An information-theoretic extension to structural VAR modelling

Nikolaus A. Siegfried*

March 2002

Preliminary Version
Please Do Not Quote

*Hamburg University, Department of Economics, Von Melle Park 5, 20146 Hamburg, Germany, nas29@columbia.edu Fon: (+49)40-42838-6407, Fax: (+49)40-42838-5546

An information-theoretic extension to structural VAR modelling

Abstract

This paper discusses techniques for estimating structural vector autoregressions. Especially when monetary policy shocks are estimated, VAR residuals turn out to be leptokurtic. It is argued that this is no coincidence but follows directly from the properties of monetary policy decisions.

The paper proceeds to suggest an independent components estimator (ICE) that works well with leptokurtic residuals. Furthermore, the ICE permits a closer link between theory and estimation because it avoids informal imposition of zero restrictions. Using the exercises by Blanchard & Quah (1989) and Christiano, Eichenbaum & Evans (1999), the new estimator is demonstrated and contrasted with current modelling techniques.

Keywords: Structural Vector Autoregressions, Information Theory, Monetary Transmission Mechanism

JEL classification: C13, C32, E51

1 Introduction

Structural vector autoregression (SVAR) models have become a major tool of analysing macroeconomic time series, especially for investigating the effects of monetary policy on the economy, or the "monetary transmission mechanism". In a SVAR, a group of economic variables is interpreted as being driven exclusively by unobservable economic shocks. In order to recover these shocks from the data, the researcher estimates a VAR by maximum likelihood and decomposes the residuals into economically meaningful *structural shocks*. This paper discusses econometric techniques for estimating SVARs, especially the role of long-run restrictions. I also suggest a new technique that extends the current method of SVAR modelling for the case of non-normal residuals.

The decomposition of the residuals is based on the assumption of normality of the unobserved structural shocks. In this special case the second-order condition of no correlation, or orthogonality, implies independence. Therefore, the SVAR approach imposes orthogonality on the structural shocks. However, although independence implies orthogonality, the converse is only true for the special case of a Gaussian distribution. Frequently, presentations of the SVAR technique gloss over this nuance (e.g. Lütkepohl 1993, p. 48). Especially the monetary policy shocks many researchers are interested in frequently turn out to be leptokurtic. Using the exercise of Christiano et al. (1999) as an example I demonstrate and give an intuition for this finding.

After orthogonalising the residuals, additional assumptions are needed for identification. On the one hand, the number of structural shocks is assumed to be equal to the number of variables in the system. On the other hand, additional neutrality assumptions about the impulse responses, i.e. the dynamic responses of the economic variables to the unobservable structural shocks, are imposed. For this purpose, the researcher uses economic theory to restrict the impulse responses of the shocks. Early studies used to rely on contemporary restrictions, e.g. that a monetary policy shock has no immediate influence on output. This method is described in section 2.

In their seminal paper, Blanchard & Quah (1989) have extended the SVAR approach so that now long-run restrictions, e.g. long-run neutrality of monetary policy or a long-run vertical Phillips curve, can be incorporated. In the wake of increased interest in cointegration, Stock & Watson (1988) have shown how cointegration affects the structural shocks. Section 3 illustrates the influence of long-run restrictions and cointegration on SVAR modelling.

An advantage of long-run restrictions is that they are interpretable as equilibrium conditions. Since economic theory focuses on equilibria, long-run restrictions provide a convenient way of linking economic theory and econometrics. This has gained them popularity among macroeconometricians. However, long-run restrictions have been criticized by Faust & Leeper (1997) for statistical reasons. Moreover, informal short-run restrictions remain necessary to identify the structural shocks completely. Macroeconomic modelling and its difficulties using SVAR are studied in section 4.

In section 5, I introduce independent component analysis, a technique from the engineering literature (e.g Bell & Sejnowski 1995), to econometrics. Independent component analysis uses higher order statistics to disentangle non-normal structural shocks. For non-Gaussian shocks, independence is sufficient to separate the shocks, which makes additional assumptions unnecessary. This permits identification of independent structural shocks without recourse to informal short-run restrictions. In order to make it a tool for econometric estimation based on a theoretical model, I extend the technique to incorporate linear (long-run) restrictions, and develop an independent components estimator (ICE) for SVAR models.

I make the technique more accessible by comparing it to another approach of avoiding informal short-run restrictions, the "agnostic" approach, pioneered by Faust (1998) and Uhlig (1999). It seeks to completely renounce on theory-based restrictions but focuses explicitly on the subjective restrictions of the researcher, who imposes fairly general assumptions by defining "plausibility windows" for the impulse responses. In section 6, I discuss the possibilities and limits of this approach, and compare it with the ICE.

In section 7, I argue why monetary policy shocks should be distributed logistically, and demonstrate the use of the ICE for identifying monetary policy shocks in the Christiano et al. (1999) framework. A final section concludes.

2 Structural Vector Autoregressions

In the dynamic simultaneous equations approach, a structural model of the economy can be written as the relationship between an n-dimensional vector X_t of economic variables and past realizations of that vector. Assuming for notational convenience that deterministic terms like constants, trends or dummies are concentrated out, the economy is described by

$$\Xi(L)X_t = \varepsilon_t, \tag{1}$$

where $\varepsilon_t \sim iid(0, \Sigma_\varepsilon)$ is a vector of structural shocks. The lag operator L lags the time index of a variable x_t by one period, or $L^s x_t = x_{t-s}$, where s is an integer. The term

$\Xi(L) = \Xi_0 + \Xi_1 L + \ldots + \Xi_q L^q$ is an n-dimensional lag polynomial of order q, i.e. the Ξ_i are matrices.

In the Cowles Commission approach, constraints on the elements of the matrices Ξ_i, $i = 1, \ldots, q$, would be used to identify the system.[1] In contrast, SVAR models restrict the covariance matrix, Σ_ε, the contemporaneous or the long-run influence of the structural shocks on the variables.

Usually, ε_t is assumed to be n-dimensional, i.e. the number of shocks is equal to the number of endogenous variables in the model. Sims (1988, p. 309) points out that this restriction is a very strong one, which is not based on economic reasons but a mere technicality. For the moment let ε_t be of dimension $n \times 1$, so that the matrix $\mathcal{E} = [\varepsilon_{it}], i = 1 \ldots n, t = 1 \ldots T$.

In a SVAR model, the shock vector ε_t is regarded as a vector of independent structural disturbances, e.g. a monetary shock, a demand shock, or a supply shock. To recover them econometrically, write the problem in estimable form as a VAR without cross-restrictions on the error terms,

$$X_t = A(L)X_t + u_t, \qquad u \sim N(0, \Sigma_u), \tag{2}$$

where $A(L) = A_1 L + \ldots + A_p L^p$ is a lag polynomial of order p. This reduced form can be estimated by maximum likelihood. Assume that $|\mathbf{I} - A(z)| = 0$ implies that $|z| > 1$ or $z = 1$, excluding explosive and seasonal roots. Furthermore, the elements of X_t are assumed to be integrated of at most order one. This is true for most macroeconomic data series, the only prominent exceptions being nominal money and the price level.[2] In this case, Engle & Granger (1987) have suggested to write the process as a vector error correction model (VECM)

$$\Delta X_t = \Pi X_{t-1} + H(L)\Delta X_t + u_t \tag{3}$$

where $\Delta = 1 - L$ is the first difference operator. $H(L)$ is a lag polynomial of order $p-1$ with elements $H_i = -\sum_{j=i+1}^{p} A_j$, $i = 1, \ldots, p-1$. Consequently, the cointegration matrix is $\Pi = A(1) - \mathbf{I}_n$. To consider the properties of Π, let $\Gamma(L) = \mathbf{I}_n - H(L)$ and rewrite equation (3) as

$$\Pi X_{t-1} = \Gamma(L)\Delta X_t - u_t. \tag{4}$$

Since the right hand variables are all stationary, ΠX_{t-1} must also be stationary. Denote the rank of the matrix Π by $\mathrm{rk}(\Pi) = r \in [0, n]$. Following Lütkepohl (1996, 4.3.1, Result

[1] For a discussion of the Cowles Commission approach, cf. Favero (2001).

[2] Examples for I(2) variables and the more complicated I(2) analysis are discussed in Johansen (1995, p. 132ff) or Juselius (1998).

(7)), Π can be decomposed into $\Pi = \alpha\beta'$, where α and β are matrices of dimension $n \times r$ and $\text{rk}(\alpha) = \text{rk}(\beta) = r$.[3] Since α is a constant matrix, $Z_t = \beta' X_t$ defines a set of r stationary linear combinations of X_t.

An interpretation of the reduced form is difficult because the matrix $A(L)$ involves many parameters. Therefore, the economic interpretation of VAR models is conducted using the impulse responses. In order to find the impulse responses, the VAR needs to be rewritten in moving average (MA) form, so that the endogenous variables only depend on the residual series,

$$\Delta X_t = C(L)u_t. \tag{5}$$

To find this representation, notice that $\text{rk}(\Pi) = n$ implies that $|\mathbf{I} - A(1)| \neq 0$, i.e. the process is stable without differencing. Conversely, $r = 0$ implies that $A(1) = \mathbf{I}$ so that the process is a VAR of order $p-1$ in first differences.[4] In both cases, (4) is invertible, and the canonical moving average representation is given by

$$C(L) = (\Gamma(L) - \Pi L(1-L)^{-1})^{-1}.$$

The problem with this representation is that the u are mutually dependent. Hence, the response of vector X to a given shock contains a mixture of information about all shocks. To avoid this, the shocks are remapped into another set of independent structural shocks,

$$u_t = S\varepsilon_t. \tag{6}$$

In the case of Gaussian shocks, orthogonality implies independence. Assuming normality of the unobservable structural shocks, the residuals may then be mapped into structural shocks. Normalising the variance to one and imposing orthogonality across the structural shocks, i.e. $\Sigma_\varepsilon = \mathbf{I}_n$, it follows from (6) that

$$\Sigma_u = SS',$$

[3] All matrix results from Lütkepohl are reproduced for convenience in appendix C.

[4] This can also be seen by rewriting (3) as

$$\Delta X_t = \tilde{H}(L)\Delta X_t + \Pi X_{t-p} + u_t,$$

where

$$\tilde{H}(L) = (p-1)\Pi + H(L)$$

is also of order $p-1$. Consequently, $\Pi = 0$ implies that ΔX_t has a stable VAR representation of length $p-1$.

which imposes $n(n+1)/2$ restrictions on the elements of S. Equating (5) and (1) yields

$$S = \Xi(L)^{-1}(1-L)C(L)^{-1}. \tag{7}$$

The impulse response of X_{t+i} to a structural shock ε_t is given by Ξ_i^{-1} because S is independent of t. With (7) and (6) we need $n(n-1)/2$ additional restrictions to recover S and ε_t from the VAR residuals u_t. Linear restrictions on S can be given implicitly in the form

$$\Phi' \mathrm{vec}\,(S) = \chi_1, \tag{8}$$

where the operator $\mathrm{vec}\,(S)$ stacks the columns of S.[5] The matrix Φ has dimension $n^2 \times a$, and imposes $a < n^2$ linear constraints on S. In economic applications, these are usually zero constrinats, and hence χ_1 is a vector of zeros. However, these restrictions are hard to implement because Φ is a matrix of dimension $n^2 \times a$ with full column rank, so that $\Phi\Phi'$ is singular. Only rewriting them in explicit form permits the estimation of S subject to the restrictions.

Denote the orthonormal basis for the null space of Φ by $\Phi_\perp$, where $\Phi_\perp$ is a matrix with full rank such that $\Phi'\Phi_\perp = \mathbf{0}$ and $\Phi_\perp\Phi_\perp' = \mathbf{I}$.[6] The relation between the implicit and the explicit representation is then seen by noting that any matrix can be rewritten as a projection into Φ and $\Phi_\perp$ (e.g. Bronstein et al. 1997, p. 570). Therefore, the projection of vec(S) into the a-dimensional column space of Φ has a complementary projection into the orthogonal column space, which has dimension $n^2 - a$.

The projection of vec(S) into the column spaces of Φ and $\Phi_\perp$ is given by

$$\begin{aligned} \mathrm{vec}\,(S) &= (P_{\Phi_\perp} + P_\Phi)\mathrm{vec}\,(S) \\ &= \Phi_\perp(\Phi_\perp'\Phi_\perp)^{-1}\Phi_\perp'\mathrm{vec}\,(S) + \Phi(\Phi'\Phi)^{-1}\Phi'\mathrm{vec}\,(S) \\ &= \Phi_\perp(\Phi_\perp'\Phi_\perp)^{-1}\Phi_\perp'\mathrm{vec}\,(S) + \Phi(\Phi'\Phi)^{-1}\chi_1 \\ &= \Phi_\perp\gamma + \Phi(\Phi'\Phi)^{-1}\chi_1. \end{aligned} \tag{9}$$

Note that in contrast to $\Phi\Phi'$, $\Phi'\Phi$ is a full rank matrix. Since Φ and χ_1 are known, an estimate of γ can be used to find S under the restrictions (8).

If the unobserved structural shocks are the outcome of a single decision variable like monetary policy shocks, the shocks can be expected to be leptokurtic. Central limit tendencies, which are present in aggregate shocks like aggregate demand or supply

[5] I give a more precise definition in appendix C

[6] $\Phi_\perp$ is not necessarily unique. However, any representation will do. Sargan (1988, pp. 34) shows one possibility to find one representation of the orthonormal basis for the null space of a matrix.

shocks, are not at play. This result is reinforced because central bankers know that their decisions are perceived by the markets as a signal, which must therefore be as clear as possible. In other words, monetary policy could best be characterised by tranquility interrupted by "occasional panic" (James Duesenberry in Leeper, Sims & Zha 1996, p. 76). Finally, it is well known from the financial markets literature that interest rate series display high kurtosis. Given that monetary policy shocks are closely linked to interest rate series it comes as no surprise that they are also leptokurtic.

Papers that investigate the monetary transmission mechanism either do not reort tests for normality of the residuals or they find leptokurtic residuals for the interest rate series (e.g. Christiano et al. 1999, Juselius 1998, Coenen & Vega 1999, Vlaar & Schuberth 1998). If this is the case, one can expect that at least one of the identified structural shocks will also be non-normal. As an example, consider the estimated monetary policy shocks from Christiano et al. (1999) (henceforth CEE).[7] They use quarterly US data for output, the output deflator, commodity prices, the Federal Funds rate, nonborrowed reserves, total reserves and the monetary aggregate M1. A monetary policy shock is identified as one which has no contemporaneous impact on the real variables output, the deflator and commodity prices, but is predetermined for the monetary variables except the one considered the monetary policy instrument. In the first model (FF), the Federal Funds rate is the monetary policy instrument, in the second model (NBR) non-borrowed reserves.[8]

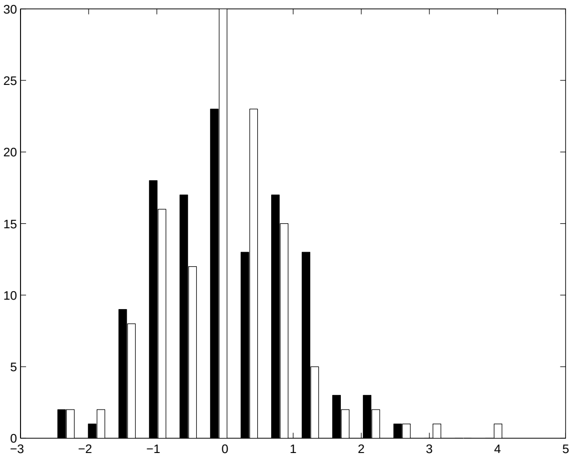
Testing for kurtosis (Lütkepohl 1993, p.153) indicates that the residuals from the VARs are multivariate leptokurtic (p-values 0.00 for both models). Therefore, the assumption of a normal distribution is not warranted, and orthogonalising will result in structural shocks which are not independent. Also for the identified shocks, the distribution is highly leptokurtic with a p-value of 0.00 (0.00). Figure 1 shows the histogram of the identified monetary policy shocks.

An alternative interpretation to the one suggested above would be that the structural shocks are in fact normally distributed and that the largest values are outliers. However, following the suggestions by Rudebusch (1998, p. 929) and Sims (in Rudebusch 1998, p. 934) to look at the shock series demonstrates that these "outliers" are exactly the observations of interest, i.e. the most extreme monetary policy shocks. Furthermore, it does not solve the problem: It is well-known that the Pearson correlation coefficient

[7] I would like to thank Charles Evans for generously sharing the data and RATS programmes to generate the monetary shocks.

[8] In the lingo of VAR modeling, the "ordering" of the variables for the Choleski decomposition in the FF (NBR) model is output, deflator, commodity prices, FF, NBR, total reserves, M1 (output, deflator, commodity prices, NBR, FF, total reserves, M1).

Figure 1: Histogram of Monetary Policy Shocks identified by CEE



Note: Monetary policy shocks assuming the Federal Funds rate (dark) or non-borrowed reserves (white) as the Central Bank's instrument

is a non-robust measure for independence. Therefore, one influential observation will change the estimated structural shocks considerably (e.g. Huber 1981).

3 Long-run restrictions and cointegration

If $0 < r < n$ in equation (3), the system is called *cointegrated* with rank r. Estimating in levels would imply a loss in efficiency because some restrictions are ignored. In first differences, the model would even be misspecified as invalid zero restrictions are imposed. In this case, the process is not easily invertible due to the reduced rank of Π. However, the fundamental MA representation of a cointegrated VAR can be written as

$$\Delta X_t = C(1)u_t + C^*(L)\Delta u_t, \tag{10}$$

where $C(1)$ is a matrix of rank $n-r$ that represents the common trend space, and $C^*(L)$ has rank r and lies in the cointegration space (cf. Engle & Granger 1987, Johansen 1991). Writing the MA representation (10) explicitly is a first step towards the structural representation of the reduced VAR under cointegration restrictions.

The solution is achieved by finding a lag polynomial with the same roots as $\mathbf{I}_n - A(L)$ but the unit root removed. Define the $n \times (n-r)$ matrix $\beta_\perp$ as an orthonormal basis for the null space of β. To simplify notation, let $\bar{\beta} = \beta(\beta'\beta)^{-1}$, so that the projection into the column space of β is given by $P_\beta = \bar{\beta}\beta' = \beta\bar{\beta}'$, and $\beta'\bar{\beta} = \bar{\beta}'\beta = \mathbf{I}_n$.

In analogy to the reasoning leading to equation (9), project ΔX_t into β and $\beta_\perp$. Using the stationary variable $Z_t = \beta' X_t$ and defining $\Theta_t = \beta_\perp' \Delta X_t$ yields

$$\begin{aligned}
\Delta X_t &= (P_{\beta_\perp} + P_\beta)\Delta X_t \\
&= \left(\beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp' + \beta(\beta'\beta)^{-1}\beta'\right)\Delta X_t \\
&= \bar{\beta}_\perp \beta_\perp' \Delta X_t + \bar{\beta}\beta'\Delta X_t \\
&= \bar{\beta}_\perp \Theta_t + \bar{\beta}\Delta Z_t. \qquad (11)
\end{aligned}$$

Using this result, one may rearrange (3) and write in matrix form

$$u_t = \begin{bmatrix} \Gamma(L)\bar{\beta}(1-L) - \alpha L, & \Gamma(L)\bar{\beta}_\perp \end{bmatrix} \begin{bmatrix} Z_t \\ \Theta_t \end{bmatrix}. \qquad (12)$$

Premultiplying (12) by $\bar{\alpha}'$ and $\bar{\alpha}_\perp'$ and stacking the results gives for $L \neq 1$

$$\begin{bmatrix} \bar{\alpha}' \\ \bar{\alpha}_\perp' \end{bmatrix} u_t = \begin{bmatrix} \bar{\alpha}' \\ \bar{\alpha}_\perp' \end{bmatrix} \begin{bmatrix} \Gamma(L)(1-L) - \alpha\beta' L \end{bmatrix} \begin{bmatrix} \bar{\beta}, & \bar{\beta}_\perp(1-L)^{-1} \end{bmatrix} \begin{bmatrix} Z_t \\ \Theta_t \end{bmatrix} \qquad (13)$$

or without restrictions on L,

$$\begin{aligned}
\begin{bmatrix} \bar{\alpha}' \\ \bar{\alpha}_\perp' \end{bmatrix} u_t &= \begin{bmatrix} \bar{\alpha}'\Gamma(L)\bar{\beta}(1-L) - \mathbf{I}_r L & \bar{\alpha}'\Gamma(L)\bar{\beta}_\perp \\ \bar{\alpha}_\perp'\Gamma(L)\bar{\beta}(1-L) & \bar{\alpha}_\perp'\Gamma(L)\bar{\beta}_\perp \end{bmatrix} \begin{bmatrix} Z_t \\ \Theta_t \end{bmatrix} \\
&\equiv \tilde{A}(L) \begin{bmatrix} Z_t \\ \Theta_t \end{bmatrix}, \qquad (14)
\end{aligned}$$

where the levels of X_t are removed when multiplying by $\bar{\alpha}_\perp'$. Inspection of (13) and (14) shows that the polynomial $\tilde{A}(z)$ has the same roots as $\Gamma(z)(1-z) - \Pi z$ except for $z = 1$. Indeed, evaluating $\tilde{A}(z)$ at $z = 1$ gives

$$\tilde{A}(1) = \begin{bmatrix} -\mathbf{I}_r & \bar{\alpha}'\Gamma(1)\bar{\beta}_\perp \\ \mathbf{0} & \bar{\alpha}_\perp'\Gamma(1)\bar{\beta}_\perp \end{bmatrix} \qquad (15)$$

which is invertible if $|\alpha_\perp'\Gamma(1)\beta_\perp| \neq 0$ using Lütkepohl (1996, 4.3.1, Result (10)). This is guaranteed if the variables are integrated of at most order one. The MA representation in first differences is therefore given by

$$\Delta X_t = [\bar{\beta}(1-L), \bar{\beta}_\perp] \begin{bmatrix} Z_t \\ \Theta_t \end{bmatrix} = [\bar{\beta}(1-L), \bar{\beta}_\perp]\tilde{C}(L) \begin{bmatrix} \bar{\alpha}' \\ \bar{\alpha}_\perp' \end{bmatrix} u_t \equiv C(L)u_t, \qquad (16)$$

where $\tilde{C}(z) = \tilde{A}(z)^{-1}$. Note that any matrix $C(L)$ can be rewritten by rearranging terms,

$$C(L) = C(1) + (1-L)C^*(L). \qquad (17)$$

Since $C(1)$ does not contain the lag operator, it describes the permanent influence of u_t on X_t. Using the information that $\tilde{C}(1) = \tilde{A}(1)^{-1}$, the long-run impact is given by

$$\begin{aligned} C(1) &= [0, \bar{\beta}_\perp]\tilde{A}(1)^{-1} \begin{bmatrix} \bar{\alpha}' \\ \bar{\alpha}'_\perp \end{bmatrix} \\ &= \bar{\beta}_\perp(\bar{\alpha}'_\perp \Gamma(1)\bar{\beta}_\perp)^{-1}\bar{\alpha}'_\perp \\ &= \beta_\perp(\alpha'_\perp \Gamma(1)\beta_\perp)^{-1}\alpha'_\perp, \end{aligned} \tag{18}$$

and the short run impact is therefore given by $(1 - L)C^*(L)$.

Cointegration relations divide the projection space into a common trends space and a cointegration space. By Lütkepohl (1996, 4.3.3, Result (2a)), the rank of $C(1)$ is at most $n - r$. This can also be seen by inserting (17) in (16) and premultiplying by β'. The r-dimensional MA representation,

$$(1 - L)Z_t = (\beta' C(1) + (1 - L)\beta' C^*(L))\, u_t, \tag{19}$$

and the fact that u_t and Z_t are stationary imply that u_t cannot influence ΔZ_t and hence $\beta' C(1) = 0$. Therefore, $C(1)$ must have a null space containing at least all r cointegrating vectors contained in β'. Together with $\Delta X_t = C(L)u_t$ from equation (16), this implies that all shocks in the cointegrating space die out in the long run. On the other hand, the common trends literature (e.g. King, Plosser, Stock & Watson 1991) points out that the innovations lying in the common trend space will each have a permanent impact on at least one economic variable. Equation (19) implies that $\beta' C^*(L)$ is stationary, which in turn means that $C^*(L)$ has at most $\text{rk}(\beta) = r$. Given that $C(L)$ has full rank and that $\text{rk}(C(1)) \leq n - r$, the decomposition implies by Lütkepohl (1996, 4.3.3, Result (3)) that the common trends space has $\text{rk}(C(1)) = n - r$, and $\text{rk}(C^*(L)) = r$ is the cointegrating rank (cf. also Johansen (1995), Theorem 4.2. or Engle & Granger (1987)).

For the structural VAR analysis, these results imply that cointegration imposes additional zero restrictions on the structural impulse responses. Second, it implies that some additional restrictions cannot be long-run. The matrix $C(1)$ is not uniquely identified since the matrix $\alpha\beta'$ is observationally equivalent to $\alpha F F^{-1}\beta'$, where F is any invertible matrix with dimensions $r \times r$. Since $\text{rk}(\alpha\beta') = r$, for each line r restrictions on α or β, or nr, are needed to identify $C(1)$. The problem then is to decompose $\Delta X_t = C(L)S\varepsilon_t$. Given that r elements of ε_t do not have a permanent influence on X_t, the structural long-run matrix, $C(1)S$, must have a block of $n \times r$ zeros. Given that $\text{rk}(C(1)) = n - r$, this implies $r(n - r)$ independent restrictions. To achieve identification in the common trend space, $(n - r)(n - r - 1)/2$ additional contemporaneous or long-run restrictions are needed. The shocks in the cointegration space are identified by $r(r-1)/2$ restrictions on the contemporaneous impact only because all of them die out in the long-run. Hence, the identification problem is reduced by the $r(n-r)$ zero restrictions on the long-run impulse responses. At the same time, it means that shocks in the cointegration space can only be disentangled by short-run restrictions. These restrictions cannot come from economic equilibrium theory but must be implemented informally. Therefore, this identification scheme implies that at least $r(r-1)/2$ restrictions do not evolve from economic theory.

Economic theory usually provides zero (i.e. $\chi_1 = \mathbf{0}$) restrictions. The permanent impact of the structural errors on the dependent variable is given by $\Phi = \text{vec}\,(C(1))$. Hence, if the j-th innovation is supposed to have no long-run effect on the k-th variable, this is expressed as $c(1)_{k,j}s_{k,j} = 0$, where $c(1)$ is an element of $C(1)$. Contemporaneous restrictions take the form $s_{k,j} = 0$.

Hence, the procedure is to estimate the lag-polynome $C(L)$ via maximum likelihood from the data, estimate the rank of Π via the well-known Johansen procedure and use economic reasoning to restrict the α or β matrices in equation (18). The remaining task is to decompose u_t into the structural shocks ε_t and the projection matrix S. This can in part be achieved by imposing long-run restrictions on S in equation (8) from a theoretical model, and in part by informal short-run restrictions. The estimate of γ can then be used to find S. Finally, the impulse response function $\Xi(L)$ results from equation (7).

4 Macroeconomic modelling and the Faust & Leeper (1997) critique

A commonly used method of SVAR modelling is to include macroeconomically relevant variables like interest rates, output, or prices in the VAR without reference to an explicit model (e.g. Gerlach & Smets 1995). This is problematic because the restrictions used to identify the structural shocks strongly influence the resulting impulse responses. Similarly, omitted variables may considerably change the estimation. Without an explicit model it remains obscure whether or not important variables have been omitted from the setup.

Therefore, an increasing number of researchers have developed long-run restrictions from explicit macroeconomic models (e.g. King et al. 1991, Galí 1992, Vlaar & Schuberth 1998, Hubrich & Vlaar 2000). Frequently, a money demand relationship or a Fisher effect are used as long-run relationships. However, these papers also illustrate that contemporary restrictions have to be imposed outside the theoretical model. As noted above, the shocks in the cointegration space can only be disentangled using contemporary restrictions. Therefore, assumptions like information lags are introduced, which are not strictly derived from the theoretical model but imposed informally. A frequent assumption in this vein is that changes in real variables are not perceived by central bankers within the same period (as e.g. in CEE).

Besides the need for informal restrictions, Faust & Leeper (1997) have shown in their seminal critique that recovering economic shocks from a SVAR by long-run restrictions is hazardous. The authors explore three difficulties when long-run restrictions are applied.

Their first critique is based on the fact that the estimator of the infinite impulse response, $C(1)$, is highly uncertain in finite samples. Long-run restrictions transfer this uncertainty to all parameters of the $C(L)$ polynomial. However, Faust and Leeper also indicate that this problem can be solved by imposing a lag length that is small relative to the sample size. This leads to more precise estimations of $C(1)$ and hence, more reliable inference given long-run restrictions. Since the thrust of this paper is elsewhere, I will not explore this critique further. Empirically, short VAR lag-lengths generate i.i.d. shocks and are therefore not hard to defend.

The second critique targets the number of shocks in a VAR. The structural shocks identified in a small VAR must be regarded as aggregates of economic shocks. If the true number of economic shocks exceeds the number of structural shocks identified by the SVAR procedure, long-run restrictions may commingle the economic shocks and prevent inference about them from the shocks identified by the SVAR. To check for robustness, Sims (1980) suggests to estimate larger models, Faust & Leeper (1997) suggest to estimate various small models.

Faust and Leeper's third argument attacks the orthogonality assumption. They suggest that economic shocks occurring at a frequency which is higher than the sampling rate may be observed distortedly. Consider a shock to the money stock to which the central bank reacts within days, as the ECB did on May 10, 2001. Using monthly data, the researcher will be unable to distinguish between the two shocks. However, this problem is more relevant if indeed only orthogonality among the shocks is considered. Unless this kind of events occurs frequently, density estimations are a more robust method of identifying independent shocks. The next section introduces a technique which solves some of the problems indicated.

5 The Independent Component Estimator

In this section, I propose an alternative method for deriving S instead of using the strong priors implied by zero restrictions. It has been argued that for theoretical reasons monetary policy shocks ought to be leptokurtic. Furthermore, covariance estimators are extremely sensitive to outliers. Therefore, correlation is a non-robust measure for independence even in approximately normally distributed samples. A more robust approach are density estimators (e.g. Huber 1981, p. 199f). Therefore, this section suggests a density-based approach to disentangle non-normal shocks.[9] Finally, restricting the number of economic shocks to the number of endogenous variables is theoretically unappealing.

In SVAR modelling, the residuals are perceived as a linear combination of unobservable structural shocks. The problem of identifying these shocks can therefore be described as discovering unobservable information from a linear mixture of this information.[10] This is a frequent problem in signal processing, and I adopt the solution proposed in the engineering literature. Independent component analysis has been introduced in a seminal paper by Jutten & Herault (1991) and developed rapidly since. In this part, I introduce this technique to econometrics and I suggest adjustments so that independent components can be estimated based on economic models.

Given the identification problem (6) without any restrictions on the dimensions of S and Σ_ε, the goal is to find a vector of independent structural shocks ε_t. Independence between the elements of ε_t is defined by (e.g. Hamilton 1994, p. 742)

$$f(\varepsilon_t) \quad = \quad \prod_{i=1}^{v} f_i(\varepsilon_{it}), \tag{20}$$

where $f_i(\varepsilon_{it})$ is the univariate density of ε_{it} and $f(\varepsilon_t)$ the multivariate density of ε_t. Hence, shocks will be independent if the difference between the left and the right hand of (20) is zero. The main concern about independent shocks is that the reaction of the endogenous variables to a given shock should contain *information* about one economic

[9]One may argue that the whole discussion about VAR estimation and cointegration has relied on the assumption of normal residuals. While this is true, it is well-known that maximum likelihood estimation is consistent for leptokurtic errors (cf. e.g. Hamilton 1994, p. 663 with references). For cointegration, Gonzalo (1994) and Silvapulle & Podivinsky (2000) have shown that Johansen-type cointegration tests are robust to excessive kurtosis. An information-theoretic approach to cointegration testing was introduced by Aparicio & Escribano (1998), and a non-parametric test has been suggested by Boswijk & André (2002).

[10]Given that the Journal of Econometrics has devoted a special issue (Vol. 107, Iss. 1, 2002) to information-theoretic approaches to econometrics, information theory seems to become a standard tool in econometrics.

shock only – or at least as little information as possible about the other shocks. Hence, a natural way of formulating the search for a statistically independent representation is given by the criterion of minimal mutual information.

To clarify the concept of information, consider the uncertainty, or entropy, about the realization of variable ε_{it} from the distribution $f_i(\varepsilon_{it})$. The Boltzman-Gibbs-Shannon entropy is given by the negative of the expected log-likelihood (e.g. Golan 2002)[11]

$$\mathcal{H}(\varepsilon_{it}) \quad = \quad -E[\log f_i(\varepsilon_{it})] = -\int f_i(\varepsilon_{it}) \log f_i(\varepsilon_{it}) d\varepsilon_{it}, \tag{21}$$

where limits of integration are infinite unless given explicitly.[12] In other words, entropy measures information in the sense that it describes the difficulty of predicting a random variable. Assume now that another variable ε_{jt} is observed. Using this observed variable ε_{jt} as a piece of information about the outcome of the unobserved variable ε_{it}, the guess about ε_{it}'s realization can be improved. This improvement is the entropy of ε_{it} conditional upon ε_{jt}. The conditional entropy is given by

$$\begin{aligned}\mathcal{H}(\varepsilon_{it}|\varepsilon_{jt}) &= -\int\int f_j(\varepsilon_{jt}) f_i(\varepsilon_{it}|\varepsilon_{jt}) \log f_i(\varepsilon_{it}|\varepsilon_{jt})\, d\varepsilon_{it} d\varepsilon_{jt} \\ &= -\int f_i(\varepsilon_{it},\varepsilon_{jt}) \log f_i(\varepsilon_{it}|\varepsilon_{jt})\, d\varepsilon_{t},\end{aligned} \tag{22}$$

where $f_i(\varepsilon_{it},\varepsilon_{jt}) = f_i(\varepsilon_{it}|\varepsilon_{jt}) f_i(\varepsilon_{jt})$ by Bayes' theorem. Now, let the *mutual information* between ε_{it} and ε_{jt} be the additional knowledge gained by observing ε_{jt},

$$\begin{aligned}\mathcal{I}(\varepsilon_{it},\varepsilon_{jt}) &= \mathcal{H}(\varepsilon_{it}) - \mathcal{H}(\varepsilon_{it}|\varepsilon_{jt}) \\ &= \int f_i(\varepsilon_{it},\varepsilon_{jt}) \log \frac{f_i(\varepsilon_{it},\varepsilon_{jt})}{f_i(\varepsilon_{it}) f_j(\varepsilon_{jt})}\, d\varepsilon_{t},\end{aligned} \tag{23}$$

using $f_i(\varepsilon_{it}) = \int f_i(\varepsilon_{it},\varepsilon_{jt}) d\varepsilon_{jt}$. This is the Kullback & Leibler (1951) divergence between the joint density and the factored densities (i.e. the case of ε_{it} and ε_{jt} being independent).[13] The conditional entropy may equivalently be written as

$$\mathcal{H}(\varepsilon_{it}|\varepsilon_{jt}) \quad = \quad \mathcal{H}(\varepsilon_{it},\varepsilon_{jt}) - \mathcal{H}(\varepsilon_{jt}),$$

so that

$$\mathcal{I}(\varepsilon_{it},\varepsilon_{jt}) \quad = \quad \mathcal{H}(\varepsilon_{it}) + \mathcal{H}(\varepsilon_{jt}) - \mathcal{H}(\varepsilon_{it},\varepsilon_{jt}).$$

[11] The definition of the expectation value is given e.g. by Mood, Graybill & Boes (1974, p. 153, Def. 18).

[12] To simplify notation, I use continuous-valued variables. To capture discrete-valued variables, simply substitute integrals by sums in the following.

[13] Non-parametric testing based on the Kullback-Leibler divergence has been proposed by Robinson (1991). In contrast with this paper, the procedure suggested here *searches* most independent variables given a joint distribution.

Therefore, the Kullback & Leibler divergence is a symmetric measure for mutual information between the elements of ε_t. The Kullback & Leibler divergence is zero if, and only if, the shocks are independent. In order to minimize interdependence across shocks, the Kullback & Leibler divergence has to be minimised. In appendix A, I show that in the case of exactly Gaussian distributed structural shocks minimizing the Kullback & Leibler divergence reduces to orthogonalising the residuals. Hence, this approach can be regarded a generalisation of orthogonalisation for non-normal structural shocks.

Unfortunately, both distributions in (23) are unobservable. Therefore, it is useful to look at the problem from another angle. Kullback & Leibler define the information in a sample u_t about the hypothesis that u_t comes from distribution f over the hypothesis that u_t comes from f_2 as the log-likelihood ratio, $\log[f(u_t)/f_2(u_t)]$. Let $f(u_t)$ be the true distribution and $f_2(u_t)$ the distribution estimated by assuming independent shocks, $f_2(u_t) = |S|^{-1}\Pi_{i=1}^{v} f_i(\varepsilon_{it})$.[14] The expected information that can be used to discern between the two distributions is given by

$$\mathcal{I}\left(f(u_t), |S|^{-1}\prod_{i=1}^{v} f_i(\varepsilon_{it})\right) = \int f(u_t)\log\frac{f(u_t)}{|S|^{-1}f(S^{-1}u_t)}\,du_t. \qquad (24)$$

Kullback & Leibler show that this measure is zero if and only if the densities coincide, and positive otherwise. Intuitively, in the case of coinciding densities no sample can be informative for discriminating between the densities. Another desirable property of this measure for information is that it is additive for independent events, i.e. n events contain n times the information contained in one event. If the events are not independent, the information conditional upon the other events is additive. Finally, statistical processing of the data, like grouping, does not add information. In contrast, statistical processing of the data reduces information unless a sufficient statistic is used. Apparently, $\mathcal{I}$ is a robust measure of the difference between densities. Minimising (24) approximates the estimated densities to the observed density of the VAR residuals. Transformation of variables shows that (24) is the same as (23). Hence, minimising (24) results in structural shocks which are least interdependent.

In appendix B, I give a minimisation algorithm. Specifically, I show that the stochastic gradient ascent is given by

$$\Delta S = -S'^{-1} + \varphi(\varepsilon_t)\varepsilon_t' S'^{-1},$$

where

[14] The latter expression evolves immediately from the distributions of functions of variables, e.g. (Mood et al. 1974, p. 205, Theorem 18): If $y = g(x)$, then under certain regularity conditions $f_Y(y) = |J|f_X(g^{-1}(y))$, where $|J|$ is the determinant of the Jordan matrix.

$$\varphi(\varepsilon_t) = \frac{\frac{\partial f(\varepsilon_t)}{\partial \varepsilon_t}}{f(\varepsilon_t)} = \left[\frac{\frac{\partial f_1(\varepsilon_{1t})}{\partial \varepsilon_{1t}}}{f_1(\varepsilon_{1t})}, \ldots, \frac{\frac{\partial f_n(\varepsilon_{nt})}{\partial \varepsilon_{nt}}}{f_n(\varepsilon_{nt})}\right]'. \tag{25}$$

The marginal distribution of these shocks $f_i(\varepsilon_{it})$ can either be based on prior assumptions about the distribution of these shocks[15] or estimated from shock series identified by other methods, as in Romer & Romer (1989), Boschen & Mills. (1995), or Maier (2000) for monetary policy shocks.

Appendix B also demonstrates how the estimator can be extended to incorporate linear restrictions. This permits incorporating strong priors, e.g. about impulse responses. The following sections give some intuition for applications of the estimator.

6 The Blanchard & Quah (1989) Case and the agnostic approach

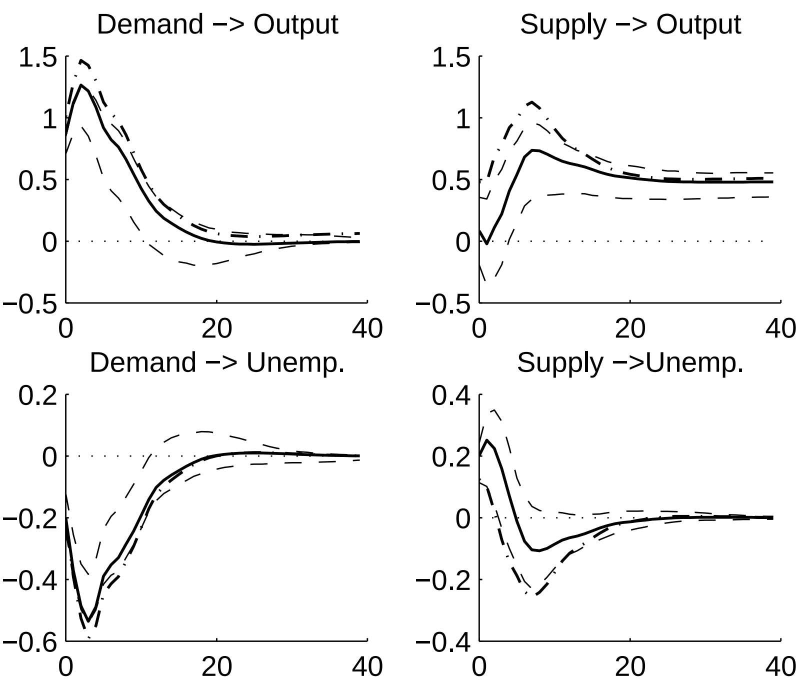
In order to get a graphical idea of the difference between the procedures, consider the exercise by Blanchard & Quah (1989, henceforth BQ) as an example. Although it is not concerned with monetary policy shocks it is useful because in this paper, BQ laid the foundations of long-run SVAR modelling so heavily criticized by Faust & Leeper (1997). Furthermore, it is a bivariate case and therefore graphically more intuitive than higher-dimensional examples. In a simple bivariate model of output growth and unemployment, the authors identify two shocks, a demand shock which does not influence output in the long run and a catch-all shock, termed supply shock. Building upon this assumption, they recover impulse response functions to supply and demand shocks. They also consider the complement of the impulse responses, and identify output and unemployment fluctuations due to demand shocks. Using quarterly US data for the period 1948:1 until 1988:4 from Weber (1995), I replicate their model. To make the results comparable, I follow BQ and allow for a secular trend in unemployment as well as a discrete change in output growth in 1974:1. This reproduces their results quite closely.

In a second step, I reestimate the model using the ICE. In contrast with BQ, I use a weaker prior and regard the shock with the *smaller* long-run impact on output as the demand shock. Figure 2 shows the impulse response functions for the BQ and the ICE approaches. The shape of the impulse responses from the ICE are rather similar to the BQ scheme. To see whether the differences are significant, I bootstrapped confidence intervals for the BQ model. Except for the more pronounced short-run effects, the

[15]For example, I have argued in section 2 why monetary policy shocks should be leptokurtic.

impulse responses from the ICE are always within one standard error bands of the BQ impulse responses. Not even the long-run effect on output differs even though BQ's prior is very strong at this point.

Figure 2: Impulse Response Functions

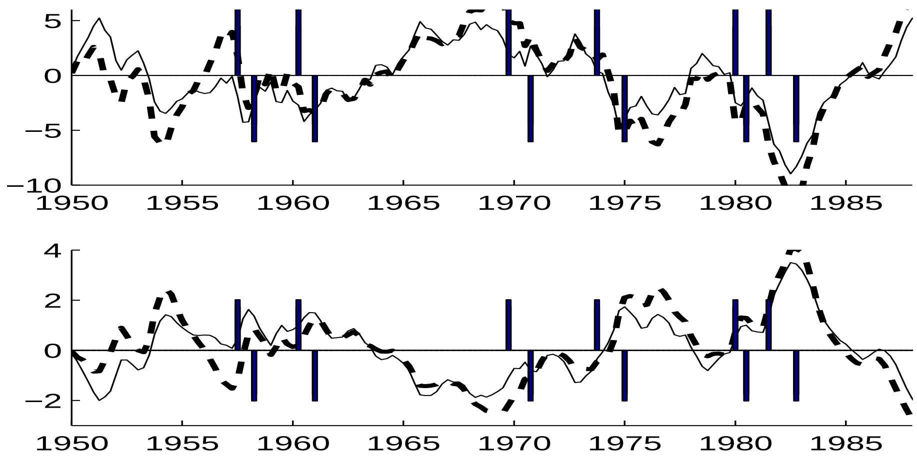


Note: Impulse Response Functions of Output and Unemployment according to the Blanchard & Quah (1989) restrictions (solid) and the ICE (dashed). Data are from Weber (1995). Thin dashed lines are one standard error bands for the Blanchard & Quah (1989) scheme.

The similarities become even clearer looking at the fluctuations due to demand. Figure 3 gives the impression that fluctuations are attributed to demand shocks almost exactly as in the BQ paper. The structural demand (supply) shock series from the two identification procedures are also highly correlated with a coefficient of 0.89 (0.99). However, it is also noteworthy that in contrast with the BQ shocks which are uncorrelated by construction, the structural shocks identified by the ICE are correlated with a coefficient of -0.57.

This example indicates how the robustness of long-run restrictions can be tested without recourse to larger models or different specifications. Using an approach that relies on robust estimators of independence one may generate impulse responses, reconstruct fluctuations and the shock series. The similarity between these and the ones found by strong priors on long-run neutrality can be used as a measure for the plausibility of the restrictions. In contrast with the problems indicated by Faust & Leeper (1997),

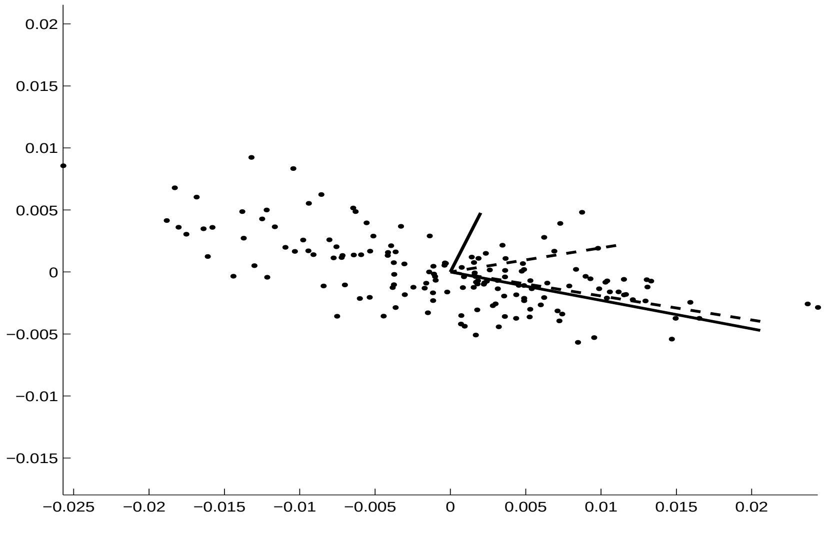
Figure 3: Fluctuations due to Demand Shocks



Note: Fluctuations due to Demand shocks of Output (upper panel) and Unemployment (lower panel) according to the BQ (solid) and ICE (dashed) schemes. NBER peaks and troughs are indicated by bars. Data are from Weber (1995)

long-run restrictions have generated robust results in this case.

Figure 4: Projection Vectors of the Residuals

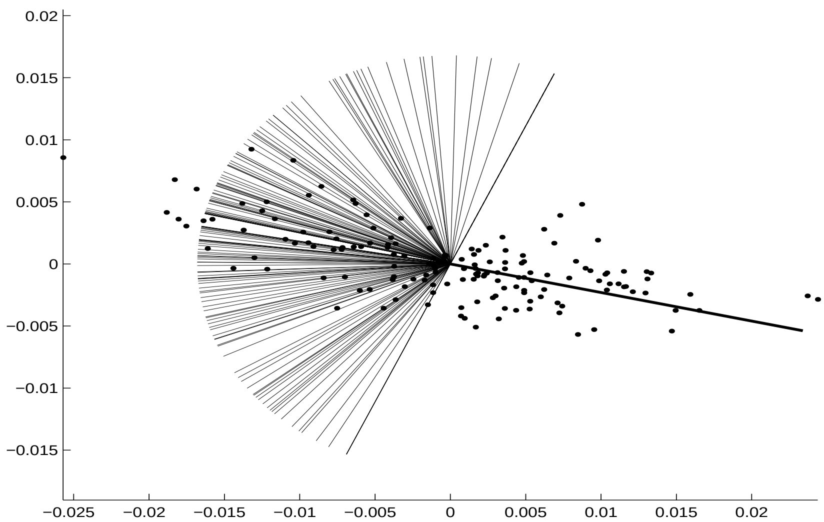


Note: VAR residuals and the projection vectors of BQ (solid) and using the ICE (dashed).

Additionally, this example can be used to gain some intuition for the difference between SVAR estimation and the ICE. The decomposition of the errors can be viewed as a projection of the VAR residuals on a two dimensional space. Figure 4 show the vectors onto which the BQ method and the ICE project the residuals. The direction of the main variance is given by the vectors going to the South-East. In the BQ scheme, this vector is given by the assumption of zero correlation. The second vector is given by the restriction on the long-run impulse responses of the demand shock. In contrast, the ICE identifies the second main direction of variance.[16]

This representation can also be used to show how the ICE differs from the agnostic approach, pioneered by Faust (1998) and Uhlig (1999). In the agnostic approach, the researcher imposes fairly general assumptions by defining "plausibility windows" for the impulse responses. After imposing orthogonality on the structural shocks, $\Sigma_u = SS' = SFF^{-1}S'$ for any invertible matrix F. All possible F-matrices are then tested given plausible impulse responses. One convenient way of representing all possible F matrices is by Givens matrices, or orthogonal rotations of S (e.g. Lütkepohl 1996, p. 246). However, the approach is restricted by imposing orthogonality first. Figure 5 demonstrates this method applied to the Blanchard & Quah (1989) data. Since orthogonality is maintained in this approach, the vector towards the South-East remains fixed. Hence, rotating the matrix S only rotates the short vector from figure 4, which points towards North-West, and leaves the long arm unchanged. In contrast, the ICE rotates both vectors and offers therefore genuinely new projections of the residuals.

Figure 5: Rotated Projection Vectors



Note: Rotated projection vectors of BQ using the agnostic approach

[16]In this, it is similar to principal component analysis. However, the ICE differs from principal components because the projection vectors are not necessarily orthogonal.

Kieler & Saarenheimo (1998) have applied the agnostic approach to find differences between the national monetary transmission mechanisms in EMU member countries. The approach is useful to show that differences are hard to spot if a sufficiently lax prior is imposed on the impulse responses. It is, however, inoperative to identify differences because it imposes the same set of restrictions on impulse response functions in all countries. In contrast, the ICE imposes distributional restrictions, and permits to use either no assumptions about the impulse responses or use priors on impulse response functions from equilibrium assumptions only.

7 The ICE and monetary policy modelling

It has been argued previously (cf. e.g. Rudebusch 1998), that central bank behaviour is not well represented by random shocks. In addition, I have shown in section 2 that identified monetary policy shocks are, and indeed should be leptokurtic. To disentangle non-normal shocks by the ICE, one needs to assume an alternative distribution for these shocks.

In order to find a suitable distribution, consider the procedure during a central bank meeting. Committee members meet, and express different opinions about the state of the economy and the required stance of monetary policy. Assume that the stance of monetary policy can be described quantitatively. The meeting terminates with a monetary policy decision, which will be a weighted average of the most extreme opinions. Simplifying this procedure, the decision is approximately the average of the most extreme positions.[17] Although a simplification in itself, this representation of monetary policy decisions is probably more realistic than random shocks. Asymptotically, monetary policy shocks will then be distributed logistically, which is more leptokurtic than the normal distribution. Therefore, this conception about the decision-making process can explain why monetary policy shocks are leptokurtic. In addition, the logistic distribution can be extended to accommodate more or less kurtosis and skewness, as described in appendix B.

A logistic prior distribution for monetary policy shocks is not only theoretically more appealing and captures leptokurtosis, but it also permits to renounce on strong prior

[17] One may object that in this procedure, "extremist" central bankers will express excessive opinions during the meetings in order to tilt the scale towards their true opinion. This is unlikely because taking averages from the most extreme opinions is not the official procedure of decision-making. Furthermore, the results would remain unchanged if all extremists started with equally exaggerated claims. Finally, extremists with this kind of behaviour would quickly loose credibility within the committee.

Table 1: Relative Contemporaneous IRs to shocks from the ICE

	i	ii	iii	iv	v	vi	vii
Y	0.06	0.08	0.20	0.22	1	0.70	0.23
P	0.91	1	0.20	0.31	0.39	0.00	0.02
$PCOM$	0.68	0.70	1	0.33	0.19	0.34	0.27
NBR	0.02	0.01	0.02	0.13	0.03	0.13	1
FF	0.12	0.11	0.14	0.69	0.32	1	0.77
TR	0.09	0.10	0.21	0.18	0.20	1	0.91
$M1$	0.17	0.28	0.60	0.64	0.25	1	0.65
$FF - shock$	0.16	0.24	0.05	0.41	0.55	0.51	0.47
$NBR - shock$	0.06	0.07	0.15	0.18	0.21	0.02	0.96

Note: Absolute impulse responses of the variables in the left column to the shocks numbered i-vii. The responses are measured relative to the maximal response of this variable to any shock. The lower part gives correlations with the CEE monetary policy shocks.

assumptions about short-run neutrality of impulse responses.[18] As an example, reconsider the CEE VAR model using the logistic distribution.[19] Since logistic shocks are completely identified by independence, I only need prior information to determine which one of the seven shocks is the monetary policy shock. CEE assume in their identification procedure that the price level, commodity prices, and GDP are known to the central bank at the time of the monetary policy decision, but they admit that since GDP data are only available with a lag the assumption of GDP being predetermined to the monetary policy decision is arguable. In contrast, sluggish price reaction to monetary policy shocks is a recurrent finding of VARs about monetary policy. Following this, I assume that a monetary policy shock is the one with the smallest contemporaneous impact on the two price variables. Since GDP reacts quicker and is probably unknown at the time of the central bank's decision, I do not consider the impact of the shocks on GDP as decisive. Because all variables are in levels, the numbers are not interpretable. Therefore, Table 7 shows the contemporaneous impulse response relative to the largest contemporaneous response to any shock.

The shocks vi and vii are most consistent with these priors, and are therefore candidates for the monetary policy shock. Especially the contemporaneous reaction of the price

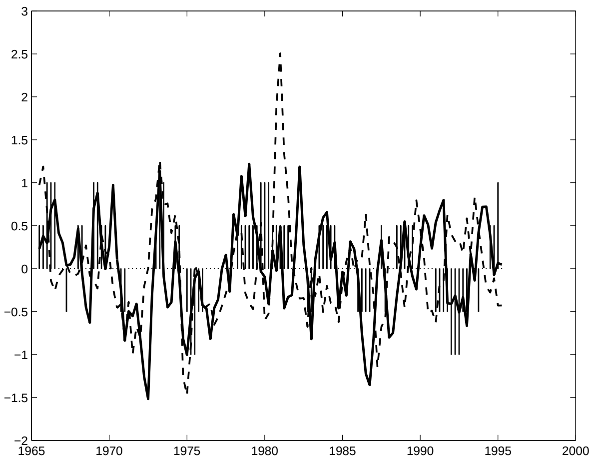
[18] In appendix B, I have adapted the estimation procedure so that linear constraints may be included. This even permits to combine strong priors from theoretical models with weak priors from informal evidence.

[19] In a complete model setup, one could model the monetary shocks by a logistic distribution and aggregate shocks as normal, where central limit tendencies are more important.

level is considerably larger to any of the shocks i to v. Both candidate shocks have a strong contemporaneous impact on the monetary variables.

Relative to the strongest shock, number vii has only 2 and 27 % influence on the price variables. It is noteworthy that is has immediate positive influence on output. However, since causality is unclear at this point, this could also indicate that the Fed has reacted to strong output growth with contractive monetary policy. Shock number vii is highly correlated (0.96) with CEE's NBR shock. It has by far the strongest contemporaneous impact on non-borrowed reserves and sizeable influence on total reserves and the Federal Funds rate. Therefore, it is a good candidate for the monetary policy shock.

Figure 6: History of shocks vi and vii



Note: 3-quarter smoothed shock series vi (solid) and vii (dashed) with Boschen-Mills narrative measure of monetary policy superimposed.

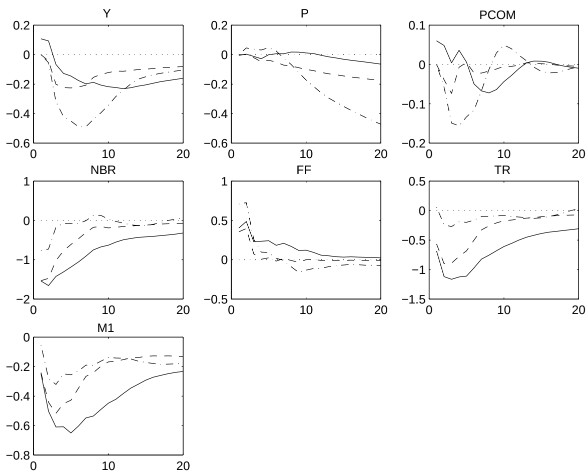
Similar arguments apply to shock number vi, except that it is less correlated with CEE's shocks. However, the impulse responses for shock vi are implausible: a shock that raises the Federal Funds rate *increases* total reserves, M1 and prices. Taken together, this evidence suggests that number vi is a money demand rather than a monetary policy shock. To verify this, look at the shock history in Figure 6. As a measure of the monetary policy stance, I have superimposed the negative of the Boschen-Mills narrative index.[20] The visual impression is that shock vii (solid) coincides better with the Boschen-Mills index. Its correlation with the index is 0.42, significantly higher than the 0.19 correlation between shock vi and the index (p-value 2.8). Finally, note that in 1980-81, the economy grew from a NBER though in 1980 (3) to a peak in 1981 (2).

[20]The series has been developed in Boschen & Mills. (1995). I use the data from Choi (1999).

Therefore, the large peak, which also appears in CEE's FF shock series, could indeed be a money demand shock, rather than a monetary policy shock.

Following the evidence that shock number vii is the monetary policy shock, consider the impulse repsonses im Figure 7: also in the long-run, the monetary policy shock has a stronger influence on non-borrowed and total reserves, as well as on the monetary aggregate than either of CEE's monetary policy shocks. In contrast, the price effects are much weaker than estimated by the CEE models. This supports the finding in CEE's paper that the price reaction is insignificant. Finally, it is noteworthy that the "price puzzle", i.e. a short-term rise in prices after a monetary policy shock, arises only with respect to commodity prices.[21]

Figure 7: Impulse Responses to a Monetary Policy Shock



Note: Impulse responses to a monetary policy shock in CEE's NBR (dashed) and FF (dash-dot) models , and the ICE (solid)

The discussion above suggests that identifying monetary policy shocks by a logistic distribution and weaker priors than usual, leads to a different interpretation of the shocks. In contrast with CEE, who interpreted them as two monetary policy shocks viewed from different angles, the two shocks now appear to be a money demand and a money supply shock.

Even if the view is maintained that monetary policy shocks are not leptokurtic but normally distributed with strong outliers, the ICE is an improvement over simple orthogonalisation. Renouncing on the orthogonality assumption makes the estimation more robust to outliers or non-normal distributions. Without recourse to assumptions

[21] Incidentally, these had become part of standard monetary VARs precisely to *avoid* the price puzzle.

about predetermination of variables, the estimator yields a unique representation of the endogenous variables in terms of independent shocks. These must only be named suitably. If less kurtosis is desired, he generalised logistic distribution (cf. Appendix B) can be used as a good approximation to the Gaussian. In fact, even decreasing the prior excess kurtosis to 0.07 did not alter the results.

In more recent contributions, independent component analysis has been extended to deal with a larger number of structural shocks than observed variables (Lewicki & Sejnowski 2000). This could be regarded as a remedy for the Faust & Leeper (1997) critique regarding the small number of shocks identified by SVAR models. However, this extension is beyond the scope of this paper and will not be pursued here.

8 Conclusion

In this paper, I have discussed SVAR modelling techniques including cointegrated models. I have focused on the identification scheme for discovering structural shocks from the VAR residuals. Two problems arise: the VAR residuals are often leptokurtic, and some of the restrictions on impulse response functions cannot be derived from a model. I argue that financial time series are necessarily involved in estimating monetary policy shocks, and therefore residuals can reasonably be expected to be leptokurtic. I argue that logistically distributed shocks are a better approximation to monetary policy shocks than Gaussian random shocks. I have therefore introduced Independent Component Analysis as a new technique to the econometric literature. The ICE works under leptokurtic distributional assumptions, and it avoids the necessity for zero restrictions on the contemporaneous impact of structural shocks. I use the exercises by Blanchard & Quah (1989) and Christiano et al. (1999) to demonstrate the new technique and contrast it with current modelling techniques.

A Orthogonality and Independence

This appendix demonstrates the link between orthogonality and the independence measure given by the Kullback & Leibler divergence. In the main text, it has been shown that the Kullback & Leibler divergence is a measure of independence. It is shown that in the specific case of multinormal variables it is equivalent to orthogonalising the variables.

Theorem 1 *In the case of multinormal variables $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_v)'$, the statements*

(a) the variables are mutually independent,

(b) the variables are orthogonal,

(c) the Kullback & Leibler divergence between the univariate distributions and the multinormal is zero are equivalent.

Proof. The equivalence of (a) and (c) has been shown in the main text. I show here that minimising the Kullback & Leibler divergence and minimising correlation is equivalent. Assume that the distribution of ε is multivariate Gaussian with mean zero.[22] Denote the univariate distributions of (unobserved) independent normal shocks by $f_i(\varepsilon_i)$ and the (observed) multivariate normal distribution by $f(\varepsilon)$. The distribution of the independent ε_i is given by $\prod_{i=1}^{v} f_i(\varepsilon_i) \sim \mathcal{N}(0, \sigma \mathbf{I})$, where σ is the univariate variance. Assuming that the multivariate covariance matrix, Σ_ε, is non-degenerate (i.e. positive definite), the Kullback & Leibler (1951) divergence is

$$\begin{aligned}
\mathcal{D}\left(f(\varepsilon) || \prod_{i=1}^{v} f_i(\varepsilon_i)\right) &= \int f(\varepsilon) \log \frac{f(\varepsilon)}{\prod_{i=1}^{v} f_i(\varepsilon_i)}\, d\varepsilon \\
&= \int f(\varepsilon) \log \left\{ \frac{(2\pi)^{-v/2} |\sigma \mathbf{I}|^{1/2}}{(2\pi)^{-v/2} |\Sigma_\varepsilon|^{1/2}} e^{1/2\varepsilon'(\sigma\mathbf{I})^{-1}\varepsilon - 1/2\varepsilon'\Sigma_\varepsilon^{-1}\varepsilon} \right\} d\varepsilon \\
&= \int f(\varepsilon) \left(\log \frac{|\sigma\mathbf{I}|^{1/2}}{|\Sigma_\varepsilon|^{1/2}} + \frac{1}{2}(\mathrm{tr}(\sigma\mathbf{I})^{-1}\varepsilon\varepsilon' - tr\Sigma_\varepsilon^{-1}\varepsilon\varepsilon') \right) d\varepsilon \\
&= \frac{1}{2} \int f(\varepsilon) \left(\log \frac{|\sigma\mathbf{I}|}{|\Sigma_\varepsilon|} + \mathrm{tr}((\sigma\mathbf{I})^{-1} - \Sigma_\varepsilon^{-1})\varepsilon\varepsilon' \right) d\varepsilon \\
&= \frac{1}{2} \log \frac{|\sigma\mathbf{I}|}{|\Sigma_\varepsilon|} \int f(\varepsilon) d\varepsilon + \frac{1}{2} \int f(\varepsilon) \mathrm{tr}((\sigma\mathbf{I})^{-1} - \Sigma_\varepsilon^{-1})\varepsilon\varepsilon' d\varepsilon \\
&= \frac{1}{2} \log \frac{|\sigma\mathbf{I}|}{|\Sigma_\varepsilon|} + \left[-(2\pi)^{-v/2} \Sigma_\varepsilon^{-1/2} \Sigma_\varepsilon \varepsilon' \varepsilon e^{-1/2\varepsilon'\Sigma_\varepsilon^{-1}\varepsilon} \right]_{-\infty}^{\infty} \\
&\quad + \frac{1}{2} tr\left(\Sigma_\varepsilon((\sigma\mathbf{I})^{-1} - \Sigma_\varepsilon^{-1}) \right)
\end{aligned}$$

[22] The zero mean assumption is for simplifying the exposition only; the results below also hold for non-zero means. Looking at the empirical side, SVAR analysis emphasises this assumption about the structural shocks.

$$
\begin{aligned}
&= \frac{1}{2}\log\frac{|\sigma\mathbf{I}|}{|\Sigma_\varepsilon|} + \frac{1}{2}tr\left(\Sigma_\varepsilon((\sigma\mathbf{I})^{-1} - \Sigma_\varepsilon^{-1})\right) \\
&= \frac{1}{2}\log\frac{|\sigma\mathbf{I}|}{|\Sigma_\varepsilon|} - \frac{v}{2} + \frac{1}{2}\mathrm{tr}(\Sigma_\varepsilon(\sigma\mathbf{I})^{-1}), \qquad (26)
\end{aligned}
$$

where $\mathrm{tr}(X)$ denotes the trace of the matrix X. Consider minimising this expression: The first term is minimised if the covariance is minimal, the last term is minimised if the variance is minimal. Since the distribution is multivariate Gaussian, the variance is unaffected by the direction of projection. Therefore, in the case of Gaussian shocks minimising the covariance is tantamount to minimising the Kullback & Leibler divergence.

Minimising this expression yields

$$
\begin{aligned}
\min\left(\frac{1}{2}\log\frac{|\sigma\mathbf{I}|}{|\Sigma_\varepsilon|} - \frac{v}{2} + \frac{1}{2}\mathrm{tr}(\Sigma_\varepsilon(\sigma\mathbf{I})^{-1})\right) &= \frac{1}{2}\min\left(\log|\sigma\mathbf{I}\Sigma_\varepsilon^{-1}| + \mathrm{tr}(\Sigma_\varepsilon(\sigma\mathbf{I})^{-1})\right) - \frac{v}{2} \\
&= \frac{1}{2}\min\left(\mathrm{tr}(\Sigma_\varepsilon(\sigma\mathbf{I})^{-1}) - \log|\Sigma_\varepsilon(\sigma\mathbf{I})^{-1}|\right) - \frac{v}{2} \\
&= \frac{v}{2} - \frac{v}{2} \\
&= 0,
\end{aligned}
$$

where I use Lütkepohl (1996, 4.2.1, Results (4a) and (7d)). The second line uses Lemma A.6 from Johansen (1995), which also implies that at the minimum $\Sigma_\varepsilon = \sigma\mathbf{I}$. In other words, the minimisation rule just decorrelates the shocks. At the minimum, the densities of the univariate independent shocks and the multinormal coincide, i.e. the shocks are independent. □

B Optimisation and Asymptotic Properties of the Independent Shock Estimator

This appendix derives the algorithm for the independent components estimator (24). I give the stochastic gradient ascent and the Hessian, so that assymptotic confidence bounds can be computed. I also give the specific results for the logistic distribution. Finally, I calculate the estimator for S under linear constraints. I frequently use rules for matrix differential calculus from Lütkepohl (1996); these are referred to as *Result*. They are reproduced for convenience in appendix C. In this appendix,the Kullback & Leibler divergence is used to measure the quality of the approximation.

Theorem 2 (Independent Shock Estimator) *In representation (6) for the observed residuals u_t, the dependence between the structural shocks ε_t is minimised by following the negative of the gradient for S,*

$$\mathcal{G} \quad = \quad S'^{-1} - \varphi(\mathcal{E})\mathcal{E}'S'^{-1}, \tag{27}$$

where

$$\varphi(\mathcal{E}) \quad = \quad \begin{bmatrix} \left.\frac{\frac{\partial f_1(\varepsilon_1)}{\partial \varepsilon_1}}{f_1(\varepsilon_1)}\right|_{\varepsilon_1=\varepsilon_{11}} & \cdots & \left.\frac{\frac{\partial f_1(\varepsilon_1)}{\partial \varepsilon_1}}{f_1(\varepsilon_1)}\right|_{\varepsilon_1=\varepsilon_{1T}} \\ \vdots & & \vdots \\ \left.\frac{\frac{\partial f_v(\varepsilon_v)}{\partial \varepsilon_v}}{f_v(\varepsilon_v)}\right|_{\varepsilon_v=\varepsilon_{v1}} & \cdots & \left.\frac{\frac{\partial f_v(\varepsilon_v)}{\partial \varepsilon_v}}{f_v(\varepsilon_v)}\right|_{\varepsilon_v=\varepsilon_{vT}} \end{bmatrix}.$$

Proof. The information usable to discern between the estimated and the observed densities is given in (24). It has been shown in the main text (p. 14) that minimising this information is tantamount to minimising the dependence between the structural shocks.

In order to minimise (24), decompose the information into marginal and conditional entropies,

$$\begin{aligned} \mathcal{I}\left(f(u_t), |S|^{-1}\prod_{i=1}^{v} f_i(\varepsilon_{it})\right) &= \mathcal{H}(u_t) - \mathcal{H}\left(u_t \,\middle|\, |S|^{-1}\prod_{i=1}^{v} f_i(\varepsilon_{it})\right) \\ &= -E[\log f(u_t)] - E\left[\log |S|^{-1}\prod_{i=1}^{v} f_i(\varepsilon_{it})\right]. \end{aligned} \tag{28}$$

Differentiating with respect to S, and considering that $f(u_t)$ does not depend on S yields the gradient,[23]

$$
\begin{aligned}
\mathcal{G} &= -E\left[\frac{\partial \log\left(|S|^{-1}\prod_{i=1}^{v} f_i(\varepsilon_{it})\right)}{\partial S}\right] \\
&= -\frac{\partial \log(|S|^{-1})}{\partial S} - E\left[\frac{\partial\left(\sum_{i=1}^{v}\log f_i(\varepsilon_{it})\right)}{\partial S}\right] \\
&= -|S^{-1}|^{-1}(-|S|^{-1})(S')^{-1} - E\left[\sum_{i=1}^{v}\left(f_i(\varepsilon_{it})^{-1}\frac{\partial f_i(\varepsilon_{it})}{\partial S}\right)\right] \\
&= S'^{-1} - E\left[\sum_{i=1}^{v}\left(f_i(\varepsilon_{it})^{-1}\frac{\partial f_i(\varepsilon_{it})}{\partial S}\right)\right] \\
&= S'^{-1} - E\left[\sum_{i=1}^{v}\left(\left(f_i(\varepsilon_{it})^{-1}\frac{\partial f_i(\varepsilon_{it})}{\partial \varepsilon_{it}}\right)\sum_{i=1}^{v}\frac{\partial \varepsilon_{it}}{\partial S}\right)\right] \\
&= S'^{-1} - E\left[\sum_{i=1}^{v}\frac{\frac{\partial(f_i(\varepsilon_{it}))}{\partial \varepsilon_{it}}}{f_i(\varepsilon_{it})}\varepsilon'_{it}S'^{-1}\right]
\end{aligned}
$$

where the third line uses *Result* 10.3.3 (5b). Replacing theoretical distributions with empirical values yields equation (27). The information in (28) is minimised by iteratively improveing the estimate of S using $\Delta S = \mathcal{G}$ until convergence. A local minimum is achieved if the gradient is zero. □

Notes:

(a) In practice, S and $\mathcal{E} = S^{-1}\mathcal{U}$ must be replaced by estimates.

(b) Since the ICE is closely connected with principal components, a conjecture is that a good starting point for the iterative algorithm is $S_1 = \sqrt{\lambda_{\hat{\Sigma}_u}}\mathcal{V}'_{\hat{\Sigma}_u}$, where $\lambda_{\hat{\Sigma}_u}$ is a diagonal matrix containing the eigenvalues of the estimated residual covariance matrix and $\mathcal{V}_{\hat{\Sigma}_u}$ is the corresponding matrix of eigenvectors.

(c) Stability of the algorithm can be improved using the natural gradient (Amari 1996),

$$\mathcal{G}^n = \mathcal{G}S'S.$$

Theorem 3 (Hessian Matrix) *If S is estimated by the algorithm from Theorem 2, the Hessian for S is*

$$
\begin{aligned}
\mathcal{H} &= (S^{-1}\mathcal{E}\otimes \mathbf{I}_v)\frac{\partial \mathrm{vec}\,(\varphi(\mathcal{E}))}{\partial \mathrm{vec}\,(\mathcal{E})'}(\mathcal{E}'\otimes S^{-1}) \\
&\quad + \left(S^{-1}\otimes \varphi(\mathcal{E})\mathcal{E}'S' + \mathbf{I}_v \otimes \{\varphi(\mathcal{E})\mathcal{E}' - \mathbf{I}_v\}\right) K_{vv}\left(S'^{-1}\otimes S^{-1}\right) \qquad (29)
\end{aligned}
$$

[23]It is easier to maximise the expression with respect to S^{-1}. This is done in the engineering literature. However, since economic theory provides for linear (long-run) restrictions on S, I derive the result for S.

where K_{vv} is a $v^2 \times v^2$ commutation matrix such that $K_{vv}vec(S)$ =vec(S'), the symbol $\otimes$ denotes the Kronecker product and $\frac{\partial \text{vec}(\varphi(\mathcal{E}))}{\partial \text{vec}(\mathcal{E})'}$ is the diagonal matrix

$$\frac{\partial \text{vec}\,(\varphi(\mathcal{E}))}{\partial \text{vec}\,(\mathcal{E})'} = \begin{bmatrix} \frac{\partial \varphi_1(\varepsilon_1)}{\partial \varepsilon_1}\Big|_{\varepsilon_1=\varepsilon_{11}} & & & 0 \\ & \frac{\partial \varphi_2(\varepsilon_2)}{\partial \varepsilon_2}\Big|_{\varepsilon_1=\varepsilon_{21}} & & \\ & & \ddots & \\ 0 & & & \frac{\partial \varphi_v(\varepsilon_v)}{\partial \varepsilon_v}\Big|_{\varepsilon_v=\varepsilon_{vT}} \end{bmatrix}.$$

Proof. Applying *Results* 10.3.1 (7) and 10.6 (1b) to Theorem 2 yields

$$\mathcal{H} = \frac{\partial \text{vec}\,(\mathcal{G})}{\partial \text{vec}\,(S)'} = -(S^{-1} \otimes S'^{-1})K_{vv} - \frac{\partial \text{vec}\,(\varphi(\mathcal{E})\mathcal{E}'S'^{-1})}{\partial \text{vec}\,(S)'},$$

where $\varphi(\mathcal{E})$ is defined in Theorem 2. Because $\mathcal{U} = S\mathcal{E}$, one may write equivalently $\varphi(S,\mathcal{U})$. *Results* 7.2 (6), 10.3.1 (4) and 10.7 (2) imply

$$\begin{aligned}
\mathcal{H} &= -(S^{-1} \otimes S'^{-1})K_{vv} - \frac{\partial \text{vec}\,(\varphi(S,\mathcal{U})\mathcal{U}'S'^{-2})}{\partial \text{vec}\,(\varphi)'}\frac{\partial \text{vec}\,(\varphi(S,\mathcal{U}))}{\partial \text{vec}\,(S)'} \\
&\quad - \frac{\partial \text{vec}\,(\varphi(S,\mathcal{U})\mathcal{U}'S'^{-2})}{\partial \text{vec}\,(S'^{-1})'}\frac{\partial \text{vec}\,(S'^{-1})}{\partial \text{vec}\,(S)'} \\
&= -(S^{-1} \otimes S'^{-1})K_{vv} - (S^{-2}\mathcal{U} \otimes \mathbf{I}_v)\frac{\partial \text{vec}\,(\varphi(S,\mathcal{U}))}{\partial \text{vec}\,(S^{-1}\mathcal{U})'}\frac{\partial \text{vec}\,(S^{-1}\mathcal{U})}{\partial \text{vec}\,(S^{-1})'}\frac{\partial \text{vec}\,(S^{-1})}{\partial \text{vec}\,(S)'} \\
&\quad + (\mathbf{I}_v \otimes \varphi(S,\mathcal{U})\mathcal{U}')\left(S^{-1} \otimes \mathbf{I}_v + \mathbf{I}_v \otimes S'^{-1}\right)\left(S^{-1} \otimes S'^{-1}\right)K_{vv} \\
&= -(S^{-1} \otimes S'^{-1})K_{vv} + (S^{-1}\mathcal{E} \otimes \mathbf{I}_v)\frac{\partial \text{vec}\,(\varphi(\mathcal{E}))}{\partial \text{vec}\,(\mathcal{E})'}(\mathcal{U}' \otimes \mathbf{I}_v)(S'^{-1} \otimes S^{-1}) \\
&\quad + (\mathbf{I}_v \otimes \varphi(S,\mathcal{U})\mathcal{U}')\left(S^{-1} \otimes \mathbf{I}_v + \mathbf{I}_v \otimes S'^{-1}\right)\left(S^{-1} \otimes S'^{-1}\right)K_{vv}
\end{aligned}$$

by *Results* 10.4.1 (4), 10.4.1 (3), 10.5.1 (1a) and 10.6 (1). Finally, using *Results* 2.4 (4) and (5) and 9.2.2 (5a) gives

$$\begin{aligned}
\mathcal{H} &= (S^{-2}\mathcal{U} \otimes \mathbf{I}_v)\frac{\partial \text{vec}\,(\varphi(\mathcal{E}))}{\partial \text{vec}\,(\mathcal{E})'}(\mathcal{U}'S'^{-1} \otimes S^{-1}) \\
&\quad + \left(S^{-1} \otimes \varphi(S,\mathcal{U})\mathcal{U}' + \mathbf{I}_v \otimes \varphi(S,\mathcal{U})\mathcal{U}'S'^{-1} - \mathbf{I}_v \otimes \mathbf{I}_v\right)K_{vv}\left(S'^{-1} \otimes S^{-1}\right)
\end{aligned}$$

Rewriting gives equation (29). □

Note: Since the problem is to minimise the distance, the Hessian must be positive (semi-) definite at the optimal S.

To make these results applicable, the distribution $f_i(\varepsilon_{it})$ must be given or estimated. In section 7, I have argued that shocks from single decisions can be modeled as averaged extreme opinions. Asymptotically, the distribution is then logistic, which is given by the cumulative distribution

$$F(\varepsilon_{it}) \quad = \quad (1+\exp(-\varepsilon_{it}))^{-1},$$

which implies the distribution function

$$f(\varepsilon_{it}) = \partial F(\varepsilon_{it})/\partial\varepsilon_{it} = F(\varepsilon_{it})(1-F(\varepsilon_{it})).$$

This distribution has some excess kurtosis with respect to the Gaussian of 1.2.

Corollary 1 (Independent Shocks Estimator for Logistic Distribution) *If the structural shocks ε_t are distributed logistically, the dependence between them in representation (6) is minimised by following the negative of the gradient*

$$\mathcal{G}^n \quad = \quad \left(-\mathbf{I}_v + \left(1 - \frac{2}{1+\exp(-\mathcal{E})}\right)\mathcal{E}'\right) S. \tag{30}$$

Proof. For the logistic distribution, the term is $\varphi(\mathcal{E}) = \frac{\exp(-\mathcal{E})-1}{\exp(-\mathcal{E})+1}$. Applying Theorem 2 to the logistic distribution, the gradient is

$$\mathcal{G} \quad = \quad -S'^{-1} + \left(1 - \frac{2}{1+\exp(-\mathcal{E})}\right)\mathcal{E}'S'^{-1}.$$

Hence, equation (30) gives the natural gradient. □

Corollary 2 (Hessian for Logistic Distribution) *If the independent shocks are distributed logistically, the Hessian for S is*

$$\begin{aligned}\mathcal{H} \quad = \quad & -2(S^{-1}\mathcal{E}\otimes\mathbf{I}_v)\varphi_l(\mathcal{E}'\otimes S^{-1}) \\ & + \left(S^{-1}\otimes\frac{\exp(-\mathcal{E})-1}{1+\exp(-\mathcal{E})}\mathcal{E}'S' + \mathbf{I}_v\otimes\left\{\frac{\exp(-\mathcal{E})-1}{1+\exp(-\mathcal{E})}\mathcal{E}' - \mathbf{I}_v\right\}\right) K_{vv}\left(S'^{-1}\otimes S^{-1}\right),\end{aligned}$$

where φ_l is a diagonal matrix with typical element

$$\frac{\exp(-\varepsilon_{ij})}{(1+\exp(-\varepsilon_{ij}))^2} \qquad i = 1\ldots v, j = 1\ldots T.$$

Proof. This is just an application of Theorem 3.

Theorem 4 (Linear Constraints) *If the independent shocks are distributed logistically, the optimal matrix S under linear constraints is given by (9), where γ is optimised by following the negative of the natural gradient*

$$\mathcal{G}_\gamma \quad = \quad \left\{\mathbf{I}_v\otimes\left(-\mathbf{I}_v + \left\{1 - \frac{2}{1+\exp(-\tilde{\mathcal{E}})}\right\}\tilde{\mathcal{E}}'\right)\right\}\Phi_\perp\gamma, \tag{31}$$

where $\tilde{\mathcal{E}} = \tilde{S}^{-1}\mathcal{U}$ and $\operatorname{vec}\left(\tilde{S}\right) = \Phi_\perp\gamma + \Phi(\Phi'\Phi)^{-1}\chi_1$.

Proof. It has been shown in the main text that equation (9) is a valid representation of S under linear constraints. To show the optimality of γ, consider that

$$\mathcal{G}_\gamma = \frac{\partial \mathcal{H}}{\partial \gamma} = \frac{\partial \mathcal{H}}{\partial \mathrm{vec}\,(S)'} \frac{\partial \mathrm{vec}\,(S)'}{\partial \gamma}.$$

Then vectorizing (29), using *Result* 7.2.(5) and inserting equation (9) yields the required result. □

It is noteworthy that the logistic distribution is less restricted than the Gaussian normal in that it can be extended according to the researcher's prior beliefs to capture more (less) kurtosis by using

$$f_i(\varepsilon_{it}) = \frac{1}{B(\tau,\tau)} F_i(\varepsilon_{it})^\tau (1 - F_i(\varepsilon_{it}))^\tau, \tag{32}$$

where $B(\cdot,\cdot)$ is the Beta function (e.g. Mood et al. 1974, p. 535) and $0 \leq \tau \leq \infty$ smaller (larger) than one. The kurtosis is given by

$$\frac{K_4}{[K_2(X)]^2} = \frac{\Psi'''(\tau)}{2\left(\Psi'(\tau)\right)^2},$$

where Ψ is the polygamma function (Wu, Hung & Lee 2000).

This expression is decreasing in τ and assumes its maximum at $\tau = 0$, where the excess kurtosis is 3. At $\tau = \infty$ the excess kurtosis is zero.

Allowing for different exponents in (32), the distribution even accommodates skewness. Since the generalised logistic distribution depends on only one (with skewness two) parameters, it may be estimated from shocks identified by other methods (cf. page 15) or even used for an iterative procedure, rather than be given exogenously.

C Matrix Results used

This appendix reproduces the matrix results from Lütkepohl (1996) referred to in the main text nad gives an expression for the vec operator.

2.4 (3)	$A(m \times n), B(p \times q), C(r \times s) : A \otimes (B \otimes C) = (A \otimes B) \otimes C = A \otimes B \otimes C$
2.4 (4)	$A, B(m \times n), C, D(p \times q) : (A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D$
2.4 (5)	$A(m \times n), B(p \times q), C(n \times r), D(q \times s) : (A \otimes B)(C \otimes D) = AC \otimes BD$
4.2.1 (4a)	$A, B(m \times m) : \det(AB) = \det(A)\det(B)$
4.2.1 (7d)	$\det(A^{-1}) = (\det A)^{-1}$ if A is nonsingular
4.3.1 (7)	$A(m \times n) : \mathrm{rk}(A) = r \Rightarrow \exists\ B(m \times r)$ and $C(r \times n)$ such that $A = BC$
4.3.1 (10)	$A(m \times n), B(m \times m), C(n \times n) : B, C$ nonsingular $\Rightarrow \mathrm{rk}(BAC) = \mathrm{rk}(A)$
4.3.3 (2a)	$\mathrm{rk}(AB) \leq \min\{\mathrm{rk}(A), \mathrm{rk}(B)\}$
4.3.3 (3)	$A, B(m \times n) : \mathrm{rk}(A + B) \leq \mathrm{rk}(A) + \mathrm{rk}(B)$
7.2 (1)	$A, B(m \times n) : \mathrm{vec}\,(A \pm B) = \mathrm{vec}\,(A) \pm \mathrm{vec}\,(B)$
7.2 (5)	$A(m \times n), B(n \times p) : \mathrm{vec}\,(AB) = (\mathbf{I}_p \otimes A)\mathrm{vec}\,(B)$
7.2 (6)	$A(m \times n), B(n \times r), C(r \times s) : \mathrm{vec}\,(ABC) = (C' \otimes A)\mathrm{vec}\,(B)\,.$
9.2.2 (5a)	$A(m \times n), B(p \times q) : K_{pm}(A \otimes B) = (B \otimes A)K_{qn}$
10.2.1 (2)	$x(m \times 1), f(x), g(x)$ real valued functions, c_1, c_2 real scalars: $\frac{\partial[c_1 f(x) + c_2 g(x)]}{\partial x} = c_1 \frac{\partial f(x)}{\partial x} + c_2 \frac{\partial g(x)}{\partial x}$
10.3.1 (4)	$X(m \times n), f(X), g(X), h(X)$ real valued functions: $\frac{\partial f(X)g(X)h(X)}{\partial X} = f(X)g(X)\frac{\partial h(X)}{\partial X} + f(X)h(X)\frac{\partial g(X)}{\partial X} + g(X)h(X)\frac{\partial f(X)}{\partial X}$
10.3.1 (7)	$X(m \times n), f(X)$ a real valued function: $\mathrm{vec}\left(\frac{\partial f(X)}{\partial X}\right) = \frac{\partial f(X)}{\partial \mathrm{vec}(X)}$
10.3.3 (5b)	$X(m \times m)$ nonsingular: $\frac{\partial \det(X^{-1})}{\partial X} = -(\det X)^{-1}(X')^{-1}$
10.4.1 (3)	$X(m \times n), A(p \times m), B(n \times q) : \frac{\partial \mathrm{vec}(AXB)}{\partial \mathrm{vec}(X)'} = B' \otimes A$
10.4.1 (4)	$X(m \times n), A(p \times n), B(m \times q) : \frac{\partial \mathrm{vec}(AX'B)}{\partial \mathrm{vec}(X)'} = (B' \otimes A)K_{mn}$
10.5.1 (1a)	$X(m \times m) : \frac{\partial \mathrm{vec}(X^2)}{\partial \mathrm{vec}(X)'} = X' \otimes \mathbf{I}_m + \mathbf{I}_m \otimes X$
10.6 (1a)	$X(m \times m)$ nonsingular: $\frac{\partial \mathrm{vec}(X^{-1})}{\partial \mathrm{vec}(X)'} = -X'^{-1} \otimes X^{-1}$
10.6 (1b)	$X(m \times m)$ nonsingular: $\frac{\partial \mathrm{vec}(X'^{-1})}{\partial \mathrm{vec}(X)'} = -(X^{-1} \otimes X'^{-1})K_{mm}$
10.7 (2)	$X(m \times n), Y(X)(p \times q), Z(Y)(r \times s) : \frac{\partial \mathrm{vec}(Z(Y(X)))}{\partial \mathrm{vec}(X)'} = \frac{\partial \mathrm{vec}(Z(Y))}{\partial \mathrm{vec}(Y)'}\frac{\partial \mathrm{vec}(Y(X))}{\partial \mathrm{vec}(X)'}$

An algebraic representation of the vec operator is $S(m \times n)\ :\ \mathrm{vec}\,(S) = \mathbf{1}_{mn} \odot (\mathbf{1}_n \otimes S \otimes \mathbf{1}'_m)\,\mathbf{1}_{mn}$, where $\odot$ is the Hadamard, or elementwise, product.

References

Amari, S.-i. (1996), 'Natural gradient works efficiently in learning', *Neural Computation* **10**, 251 – 276.

Aparicio, F. M. & Escribano, A. (1998), 'Information-theoretic analysis of serial dependence and cointegration', *Studies in Nonlinear Dynamics and Econometrics* **3**(3), 119 – 140.

Bell, A. J. & Sejnowski, J. (1995), 'An information-maximisation approach to blind separation and blind deconvolution', *Neural Computation* **7**(6), 1004 – 1034.

Blanchard, O. J. & Quah, D. (1989), 'The dynamic effects of aggregate demand and supply disturbances', *American Economic Review* **79**, 655–673.

Boschen, J. & Mills., L. (1995), 'The relation between narrative and money market indicators of monetary policy', *Economic Inquiry* **33**(1), 24 – 44.

Boswijk, H. P. & André, L. (2002), 'Semi-nonparametric cointegration testing', *Journal of Econometrics* p. forthcoming.

Bronstein, I. N., Semendjajew, K. A., Musiol, G. & Mühlig, H. (1997), *Taschenbuch der Mathematik*, 3. edn, Harri Deutsch, Thun and Frankfurt.

Choi, W. G. (1999), 'Asymmetric monetary effects on interest rates across monetary policy stances', *Journal of Money, Credit and Banking* **31**(3), 386 – 416.

Christiano, L. J., Eichenbaum, M. & Evans, C. L. (1999), Monetary policy shocks: What have we learned and to what end?, *in* J. Taylor & M. Woodford, eds, 'Handbook of Macroeconomics', Elsevier, Amsterdam.

Coenen, G. & Vega, J.-L. (1999), The demand for M3 in the euro area, Working Paper 6, European Central Bank, Frankfurt.

Engle, R. F. & Granger, C. W. J. (1987), 'Co-integration and error correction: Representation, estimation, and testing', *Econometrica* **55**(2), 251–276.

Faust, J. (1998), 'The robustness of identified VAR conclusions about money', *Carnegie-Rochester Conference Series on Public Policy* **49**, 207 – 244.

Faust, J. & Leeper, E. M. (1997), 'When do long-run identifying restrictions give reliable results', *Journal of Business and Economic Statistics* **15**(3), 345 – 353.

Favero, C. A. (2001), *Applied Macroeconometrics*, Oxford University Press, Oxford.

Galí, J. (1992), 'How well does the IS-LM model fit postwar US-data?', *Quarterly Journal of Economics* **107**, 709 – 738.

Gerlach, S. & Smets, F. (1995), The monetary transmission mechanism: Evidence from the G-7 countries, Working Paper 26, Bank for International Settlements, Basle.

Golan, A. (2002), 'Information and entropy econometrics – editor's view', *Journal of Econometrics* **107**, 1– 15.

Gonzalo, J. (1994), 'Five alternative methods of estimating long-run equilibrium relationships', *Journal of Econometrics* **60**, 203 – 233.

Hamilton, J. (1994), *Time Series Analysis*, PUP, Princeton.

Huber, P. J. (1981), *Robust statistics*, Wiley, New York, NY.

Hubrich, K. & Vlaar, P. J. G. (2000), Germany and the euro area: Differences in the transmission process of monetary policy, Research Memorandum 613, De Nederlandsche Bank, Amsterdam.

Johansen, S. (1991), 'Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models', *Econometrica* **59**, 1551–1580.

Johansen, S. (1995), *Likelihoo-Based Inference in Cointegrated Vector Auto-Regressive models*, Oxford University Press, Oxford, New York.

Juselius, K. (1998), 'A structured VAR for Denmark under changing monetary regimes', *Journal of Business and Economic Statistics* **16**(4), 400 – 411.

Jutten, C. & Herault, J. (1991), 'Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture', *Signal Processing* **24**, 1 – 10.

Kieler, M. & Saarenheimo, T. (1998), Differences in monetary policy transmission? A case not closed, Economic Paper 132, European Commission, Bruxelles.

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

Kullback, S. & Leibler, R. A. (1951), 'On information and sufficiency', *Annals of Mathematical Statistics* **22**, 79 – 86.

Leeper, E., Sims, C. A. & Zha, T. (1996), 'What does monetary policy do?', *Brookings Papers on Economic Activity* **2**, 1 – 78.

Lewicki, M. S. & Sejnowski, T. J. (2000), 'Learning overcomplete representations', *Neural Computation* **12**(2), 337 – 365.

Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, second edn, Springer, Berlin.

Lütkepohl, H. (1996), *Handbook of Matrices*, Wiley, Chichester, New York.

Maier, P. (2000), 'Pressure on the Bundesbank?', *Kredit und Kapital* **33**(4), 1 – 30.

Mood, A., Graybill, F. A. & Boes, D. C. (1974), *Introduction to the Theory of Statistics*, McGraw-Hill, London.

Robinson, P. M. (1991), 'Consistent nonparametric entropy-based testing', *Review of Economic Studies* **58**, 437 – 453.

Romer, C. & Romer, D. (1989), 'Does monetary policy matter? a new test in the spirit of Friedman and Schwartz', *NBER Macroeconomics Annual* pp. 121 – 183.

Rudebusch, G. D. (1998), 'Do measures of monetary policy in a VAR make sense?', *International Economic Review* **39**(4), 907 – 31.

Sargan, D. (1988), *Lectures on Advanced Econometric Theory*, Basil Blackwell, Oxford.

Silvapulle, P. S. & Podivinsky, J. M. (2000), 'The effect of non-normal disturbances and conditional heteroskedasticity on multiple cointegration tests', *Journal of Statistical Computation and Simulation* **65**(2), 173 – 189.

Sims, C. A. (1980), 'Macroeconomics and reality', *Econometrica* **48**, 1 – 48.

Sims, C. A. (1988), Identifying policy effects, *in* R. C. Bryant, ed., 'Empirical Macroeconomics for Interdependent Economies', Brookings Institution, Washington DC, pp. 305 – 321.

Stock, J. H. & Watson, M. W. (1988), 'Testing for common trends', *Journal of the American Statistical Association* **83**(404), 1097 – 1107.

Uhlig, H. (1999), What are the effects of monetary policy on output? Results from an agnostic identification procedure, Discussion Paper 9928, CentER, Tilburg University.

Vlaar, P. J. G. & Schuberth, H. (1998), Monetary transmission and controllability of money in Europe, Research Memorandum 544, De Nederlandsche Bank, Amsterdam.

Weber, C. E. (1995), 'Cyclical output, cyclical unemployment and Okun's coefficient: A new approach', *Journal of Applied Econometrics* **10**(4), 433 – 445.

Wu, J.-W., Hung, W.-L. & Lee, H.-M. (2000), 'Some moments and limit behaviors of the generalized logistic distribution with applications', *Proceedings of the National Science Council, ROC(A)* **24**(1), 7 – 14.