EUI WORKING PAPERS

Structural Vector Autoregressions
with Nonnormal Residuals

MARKKU LANNE AND
HELMUT LÜTKEPOHL

EUROPEAN UNIVERSITY INSTITUTE

Department of Economics

Structural Vector Autoregressions with Nonnormal Residuals

MARKKU LANNE AND HELMUT LÜTKEPOHL

Structural Vector Autoregressions with Nonnormal Residuals[1]

Markku Lanne

European University Institute,
University of Jyväskylä, RUESG and HECER
School of Business and Economics,
P.O. Box 35, FIN-40014 University of Jyväskylä, FINLAND

and

Helmut Lütkepohl

Department of Economics, European University Institute
Via della Piazzuola 43, I-50133 Firenze, ITALY

Abstract. In structural vector autoregressive (SVAR) models identifying restrictions for shocks and impulse responses are usually derived from economic theory or institutional constraints. Sometimes the restrictions are insufficient for identifying all shocks and impulse responses. In this paper it is pointed out that specific distributional assumptions can also help in identifying the structural shocks. In particular, a mixture of normal distributions is considered as a plausible model that can be used in this context. Our model setup makes it possible to test restrictions which are just-identifying in a standard SVAR framework. In particular, we can test for the number of transitory and permanent shocks in a cointegrated SVAR model. The results are illustrated using a data set from King, Plosser, Stock and Watson (1991) and a system of US and European interest rates.

Key Words: Mixture normal distribution, cointegration, vector autoregressive process, vector error correction model, impulse responses

JEL classification: C32

---

1 Introduction

Structural vector autoregressive (SVAR) models are a standard tool for empirical economic analysis. The basic underlying model is usually a vector autoregressive (VAR) or vector error correction (VEC) model in reduced form. If one wants to use these models for impulse response analysis, structural information is required to identify the relevant shocks and impulse responses. Such information usually comes from economic theory or from structural and institutional knowledge related to a specific model setup or the variables involved. If some of the variables are integrated or cointegrated, long-run restrictions may also be derived from the cointegration properties of the data. In some cases there is not enough information from such sources, however, to fully identify all shocks and impulse responses. In that case different plausible restrictions are sometimes considered and the robustness of the main results with respect to uncertain identifying assumptions may be investigated.

In this study it is argued that distributional assumptions can be helpful in identifying shocks of interest and in particular they may be substituted for missing identifying information from other sources. We use an idea put forward by Lanne and Saikkonen (2005) in the context of multivariate GARCH models and consider a mixture of normal distributions for the error terms of our basic model. In VAR analyses it is not uncommon that nonnormal residuals are found. Thus, it is plausible to specify more general distributions explicitly. A mixture of normal distributions is plausible, for instance, if there are different regimes operating within the sample period, one with a smaller and one with a larger variance. For example, if there is a period with high volatility or if there are some outliers which may be generated by a different distribution than the remaining observations such a model is appealing. Another example may be a system which reacts differently in expansionary periods and recessions.

In the following, conditions will be discussed which ensure identification of shocks if they have a mixture of two normal distributions. We will also discuss how such identifying restrictions can be combined with restrictions from other sources. Thereby it becomes possible to test restrictions which are just-identifying in the usual SVAR framework. For example, if cointegrated systems are considered, the number of shocks with transitory effects is often assumed to be identical to the number of cointegration relations and, accordingly, the number of permanent shocks equals the number of common trends. In our framework such an assumption can be checked by testing the implied restrictions even if the shocks are not identified in a standard SVAR model. We will discuss tests for the number of transitory and permanent

shocks in a cointegrated SVAR model and we will also consider tests of other restrictions which are just-identifying in the standard setup.

Two examples will be considered to illustrate SVAR modelling with mixture normal residuals. The first one is based on a well-known data set from King et al. (1991) consisting of three US macroeconomic variables. It primarily serves to illustrate some advantages of the mixed normal SVAR model relative to the standard approach. The second example considers a data set from Brüggemann and Lütkepohl (2005) consisting of two European and two US interest rates. We will be able to test some of the assumptions of the previous analysis and find that they are not supported in our modelling framework. As in the previous study, we find support for the view that US monetary policy has a stronger impact on European interest rates than vice versa.

Our study is structured as follows. In the next section the general model setup is presented and identifying restrictions are discussed. Estimation of the models is considered in Section 3 and the two examples are presented in Section 4. Extensions and conclusions are provided in Section 5. A result related to mixture normal distributions is given in the Appendix.

2 The Model Setup

2.1 The Reduced Form

Consider the following n-dimensional reduced form VAR model of order p,

$$A(L)y_t = W w_t, \tag{2.1}$$

where $A(L) = I_n - A_1 L - \cdots - A_p L^p$ is a matrix polynomial in the lag operator L with $(n \times n)$ coefficient matrices A_j $(j = 1, \ldots, p)$ and I_n denotes the $(n \times n)$ identity matrix. Here W is a nonsingular $(n \times n)$ parameter matrix and the n-dimensional error term w_t is a mixture of two serially independent normal random vectors such that

$$w_t = \begin{cases} e_{1t} \sim \mathcal{N}(0, I_n) \text{ with probability } \gamma, \\ e_{2t} \sim \mathcal{N}(0, \Psi) \text{ with probability } 1 - \gamma. \end{cases} \tag{2.2}$$

The parameter matrix Ψ is a diagonal matrix, that is, $\Psi = \text{diag}(\psi_1, ..., \psi_n)$ with positive diagonal elements ψ_j $(j = 1, \ldots, n)$. Notice that the jth component of w_t has a standard normal distribution if $\psi_j = 1$. Hence, there may be some components of w_t which do not have a mixture distribution and, in fact, $w_t \sim \mathcal{N}(0, I_n)$ if $\Psi = I_n$. In other words, a model with normal errors is a special case of our model setup. Notice also that w_t has mean zero and

covariance matrix $\gamma I_n + (1 - \gamma)\Psi$, that is, $w_t \sim (0, \gamma I_n + (1 - \gamma)\Psi)$. In our setup the mixture probability γ, $0 < \gamma < 1$, is also a parameter of the model. This kind of mixture normal distribution was used in a multivariate GARCH modelling context by Lanne and Saikkonen (2005).

In (2.1) deterministic terms are neglected for simplicity. They can be added easily to the model without affecting the essential parts of the following discussion. We have dropped them from the model because they do not have a role in structural modelling and impulse response analysis.

If some of the variables are $I(1)$, then a VEC version of the VAR process is often more useful in structural analysis because it separates the long-run from the short-run movements of the variables and therefore makes it easy to impose restrictions on the long-run behavior of some or all of the shocks. The VEC representation of the model (2.1) is (see Lütkepohl (2005, Sec. 6.3))

$$\Delta y_t = \alpha\beta'y_{t-1} + \Gamma_1\Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + Ww_t, \qquad (2.3)$$

where $\Delta = 1 - L$ is the differencing operator, β is an $(n \times r)$ cointegration matrix with cointegrating rank $r < n$, α is an $(n \times r)$ loading matrix for the cointegration relations, that is, $\alpha\beta'y_{t-1}$ is the error correction term with long-run relations $\beta'y_t$ and the Γ_j are $(n \times n)$ short-run coefficient matrices. The parameters in (2.3) may be obtained from $A(L)$ in (2.1) by rearranging terms such that $A(L) = I_n - \alpha\beta'L - \Gamma_1\Delta L - \cdots - \Gamma_{p-1}\Delta L^{p-1}$.

2.2 Structural Short-run Restrictions

The structural shocks, say ε_t, are usually defined such that they are zero mean uncorrelated random variables with unit variances, that is, $\varepsilon_t \sim (0, I_n)$. In a so-called B model setup they are related to the reduced form errors $u_t = Ww_t$ by the relation

$$u_t = B\varepsilon_t$$

(see Lütkepohl (2005, Chapter 9) for a detailed introductory account of these models). Clearly, $E(u_t u_t') = \Sigma_u = BB'$ and B is not unique in general. Thus, to identify the structural shocks, restrictions have to be imposed on B. Because the elements in B represent the instantaneous effects of the shocks on the variables, economic theory or institutional knowledge sometimes suggests zero restrictions for B. In other words, some shocks do not have an instantaneous impact on some of the variables.

A popular way to choose B is to consider a triangular matrix which may be obtained from a Choleski decomposition of Σ_u. The shocks then have a recursive structure where for a lower triangular matrix B the first shock can have an instantaneous impact on all the variables, whereas the

3

second one may only influence the second to last variables instantaneously and so on. Although such a triangular orthogonalization of the residuals may occasionally be justified on theoretical grounds, it is sometimes chosen because no firm theoretical constraints are available.

If no restrictions are present for B but a mixture of normals can be justified for the reduced form errors, local identification can be obtained according to the proposition given in the Appendix if all diagonal elements of Ψ are different. In that case, W and $\Sigma_w = E(w_t w_t') = \gamma I_n + (1 - \gamma)\Psi$ can be estimated and $B = W\Sigma_w^{-1/2}$. Local identification is all we can hope for in SVAR modelling because all signs in a column of B can always be reversed without changing the product BB'.

Occasionally there are firm restrictions which identify some but not all of the shocks. Suppose that m shocks are identified by directly restricting B or, equivalently, W and assume without loss of generality that they are placed in the first m positions of w_t. Then, for just-identification of all shocks, only the last $n - m$ components of w_t need to have mixture normal distributions and the first m elements may have a multivariate standard normal distribution. In that case

$$E(w_t w_t') = \begin{bmatrix} I_m & 0 \\ 0 & \gamma I_{n-m} + (1 - \gamma)\Psi_{n-m} \end{bmatrix}, \tag{2.4}$$

where Ψ_{n-m} is an $((n-m) \times (n-m))$ diagonal matrix with distinct diagonal elements. Thereby all structural shocks ε_t are identified by the Proposition in the Appendix. Note, however, that identification is obtained even if Ψ_{n-m} has one diagonal element which is one as long as all diagonal elements are different. If w_t has the covariance matrix given in (2.4), the first m components of ε_t are identical to the corresponding components of w_t whereas the last $n - m$ components of ε_t are obtained from the last $n - m$ components of w_t by dividing by the respective standard deviations (the diagonal elements of $[\gamma I_{n-m} + (1 - \gamma)\Psi_{n-m}]^{1/2}$).

If some of the first m components also have mixture normal distributions, the shocks are actually over-identified. Such a model nests the one implied by (2.4) and restrictions can be tested as long as the shocks remain identified. For example, the restrictions $\psi_1 = \cdots = \psi_m = 1$ can be tested. Moreover, if w_t has a general normal mixture distribution with covariance matrix $\gamma I_n + (1-\gamma)\Psi$, where all diagonal elements of Ψ are distinct, identifying restrictions for the shocks from other sources can be tested.

Although we have just discussed the B model for ease of exposition, similar considerations apply for the more general so-called AB-model of Amisano and Giannini (1997) (see also Lütkepohl (2005, Chapter 9)). In that case the structural shocks are such that $u_t = A^{-1}B\varepsilon_t$ and identifying restrictions may

be placed on both A and B. In the previous discussion, B may just be replaced by $A^{-1}B$.

2.3 Models with Cointegrated Variables and Long-run Restrictions

If some variables are integrated, the long-run effects of the structural shocks can be shown to be ΞB in the B-model, where

$$\Xi = \beta_\perp \left[\alpha'_\perp \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_\perp \right]^{-1} \alpha'_\perp. \tag{2.5}$$

The symbols $\alpha_\perp$ and $\beta_\perp$ denote orthogonal complements of α and β, respectively (see Johansen (1995) for the derivations and Lütkepohl (2005, Chapter 9) for an introductory discussion). Because α and β both have rank r, the matrix Ξ and, hence, ΞB must have rank $n - r$. It follows that at most r of the shocks may have transitory effects only and, hence, they are associated with zero columns in the long-run matrix ΞB. If such an assumption can be justified, zero restrictions on the long-run matrix may be imposed and used for identifying the structural shocks. In some cases this may even result in just-identified shocks. If there is more than one transitory shock or more than one permanent shock, additional restrictions will be needed, however. If they are not available from other sources, the mixture normal distribution may be helpful again.

Suppose that there are r transitory shocks, ε_t^t, and $n - r$ permanent shocks, ε_t^p, and they are arranged such that $\varepsilon_t' = (\varepsilon_t^{p\prime}, \varepsilon_t^{t\prime})$. Hence, $\Xi B = [\Phi_{n \times (n-r)} : 0_{n \times r}]$, where $\Phi_{n \times (n-r)}$ is an $(n \times (n - r))$ matrix. Furthermore, suppose that there are no other zero restrictions available for B and ΞB. Then identification of the shocks is obtained according to the Proposition in the Appendix if the corresponding subvectors of w_t, say w_t^p and w_t^t have independent mixture normal distributions with covariance matrices $E(w_t^p w_t^{p\prime}) = \gamma I_{n-r} + (1 - \gamma) \Psi_{n-r}$ and $E(w_t^t w_t^{t\prime}) = \gamma I_r + (1 - \gamma) \Psi_r$, respectively, where Ψ_{n-r} and Ψ_r must both have distinct diagonal elements, whereas some of the diagonal elements of Ψ_{n-r} may be the same as those of Ψ_r.

If w_t has a fully general mixture normal distribution as in (2.2), the zero constraints on ΞB are, in fact, over-identifying restrictions which can be tested. Hence, in this case we can test for the number of transitory and permanent shocks. Denoting the number of transitory shocks by r^*, we can, for example, test the null hypothesis $H_0 : r^* = r$ or, equivalently, $H_0 : \Xi B = [\Phi_{n \times (n-r)} : 0_{n \times r}]$ against the alternative hypothesis $H_1 : r^* < r$, that is, ΞB is unrestricted. Alternatively, we may test against $H_1 : r^* = r - 1$,

that is, $\Xi B = [\Phi_{n \times (n-r+1)} : 0_{n \times (r-1)}]$. We can also test a sequence of null hypotheses $H_0 : r^* = r$, $H_0 : r^* = r - 1$, ..., $H_0 : r^* = 1$ to determine the number of transitory shocks. The testing sequence stops and the number of transitory shocks is chosen accordingly if one of the null hypotheses cannot be rejected.

3 Estimation

Because a specific distribution for the reduced form error term is used, maximum likelihood (ML) is a plausible estimation method. The error term w_t has density

$$\phi_w(w_t) = \gamma (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}w_t'w_t\right\}$$
$$+ (1 - \gamma)(2\pi)^{-n/2} \det(\Psi)^{-1/2} \exp\left\{-\frac{1}{2}w_t'\Psi^{-1}w_t\right\}$$

and, neglecting the constant terms, the conditional distribution of y_t given y_{t-1}, y_{t-2}, ..., has a density

$$f_{t-1}(y_t) = \gamma \det(W)^{-1}$$
$$\times \exp\left\{-\frac{1}{2}(A(L)y_t)'(WW')^{-1}(A(L)y_t)\right\}$$
$$+ (1 - \gamma)\det(\Psi)^{-1/2}\det(W)^{-1}$$
$$\times \exp\left\{-\frac{1}{2}(A(L)y_t)'(W\Psi W')^{-1}(A(L)y_t)\right\}.$$

Collecting all the parameters in the vector ϑ, the log-likelihood function can be written as

$$l_T(\vartheta) = \sum_{t=1}^{T} \log f_{t-1}(y_t),$$

where an additive constant is dropped. If ϑ is identified, $l_T(\vartheta)$ can be maximized with standard nonlinear optimization algorithms.

For stationary processes we can appeal to standard ML theory and conclude that the estimators have the usual limiting properties, that is, they are consistent and asymptotically normal. Restrictions can be tested by likelihood ratio (LR) tests using the usual χ^2 limiting distributions if the model is identified under both the null and alternative hypotheses.

For VEC models the cointegration relations may be estimated in a first step by the Johansen (1995) reduced rank regression which is equivalent to

ML for a Gaussian model but is just a quasi ML procedure under the present mixture normal assumption. In a second step we may then condition on the estimated cointegration relations and maximize the log likelihood with respect to the other parameters. Such a procedure will be used in the next section where examples are considered.

4   Examples

In this section the previously discussed results will be illustrated with two examples. The first one uses a small system of US macro variables from King et al. (1991). The second example considers a system of four interest rate series from the US and Europe. It uses data which were analyzed earlier by Brüggemann and Lütkepohl (2005).

4.1   US Macro Model

The first example is based on quarterly US data for the period $1947Q1 -$ $1988Q4$ from King et al. (1991) for the three variables log consumption (c_t), log investment (i_t) and log private output (q_t) (all multiplied by 100).[2] These series were also used as an example by Lütkepohl (2005, Chapter 9) who treated all of them as $I(1)$ and found evidence for $r = 2$ cointegration relations. He fitted a VEC model with a constant term, cointegrating rank $r = 2$ and one lagged difference of the variables, that is, $p - 1 = 1$ in a model such as (2.3). Moreover, only one shock was assumed to have permanent effects. This restriction identifies the permanent shock while an additional restriction is required for identifying the two transitory shocks. There is no guidance from economic theory for the present system how to impose such a restriction. Therefore using some distributional assumption to obtain identification may be an option. In fact, standard Jarque-Bera nonnormality tests reject the normality hypothesis for the residuals of the present model. Hence, there is some support for using a nonnormal residual distribution.

We have used the cointegration relations obtained from a reduced rank regression and we have estimated several models with mixture normal distributions conditionally on these cointegration relations.[3] Some results are presented in Table 1. Since some of the models are over-identified, we can actually test the restrictions. Test results are presented in Table 2.

Model (1) is identified by distributional assumptions only. The estimated $\tilde{\psi}_i$'s are all different although, given their standard errors, it is not clear that

---

[2]The data are available at the website http://www.wws.princeton.edu/∼mwatson/.

[3]All computations were done with GAUSS programs.

they are actually significantly different. For illustrative purposes we will assume, however, that Model (1) is identified. Adding the restriction that the first shock is the only one with permanent effects while the second and last shocks have transitory effects only, the model is over-identified and the restriction that two shocks are transitory can be tested. The test of Model (2) versus Model (1) is the first one reported in Table 2. Based on the $\chi^2(2)$ null distribution the p-value is 0.34 and hence the restriction is clearly not rejected at common significance levels. Notice that the six zero restrictions imposed on ΞB to exclude permanent effects of the last two shocks account for two linearly independent restrictions only because the matrix ΞB has rank one (see Lütkepohl (2005, Section 9.2)). Therefore there are only two degrees of freedom in the limiting null distribution.

Under our assumptions we can also test whether one of the diagonal elements of Ψ is one and, thus, the corresponding residual has actually a normal distribution. Recall that the model is identified if all diagonal elements are distinct. Thus, anyone of the elements may be one as long as only one of them has this particular value. In Models (3) and (4) we have restricted one of the ψ_j's to one. Testing these models against Model (2), $\psi_1 = 1$ is clearly rejected, whereas $\psi_2 = 1$ is not. Given the estimates and their standard errors in Model (2), this outcome is not surprising. Note, however, that the first shock is identified by the assumption that it is the only permanent one. Thus, rejecting $\psi_1 = 1$ indicates that the permanent shock is not well modelled with a normal distribution.

The final four tests in Table 2 check zero restrictions on the instantaneous effects, that is, zero restrictions on the last two columns of B. In a classical model setup with normal residuals one such restriction is required to just-identify the two transitory shocks. Hence, such a restriction cannot be tested in that framework. In the present mixture normal setup a test becomes possible, however, because the restrictions are now over-identifying. In Model (5) the second transitory shock cannot have an instantaneous impact on consumption (the first variable in the system), that is, $b_{13} = 0$. Testing this restriction together with the constraint that the last two shocks are transitory against a model which is identified purely by the mixture normal distribution (Model (5) versus Model (1)), the restrictions clearly cannot be rejected at common test levels. Moreover, testing $b_{13} = 0$ only (Model (5) against Model (2)) the restriction is not rejected at the 5% level.

In contrast, testing that the third shock has no instantaneous effect on output, the third variable in our system ($b_{33} = 0$) is clearly rejected at a 5% level of significance. The p-values of both tests, (6) versus (1) and (6) versus (2) are clearly smaller than 0.05. Thus, utilizing the nonnormality of the residual distributions allows us to discriminate between restrictions which

cannot be tested in a classical normal SVAR setup.

It may also be of interest to compare the impulse responses implied by the different identifying assumptions for the shocks. Impulse response functions corresponding to some of the models in Table 1 are therefore depicted in Figures 1 - 3 with 95% confidence intervals. The confidence intervals are based on 2000 bootstrap replications using the method referred to as Hall's percentile method by Benkwitz, Lütkepohl and Wolters (2001). These confidence intervals are presented here because they have a better theoretical foundation than the ones more commonly used in impulse response analysis.

In Figure 1 impulse responses obtained in a standard SVAR setup with normal residuals are depicted for comparison purposes. Here the identifying restrictions are the same as in Model (5), except that the residuals are now treated as normally distributed. Hence, the restrictions are just-identifying here and they were not rejected in the mixture normal setup. The corresponding impulse responses from Model (5) with mixture normal residuals are depicted in Figure 2. Not surprisingly they look very similar to those in Figure 1 because they satisfy the same identifying long-run and short-run restrictions. They are just estimated using a different likelihood function which allows for mixture normal residuals. The main implication is that the confidence intervals around the impulse responses in Figure 2 tend to be a little wider than the corresponding ones in Figure 1. This is a reflection of the more general residual distribution underlying Figure 2. Exceptions to this general outcome are the confidence intervals for the responses of i and q to the second transitory shock which are wider under the normality assumption (see $\varepsilon^{t2} \rightarrow i$ and $\varepsilon^{t2} \rightarrow q$). In particular the responses of investment have much wider confidence intervals initially (note the difference in scales in Figures 1 and 2, respectively).

Using the mixture normal distribution the transitory shocks remain identified if the restriction $b_{13} = 0$ is removed. The resulting impulse responses based on Model (2) are depicted in Figure 3. Apart from the fact that the second transitory shock now has an instantaneous impact on consumption, nothing much has changed relative to Figure 2. This outcome is expected, of course, because the restriction $b_{13} = 0$ was not rejected by the LR test in Table 2. Notice, however, that the instantaneous response of consumption to the second transitory shock is significant when judged on the basis of the bootstrap confidence interval although it was not significant at the 5% level when Model (5) was tested against Model (2) with the LR test. Because we do not know much about the small sample properties of our statistical methods it is an advantage that we do not have to rely on these procedures and can relax the restriction $b_{13} = 0$ in our setup.

As a final remark regarding this example it may be worth emphasizing

9

again that the main purpose of the previous discussion is to illustrate the issues related to identifying shocks in structural VEC models by a mixture normal distribution. The model may not be a perfect one for the present data set because there may be some autocorrelation left in the residuals. Also, if the mixture normal distribution is in fact a good one for the residuals, the inference procedures used in specifying the model may not be the best ones. For example, results of Lucas (1997) and Boswijk and Lucas (2002) suggest that Johansen's LR tests may not be the best tools for determining the cointegrating rank in this case. Developing better instruments for other stages of a structural VEC analysis under nonnormality assumptions is beyond the scope of this paper, however.

4.2   US and European Interest Rates

Brüggemann and Lütkepohl (2005) considered euro area and US short-term and long-term interest rates and found support for both the expectations hypothesis of the term structure and the uncovered interest rate parity. More precisely, they analyzed four monthly interest rate series for the period $1985M1 - 2004M12$. The series are a euro area three months money market rate r_t^{EU}, a euro area 10-year bond rate R_t^{EU}, a US three months money market rate r_t^{US} and a US 10-year bond rate R_t^{US}.[4] Brüggemann and Lütkepohl (2005) found that all four variables are $I(1)$ whereas the two spreads $R_t^{US} - r_t^{US}$ and $R_t^{EU} - r_t^{EU}$ as well as the two parities $R_t^{US} - R_t^{EU}$ and $r_t^{US} - r_t^{EU}$ are stationary and, hence, there are three linearly independent cointegration relations in the system of four series.

Therefore Brüggemann and Lütkepohl (2005) used a VEC model for $y_t = (R_t^{US}, r_t^{US}, R_t^{EU}, r_t^{EU})'$ with a constant term, three lags of Δy_t and a cointegrating rank of $r = 3$ to investigate the impact of monetary shocks in the US and in Europe. Because the cointegrating rank is $r = 3$ they restricted three shocks to have transitory effects only. Thereby they identified the permanent shock. It is not obvious why there should be three transitory shocks, however. The fact that there are three cointegration relations implies that there can be at most three transitory shocks. Of course, there could be fewer such shocks. Moreover, there is no firm theory that suggests how to identify the transitory shocks. Because, based on standard Jarque-Bera tests, there is some evidence that the model residuals are nonnormal, using a more general distribution is plausible. Admittedly, the nonnormality may be due to ARCH effects in the present system of variables. A mixture

---

[4]The European interest rate series are constructed from German interest rates until the end of 1998 and the corresponding euro area rates afterwards. For more details on the data and their sources see the data appendix.

of normal distributions may still not be implausible here because it can represent observations coming from two regimes with different volatility which is not untypical for interest rate series. In fact, such periods with different volatility are actually present in the series under investigation here, as can be seen in Figure 4, where the first differences of all four series are plotted.

We have estimated different models with mixture normal residuals and present some results in Table 3, where the number of transitory shocks (i.e., the number of zero columns in ΞB) is denoted by r^*. Clearly, the $\tilde{\psi}_i$'s are quite different in all models, although they are not in all cases significantly different in the sense that two-standard error confidence intervals around the estimates would overlap. This is true in particular for the fully unrestricted Model (1). Thus, it is unclear whether the model is actually identified, that is, the W matrix may not be locally unique. Admittedly, we don't know much about the small sample properties of our estimators and in the following we assume again that the underlying ψ_i's are distinct so that Model (1) is identified.

Notice that some of the $\tilde{\psi}_i$'s in Model (1) are in fact quite large. Three of them are substantially larger than one. Hence, thinking of the two mixing normal distributions as representing different regimes, the regime represented by $\mathcal{N}(0, \Psi)$ is clearly one with higher volatility in three of the four components than the one represented by a standard normal $\mathcal{N}(0, I_4)$ distribution. The unnormalized w_t and the normalized ε_t shocks are identified purely by their stochastic properties in Model (1). Therefore it is difficult or impossible to interpret them as economic shocks without further assumptions or restrictions. Hence, in the following we will investigate different sets of restrictions which may help to attach economic meaning to the shocks.

When some of the shocks are restricted to be transitory as in Models (2) - (6), the $\tilde{\psi}_i$'s partly change considerably, depending on the positioning of the permanent and transitory shocks. In all the models one of the $\tilde{\psi}_i$'s is substantially larger than the others. In other words, one of the shocks has a much larger variance than the others. For example, in Model (1) $\tilde{\psi}_2 = 128.8$ which is more than 10 times larger than the second largest $\tilde{\psi}_i$. In contrast, when only one impulse is allowed to have permanent effects as in Model (2), the largest $\tilde{\psi}_i$ shifts to the first position which corresponds to the permanent shock. In other words, the unnormalized permanent shock is the one with the largest variance. In this case the difference in size of the $\tilde{\psi}_i$'s is not quite as dramatic as in Model (1) which suggests that the observations are associated with the two regimes in a different way in Models (1) and (2), respectively. This is also reflected in the fact that the estimated mixture probabilities $\tilde{\gamma}$ are quite different for the two models. We will see shortly, however, that Model (2) is rejected by the data. In fact, all models which

11

are not rejected by the data in the following tests (Models (3), (4) and (5)) have one very large $\tilde{\psi}_i$ corresponding to a transitory shock and the estimated mixture probabilities are quite similar. They are in fact identical up to two significant digits ($\tilde{\gamma} = 0.32$ for Models (1), (3), (4) and (5)). We will turn to tests of the restrictions imposed on the models now.

Based on the results in Table 3 we can test for the number of transitory shocks as explained in Section 2.3. The relevant results are given in Table 4. Both tests, $H_0 : r^* = 3$ versus $H_1 : r^* < 3$ and $H_0 : r^* = 3$ versus $H_1 : r^* = 2$ have very small p-values and clearly reject at common significance levels. In other words, three transitory shocks are clearly rejected. Testing a model with only two transitory shocks, the result is more favorable. Both alternative tests have p-values that indicate compatibility of two transitory shocks with the data in our setup.

If one accepts that there are two transitory shocks and one places them in the last positions of the vector of shocks, the discussion in Section 2.3 suggests that identification is ensured if $\psi_1 \neq \psi_2$ and $\psi_3 \neq \psi_4$. Given the estimates and their standard errors, this may well hold in Model (3). In any case, two-standard error confidence intervals around the respective estimates do not overlap.

In the lower part of Table 4 we also give results of tests of other restrictions of potential interest. Clearly, as three transitory shocks are rejected by the data, it does not make sense to test the validity of the identifying restrictions for the transitory shocks used by Brüggemann and Lütkepohl (2005). Because monetary policy shocks are sometimes thought of as being transitory (e.g., Evans and Marshall (1998)), the two transitory shocks in our system may represent monetary policy shocks in Europe and the US, respectively. Without further restrictions they could, of course, both be mixtures of such shocks. One way to associate them uniquely with one of the two currency areas would be to impose suitable restrictions. Therefore we used models with two transitory shocks and restricted the instantaneous impact of one of them to be zero for the US interest rates (Model (5)) and the other one has no instantaneous impact on the European interest rates (Model (6)). Clearly, the US monetary shock should be allowed to have an instantaneous impact on US interest rates. Thus, restricting the first transitory shock (the third shock in the ε_t vector) to have no instantaneous effect on US interest rates it cannot be the US monetary policy shock and, hence, must be the European monetary shock if indeed each of the two transitory shocks represents one of the monetary policy shocks. Consequently, the third shock in our system is viewed as the European monetary policy shock and the fourth shock is regarded as US monetary policy shock.

In Table 4 the corresponding restrictions are tested and it turns out that

the constraints imposed on the instantaneous impacts of the first transitory shock cannot be rejected at the 10% level of significance whereas the restrictions on the effects of the second transitory shock are clearly rejected at common significance levels. Thus, Model (5) is the preferred one and the first transitory shock can clearly be associated with European monetary policy. Moreover, European monetary policy shocks do not seem to affect US interest rates instantaneously. On the other hand, US monetary policy has an immediate impact on European interest rates because there is no transitory shock which affects only US interest rates instantaneously. Note, however, that in Model (5) which is not rejected by the data, $\tilde{\psi}_4$ is much larger than all other $\tilde{\psi}_i$'s (see Table 3). Thus, the last transitory shock is composed of two regimes one of which has a much larger volatility than the other one. Notice in Figure 4 that in the late 1980s the volatility in the US short-term interest rate is larger than in other periods. This larger volatility may be captured by the high variance regime of the fourth shock. In roughly the same period there is also higher volatility in the EU short-term rate. Thus, it is possible that the fourth shock captures a mixture of US and EU monetary policy. On the other hand, the increased volatility in the European interest rate may be due to increased volatility in the US. Therefore we will assume in the following that the last shock represents a US monetary policy shock.

The finding that US monetary policy may have an instantaneous impact on European interest rates but not vice versa is in line with the conclusion of Brüggemann and Lütkepohl (2005) that US interest rates have a stronger impact on European monetary policy than vice versa. In their framework they could not formally test this result, however. Although also in our framework some assumptions are necessary (e.g., we assume that monetary policy shocks are transitory), statistical tests can carry us one step further in checking restrictions that are not over-identifying in the standard framework and, hence, cannot be tested in that setting.

We have also computed the impulse responses for the two transitory shocks and show them in Figure 5 together with 95% bootstrap confidence intervals. It turns out that a contractionary EU shock leads to an increase in the US long-term rate, although not instantaneously. There is no significant reaction of the US short-term rate, however. In contrast, a contractionary US monetary shock has a significant instantaneous and longer lasting impact on the EU short-term rate and there is also a potentially significant impact on the EU long-term rate. Thus, overall the impulse responses are in line with the conclusion that US monetary policy may be more important for Europe than vice versa.

5 Conclusions and Extensions

In this paper we have used distributional assumptions for the residuals of a VAR model to identify some or all shocks to be used for an impulse response analysis. Specifically we have used a mixture of two normal distributions to obtain fully identified shocks. We have also shown how such nonnormal error distributions can be combined with restrictions from other sources to identify the shocks and impulse responses. For example, they can be combined with restrictions derived from the cointegration properties of a system of variables. Two empirical examples have been used to illustrate the virtue of the approach for applied work.

Although in practice it will be easy to justify nonnormal distributions for many econometric models, our approach has some limitations that deserve further consideration in future work. First of all, we have considered mixtures of two normal distributions only. This may be reasonable if two different regimes are a plausible assumption for the sample period. Sometimes it may be more natural to allow for more than two regimes, however, and, hence, mix more than two normal distributions. Second, we have allowed for just one mixture probability. In other words, the mixture probabilities are assumed to be identical for all components of the vector of innovations. Again this may not always be a realistic assumption. Third, although we have argued that a mixture normal distribution is often a plausible extension of the usual normal distribution, there may be alternative appealing distributions which are worth considering in this context. For example, Siegfried (2002) argues that monetary policy shocks may be leptokurtic and he considers a logistic distribution. He also presents an algorithm for estimating independent rather than just uncorrelated shocks. Generally, the potential of other distributions for identifying the shocks in a structural VAR model may be worth investigating. Finally, for the structural analysis a reduced form model has to be found which describes the data generation process well. Specification and inference procedures which account for residuals with mixture normal distributions may be worth exploring. Moreover, given the large number of parameters in full VAR models, ML estimation may be a computational challenge in some cases and it may be worth considering other estimation methods. We leave these issues for future research.

Appendix

In this appendix we present a proposition which is useful in dealing with the mixture normal distribution.

Proposition. Let $u = Ww$, where W is a nonsingular $(n \times n)$ fixed matrix and the random vector w is a mixture of two normal random vectors such that

$$w = \begin{cases} e_1 \sim \mathcal{N}(0, I_n) \text{ with probability } \gamma, \\ e_2 \sim \mathcal{N}(0, \Psi) \text{ with probability } 1 - \gamma \end{cases}$$

with $0 < \gamma < 1$ and $\Psi = \mathrm{diag}(\psi_1, ..., \psi_n)$ is a diagonal matrix with positive diagonal elements. Then the columns of W are uniquely determined up to multiplication by -1 if and only if all ψ_j are mutually different. $\square$

Proof: Because $u \sim (0, W[\gamma I_n + (1 - \gamma) \Psi] W')$, we have to show that W is unique in the covariance term up to multiplication of its columns by -1.

To show the "if" part of the proposition, let Q be a matrix such that

$$W[\gamma I_n + (1 - \gamma) \Psi] W' = WQ[\gamma I_n + (1 - \gamma) \Psi] Q'W'. \qquad (A.1)$$

We have to show that the only feasible Q matrix is a diagonal matrix with diagonal elements ± 1. Multiplying $(A.1)$ from the left by W^{-1} and from the right by its transpose gives

$$\gamma(I_n - QQ') = (1 - \gamma)(Q\Psi Q' - \Psi). \qquad (A.2)$$

This holds for all γ from the unit interval only if both sides are zero and, hence, $QQ' = I_n$. In other words, Q has to be orthogonal. Moreover, $\Psi = Q\Psi Q'$ or, equivalently, $\Psi Q = Q\Psi$. Denoting the ijth element of Q by q_{ij}, the last matrix equality means that $\psi_i q_{ij} = \psi_j q_{ij}$ and, hence, $q_{ij} = 0$ for $i \neq j$ because $\psi_i \neq \psi_j$. Consequently, Q must be a diagonal matrix with ± 1 on the diagonal because the diagonal elements of a diagonal matrix are its eigenvalues and the eigenvalues of an orthogonal matrix are all ± 1.

It is also easy to show that W is not unique if at least two of the diagonal elements of Ψ are equal. Hence, the "only if" part follows. Thereby the proposition is proven.

Data Appendix

The data used in the example of Section 4.2 are the same ones used by Brüggemann and Lütkepohl (2005). In the following we reproduce the data sources given in that article.

Monthly data for the period $1985M1 - 2004M12$ are used. Euro area interest rate series correspond to German interest rates for the period $1985M1 - 1998M12$ and to euro area interest rates for the period $1999M1 - 2004M12$. Monthly values are averages over all business days. The data are taken from the sources listed below:

1. **US long term interest rate** (R^{US}): 10-year T-Bill rate taken from FRED II database. Series ID: GS10.

2. **US short term interest rate** (r^{US}): 3-month money market rate taken from FRED II database. Series ID: CD3M.

3. **Euro area long term interest rate** (R^{EU}): $1985M1-1998M12$: 10-year government bond rate (*Umlaufsrendite*) taken from *Deutsche Bundesbank* database. Series ID: WU0004. $1999M1-2004M12$: 10-year euro area government benchmark bond yield. Source: ECB monthly bulletin, T04.07_c5.

4. **Euro area short term interest rate** (r^{EU}): $1985M1-1998M12$: 3-month money market rate taken from *Deutsche Bundesbank* database. Series ID: SU0107. $1999M1-2004M12$: 3-month money market rate (EURIBOR). Source: ECB monthly bulletin, T04.06_c3.

References

Amisano, G. and Giannini, C. (1997). *Topics in Structural VAR Econometrics*, 2nd edn, Springer, Berlin.

Benkwitz, A., Lütkepohl, H. and Wolters, J. (2001). Comparison of bootstrap confidence intervals for impulse responses of German monetary systems, *Macroeconomic Dynamics* **5**: 81–100.

Boswijk, P. and Lucas, A. (2002). Semi-nonparametric cointegration testing, *Journal of Econometrics* **108**: 253–280.

Brüggemann, R. and Lütkepohl, H. (2005). Uncovered interest rate parity and the expectations hypothesis of the term structure: Empirical results for the U.S. and Europe, *Applied Economics Quarterly* **51**: 143–154.

Evans, C. L. and Marshall, D. A. (1998). Monetary policy and the term structure of nominal interest rates: Evidence and theory, *Carnegie-Rochester Conference Series on Public Policy* **49**: 53–111.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

King, R. G., Plosser, C. I., Stock, J. H. and Watson, M. W. (1991). Stochastic trends and economic fluctuations, *American Economic Review* **81**: 819–840.

Lanne, M. and Saikkonen, P. (2005). A multivariate generalized orthogonal factor GARCH model, *HECER Discussion Paper 63*, University of Helsinki.

Lucas, A. (1997). Cointegration testing using pseudo likelihood ratio tests, *Econometric Theory* **13**: 149–169.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Siegfried, N. A. (2002). An information-theoretic extension to structural VAR modelling, *Discussion paper*, Hamburg University.

Table 1: Estimation results for US macro VEC model with cointegrating rank $r = 2$, one lagged difference and unrestricted intercept term (sample period: $1947Q1 - 1988Q4$)

Model no.	Restrictions	$\tilde{\gamma}$	$\tilde{\psi}_1$	$\tilde{\psi}_2$	$\tilde{\psi}_3$	max $l_T(\vartheta)/T$
(1)	–	0.46 (0.11)	0.77 (0.28)	0.20 (0.07)	0.15 (0.05)	−1.568
(2)	$\Xi B = [\Phi : 0_{3\times2}]$	0.47 (0.11)	0.19 (0.07)	0.79 (0.28)	0.16 (0.05)	−1.574
(3)	$\Xi B = [\Phi : 0_{3\times2}]$	0.42 (0.14)	1	0.16 (0.05)	0.64 (0.25)	−1.620
(4)	$\Xi B = [\Phi : 0_{3\times2}]$	0.47 (0.11)	0.18 (0.07)	1	0.17 (0.05)	−1.575
(5)	$\Xi B = [\Phi : 0_{3\times2}], b_{13} = 0$	0.41 (0.12)	0.27 (0.09)	0.17 (0.05)	0.57 (0.19)	−1.583
(6)	$\Xi B = [\Phi : 0_{3\times2}], b_{33} = 0$	0.45 (0.15)	0.19 (0.06)	0.26 (0.09)	0.44 (0.16)	−1.598

Note: Standard errors in parentheses obtained from the inverse Hessian of the log-likelihood function.

Table 2: LR tests based on results from Table 1

Hypotheses		LR statistic	Assumed	p-value
H_0	H_1		distribution	
(2)	(1)	2.18	$\chi^2(2)$	0.34
(3)	(2)	15.39	$\chi^2(1)$	0.00
(4)	(2)	0.43	$\chi^2(1)$	0.51
(5)	(1)	5.15	$\chi^2(3)$	0.16
(5)	(2)	2.97	$\chi^2(1)$	0.08
(6)	(1)	10.27	$\chi^2(3)$	0.02
(6)	(2)	8.09	$\chi^2(1)$	0.00

Table 3: Estimation results for interest rate VEC model with cointegrating rank $r = 3$, three lagged differences and unrestricted intercept term (sample period: $1985M1 - 2004M12$)

Model no.	Restrictions	$\tilde{\gamma}$	$\tilde{\psi}_1$	$\tilde{\psi}_2$	$\tilde{\psi}_3$	$\tilde{\psi}_4$	max $l_T(\vartheta)/T$
(1)	$r^* = 0$	0.32 (0.04)	0.60 (0.13)	128.8 (32.0)	7.80 (2.20)	5.01 (1.16)	5.1505
(2)	$r^* = 3$ ($\Xi B = [\Phi : 0_{4\times3}]$)	0.48 (0.05)	22.00 (5.19)	5.57 (1.73)	0.71 (0.16)	7.47 (1.96)	5.1146
(3)	$r^* = 2$ ($\Xi B = [\Phi : 0_{4\times2}]$)	0.32 (0.04)	0.60 (0.14)	5.39 (1.56)	6.61 (2.05)	129.5 (35.7)	5.1472
(4)	$r^* = 1$ ($\Xi B = [\Phi : 0_{4\times1}]$)	0.32 (0.04)	0.60 (0.13)	7.74 (2.17)	5.00 (1.17)	130.7 (33.5)	5.1497
(5)	$r^* = 2$, $b_{13} = b_{23} = 0$	0.32 (0.04)	0.61 (0.14)	6.72 (1.91)	5.42 (1.43)	140.5 (45.2)	5.1384
(6)	$r^* = 2$, $b_{34} = b_{44} = 0$	0.49 (0.05)	0.73 (0.17)	18.46 (4.32)	5.80 (1.78)	8.46 (2.42)	5.1094

Note: Standard errors in parentheses obtained from the inverse Hessian of the log-likelihood function.

Table 4: LR tests based on results from Table 3

Hypotheses		LR statistic	Assumed distribution	p-value
H_0	H_1			
$r^* = 3$	$r^* < 3$	17.24	$\chi^2(3)$	0.0006
$r^* = 2$	$r^* < 2$	1.57	$\chi^2(2)$	0.46
$r^* = 3$	$r^* = 2$	15.65	$\chi^2(1)$	7.63×10^{-5}
$r^* = 2$	$r^* = 1$	1.20	$\chi^2(1)$	0.27
(5)	(3)	4.22	$\chi^2(2)$	0.12
(6)	(3)	18.13	$\chi^2(2)$	0.0001
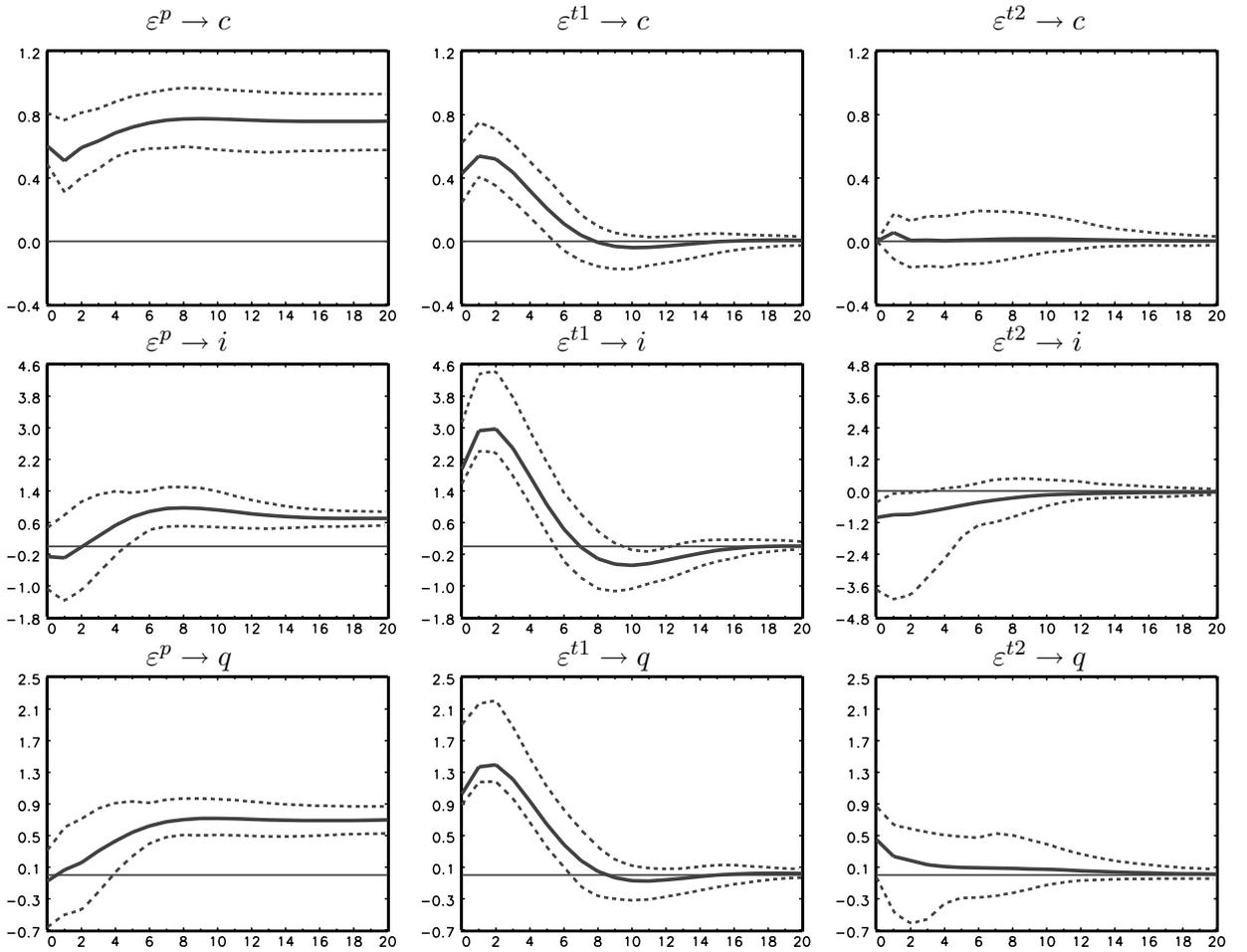
Figure 1: Impulse responses of output, consumption, and investment with 95% Hall percentile bootstrap confidence intervals based on 2000 bootstrap replications (identification: normal residuals, $\Xi B = [\Phi : 0_{3\times2}]$, $b_{13} = 0$)
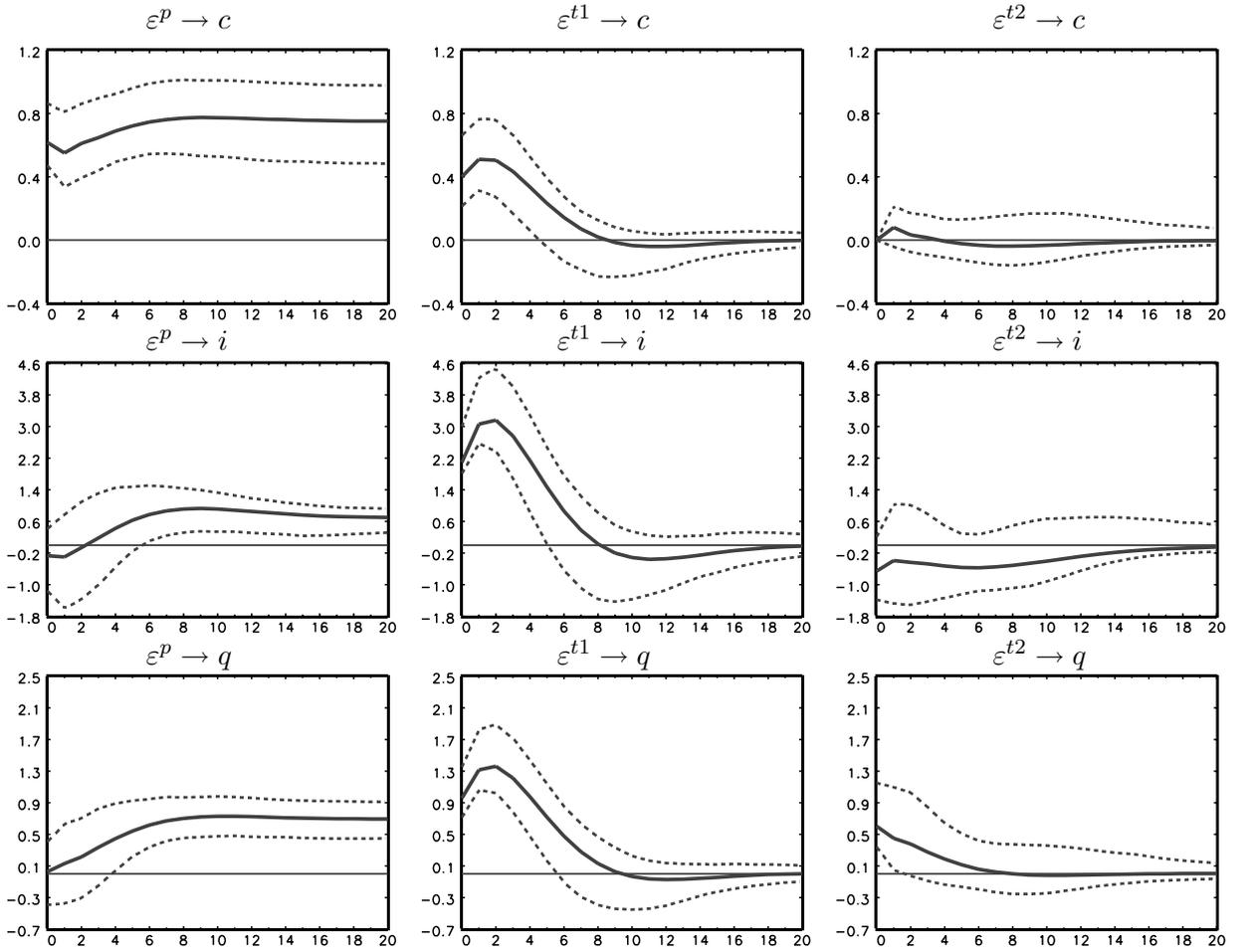
Figure 2: Impulse responses of output, consumption, and investment from Model (5) with 95% Hall percentile bootstrap confidence intervals based on 2000 bootstrap replications (identification: mixture normal distribution, $\Xi B = [\Phi : 0_{3\times 2}]$, $b_{13} = 0$)
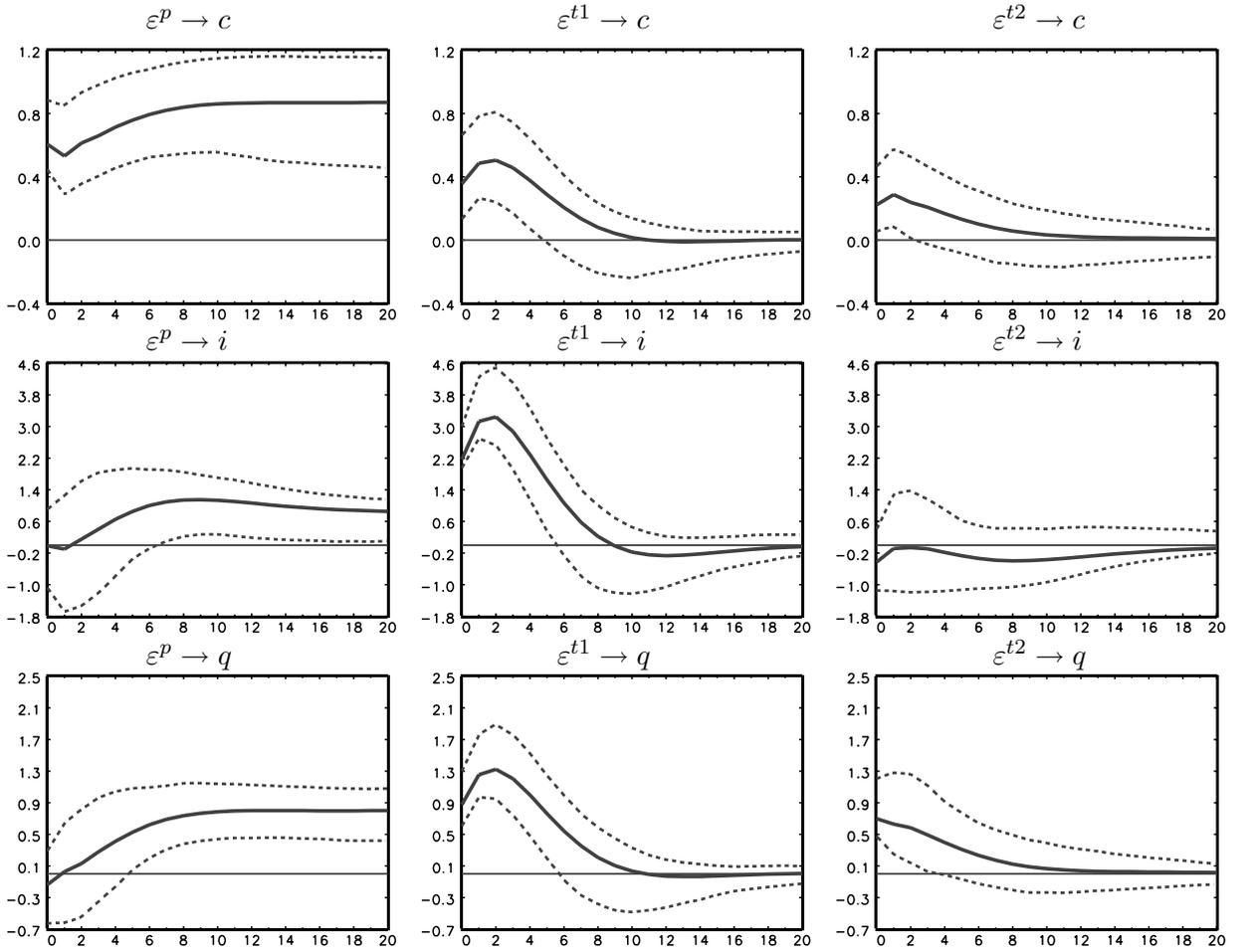
Figure 3: Impulse responses of output, consumption, and investment from Model (2) with 95% Hall percentile bootstrap confidence intervals based on 2000 bootstrap replications (identification: mixture normal residuals, $\Xi B = [\Phi : 0_{3\times 2}]$)
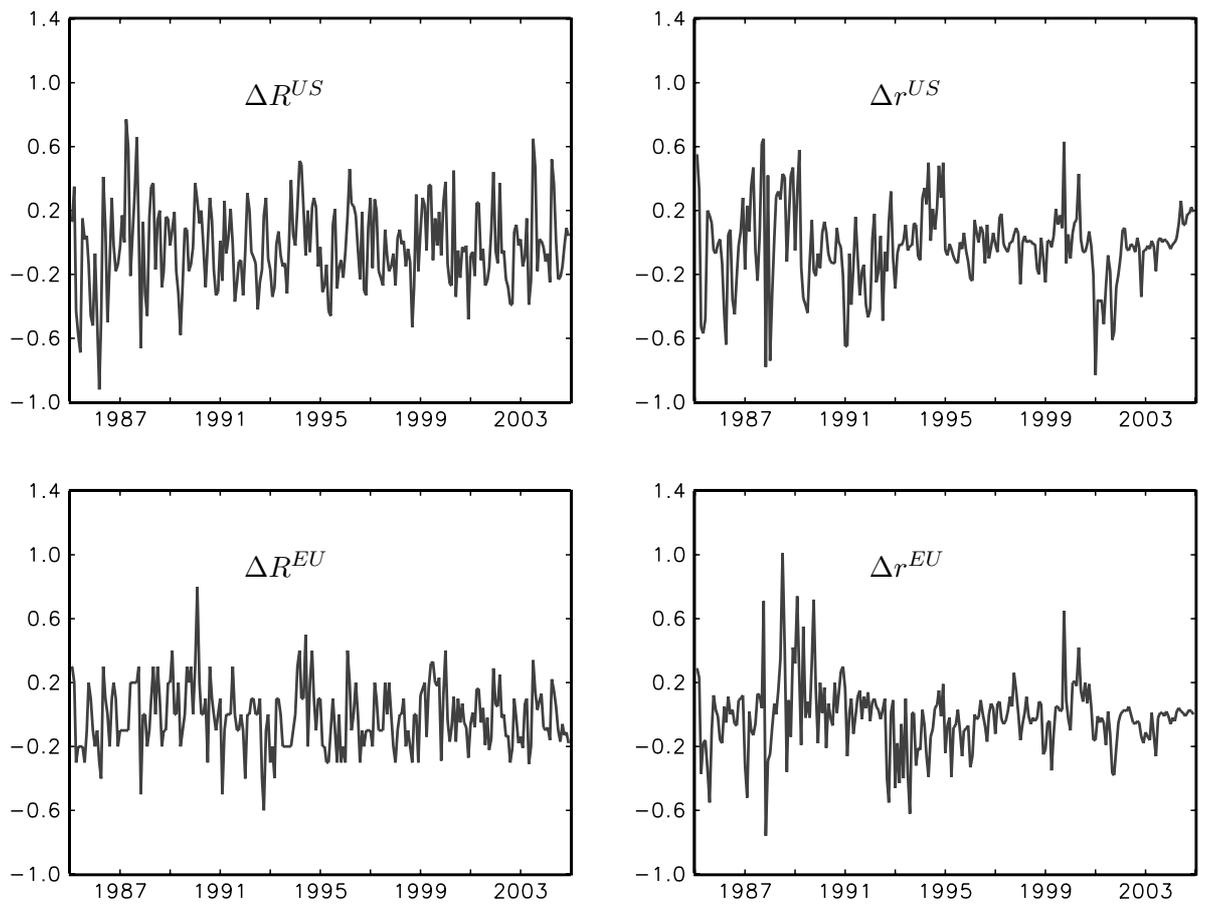
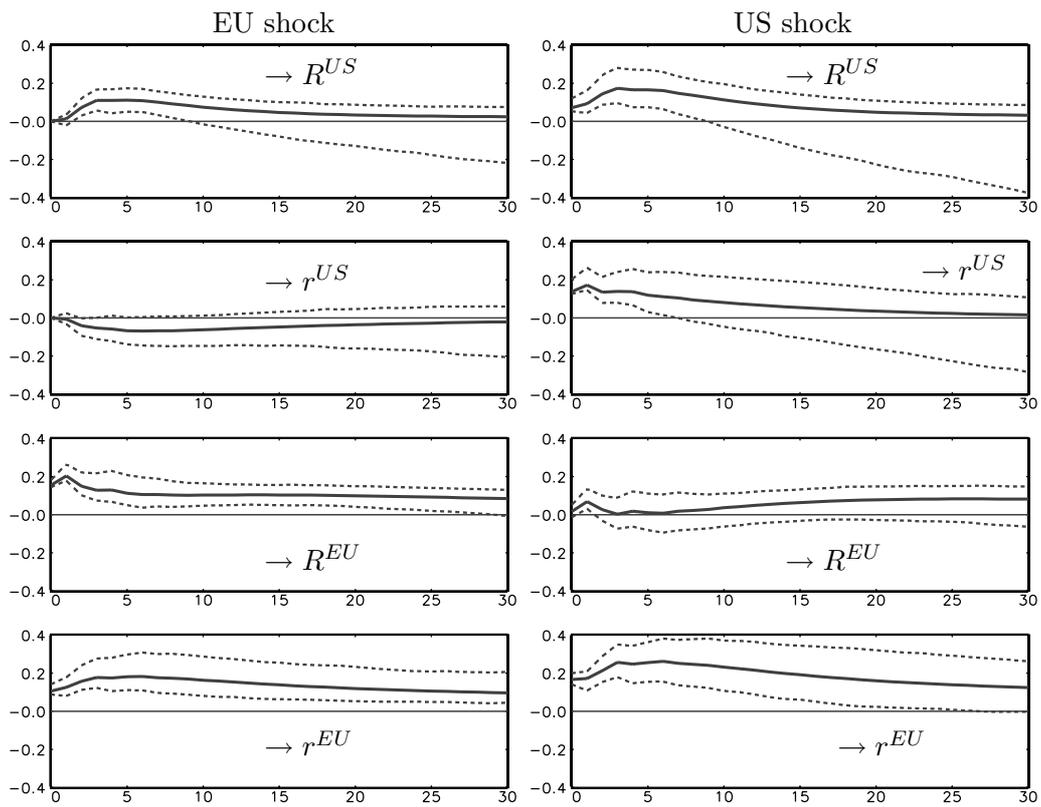Figure 4: First differences of interest rate series.

Figure 5: Responses of interest rates to EU and US monetary shocks based on Model (5) from Table 3 with 95% Hall percentile bootstrap confidence intervals based on 2000 bootstrap replications.