimated by minimizing the objective function

$$J = \begin{pmatrix} \text{vech}_1 \\ \vdots \\ \text{vech}_M \end{pmatrix}' W \begin{pmatrix} \text{vech}_1 \\ \vdots \\ \text{vech}_M \end{pmatrix}, \tag{14.3.7}$$

where

$$\text{vech}_1 = \text{vech}\left(\frac{1}{\#\mathcal{T}_1} \sum_{t \in \mathcal{T}_1} \hat{u}_t \hat{u}_t' - B_0^{-1} B_0^{-1\prime} \right),$$

$$\text{vech}_m = \text{vech}\left(\frac{1}{\#\mathcal{T}_m} \sum_{t \in \mathcal{T}_m} \hat{u}_t \hat{u}_t' - B_0^{-1} \Lambda_m B_0^{-1\prime} \right) \quad \text{for} \quad m = 2, \dots, M,$$

and W is a positive definite weighting matrix. For example, W may be constructed as a block-diagonal matrix from the inverse covariance matrices of the $\text{vech}(\hat{u}_t \hat{u}_t')$ as

$$W = \begin{bmatrix} W_1 & & 0 \\ & \ddots & \\ 0 & & W_M \end{bmatrix},$$

with

$$W_m = \left(\frac{1}{\#\mathcal{T}_m} \sum_{t \in \mathcal{T}_m} \left(\text{vech}(\hat{u}_t \hat{u}_t') - \overline{\text{vech}(\hat{u}\hat{u}')} \right) \right.$$
$$\left. \times \left(\text{vech}(\hat{u}_t \hat{u}_t') - \overline{\text{vech}(\hat{u}\hat{u}')} \right)' \right)^{-1}$$

and

$$\overline{\text{vech}(\hat{u}\hat{u}')} = \frac{1}{\#\mathcal{T}_m} \sum_{t \in \mathcal{T}_m} \text{vech}(\hat{u}_t \hat{u}_t') \quad \text{for} \quad m = 1, \dots, M$$

(see, e.g., Ehrmann, Fratzscher, and Rigobon (2011) and Wright (2012) for applications of GMM estimators for related models).

The GMM and ML estimators of the identified parameters have standard asymptotic properties under common assumptions. In fact, Rigobon (2003) shows that consistent estimators of the structural parameters are obtained, even if the change points for the volatility regimes are misspecified.

Testing the Model Assumptions. In principle, one can test the crucial assumptions underlying the heteroskedastic model. Three types of tests are of particular interest. The first test addresses the question of the existence of the decomposition (14.3.2). An LR test of this null hypothesis can be constructed by comparing the maxima of the log-likelihood with and without the restriction imposed. This LR statistic under usual assumptions has an asymptotic χ^2 distribution with

$$\frac{1}{2}MK(K+1) - K^2 - (M-1)K$$

degrees of freedom under the null hypothesis (see Lanne, Lütkepohl, and Maciejowska 2010).

The second test assesses the validity of the identifying restrictions conditional on the existence of the decomposition (14.3.2). The central idea is that the estimated diagonal elements of the Λ_m matrices can be used to investigate whether the diagonal elements are distinct and, hence, whether the shocks are identified. Unfortunately, developing formal statistical tests of the relevant null hypotheses is not straightforward. To see this, consider a bivariate system with $M = 2$ such that there are two volatility regimes. Decomposing the two covariance matrices as in (14.2.2), we have $\Sigma_1 = B_0^{-1} B_0^{-1\prime}$ and $\Sigma_2 = B_0^{-1} \Lambda B_0^{-1\prime}$ with $\Lambda = \text{diag}(\lambda_1, \lambda_2)$. We are interested in testing

$$\mathbb{H}_0 : \lambda_1 = \lambda_2 \quad \text{versus} \quad \mathbb{H}_1 : \lambda_1 \neq \lambda_2.$$

Note that the elements of B_0^{-1} are not identified under $\mathbb{H}_0$, but only under $\mathbb{H}_1$ when $\lambda_1 \neq \lambda_2$. It is well known that, when there are unidentified parameters under $\mathbb{H}_0$, standard tests such as LR and Wald tests may not have their usual asymptotic χ^2 distributions. This point is not fully appreciated in this literature.

Finally, it is possible to treat conventional short-run or long-run exclusion restrictions on B_0^{-1} (or on B_0) as overidentifying restrictions in heteroskedastic VAR models. The validity of these overidentifying restrictions may be formally tested by comparing the maxima of the log-likelihood with and without the restriction imposed. This LR test under general conditions has an asymptotic

$\chi^2(n)$ distribution, where n denotes the number of overidentifying restrictions. Note that there are no formal tests of overidentifying sign restrictions, but we may assess whether the estimate of B_0^{-1} from the heteroskedastic VAR model satisfies the sign restrictions (see Lütkepohl and Netšunajev 2014).

The asymptotic properties of these estimators and test statistics may differ substantially from their small-sample properties because the model is highly parameterized and macroeconomic time series are typically relatively short. In addition, any given volatility regime may be present only for a relatively short time span. Hence, the interpretation of the statistical results may not be straightforward. Moreover, the volatility regimes were assumed to be known, which is not the case in practice. Instead there will be uncertainty regarding the change points and perhaps some pretesting or other statistical methods for determining the change points may have been used (see, e.g., Ehrmann, Fratzscher, and Rigobon 2011). All these problems do not add to the reliability of inference in this context. Still, the fact that the data may speak to the crucial issue of the identification of the shocks is a potential advantage over having no information on these issues.

A Detailed Empirical Example. To illustrate how unconditional heteroskedasticity can be used to aid in the identification of structural shocks in a structural VAR analysis we consider an example from Lanne and Lütkepohl (2014) who compare different identification schemes for the monetary policy models discussed in Christiano, Eichenbaum, and Evans (1999). The latter models were already discussed in Chapter 8. They are just-identified and their identifying assumptions cannot be tested within the conventional structural VAR framework.

Let gdp_t be the log of real GDP, p_t the log of the GDP deflator, $pcom_t$ the smoothed change in an index of sensitive commodity prices, nbr_t the log of nonborrowed reserves plus extended credit, tr_t the log of total reserves, i_t the federal funds rate, and m_t the log of M1. Thus, we are dealing with a seven-dimensional system. We follow Christiano, Eichenbaum, and Evans (1999) in using monthly data for the period 1965m7–1995m6. The sample size is $T = 360$. The reduced-form model is a VAR(12) with an intercept.

As in Section 8.4.5, the vector of variables, y_t, is partitioned as

$$y_t = \begin{pmatrix} y_{1t} \\ z_t \\ y_{2t} \end{pmatrix},$$

where y_{1t} is $K_1 \times 1$, z_t is a scalar variable, and y_{2t} is $K_2 \times 1$. The vector y_{1t} contains variables whose contemporaneous values appear in the monetary authority's information set. They are assumed to be orthogonal to the monetary policy shock. The variable z_t is the policy instrument of the monetary authority, and the variables in y_{2t} appear with a lag in the information set.

Identification of the monetary policy shock is achieved by assuming that the matrix B_0 is lower block-triangular:

$$B_0 = \begin{bmatrix} \underset{K_1\times K_1}{B_{11,0}} & 0_{K_1\times 1} & 0_{K_1\times K_2} \\ \underset{1\times K_1}{B_{21,0}} & \underset{1\times 1}{B_{22,0}} & 0_{1\times K_2} \\ \underset{K_2\times K_1}{B_{31,0}} & \underset{K_2\times 1}{B_{32,0}} & \underset{K_2\times K_2}{B_{33,0}} \end{bmatrix}.$$

The monetary policy shock is the $(K_1 + 1)^{\text{th}}$ element of w_t. The other shocks are not of direct interest and can therefore be identified arbitrarily by choosing B_0 to be lower-triangular, which means that $B_{11,0}$ and $B_{33,0}$ are restricted to be lower triangular. This choice does not affect the monetary policy shock.

Lanne and Lütkepohl (2014) compare the three alternative identification schemes already discussed in Chapter 8, Section 8.4.5. They differ by the choice of y_{1t}, z_t, and y_{2t} as follows:

FFR policy shock: $y_{1t} = (gnp_t, p_t, pcom_t)'$, $z_t = i_t$, and $y_{2t} = (nbr_t, tr_t, m_t)'$.

NBR policy shock: $y_{1t} = (gnp_t, p_t, pcom_t)'$, $z_t = nbr_t$, and $y_{2t} = (i_t, tr_t, m_t)'$.

NBR/TR policy shock: $y_{1t} = (gnp_t, p_t, pcom_t, tr_t)'$, $z_t = nbr_t$, and $y_{2t} = (i_t, m_t)'$.

There are a number of reasons for structural breaks in the model during the sample period 1965m7–1995m6. Following Christiano, Eichenbaum, and Evans (1999) and Bernanke and Mihov (1998a), Lanne and Lütkepohl (2014) consider two structural breaks in 1979m10 and 1984m2. Thus, we work with three volatility regimes. Notice that the period 1979m10-1984m2 roughly corresponds to the era of Fed chairman Paul Volcker. This period is often regarded as special from the point of view of monetary policy. Assuming that the structural breaks affect only the error variances, Lanne and Lütkepohl (2014) report the estimated relative variances given in Table 14.1.

The estimated λ_{km} in Table 14.1 are of particular interest because they contain the crucial information for the identification of the B_0 matrix. For each pair k, l either $\lambda_{k2} \neq \lambda_{l2}$ or $\lambda_{k3} \neq \lambda_{l3}$ has to hold for full identification. For example, given the standard errors in Table 14.1, $\lambda_{12} \neq \lambda_{22}$ may not hold, whereas $\lambda_{13} \neq \lambda_{23}$ may well hold. Of course, it is not clear that such results hold for all pairs of relative variances, as required for full identification. This hypothesis should ideally be tested by formal statistical procedures. Such tests are not provided in Lanne and Lütkepohl (2014) who point out, however, that even if B_0 is not fully identified by heteroskedasticity, the available information may be sufficient to test the conventional restrictions implied by the three identification schemes presented earlier. Clearly, if some of the restrictions are rejected,

Table 14.1. *Estimation Results for Parameters of VAR(12) Model with Unconditionally Heteroskedastic Errors for Sampling Period 1965m7–1995m6 and Variance Changes in 1979m10 and 1984m2*

Regime	$m = 2$		$m = 3$	
Parameter	Estimate	Std Error	Estimate	Std Error
λ_{1m}	0.8735	0.2761	0.4623	0.0801
λ_{2m}	1.2161	0.4405	0.7458	0.1284
λ_{3m}	0.9791	0.2845	0.7992	0.1397
λ_{4m}	5.2158	1.1729	0.6916	0.1179
λ_{5m}	1.5875	0.4137	1.8507	0.3110
λ_{6m}	1.7791	0.5160	1.4448	0.2481
λ_{7m}	0.6639	0.3452	0.3619	0.0757

Source: Adapted from Table 1 of Lanne and Lütkepohl (2014).

this implies that there is sufficient identifying information to enable the data to speak against the restrictions.

Table 14.2 reports test results from Lanne and Lütkepohl (2014) for the zero restrictions on B_0 implied by the three competing identification schemes. The p-values reported in the table are determined using the degrees of freedom in the asymptotic χ^2 distribution of the test statistics that would have been obtained if the null hypotheses were true and the B_0 matrix were fully identified

Table 14.2. *LR Type Tests of Identification Schemes Based on Unconditionally Heteroskedastic Models with Regime Changes in 1979m10 and 1984m2*

	$\mathbb{H}_0$	df	LR	p-value
FFR	$B_{12,0} = 0_{3\times1}$	3	2.7770	0.4273
	$B_{13,0} = 0_{3\times3}$	9	9.1872	0.4202
	$B_{23,0} = 0_{1\times3}$	3	1.5451	0.6719
	$B_{12,0} = 0_{3\times1}, B_{13,0} = 0_{3\times3}, B_{23,0} = 0_{1\times3}$	15	17.4070	0.2951
NBR	$B_{12,0} = 0_{3\times1}$	3	1.2528	0.7404
	$B_{13,0} = 0_{3\times3}$	9	28.8631	0.0007
	$B_{23,0} = 0_{1\times3}$	3	1.3433	0.7189
	$B_{12,0} = 0_{3\times1}, B_{13,0} = 0_{3\times3}, B_{23,0} = 0_{1\times3}$	15	37.7858	0.0010
NBR/TR	$B_{12,0} = 0_{4\times1}$	4	26.3575	2.6802e−5
	$B_{13,0} = 0_{4\times2}$	8	20.9426	0.0073
	$B_{23,0} = 0_{1\times2}$	2	0.6334	0.7286
	$B_{12,0} = 0_{4\times1}, B_{13,0} = 0_{4\times2}, B_{23,0} = 0_{1\times2}$	14	60.8736	8.3617e−8

Source: Adapted from Table 2 of Lanne and Lütkepohl (2014).

by heteroskedasticity. As this is not necessarily the case, the actual degrees of freedom may be smaller and, hence, also the p-values may be smaller. In other words, whenever a restriction in Table 14.2 is rejected based on its p-value, it would also be rejected if the actual number of degrees of freedom were smaller than assumed in the table. Thus, there is strong evidence that some of the restrictions associated with the NBR and NBR/TR identification schemes can be rejected. For example, $B_{13,0} = 0$ and the joint hypothesis $B_{12,0} = 0$, $B_{13,0} = 0$, $B_{23,0} = 0$ are clearly rejected for the NBR scheme at conventional significance levels. Likewise, $B_{12,0} = 0$, $B_{13,0} = 0$, and the joint hypothesis $B_{12,0} = 0$, $B_{13,0} = 0$, $B_{23,0} = 0$ are clearly rejected for the NBR/TR scheme. In other words, the only identification scheme not rejected in our framework is the FFR scheme. Hence, the changes in volatility help discriminate between these three identification schemes.

It is perhaps worth emphasizing that not rejecting the assumptions underlying the FFR scheme does not necessarily imply that this scheme is correct. Of course, if a statistical test does not reject the null hypothesis of interest, this fact can always be due to low power and does not necessarily confirm the validity of the null hypothesis. In the current example, it is even more problematic to view a non-rejection as support for the FFR model. The reason is that a non-rejection may simply reflect the absence of overidentifying information, since the B_0 matrix may not be fully identified by heteroskedasticity.

Other Empirical Examples. Similar models with extraneously assigned volatility regimes have also been used in a number of other studies. For example, Rigobon (2003) applies this approach to investigate the relationship between returns on sovereign bonds in Argentina, Brazil, and Mexico, using daily bond returns for the period January 1994 to December 2001. Identification by heteroskedasticity may seem appealing in this application because bond yields are simultaneously determined, rendering conventional exclusion restrictions clearly inappropriate, but the resulting shock and impulse response estimates have no obvious economic interpretation.

A similar problem of interpretation arises in Ehrmann, Fratzscher, and Rigobon (2011) who rely on identification through heteroskedasticity when investigating the linkages between the U.S. and euro area money, bond, and equity markets. They analyze a system of short-term interest rates, long-term bond rates, stock returns from both regions, and an exchange rate based on two-day returns for the period 1989–2008. Changes in volatility are determined based on estimates of the variance from a rolling window.

These examples illustrate a tendency in applied studies to rely on identification by heteroskedasticity to compensate for the lack of economic identification conditions. It should be clear that shocks obtained in this way need not correspond to the economically interpretable structural shocks that economists are interested in. Nor does the implied estimate of B_0 correspond to an economically meaningful structure.

There are exceptions to this rule, however. Notably, identification by heteroskedasticity may allow us to identify selected elements of B_0 in certain economic models. For example, Rigobon and Sack (2003) propose a structural model in which the slope parameter of the monetary policy reaction function with respect to stock returns can be identified by shifts in the variance of shocks to investor risk preferences relative to monetary policy shocks. They distinguish between four different volatility regimes that are determined by 30-day rolling variance estimates. Their analysis is based on daily U.S. data from March 1985 to December 1999. They find that monetary policy reacts to stock returns. For a related application see also Rigobon and Sack (2004).

Finally, Wright (2012) is another study that bridges the gap between statistical and economic identification. Wright observes that monetary shocks have larger than usual variance on days of FOMC meetings and on days when important speeches are delivered by FOMC members. He uses daily data from November 3, 2008 to September 30, 2011 on six U.S. interest rates and identifies monetary policy shocks by utilizing the difference in the error variance of normal days and days with larger variance. Wright is only interested in the monetary policy shock and, hence, in partial identification of the system. The assumption that only the volatility of the monetary policy shock changes on days of larger variance allows him to label the identified shock as a monetary policy shock.

Even if identification by heteroskedasticity does not suffice to estimate the parameters of the underlying economic model, we may use the model estimates obtained under these assumptions to test the validity of alternative structural VAR models based on conventional identifying restrictions. For example, Lanne and Lütkepohl (2008) use identification by heteroskedasticity to assess the validity of several models for the U.S. money market that have been considered in the literature. Their results are in line with those discussed in our earlier detailed example.

14.3.2 Structural VAR Models with Markov Switching in the Variances

In practice, the timing of the volatility changes is rarely known. Typically it has to be determined from the data. One approach to modeling volatility changes as an endogenous process was proposed by Lanne, Lütkepohl, and Maciejowska (2010). They model volatility as state dependent with the state of the system evolving according to a Markov process. The methodology for using this approach in the structural VAR framework is partly due to Herwartz and Lütkepohl (2014). We first present the model setup and then discuss identification conditions, statistical inference, and applications.

Model. It is assumed that the volatility changes are driven by a discrete Markov process s_t, $t \in \mathbb{Z}$, with M states. In other words, $s_t \in \{1, \dots, M\}$. The transition probabilities between the states are

$$p_{ij} = \mathbb{P}(s_t = j | s_{t-1} = i), \quad i, j = 1, \dots, M.$$

Lanne, Lütkepohl, and Maciejowska (2010) postulate that the conditional distribution of u_t, given the state s_t, is normal, i.e.,

$$u_t | s_t \sim \mathcal{N}(0, \Sigma_{s_t}), \tag{14.3.8}$$

where all Σ_m, $m = 1, \dots, M$, are distinct, i.e., $\Sigma_m \neq \Sigma_n$ for $m \neq n$. The assumption of conditional normality allows one to employ ML estimation. Note that the implied unconditional distribution of u_t is not Gaussian in general and can capture a wide range of non-Gaussian distributions.

We emphasize that only the error covariance matrices depend on the Markov process, whereas the VAR coefficients are assumed to be time-invariant. VAR models in which the VAR coefficients change over time according to a Markov process have been considered by a number of authors (see, e.g., Rubio-Ramírez, Waggoner, and Zha 2005; Sims and Zha 2006b; Sims, Waggoner, and Zha 2008). Such models are discussed in Chapter 18. In this chapter, we consider VAR models in which only the error covariance structure changes. More precisely, a model with VAR lag order p and M volatility states is denoted as an MS(M)-VAR(p) model, where MS stands for Markov switching.

Although there is only a finite number of volatility states, the model can mix these states by assigning probabilities strictly between zero and one to the states in any particular period t. Thus, the model may capture gradual transitions from one state to another, and may be interpreted as a model with a continuum of states. The state covariance matrices $\Sigma_1, \dots, \Sigma_M$ are used for the identification of shocks, as in the case of known change points. In other words, the state covariance matrices are decomposed as in (14.3.2) or (14.3.3).

Identification. Uniqueness of the B_0^{-1} or B_0 matrices and, hence, of the structural shocks holds under the same conditions as for the case of known change points. The crucial condition is that there is enough heterogeneity in the volatility changes. In the literature, the variance normalization implied by the decomposition (14.3.2) rather than (14.3.3) is typically used. In other words, we normalize the variances of the structural shocks in the first state to be one. Thus the λ_{km} in the other states can be interpreted as variances relative to the first state. If they satisfy the condition

$$\forall k, l \in \{1, \dots, K\} \text{ with } k \neq l, \exists\, m \in \{2, \dots, M\} \text{ such that } \lambda_{km} \neq \lambda_{lm}, \tag{14.3.9}$$

then B_0^{-1} is unique up to column sign changes and column permutations. This condition must be satisfied for all of the elements of B_0 to be identified by the volatility changes.

Estimation and Inference. The statistical analysis of the MS-VAR model involves the choice of the number of volatility states and the estimation of the parameters of the overall model including the transition probabilities of the Markov process, the diagonal elements of the Λ_m matrices, and the B_0^{-1} or B_0 matrix.

Under assumption (14.3.8), which assigns a conditional normal distribution to the reduced-form errors, the log-likelihood function can be set up as

$$\log l(\boldsymbol{\alpha}, B_0^{-1}, \boldsymbol{\lambda}, P|\mathbf{y}) = \sum_{t=1}^{T} \log\left(\sum_{m=1}^{M} \mathbb{P}(s_t = m|Y_{t-1}) f(y_t|s_t = m, Y_{t-1})\right), \quad (14.3.10)$$

where $\boldsymbol{\lambda}$ is the vector of all diagonal elements of $\Lambda_2, \dots, \Lambda_M$, P is the matrix of transition probabilities, $\mathbf{y}$ is the full sample, $Y_{t-1} \equiv (y'_{t-1}, \dots, y'_{t-p})'$ and

$$f(y_t|s_t = m, Y_{t-1}) = (2\pi)^{-K/2} \det(\Sigma_m)^{-1/2} \exp\left\{-\frac{1}{2} u_t' \Sigma_m^{-1} u_t\right\}.$$

The maximization of this log-likelihood function involves a highly nonlinear optimization problem that poses a number of computational challenges, even when the parameters are identified. Herwartz and Lütkepohl (2014) adopt an expectation-maximization (EM) algorithm from Krolzig (1997) for estimating this model.

They also discuss a number of problems that have to be addressed in maximizing the likelihood or log-likelihood. In particular, they point out that the likelihood function is strictly speaking unbounded and has many local maxima, the largest of which implies the preferred estimate. Moreover, the variances have to remain strictly positive and the covariance matrices positive definite in each stage of the iterative optimization algorithm. Moreover, the same ordering of the shocks has to be enforced in each iteration step. In other words, the λ_{km} have to be ordered in some way.

In practice, the ML-estimation algorithm works reliably only for reasonably small models with a small number of variables K, a small number of volatility states M, and a moderate number of lags p. If the conditional normality of the errors underlying assumption (14.3.8) does not hold, the estimation procedure may be interpreted as a quasi-ML procedure. Unfortunately, as shown by Campbell (2002), quasi-ML estimators for MS models may be inconsistent. Thus, the method has to be used with caution if the assumption of conditional normality cannot be justified. Unless a specific alternative conditional distribution can be justified, working with ML estimators based on other distributions may be preferable.

Given that the ML parameter estimators have standard asymptotic properties, inference on the $\boldsymbol{\lambda}$ parameters can in principle be conducted as in the case of known volatility change points. In other words, the data can be used to learn about the validity of the identification conditions. Unfortunately, given the problems with unidentified parameters under the null hypothesis mentioned in Section 14.3.1, the asymptotic properties of standard LR and Wald tests are currently unknown in general. The χ^2 distributions that have been used in this context in some of the related literature are not likely to be valid.

If the shocks can be identified by the volatility structure, conventional identifying restrictions become overidentifying in this model and can be tested, as discussed earlier. For example, zero restrictions on the impact effects or the long-run effects can be tested by LR or Wald tests conditional on the MS-VAR model specification.

It still remains to be shown how the number of volatility states in MS-VAR models may be chosen. Standard model selection criteria can be used for that purpose. Based on simulation experiments, Psaradakis and Spagnolo (2003, 2006) conclude that this approach may not be very reliable, however. Testing models with different numbers of states against the data is not easy either. The usual tests have nonstandard properties because of unidentified parameters under the null hypothesis (e.g., Hansen 1992; Garcia 1998).

After identifying the structural shocks, the model can be used for structural analysis. As discussed in earlier chapters, impulse responses are the standard tool for this purpose. Confidence intervals or joint confidence bands are usually generated with bootstrap methods. Because even a single estimation of the structural MS-VAR model is difficult, it is clear that standard bootstrap approaches are computationally challenging in this context. To mitigate the computational problems, Herwartz and Lütkepohl (2014) proposed a fixed-design wild bootstrap procedure.

An alternative to the use of frequentist methods of estimation and inference for heteroskedastic MS-VAR models is Bayesian methods. The estimation of structural MS-VAR models is discussed in Sims, Waggoner, and Zha (2008), Sims and Zha (2006b), and Rubio-Ramírez, Waggoner, and Zha (2005). These methods are discussed in Chapter 18, where more general MS-VAR models are considered that allow for time-varying slope coefficients as well. Bayesian methods for the specific MS-VAR models of this section, in which only the innovation covariance matrices vary, are developed in Kulikov and Netšunajev (2013) for the case of just-identified models and in Woźniak and Droumaguet (2015) for partially identified and overidentified models.

Empirical Examples. Lütkepohl and Velinov (2016) investigate to what extent stock prices reflect their underlying economic fundamentals, building on work by Velinov and Chen (2015). According to the dividend discount model, an asset's price is the sum of its expected future discounted dividends. Since these

Table 14.3. *Estimates and Standard Errors of Relative Variances of MS(3)-VAR(2) Model for* $y_t = (\Delta q_t, \Delta r_t, \Delta s_t)'$

Regime	$m = 2$		$m = 3$	
Parameter	Estimate	Standard Error	Estimate	Standard Error
λ_{1m}	0.267	0.094	0.845	0.252
λ_{2m}	0.277	0.089	3.777	1.004
λ_{3m}	3.564	1.051	14.878	4.117

Source: Extracted from Table 1 of Lütkepohl and Velinov (2016).

dividends are related to real economic activity, this variable may serve as a proxy for the fundamentals.

For illustrative purposes, consider a three-dimensional model for the U.S. economy. Let $y_t = (\Delta gdp_t, \Delta r_t, \Delta sp_t)'$, where Δgdp_t is the growth rate of real GDP, Δr_t is the change in a real interest rate, and Δsp_t denotes real stock returns. We use quarterly data for the period 1947q1–2012q3.[3] The model is set up in first differences because tests provide no evidence for cointegration. Related models have been used by a number of authors (see, e.g., Lee 1995; Rapach 2001; Binswanger 2004; and Lanne and Lütkepohl 2010).

Based on model selection criteria, Lütkepohl and Velinov (2016) choose an MS(3)-VAR(2) model. An LR test does not reject the existence of the decomposition (14.3.2) of the three covariance matrices. The crucial identification condition is linked to the diagonal elements of the Λ_m matrices, which have to be sufficiently heterogeneous. The estimated quantities are shown in Table 14.3. Note that in a model with three volatility states we have to investigate for each pair k, l with $k \neq l$ whether $\lambda_{k2} \neq \lambda_{l2}$ or $\lambda_{k3} \neq \lambda_{l3}$. Thus, ideally one would want to test null hypotheses

$$\mathbb{H}_0 : \lambda_{k2} = \lambda_{l2}, \ \lambda_{k3} = \lambda_{l3}$$

for all pairs k, l. If all null hypotheses can be rejected, we have statistical evidence that the identification condition (14.3.9) holds. In the absence of suitable asymptotically valid tests, the estimated λ_{im} and their standard errors help determine whether assuming full identification by changes in volatility is reasonable in this model.

Next we focus on the support for conventional identifying restrictions for this class of models. The central question of interest is how important shocks to fundamentals are in driving stock prices. Suppose that there are two nonfundamental shocks and one fundamental shock. In a conventional analysis these structural shocks may be identified by imposing restrictions on the long-run

[3] See Velinov (2013) for the details on the data.

effects such that the long-run multiplier matrix can be written as

$$\Upsilon = \begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \end{bmatrix}. \tag{14.3.11}$$

Each column in Υ corresponds to the effects of a specific shock on the level of each of the three variables (see Chapter 10). Thus, the first shock can have permanent effects on all variables and is hence classified as a fundamental shock. Notably its potential permanent effect on GDP distinguishes this shock from the other two shocks which are not regarded as fundamental. The second shock captures all other components that may have long-run effects on the real interest rate and does not have a specific economic interpretation. The last shock is not allowed to have long-run effects on any variable but stock prices.

These long-run identifying restrictions can be treated as overidentifying restrictions in the MS-VAR model. Lütkepohl and Velinov (2016) report a p-value of 0.207 for the LR test of the null that Υ is lower triangular. They cannot reject this restriction at conventional significance levels.

Since the long-run identification restrictions are not rejected, one could impose them in addition to the MS-VAR structure in estimating the structural impulse responses. This modification would require estimating (14.3.10) subject to the additional restrictions on the elements of Υ (see also Chapter 11 for a similar approach in a simpler context). Alternatively, one could proceed by dropping the MS-VAR structure and fitting the structural VAR model subject to the long-run restrictions only. The point estimates of the structural impulse responses from this model typically remain consistent even in the presence of unmodeled conditional (or for that matter unconditional) heteroskedasticity in the error term. Only the construction of the impulse response confidence intervals may have to be modified to allow for unmodeled volatility.

There are also a number of other studies that use the structural MS-VAR model. Lanne, Lütkepohl, and Maciejowska (2010) consider a small U.S. macroeconomic model due to Primiceri (2005) consisting of inflation, unemployment, and an interest rate. They find that previously used identification restrictions are not supported by the data when changes in volatility are taken into account. They also consider a model of the U.S. economy based on Sims, Waggoner, and Zha (2008) for log GDP, inflation, and a short-term interest rate. Using quarterly data from 1959q1–2005q4 they find support for conventional identification restrictions, as used by Sims, Waggoner and Zha.

Herwartz and Lütkepohl (2014) consider a model from Peersman (2005) that investigates the causes of the early millennium economic slowdown. Peersman uses a four-variable system consisting of the price of oil, output, a consumer price index, and a short-term interest rate. He imposes zero restrictions on the impact effects and the long-run effects of the shocks for identification. Using quarterly U.S. data from 1980q1 to 2002q2 and models with two and three volatility states, Herwartz and Lütkepohl find that some, but not all of the identifying restrictions used by Peersman are in line with the data, when changes in volatility are accounted for. For a related analysis of the Peersman model see also Chapter 11.

Lütkepohl and Netšunajev (2014) illustrate how the structural MS-VAR approach can be used for testing models identified by sign restrictions. They do not find evidence against the identifying restrictions used by Kilian and Murphy (2012), thereby supporting the findings of the latter article. Finally, Netšunajev (2013) reconsiders alternative approaches to identifying technology shocks based on the structural MS-VAR methodology. His analysis sheds new light on the conflicting evidence on the impact on hours worked.

The MS-VAR model has appeal because it lets the data assign the observations to different volatility regimes. Unlike the model in Section 14.3.1, the MS-VAR model does not require the user to know when volatility shifts occurred. Moreover, its identification conditions can in principle be tested. Finally, the model may be estimated by Gaussian ML, if the model is small enough, allowing the use of existing results for Gaussian ML estimators of Markov-switching models. The main drawback of the model is that for larger models with long lag orders, for a larger set of variables, or for many different volatility regimes, estimation is difficult and potentially unreliable. Moreover, even for small models that can be estimated easily, little is known about the asymptotic and finite-sample properties of the proposed methods for constructing confidence intervals about structural impulse responses, for example.

14.3.3 Structural VAR Models with Smooth Transitions in the Variances

Model Setup. Another model with endogenously changing volatility postulates a smooth change in the unconditional error covariance matrix (see Lütkepohl and Netšunajev 2015). More specifically, the change in the covariance structure is modeled as a smooth transition from a volatility regime characterized by a positive definite covariance matrix Σ_1 to a regime with a different positive definite covariance matrix Σ_2. The transition is described by a smooth transition function $G(\gamma, c, s_t)$ that depends on parameters γ and c as well as a transition variable s_t. Lütkepohl and Netšunajev use a logistic transition function,

$$G(\gamma, c, s_t) = (1 + \exp[-\exp(\gamma)(s_t - c)])^{-1}. \tag{14.3.12}$$

Clearly, $0 < G(\gamma, c, s_t) < 1$. Since $\exp(\gamma) > 0$ for positive and negative values of γ, $G(\gamma, c, s_t)$ is close to zero when s_t is much smaller than c and close to one if s_t is much larger than c (see also Chapter 18). The reduced-form error covariance matrix is specified as

$$\mathbb{E}(u_t u_t') = (1 - G(\gamma, c, s_t))\Sigma_1 + G(\gamma, c, s_t)\Sigma_2. \tag{14.3.13}$$

This matrix is positive definite because it is a convex combination of two positive definite matrices, Σ_1 and Σ_2. Given that $G(\gamma, c, s_t)$ is a continuous function, there is a continuum of potential covariance matrices, corresponding to an infinite number of volatility states. The number of states depends on the transition variable s_t. A transition variable may assign different volatility regimes to different parts of the sample.

Extensions of this smooth-transition VAR (ST-VAR) model to more transition terms are possible in principle, but are more difficult to handle. Such extensions may not be necessary, if a suitable transition variable can be found.

Identification. Identification conditions for structural shocks are linked to the two limiting covariance matrices Σ_1 and Σ_2. Choosing B_0^{-1} such that

$$\Sigma_1 = B_0^{-1} B_0^{-1\prime} \quad \text{and} \quad \Sigma_2 = B_0^{-1} \Lambda_2 B_0^{-1\prime}, \tag{14.3.14}$$

where $\Lambda_2 = \text{diag}(\lambda_{12}, \ldots, \lambda_{K2})$ is a diagonal matrix with positive diagonal elements, B_0^{-1} is unique up to column sign and column permutation, if the diagonal elements of Λ_2 are all distinct. Under this condition the structural shocks $w_t = B_0 u_t$ are identified and their covariance matrix is diagonal,

$$\mathbb{E}(w_t w_t') = (1 - G(\gamma, c, s_t)) I_K + G(\gamma, c, s_t)\Lambda_2.$$

Estimation and Inference. If u_t is normally distributed, the log-likelihood function of the model is

$$\log l = \text{constant} - \frac{1}{2} \sum_{t=1}^{T} \log(\det(\Sigma_t)) - \frac{1}{2} \sum_{t=1}^{T} u_t' \Sigma_t^{-1} u_t, \tag{14.3.15}$$

where $\Sigma_t = \mathbb{E}(u_t u_t')$. Maximizing this function with respect to the model parameters requires iterative optimization techniques.

For a given transition function, however, the maximization of the log-likelihood with respect to the remaining parameters is straightforward. Lütkepohl and Netšunajev (2015) propose using a grid search over γ and c. They point out that, for the logistic transition function, the effective range of these parameters depends on the transition variable, but is limited to a bounded range, outside of which the transition function will not change noticeably any more. The range is chosen such that the values of the transition function cover the unit interval. In a first round a slightly wider grid can be used, which is then refined in a second round in the neighbourhood of the values maximizing the log-likelihood function in the first round.

For given values of the transition parameters γ and c, Lütkepohl and Netšunajev (2015) propose to estimate the other parameters by iterating the following steps:

Step 1: Given starting values of the reduced-form VAR parameters $\nu, A_1, \dots, A_p$, the structural parameters B_0^{-1} and Λ_2 are estimated by maximizing the log-likelihood function (14.3.15) using nonlinear maximization.

Step 2: Given the structural parameter estimates from step 1, the parameters of the reduced-form VAR model are reestimated. For given structural parameters B_0^{-1} and Λ_2 the model is linear in $\nu, A_1, \dots, A_p$. Hence, the vectorized VAR coefficients $\boldsymbol{\alpha} = \text{vec}[\nu, A_1, \dots, A_p]$ can be estimated by generalized least squares,

$$\widehat{\boldsymbol{\alpha}} = \left[(Z \otimes I_K)\boldsymbol{\Sigma}^{-1}(Z' \otimes I_K)\right]^{-1}(Z \otimes I_K)\boldsymbol{\Sigma}^{-1}\mathbf{y},$$

where

$$\boldsymbol{\Sigma}^{-1} \equiv \begin{bmatrix} \mathbb{E}(u_1 u_1')^{-1} & & 0 \\ & \ddots & \\ 0 & & \mathbb{E}(u_T u_T')^{-1} \end{bmatrix}$$

is a $KT \times KT$ block-diagonal covariance matrix, $\mathbf{y} \equiv \text{vec}[y_1, \dots, y_T]$ is a $KT \times 1$ data vector, and the t^{th} column of the $(1 + Kp) \times T$ data matrix Z is $Z_{t-1} \equiv (1, y_{t-1}', \dots, y_{t-p}')'$.

The resulting estimates of $\nu, A_1, \dots, A_p$ are then used as inputs in step 1. These steps are iterated until convergence.

This estimation procedure is technically straightforward and numerically robust even for larger VAR models, provided the transition variable is chosen such that the transition parameters and the matrices Σ_1 and Σ_2 are identified. Note that γ and c are not identified if $\Sigma_1 = \Sigma_2$. Thus, the two volatility regimes associated with these two covariance matrices must be clearly distinct. If identification of the volatility model is ensured, the asymptotic properties of the parameter estimators are standard, even if u_t is not normally distributed. In that case, the estimators obtained by maximizing the Gaussian likelihood function may be interpreted as quasi-ML estimators.

The identification of B_0 or B_0^{-1}, in principle, may be assessed based on the asymptotic properties of the diagonal elements of Λ_2. The null hypothesis of no identification is equivalent to

$$\mathbb{H}_0 : \lambda_{k2} = \lambda_{l2}$$

for all $k, l \in \{1, \dots, K\}$ with $k \neq l$. General asymptotically valid tests are currently not available. As in the previous models, if the identification conditions are satisfied and all diagonal elements of Λ_2 are distinct, any further restrictions imposed on B_0^{-1} become overidentifying and can be tested using the LR test discussed earlier.

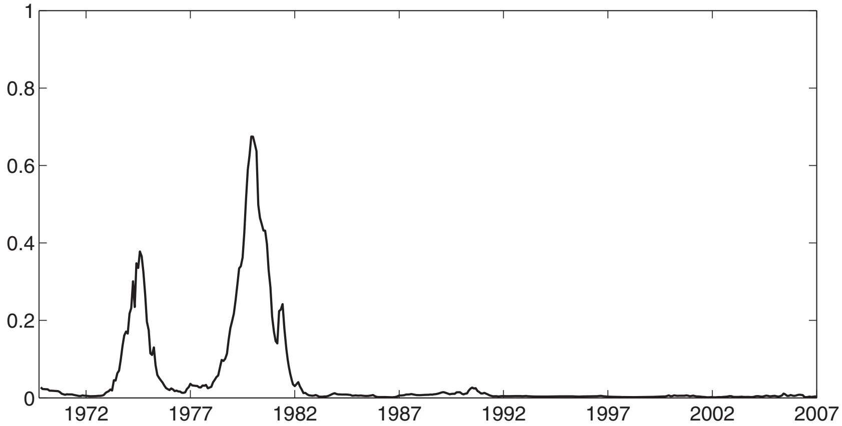
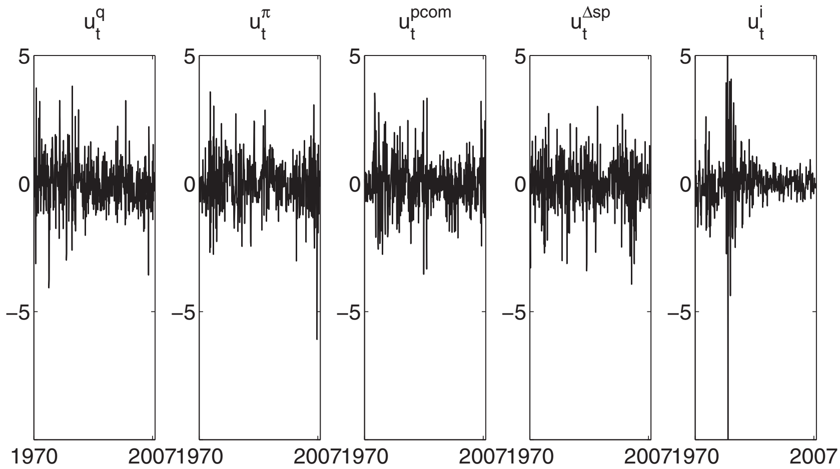


Figure 14.1. Transition function of ST-VAR(3) model with transition variable $s_t = \pi_{t-2}$.
Source: Lütkepohl and Netšunajev (2015).

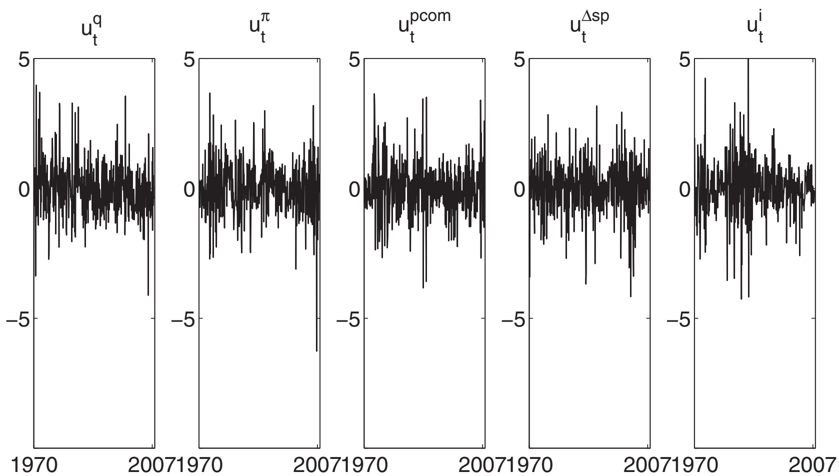
Empirical Examples. We consider an example from Lütkepohl and Netšunajev (2015) who investigate the interdependence between U.S. monetary policy and stock prices. They reconsider a study of Bjørnland and Leitemo (2009). Let $y_t = (q_t, \pi_t, pcom_t, \Delta sp_t, i_t)'$, where q_t is the linearly detrended log of an industrial production index, π_t is the annual change in the log of consumer prices scaled by 100, $pcom_t$ is the annual change in the log of the World Bank (non energy) commodity price index scaled by 100, Δsp_t is the monthly returns of the real S&P500 stock price index deflated by the consumer price index to measure the real stock prices, and i_t is the federal funds rate.

Lütkepohl and Netšunajev (2015) use monthly data for the period 1970m1-2007m6 and fit an ST-VAR(3) model with a logistic transition function and two different transition variables. We focus on the results obtained when lagged inflation is used as the transition variable. More precisely, the transition variable is $s_t = \pi_{t-2}$. Lütkepohl and Netšunajev justify this choice by observing that inflation is an important driving force of monetary policy and show that model selection criteria favor the second lag over the first.

They estimate the transition function shown in Figure 14.1 which indicates two periods of higher volatility, one in the mid-1970s and the other one around 1980. Clearly, when the transition function is close to zero, the error variance is close to Σ_1. For the periods in the mid-1970s and around 1980, the error variance is a convex combination of Σ_1 and Σ_2 with higher weight on the latter matrix for larger values of the transition function. It is also apparent in Figure 14.1 that the volatility in the two periods of higher volatility is quite different. The weights assigned to Σ_1 and Σ_2 are quite different in the two periods. This example shows the flexibility of this simple model.



(a) Standardized residuals of VAR(3) model



(b) Standardized residuals of ST-VAR(3) model, $s_t = \pi_{t-2}$

Figure 14.2. Residuals of the VAR(3) and ST-VAR(3) models.
Source: Lütkepohl and Netšunajev (2015).

In Figure 14.2 the standardized residuals of a reduced-form VAR(3) model that does not account for heteroskedasticity and the reduced-form ST-VAR(3) model are shown. The residuals are standardized by dividing them by the estimated standard deviation. For example, for the ST-VAR(3) model the k^{th} estimated residual, $\hat{u}_{kt}$, is divided by the square root of the k^{th} diagonal element of the estimated Σ_t covariance matrix. The standardized residuals of

Table 14.4. *Estimates of Relative Variances of Structural ST-VAR(3) Model with Transition Variable* $s_t = \pi_{t-2}$

Parameter	Estimate	Std.Dev.
λ_{12}	0.899	0.957
λ_{22}	2.739	1.608
λ_{32}	4.176	2.357
λ_{42}	8.091	3.108
λ_{52}	299.562	72.414

Source: Extracted from Table 2 of Lütkepohl and Netšunajev (2015).

the ST-VAR(3) model are seen to be overall more homogeneous throughout the sample than those of the standard VAR(3) model. Thus, the smooth-transition model captures the changes in the volatility at least to some extent.

Since the relative variances in Λ_2 are of central importance for the identification of the shocks, they are presented in Table 14.4 in increasing order. Except for the first element, all the estimated λ_{k2} are clearly larger than 1. Hence, the regime associated with Σ_2 represents a high volatility regime, confirming our previous interpretation of the transition function.

Full identification of all shocks by heteroskedasticity is achieved if all diagonal elements of Λ_2 are distinct. Based on the results in Table 14.4, it is difficult to argue that this condition is satisfied. Thus, we may not have a fully identified set of shocks. We may still use this model to test the validity of conventional identifying restrictions, however, because, under the null that these restrictions are valid, the model will be overidentified and the LR test will retain its asymptotic χ^2 distribution. The only difference is that the asymptotic distribution of the test statistic may have fewer degrees of freedom.

Bjørnland and Leitemo (2009) placed the stock price and monetary policy shocks last in the vector of structural shocks such that

$$w_t = (w_{1t}, w_{2t}, w_{3t}, w_{4t}^{sp}, w_{5t}^{m})',$$

where w_{4t}^{sp} and w_{5t}^{m} denote the stock market and monetary shocks, respectively. They used a combination of short-run and long-run identifying restrictions for the structural shocks. More precisely, they imposed the following restrictions on the matrix of impact effects, B_0^{-1}, and the matrix of long-run effects, $\boldsymbol{\Upsilon}$:

$$B_0^{-1} = \begin{bmatrix} * & 0 & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Upsilon} = \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & 0 \\ * & * & * & * & * \end{bmatrix}. \quad (14.3.16)$$

 https://doi.org/10.1017/9781108164818.015

Table 14.5. *Tests for Identifying Restrictions in Structural ST-VAR Models*

$\mathbb{H}_0$	$\mathbb{H}_1$	df	LR statistic	p-value
R1	unrestricted B_0^{-1}, $\boldsymbol{\Upsilon}$	10	25.811	0.004
R1	R2	1	7.494	0.006

Source: Extracted from Table 3 of Lütkepohl and Netšunajev (2017).

As before, asterisks indicate unrestricted elements and zeros denote elements restricted to zero. Thus, the stock market and monetary shocks are both assumed to have no impact effect on industrial production (q_t), inflation (π_t), and commodity prices ($pcom_t$). Moreover, the zero entry in the last column of $\boldsymbol{\Upsilon}$ means that monetary shocks have no long-run effects on the stock market. Since the first three shocks are not of interest for their analysis, Bjørnland and Leitemo (2009) identified them arbitrarily by the zero restrictions in the second and third columns of B_0^{-1}. Their identification does not have an impact on the last two shocks which are of central importance for the analysis (see Chapter 10).

Following Lütkepohl and Netšunajev (2015), we focus on the set of restrictions:

R1: Bjørnland-Leitemo identification with restrictions as in (14.3.16).
R2: Only B_0^{-1} restricted as in (14.3.16),

where the restrictions in R2 are a subset of the restrictions in R1. They use LR tests to test R1 against an unrestricted ST-VAR model and against R2. The p-values in Table 14.5 are smaller than 5%. The first test shows that the Bjørnland-Leitemo restrictions are rejected in favor of an unrestricted model. The second test reveals that imposing the additional long-run restriction on $\boldsymbol{\Upsilon}$ is problematic.

Despite the limited identifying information from heteroskedasticity, both null hypotheses are rejected. Of course, in that situation it is not justified to use these restrictions for impulse response analysis. On the other hand, using the shocks obtained by heteroskedasticity is not justified either, because, as we have seen before, they are not fully identified and because they have no natural economic interpretation. Thus, in this model it is difficult to disentangle the effects of stock market and monetary shocks. The only firm conclusion is that the previously found results have little empirical support when changes in volatility are taken into account.

14.3.4 Structural VAR Models with GARCH Errors

Yet another approach to modeling the conditional volatility of the VAR errors relies on a multivariate GARCH process for the VAR innovations. The

GARCH model is relevant for empirical work, given that the conditional heteroskedasticity of many economic and financial time series is well approximated by GARCH processes. Unlike models of discrete volatility changes, the GARCH model usually does not provide a clear partitioning of the sample period into periods of high and low volatility.

Model. In this setup, the structural shocks are assumed to be orthogonal and their conditional variances are modeled by individual GARCH(1, 1) processes. It is assumed that the conditional covariance matrix of u_t, given information up to period $t-1$, $\mathcal{F}_{t-1}$, is

$$\mathbb{E}(u_t u_t' | \mathcal{F}_{t-1}) = B_0^{-1} \Sigma_{t|t-1} B_0^{-1\prime}, \tag{14.3.17}$$

where $\Sigma_{t|t-1} = \text{diag}(\sigma^2_{1,t|t-1}, \dots, \sigma^2_{K,t|t-1})$ is a diagonal matrix. The individual conditional variances of the structural shocks $w_t = B_0 u_t$ are assumed to have a GARCH(1,1) structure of the form

$$\sigma^2_{k,t|t-1} = (1 - \gamma_k - g_k) + \gamma_k w^2_{k,t-1} + g_k \sigma^2_{k,t-1|t-2}, \quad k = 1, \dots, K, \tag{14.3.18}$$

where $\gamma_k, g_k \geq 0$. Higher-order GARCH processes can be considered in principle. This is typically not done in practice, however. The GARCH processes in (14.3.18) are set up such that the unconditional variances of the structural shocks are 1. Hence, $\mathbb{E}(w_t w_t') = I_K$ and the unconditional covariance matrix of the reduced-form errors is

$$\mathbb{E}(u_t u_t') = \Sigma_u = B_0^{-1} B_0^{-1\prime}.$$

A multivariate GARCH model of this type was proposed by van der Weide (2002) under the name of generalized orthogonal GARCH (GO-GARCH). It has been used by Normandin and Phaneuf (2004) and others for structural VAR analysis. There are also studies using the VAR-GARCH approach based on other types of GARCH models (e.g., Weber 2010; Strohsal and Weber 2015). We focus on the GO-GARCH setup because it has advantages in specifying and testing the identifying restrictions.

Identification. Sentana and Fiorentini (2001) and Milunovich and Yang (2013) provide identification conditions for this type of model. Let G be a $T \times K$ matrix with k^{th} column consisting of the conditional variances of the k^{th} structural shock, $(\sigma^2_{k,1|0}, \dots, \sigma^2_{k,T|T-1})'$. Sentana and Fiorentini (2001) show that B_0^{-1} is identified up to column signs and permutations, if G has full column rank. This condition is easy to verify in the unlikely case that the true conditional variances are known.

Milunovich and Yang (2013) show that, equivalently, identification requires at least $K-1$ of the GARCH processes being nontrivial in that $\gamma_k \neq 0$ for at least $K-1$ of the K processes in (14.3.18). In other words, identification requires sufficient heterogeneity in the conditional variances. To achieve full identification, at most one of the structural shocks may be homoskedastic. The latter condition may be investigated by formal statistical tests.

Since the GARCH structure offers additional identifying information, one may even give up the requirement of instantaneously uncorrelated shocks and replace it by other assumptions, as discussed in Section 14.4.2. For example, Weber (2010) assumes constant conditional correlations instead.

Estimation and Inference. To allow explicitly for the possibility that only $r < K$ structural shocks have a nontrivial GARCH(1, 1) structure, we write the volatility model as

$$u_t = B_0^{-1} \begin{bmatrix} \Lambda_{t|t-1}^{1/2} & 0 \\ 0 & I_{K-r} \end{bmatrix} \varepsilon_t, \tag{14.3.19}$$

where $\varepsilon_t \overset{iid}{\sim} (0, I_K)$ and

$$\Lambda_{t|t-1} = \begin{bmatrix} \sigma_{1,t|t-1}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{r,t|t-1}^2 \end{bmatrix} \tag{14.3.20}$$

is an $r \times r$ diagonal matrix with nontrivial univariate GARCH(1,1) processes on the diagonal. Thus, the structural shocks are

$$w_t = \begin{bmatrix} \Lambda_{t|t-1}^{1/2} & 0 \\ 0 & I_{K-r} \end{bmatrix} \varepsilon_t$$

and, conditionally on $\mathcal{F}_{t-1}$, the reduced-form errors have a distribution with zero mean and conditional covariance matrix

$$\Sigma_{t|t-1} = B_0^{-1} \begin{bmatrix} \Lambda_{t|t-1} & 0 \\ 0 & I_{K-r} \end{bmatrix} B_0^{-1\prime}.$$

Postulating a normal distribution for ε_t, Lanne and Saikkonen (2007) show how the log-likelihood function can be set up such that ML estimation of the model becomes feasible. They use a polar decomposition of B_0^{-1} defined as a decomposition such that $B_0^{-1} = CR$, where C is a symmetric, positive definite $K \times K$ matrix and $R = [R_1, R_2]$ is an orthogonal $K \times K$ matrix partitioned such that R_1 is $K \times r$ and R_2 is $K \times (K-r)$. Note that

$$\Sigma_u = B_0^{-1} B_0^{-1\prime} = CRR'C' = CC.$$

Thus, C is the unique square root matrix of Σ_u. Lanne and Saikkonen (2007) decompose the conditional covariance matrices as

$$\Sigma_{t|t-1} = \Sigma_u + CR_1(\Lambda_{t|t-1} - I_r)R_1'C$$

and show that

$$u_t' \Sigma_{t|t-1}^{-1} u_t = u_t' \Sigma_u^{-1} u_t + u_t' C^{-1} R_1 (\Lambda_{t|t-1}^{-1} - I_r) R_1' C^{-1} u_t.$$

The Gaussian log-likelihood function is

$$\log l = \sum_{t=1}^{T} \log f_{t|t-1}(u_t).$$

Using the previously given decomposition, the conditional densities have the form

$$\begin{aligned} f_{t|t-1}(u_t) &= (2\pi)^{-K/2} \det(\Sigma_{t|t-1})^{-1/2} \exp\left(-\frac{1}{2} u_t' \Sigma_{t|t-1}^{-1} u_t\right) \\ &= (2\pi)^{-K/2} \det(\Sigma_u)^{-1/2} \exp\left(-\frac{1}{2} u_t' \Sigma_u^{-1} u_t\right) \prod_{k=1}^{r} \sigma_{k,t|t-1}^{-1} \\ &\quad \times \exp\left(-\frac{1}{2} u_t' C^{-1} R_1 (\Lambda_{t|t-1}^{-1} - I_r) R_1' C^{-1} u_t\right). \end{aligned}$$

If there are just r nontrivial GARCH components, the matrix R_2 from the polar decomposition of B_0^{-1} is not identified. The log-likelihood does not depend on R_2, however, but only on the identified parameters $\nu, A_1, \ldots, A_p, C, R_1$, and the GARCH parameters γ_k, g_k $(k = 1, \ldots, r)$. Hence, for given r, the model can be estimated.

Of course, estimation of high-dimensional multivariate GARCH models by iterative optimization is not easy. In the present situation, the log-likelihood function factors into a component that is a function of C, and a component that is a function of C, R_1 and $\Lambda_{t|t-1}$. Therefore the estimation can be broken down into two steps. First, C is estimated as the square root matrix of an estimator of Σ_u, for example, based on the residuals from the LS estimation of the VAR part. In the second step the rows of R_1 and the GARCH equation parameters are estimated conditional on the estimated C and separately for $k = 1, \ldots, r$. Each equation $k + 1$ is estimated conditional on the previously estimated equation k. Thus, effectively only univariate GARCH models have to be estimated in this step. Details of this procedure can be found in Section 4 of Lanne and Saikkonen (2007). This approach results in inefficient estimators, but can be used to obtain initial values for a full likelihood maximization.

Clearly, the choice of r is crucial in this procedure. It is also important for the identification of the structural shocks. If $r < K - 1$ the matrix B_0^{-1} and, hence, the structural shocks w_t are not fully identified by the GARCH structure. Since in many macroeconomic studies variables are involved that may not be conditionally heteroskedastic, it cannot be taken for granted that r is at least equal to $K - 1$. In other words, it is important to have formal statistical procedures for exploring the number of nontrivial GARCH components that drive the volatility changes in the errors.

Suitable tests based on proposals by Lanne and Saikkonen (2007) are discussed in Lütkepohl and Milunovich (2016). For a given $r_0 < K$ the following pair of hypotheses is tested:

$$\mathbb{H}_0 : r = r_0 \quad \text{versus} \quad \mathbb{H}_1 : r > r_0.$$

Test statistics may be constructed as follows. Under $\mathbb{H}_0$, the $K \times r_0$ submatrix consisting of the first r_0 columns of B_0^{-1} can be estimated consistently. Note also that

$$B_0 = R'C^{-1} = \begin{bmatrix} R_1' \\ R_2' \end{bmatrix} C^{-1}.$$

Thus, the last $K - r$ components of w_t are $w_{2t} = R_2' C^{-1} u_t$.

Lanne and Saikkonen (2007) propose to estimate R_2 as an orthogonal complement of an estimator $\widehat{R}_1$ of R_1. More precisely,

$$\widehat{R}_2 = \widehat{R}_{1\perp}(\widehat{R}_{1\perp}' \widehat{R}_{1\perp})^{-1/2}.$$

Clearly, since R_2 is not identified, this estimator may not be consistent for R_2, but estimates some linear transformation of R_2. Thus, the estimator

$$\widehat{w}_{2t} = \widehat{R}_2' \widehat{C}^{-1} \widehat{u}_t,$$

is not necessarily an estimator of w_{2t}, but of some linear transformation of w_{2t}. However, if w_{2t} is not driven by GARCH components, the same is true of any linear transformation. Hence, tests for remaining GARCH in w_{2t} can be based on $\widehat{w}_{2t}$.

Under $\mathbb{H}_0$, the last $K - r_0$ components of w_t do not exhibit conditional heteroskedasticity. Therefore Lanne and Saikkonen (2007) propose test statistics based on the autocovariances of the univariate mean-adjusted sum of squared components of $\widehat{w}_{2t}$,

$$\xi_t = \widehat{w}_{2t}' \widehat{w}_{2t} - T^{-1} \sum_{i=1}^{T} \widehat{w}_{2i}' \widehat{w}_{2i},$$

and a related vector quantity consisting of the mean-adjusted squares and cross-products of the $\widehat{w}_{2t}$ components,

$$\vartheta_t = \text{vech}(\widehat{w}_{2t} \widehat{w}_{2t}') - T^{-1} \sum_{i=1}^{T} \text{vech}(\widehat{w}_{2i} \widehat{w}_{2i}').$$

The test statistics are

$$Q_1(H) = T \sum_{h=1}^{H} [\tilde{\gamma}(h)/\tilde{\gamma}(0)]^2, \tag{14.3.21}$$

where

$$\tilde{\gamma}(h) = T^{-1} \sum_{t=h+1}^{T} \xi_t \xi_{t-h},$$

and

$$Q_2(H) = T \sum_{h=1}^{H} \text{tr}[\tilde{\Gamma}(h)' \tilde{\Gamma}(0)^{-1} \tilde{\Gamma}(h) \tilde{\Gamma}(0)^{-1}], \qquad (14.3.22)$$

where

$$\tilde{\Gamma}(h) = T^{-1} \sum_{t=h+1}^{T} \vartheta_t \vartheta_{t-h}' \quad \text{for} \quad h = 0, 1, \dots$$

Lanne and Saikkonen (2007) show that $Q_1(H)$ and $Q_2(H)$ have asymptotic χ^2 distributions with H and $H(K - r_0)^2(K - r_0 + 1)^2/4$ degrees of freedom, respectively, under the null hypothesis.

Lütkepohl and Milunovich (2016) also propose a closely related LM type test based on the auxiliary model

$$\eta_t = \delta_0 + D_1 \eta_{t-1} + \cdots + D_H \eta_{t-H} + e_t, \qquad (14.3.23)$$

where $\eta_t = \text{vech}(\widehat{w}_{2t} \widehat{w}_{2t}')$ and e_t is an error term. The null hypothesis $\mathbb{H}_0 : r = r_0$ is equivalent to

$$\mathbb{H}_0 : D_1 = \cdots = D_H = 0$$

and the corresponding LM statistic is

$$LM(H) = \frac{1}{2} T(K - r_0)(K - r_0 + 1) - T \text{tr}[\tilde{\Sigma}_e \tilde{\Gamma}(0)^{-1}], \qquad (14.3.24)$$

where $\tilde{\Sigma}_e$ is the estimated residual covariance matrix from model (14.3.23). The statistic $LM(H)$ has the same asymptotic χ^2 distribution as $Q_2(H)$.

Lütkepohl and Milunovich (2016) compare the tests based on $Q_1(1)$, $Q_2(1)$, and $LM(1)$ in a simulation study and find that the performance of the tests very much depends on the persistence of the underlying GARCH processes (measured by $\gamma_k + g_k$) and the sample size. None of the tests uniformly dominates its competitors in terms of size and power. In fact, even for moderate sample sizes the power of the tests is very low, if the underlying GARCH processes are not very persistent. This result implies in particular that sequential testing procedures for determining the number of nontrivial GARCH components r are not very reliable and tend to end up with estimates of r that are lower than the true r_0. A plausible sequential testing procedure for the true number of GARCH components, r_0, tests null hypotheses $\mathbb{H}_0 : r = i$ for $i = 1, 2, \dots,$ until the first hypothesis is rejected. One may be tempted to test a reverse sequence of hypotheses, $\mathbb{H}_0 : r = j$, $j = K - 1, K - 2, \dots$. Such a test is problematic, however, because the true rank may be lower than the one specified under $\mathbb{H}_0$. A test of full identification can be set up by specifying $\mathbb{H}_0 : r = K - 2$. If that null hypothesis is rejected against $\mathbb{H}_1 : r > K - 2$, full identification is supported by the data because this indicates that the number r of nontrivial GARCH components is at least $K - 1$.

A drawback of the VAR-GARCH model is that the log-likelihood function is difficult to maximize globally for large models. The procedure of Lanne and Saikkonen (2007) described earlier can be applied, however, to obtain feasible first-stage estimates which can be used as starting values in the maximization.

Empirical Examples. Bouakez and Normandin (2010) examine the importance of U.S. monetary policy shocks for exchange rate dynamics using structural VAR-GARCH models. They consider eight-dimensional models for the U.S. and the G7 countries. They assume that seven univariate GARCH components drive the multivariate GARCH process and that the GARCH models fully identify the structural shocks. Based on this assumption, they conduct a number of tests of conventional identifying restrictions. Lütkepohl and Milunovich (2016) in turn test the statistical identifying assumptions underlying the VAR-GARCH model used by Bouakez and Normandin (2010).

For illustrative purposes we focus on one example from Lütkepohl and Milunovich (2016) based on an 8-dimensional model for the United States and Canada. Let $y_t = (q_t, p_t, pcom_t, nbr_t, tr_t, i_t, d_t, ex_t)'$, where q_t is the log of a U.S. industrial production index, p_t is the log of the U.S. consumer price index, $pcom_t$ is the log of a world export commodity price index, nbr_t is the log of U.S. nonborrowed reserves, tr_t is the log of U.S. total reserves, i_t is the federal funds rate, d_t is the difference between the Canadian short-term interest rate and the U.S. three-months Treasury bill rate, and ex_t is the log of the exchange rate (U.S. dollars per one Canadian dollar).

In testing conventional identifying restrictions, the question arises of how many GARCH components drive the volatility changes in this system. Financial variables such as i_t and ex_t are likely to exhibit GARCH dynamics, but some other variables may not, so identification cannot be taken for granted. Tests of the validity of the GARCH-based identification may be conducted based on the statistics (14.3.21), (14.3.22), and (14.3.24).

Using monthly data for the period 1982m11–2004m10, Lütkepohl and Milunovich (2016) obtain the test results presented in Table 14.6. Of particular interest is testing

$$\mathbb{H}_0 : r = 6 \quad \text{versus} \quad \mathbb{H}_1 : r > 6.$$

Rejecting that null hypothesis would be strong evidence of the full statistical identification of the shocks by the GARCH structure. However, the p-values of all three test statistics are greater than 0.6 and, hence, do not reject $\mathbb{H}_0$ at

Table 14.6. *p-Values for the Identification Tests Applied to U.S./Canadian Data from Bouakez and Normandin (2010)*

Test	$\mathbb{H}_0 : r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$
$LM(1)$	0.000	0.027	0.027	0.150	0.390	0.689	0.528
$Q_1(1)$	0.174	0.183	0.046	0.139	0.214	0.617	0.445
$Q_2(1)$	0.000	0.024	0.029	0.127	0.335	0.632	0.448

Source: Extracted from Table 9 of Lütkepohl and Milunovich (2016).

conventional significance levels. In fact, in Table 14.6 the last null hypothesis that can be rejected at a 5% significance level is $\mathbb{H}_0 : r = 3$. In other words, the tests provide evidence that there may be four GARCH components driving the eight variables. Even if one accounts for low power, it is difficult to argue that there is evidence for more than five GARCH components. Thus, we do not have support for the hypothesis of a fully identified model. Consequently, tests of conventional identifying restrictions have to be interpreted with caution. If such restrictions are rejected within a model with four or five GARCH components this is, of course, strong evidence against their validity. A failure to reject conventional exclusion restrictions, in contrast, could simply be due to not having enough identifying restrictions.

In another example of the structural VAR-GARCH approach, Normandin and Phaneuf (2004) compare identification schemes for monetary policy shocks based on U.S. monthly data for 1982m11–1998m12. They find that some of the assumptions that have been used in conventional analyses are rejected in the VAR-GARCH setup. However, they do not use formal identification tests and, hence, their results have to be interpreted with caution.

There are a number of other studies that utilize the VAR-GARCH model for structural analysis. Examples are Bouakez, Essid, and Normandin (2013) who study the effects of monetary policy shocks on the stock market and Bouakez, Chihi, and Normandin (2014) who examine the effects of fiscal policy shocks on the U.S. economy.

14.4 Alternative Approaches Using Heteroskedasticity

The approach of using heteroskedasticity in the errors of structural VAR models may be extended in several directions. First, as mentioned before, changes in the volatility of the shocks may coincide with changes in the impact effects of the shocks. Second, heteroskedasticity can also be used to dispense with the assumption of instantaneously uncorrelated shocks, albeit at the cost of the resulting shocks no longer being structural in the sense specified in Chapter 1. These two extensions are discussed next.

14.4.1 Time-Varying Instantaneous Effects

Bacchiocchi, Castelnuovo, and Fanelli (2013) and Bacchiocchi and Fanelli (2015) consider the possibility of time-varying instantaneous effects of the shocks. If there are two volatility states with covariance matrices Σ_1 and Σ_2, respectively, then the decomposition

$$\Sigma_1 = GG' \quad \text{and} \quad \Sigma_2 = (G + C)(G + C)' \tag{14.4.1}$$

may be used to obtain the structural shocks from $w_t = G^{-1}u_t$, if a given t corresponds to the first volatility regime, and, alternatively, from $w_t = (G + C)^{-1}u_t$, if that t belongs to the second volatility regime. Thus, G and $G + C$ are the impact effects matrices corresponding to the first and second regime, respectively. In other words, C represents the change in the impact effects of the structural shocks from regime 1 to regime 2.

Clearly, the relations in expression (14.4.1) are insufficient to identify all the elements of the two $K \times K$ matrices G and C. Therefore further identifying assumptions or restrictions are required. For example, both matrices may be assumed to be lower triangular. This problem is not alleviated if there are more than two volatility regimes. If the impact effects are allowed to vary freely across the volatility regimes, heteroskedasticity is not helpful for the identification of the shocks, because each additional volatility regime increases the number of restrictions required for full identification.

If such assumptions are easy to justify in some context, then using this approach may be useful. Note, however, that a change in the impact effects results in more general changes in the impulse response functions, even if the reduced-form VAR parameters are not regime dependent. Such a model is quite restrictive, because it will often be difficult to argue that the impact effects of the shocks are regime dependent, but the reduced-form parameters apart from the covariances do not change. If the reduced-form parameters are allowed to vary more generally, then the model becomes a time-varying coefficient model. Such models are discussed in more detail in Chapter 18. They are not considered further in the current chapter.

14.4.2 Correlated Shocks

Weber (2010) and Strohsal and Weber (2015) suggest that allowing for instantaneously correlated shocks in the structural model makes sense when working with low-frequency macroeconomic data. They note that if there are enough volatility regimes, the assumption of uncorrelated shocks can be replaced by some other assumption that limits the number of structural parameters. For example, if there are M volatility regimes, one could postulate a regime-invariant correlation structure and decompose the covariance matrices as

$$\Sigma_m = G\Lambda_m^{1/2} R \Lambda_m^{1/2} G', \quad m = 1, \ldots, M, \tag{14.4.2}$$

where R is a correlation matrix that is invariant across the volatility regimes and the Λ_m are again diagonal matrices with positive diagonal elements. If the structural shocks are obtained as $w_t = G^{-1}u_t$, they have covariance matrices

$$\mathbb{E}(w_t w_t') = G^{-1}\Sigma_m G^{-1\prime} = \Lambda_m^{1/2} R \Lambda_m^{1/2},$$

if t belongs to the m^{th} volatility regime. Thus, the diagonal elements of Λ_m are seen to be the variances of w_t in regime m.

If M is large enough and the covariance matrices satisfy suitable conditions, R and G can be recovered from equation (14.4.2), in which case the shocks are statistically identified. A drawback of this setup is that a decomposition such as that in (14.4.2) may not be easy to justify from an economic point of view and is unhelpful in answering the type of questions commonly analyzed using structural VAR models. Therefore, we do not elaborate on these models.

14.5 Identification by Non-Gaussianity

Identification by heteroskedasticity uses statistical properties of the data for identification. There have also been attempts to use other statistical properties of the data for this purpose. Notably, if the errors, u_t, of the reduced-form VAR process are not Gaussian, this feature can be used for identification. Studies that consider this device include Siegfried (2002), Lanne and Lütkepohl (2010), Moneta, Entner, Hoyer, and Coad (2013), Gourieroux and Monfort (2014), Lanne, Meitz, and Saikkonen (2017), and Herwartz (2015).

14.5.1 Independent Shocks

Gourieroux and Monfort (2014) state a mathematical result that can be used for uniquely identifying a set of structural shocks in a stationary VAR model with iid shocks. Related results are also provided in Lanne, Meitz, and Saikkonen (2017). Suppose $w = (w_1, \ldots, w_K)'$ and $w^* = (w_1^*, \ldots, w_K^*)'$ are both K-dimensional random vectors with independent components, and that at most one component is normally distributed, whereas all other components have nonnormal distributions. If there exists a nonsingular matrix C such that $w^* = Cw$, then $w_k^* = \gamma_k w_{\pi(k)}$, $k = 1, \ldots, K$. Here $\pi(\cdot)$ denotes a permutation of the numbers $1, \ldots, K$. In other words, the k^{th} component of w^* is a multiple of one of the components of w. Thus, if all components of w and w^* have unit variance such that both vectors have an identity covariance matrix, then C must be a permutation matrix that merely reorders the components of w, possibly changing their signs in the process.

This result implies that in a non-Gaussian environment, identification can be obtained by insisting that the structural shocks be stochastically independent rather than just uncorrelated. More precisely, if in a VAR analysis the structural shocks are obtained by a linear transformation of the reduced-form errors, $w_t = B_0 u_t$, and the structural shocks are mutually stochastically independent, have variance one, and at most one of them is normally distributed, then B_0 is unique except for row permutation and row sign changes. This follows by noting that any other K-dimensional random vector with independent components that is obtained by a linear transformation of w_t must be just a reordering of the components of w_t, possibly with reversed sign. Hence, the only linear transformations that preserve the independence of the components are of the form $\mathcal{P}B_0$, where $\mathcal{P}$ is a permutation matrix that permutes the rows of B_0. Hence, $\mathcal{P}^{-1}$ is also a permutation matrix and $B_0^{-1}\mathcal{P}^{-1}$ is the matrix B_0^{-1} with permuted columns. In other words, the matrix of impact effects is unique apart from column permutations and column sign changes.

Of course, this discussion assumes that there exists a linear transformation of the reduced-form residuals that results in a set of independent structural shocks. As illustrated later in this chapter, there are counterexamples of non-Gaussian processes in which no linear transformation results in independent shocks. Thus, we need to consider both linear and nonlinear transformations. The question then arises of how that transformation can be found. Siegfried (2002) and Gourieroux and Monfort (2014) point out that a method called independent component analysis (ICA) can be helpful in this context. Further references include Jutten and Herault (1991), Hyvärinen, Karhunen, and Oja (2001), and Stone (2004). Moneta, Entner, Hoyer, and Coad (2013) use this method for identifying mutually independent shocks.

An alternative procedure that restricts attention to linear transformations is proposed by Herwartz (2015). The premise is that structural shocks may be obtained by a linear transformation of the reduced-form shocks. Herwartz examines alternative rotations of the orthogonalized reduced-form shocks. He then tests the mutual independence of the elements of the implied vector of structural shocks for each rotation. Finally, he chooses the rotation that maximizes the p-value of the test. Using the rotation with the largest p-value amounts to choosing the structural errors that are least dependent. A possible drawback of this approach is that rotating uncorrelated shocks does not necessarily lead to independent innovations. In other words, the independent components may not be linear transformations of the uncorrelated residuals. An example of a model where independent errors cannot be determined by linear transformations of the reduced-form errors is considered in Section 14.5.2.

Herwartz and Plödt (2016) use the methodology proposed by Herwartz (2015) to investigate the dynamics in the market for crude oil. As in Kilian (2009) and Kilian and Murphy (2012), the variables are the change in global crude oil production, a measure of global real economic activity, and the real price of oil, but the data definitions and sample period differ from those used in Kilian (2009). Exploiting the nonnormality of the data and applying the procedure of Herwartz (2015), they find shocks that imply impulse responses similar to those in Kilian (2009). Hence, they label the shocks accordingly as an oil supply shock, an aggregate demand shock, and an oil-market specific demand shock.

Yet another procedure for exploiting non-Gaussianity has been discussed in Lanne, Meitz, and Saikkonen (2017). They propose an ML estimator of the transformation matrix that delivers independent errors w_t, conditional on additional assumptions on the distribution of u_t. For identification, Lanne et al. do not assume that the marginal error processes w_{kt} are serially independent, but just uncorrelated. In other words, $\mathbb{E}(w_{kt}w_{k,t+j}) = 0$ for $j \neq 0$, but w_{kt} and $w_{k,t+j}$ may not be independent. Thereby, their model allows, for example, for GARCH in the marginal processes w_{kt}.

Of course, as in the discussion of identification by heteroskedasticity, the shocks obtained in this way are based on purely mathematical properties of the underlying distributions. They will not have any economic interpretation, in general, without additional economic information. However, this framework may, in principle, be used for testing conventional identifying restrictions from models that do not exploit the nonnormality of the data. In the absence of such tests, one may still be able to infer the economic interpretation of these estimates, if the impulse responses of interest obtained using this procedure match the impulse response estimates from the corresponding VAR model identified by conventional restrictions.

14.5.2 Uncorrelated Shocks

Lanne and Lütkepohl (2010) use a quite different approach to take advantage of the nonnormality of the observations. They assume that the reduced-form residuals have a mixture normal distribution, i.e.,

$$u_t \sim \begin{cases} \mathcal{N}(0, \Sigma_1) & \text{with probability } \gamma, \\ \mathcal{N}(0, \Sigma_2) & \text{with probability } 1-\gamma, \end{cases} \tag{14.5.1}$$

where $0 < \gamma < 1$ is the mixing probability and $\Sigma_1 \neq \Sigma_2$. The class of normal mixture distributions is very broad and members of this class can approximate many continuous distributions. The mixture distribution in (14.5.1) implies that the reduced-form residuals, u_t, have mean zero and a covariance matrix

$$\mathbb{E}(u_t u_t') = \gamma\Sigma_1 + (1-\gamma)\Sigma_2.$$

Using the decomposition $\Sigma_1 = B_0^{-1}B_0^{-1\prime}$ and $\Sigma_2 = B_0^{-1}\Lambda B_0^{-1\prime}$, where B_0^{-1} is a nonsingular $K \times K$ matrix and Λ is a diagonal matrix with positive diagonal elements, and choosing structural shocks $w_t = B_0 u_t$, implies that w_t has instantaneously uncorrelated components. Moreover, as discussed earlier, B_0^{-1}

is unique up to column sign changes and column permutations if the diagonal elements of Λ are all distinct. Hence, using this device, the structural shocks are identified completely analogously to the case of identification via heteroskedasticity. Actually, the model is a special case of the Markov-switching model considered in Section 14.3.2, where the state of the process in period t is independent of the states in the previous periods. In other words, if s_t denotes the Markov process that specifies the state of the system, then the transition probabilities are

$$\mathbb{P}(s_t = m) = \mathbb{P}(s_t = m|s_{t-1} = 1) = \cdots = \mathbb{P}(s_t = m|s_{t-1} = M).$$

For example, for our model with two states we obtain a transition matrix

$$\begin{bmatrix} \gamma & \gamma \\ 1-\gamma & 1-\gamma \end{bmatrix}.$$

It is worth contrasting this model with the model of Gourieroux and Monfort (2014), where independence of the structural shocks is used as the identifying device. If in the normal mixture model the structural shocks are chosen as $w_t = B_0 u_t$, they are instantaneously uncorrelated, but generally not independent. Since independence implies uncorrelatedness, it is tempting to think of them as independent as well. Notice, however, that the w_t have a normal mixture distribution,

$$w_t \sim \begin{cases} \mathcal{N}(0, I_K) & \text{with probability } \gamma, \\ \mathcal{N}(0, \Lambda) & \text{with probability } 1-\gamma. \end{cases}$$

Clearly, if the state of one component of w_t is given, in general, this fact also determines the state of the other components. Hence, two components w_{kt} and w_{jt} are stochastically dependent. This example also illustrates that the assumption of independent structural errors is more restrictive than it may seem at first sight. In fact, the assumption that independent components can be obtained by a linear transformation of the reduced-form residuals is a very special case when considering nonnormal distributions.

The statistical analysis of the normal mixture model is straightforward. A VAR model is fitted by LS and the residuals are tested for nonnormality, as discussed in Chapter 2. If normality is rejected, the model with mixed normal residuals is fitted by ML. Since we assume a specific distribution, the likelihood function is readily available which makes ML estimation a natural choice. The diagonal elements of Λ are then investigated. If suitable tests suggest that they are all distinct, we have identified the structural shocks. Although such tests do not seem to be currently available, developing them may be possible.

The use of the mixed normal distribution is appealing, if the system consists of two different states that are characterized by two normal distributions and if the underlying shocks are mutually uncorrelated in either state. If instead the mixture normal distribution is just seen as a description of a general non-normal distribution, then the specific choice of transformation matrix for deriving the structural shocks from the reduced-form errors is arbitrary. In the latter case, this approach is merely a technical device and not a technique for identifying economically interpretable structural shocks. Therefore the interpretation of the shocks is problematic.

Lanne and Lütkepohl (2010) illustrate the use of these models by two examples. One of them considers the system that we already used to illustrate the MS-VAR analysis in Section 14.3.2 consisting of U.S. real output growth, changes in the real interest rate, and real stock returns. The other example investigates interest rate linkages between the U.S. and Europe.

14.6 Discussion

In this chapter we studied the question of how statistical properties of VAR errors can deliver additional information that is sufficient for the identification of mutually uncorrelated, but not necessarily economically interpretable shocks. Our analysis focused primarily on the use of volatility changes in the data. We also briefly discussed data-based identification methods exploiting the non-Gaussianity of many economic time series.

Although mutually uncorrelated shocks identified by heteroskedasticity or non-Gaussianity alone have no economic interpretation without additional assumptions, they can be useful in assessing the validity of conventional identifying restrictions that cannot be tested against the data in standard structural VAR models. Conventional just-identifying restrictions in this setting become overidentifying restrictions that can be formally tested.

We reviewed a number of models for changes in the volatility of the data that are used in the literature. If volatility changes are known to occur at some point in time, volatility breaks can be modeled by extraneously imposing these volatility changes. This situation is rare in practice, however. It is more likely that statistical pretests are used to determine the dates of the volatility changes. In addition, often it is not clear whether the change in volatility is abrupt or smooth. In the latter case, it is preferable to work with models that allow the volatility to evolve endogenously.

We discussed three different models that endogenize the changes in the volatility of the VAR innovations: Markov-switching, smooth transition, and VAR-GARCH models. The MS-VAR model is particularly useful if there is a finite number of volatility regimes and if the transition between these regimes can be described by a Markov process whose parameters can be estimated from the data. Although this MS-VAR model is often economically and empirically plausible, its main drawback is that it is difficult to estimate when there are many variables, long VAR lags, or a large number of volatility states. An alternative model, for which robust estimation algorithms exist, describes the change in volatility by a smooth transition function. Even if this model is based on two covariance states only, it can describe complex changes in volatility. Finally, we presented a VAR-GARCH model that allows the conditional variance of the errors to evolve over time, while preserving a constant unconditional variance. Estimation of this model can also be challenging for larger models.

If there is no prior knowledge on the structure of the volatility changes, it is difficult to decide between different volatility models. Using model selection criteria for that purpose may be a useful strategy. This approach has been employed by Lütkepohl and Netšunajev (2017), but there do not seem to be formal results on the properties of these criteria. In some cases, it may be advisable to consider different models and to investigate the robustness of the empirical results of interest.

Finally, we discussed how the non-normality of many economic time series may be used as a device for generating additional identifying information about the structural shocks. It can be shown that structural shocks being non-Gaussian implies the uniqueness of these shocks, provided that the structural shocks are assumed to be independent. We emphasized that such independent shocks, in general, cannot be obtained by a linear transformation of non-Gaussian reduced-form errors. There are alternative algorithms in the literature, but at this point no consensus has emerged on how to transform the reduced-form residuals in practice. An additional question that does not appear to have been discussed in the literature to date is how strong the degree of non-Gaussianity has to be for estimators derived in this way to be reliable.