LECTURE NOTES ON STRUCTURAL VECTOR AUTOREGRESSIONS

ANDERS WARNE

1. OUTLINE

- Macroeconomic background
  - Sims (1980)
  - Stock and Watson (1988)
- Vector Autoregressions
  1. Stationarity vs. nonstationarity
  2. Structural models
  3. Dynamic experiments
  4. Estimation
  - Lütkepohl (1991), chapter 2
  - Hamilton (1994), chapter 11
  - Sims (1980)
  - Cooley and LeRoy (1985)
  - Runkle (1987)
- Cointegration and Common Trends
  - Johansen and Juselius (1990)
  - King, Plosser, Stock, and Watson (1991)
  - Mellander, Vredin, and Warne (1992)
  - Englund, Vredin, and Warne (1994)
  - Jacobson, Vredin, and Warne (1996)

These notes will not discuss estimation and inference in structural VAR's; the reader is instead adviced to consult the sources listed above. Rather, the main purpose is to explain terminology and concepts by focusing on a few simple examples. Some familiarity with ARIMA models is assumed.

DATE: First version: November 1996 (This compilation: January 2000).

Example: Let Y_t and C_t denote (the natural logarithms of) aggregate income and consumption, respectively. Consider the following version of the permanent income hypothesis (PIH) for $t = 1, 2, \dots$:

$$Y_t = Y_t^{\mathrm{p}} + v_t, \tag{2.1}$$

$$Y_t^{\mathrm{p}} = \mu_Y + Y_{t-1}^{\mathrm{p}} + u_t, \tag{2.2}$$

$$C_t = Y_t^{\mathrm{p}}, \tag{2.3}$$

where (u_t, v_t) is $iid(0, \mathrm{Diag}(\sigma_u^2, \sigma_v^2))$ and Y_0^{p} is fixed. Solving for permanent income, Y_t^{p}, in terms of Y_0^{p} and u_i we obtain

$$Y_t^{\mathrm{p}} = Y_0^{\mathrm{p}} + \mu_Y t + \sum_{i=1}^{t} u_i. \tag{2.4}$$

Hence, aggregate income and consumption are given by:

$$\begin{aligned} Y_t &= Y_0^{\mathrm{p}} + \mu_Y t + \textstyle\sum_{i=1}^{t} u_i + v_t, \\ C_t &= Y_0^{\mathrm{p}} + \mu_Y t + \textstyle\sum_{i=1}^{t} u_i. \end{aligned} \tag{2.5}$$

Note: Y_t and C_t are nonstationary since, conditional on Y_0^{p}, the mean and the variance for both variables depend on t. For example,

$$E[Y_t | Y_0^{\mathrm{p}}] = Y_0^{\mathrm{p}} + \mu_Y t, \tag{2.6}$$

$$V[Y_t | Y_0^{\mathrm{p}}] = \sigma_u^2 t + \sigma_v^2. \tag{2.7}$$

Furthermore, the following transformations of aggregate income and consumption are weakly stationary[1]

$$\Delta Y_t = \mu_Y + u_t + \Delta v_t, \tag{2.8}$$

$$\Delta C_t = \mu_Y + u_t, \tag{2.9}$$

$$C_t - Y_t = -v_t. \tag{2.10}$$

Here, $\Delta = 1 - L$ is the first difference operator and L is the lag operator, i.e. $Lx_t = x_{t-1}$. Technically, we have found that Y_t, C_t are integrated of order 1 (denoted by I(1)) and cointegrated of order (1,1) (denoted by CI(1,1)). The latter property means that a linear combination of I(1) variables is I(0) (weakly stationary).

The term *integrated* comes from the observation that, e.g., Y_t in (2.5) includes a component where we sum from 1 to t (discrete integration) over a stationary variable. Since we

[1] A time series is said to be *weakly stationary* if the first and second moments are invariant (in an absolute sense) with respect to time.

sum once over this interval we say that this component is integrated of order 1. The change in Y_t includes zero such summations and is therefore integrated of order zero.

3. Vector Autoregressions

Question 1: *What is a VAR system?*

From equation (2.8) we have that

$$C_t = \mu_Y + C_{t-1} + u_t. \tag{3.1}$$

That is, aggregate consumption is decribed by an AR(1) process. Moreover, equation (2.8) also gives us that aggregate income is related to consumption according to

$$Y_t = C_t + v_t. \tag{3.2}$$

Substituting for C_t we obtain

$$\begin{aligned} Y_t &= \mu_Y + C_{t-1} + u_t + v_t \\ &= \mu_Y + C_{t-1} + w_t. \end{aligned} \tag{3.3}$$

Collecting (3.1) and (3.3), they can be written in vector form as

$$\begin{bmatrix} Y_t \\ C_t \end{bmatrix} = \begin{bmatrix} \mu_Y \\ \mu_Y \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ C_{t-1} \end{bmatrix} + \begin{bmatrix} w_t \\ u_t \end{bmatrix}, \tag{3.4}$$

or more compactly

$$x_t = \mu + \Pi_1 x_{t-1} + \varepsilon_t, \tag{3.5}$$

a VAR(1) system for x_t.

- ε_t and x_{t-1} are uncorrelated. A consequence of this is that $\varepsilon_t = x_t - E[x_t | x_{t-1}, x_{t-2}, \ldots]$. In other words, ε_t is a *Wold innovation*, i.e. it represent the new information in x_t relative to its history.
- The covariance matrix of ε_t is given by

$$\begin{aligned} E[\varepsilon_t \varepsilon_t'] &= \begin{bmatrix} E[w_t^2] & E[w_t u_t] \\ E[w_t u_t] & E[u_t^2] \end{bmatrix} \\ &= \begin{bmatrix} \sigma_u^2 + \sigma_v^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 \end{bmatrix} \equiv \Sigma, \end{aligned} \tag{3.6}$$

This matrix is positive definite since $a'\Sigma a > 0$ for all $a \in \mathbb{R}^2 : a \neq 0$.

- The system in (3.5) is nonstationary since the individual time series, Y_t and C_t, are nonstationary.
- The system in (3.5) is a *reduced form*, i.e. neither the parameters (μ, Π_1, Σ) nor the innovations ε_t have an economic interpretation.

QUESTION 2: *What is a structural VAR system?*

Consider the VAR system

$$B_0 x_t = \gamma + B_1 x_{t-1} + \eta_t, \tag{3.7}$$

where η_t is $iid(0, \Omega)$ and Ω is positive definite.

Roughly, we shall say that (3.7) is a structural VAR(1) system if the parameters $(\gamma, B_0, B_1, \Omega)$ and/or the innovations η_t can be given an economic interpretation.

NOTE: Combining equations (3.1) and (3.2) we get

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Y_t \\ C_t \end{bmatrix} = \begin{bmatrix} 0 \\ \mu_Y \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ C_{t-1} \end{bmatrix} + \begin{bmatrix} v_t \\ u_t \end{bmatrix}, \tag{3.8}$$

where

- u_t is a shock to permanent income.
- v_t is a shock to transitory income.
- u_t and v_t are uncorrelated (independent if the joint distribution is Gaussian).

We can examine how Y and C react to permanent and transitory income shocks by using equation (2.5). For income we find that a one standard deviation shock at an arbitrary t in permanent and transitory income, respectively, implies the following responses:

$$\begin{aligned} \text{resp}(Y_{t+j} | u_t = \sigma_u, u_{t+1} = \ldots = u_{t+j} = 0) &= \sigma_u \quad \text{for all } j \geq 0, \\ \text{resp}(Y_{t+j} | v_t = \sigma_v, v_{t+1} = \ldots = v_{t+j} = 0) &= \begin{cases} \sigma_v & \text{if } j = 0, \\ 0 & \text{for all } j \geq 1. \end{cases} \end{aligned}$$

while the reactions in consumption are:

$$\begin{aligned} \text{resp}(C_{t+j} | u_t = \sigma_u, u_{t+1} = \ldots = u_{t+j} = 0) &= \sigma_u \quad \text{for all } j \geq 0, \\ \text{resp}(C_{t+j} | v_t = \sigma_v, v_{t+1} = \ldots = v_{t+j} = 0) &= 0 \quad \text{for all } j \geq 0. \end{aligned}$$

These dynamic functions are called *impulse response functions*. Notice that permanent income shocks have permanent effects on income and consumption, while transitory income shocks only have transitory effects on income and no effect on consumption.

We can also study the relative importance of the two shocks through *forecast error variance decompositions*. To derive these parameters we first note that income at $t + j$ is given by

$$Y_{t+j} = Y_0^{\mathrm{p}} + \mu_Y(t + j) + \sum_{i=1}^{t+j} u_i + v_{t+j}.$$

Hence, the expectation of Y_{t+j} conditional on current (t) and past values of x (and the parameters) is

$$E[Y_{t+j}|x_t, x_{t-1}, \dots] = Y_0^{\mathrm{p}} + \mu_Y(t + j) + \sum_{i=1}^{t} u_i.$$

Accordingly, the forecast error for all $j \geq 1$ is

$$Y_{t+j} - E[Y_{t+j}|x_t, x_{t-1}, \dots] = \sum_{i=t+1}^{t+j} u_i + v_{t+j}.$$

The variance of this random variable is given by

$$V[Y_{t+j} - E\left[Y_{t+j}|x_t, x_{t-1}, \dots\right]] = j\sigma_u^2 + \sigma_v^2.$$

The forecast error variance thus contains two parts; one part is due to permanent income shocks while the remainder is due to transitory income shocks. The share of the total forecast error variance explained by permanent income shocks is thus

$$w_{yp,j} = \frac{j\sigma_u^2}{j\sigma_u^2 + \sigma_v^2},$$

while the share explained by transitory income shocks is

$$w_{y\tau,j} = \frac{\sigma_v^2}{j\sigma_u^2 + \sigma_v^2}.$$

REMARKS:

1. $w_{yp,j}, w_{y\tau,j} \geq 0$ for all $j \geq 1$,
2. $w_{yp,j} + w_{y\tau,j} = 1$ for all $j \geq 1$, and
3. $\lim_{j\to\infty} w_{yp,j} = 1$.

Similar expressions can be derived for consumption.

In summary, to analyse the dynamic behavior of a structural VAR model, impulse response functions represent the reactions in the endogenous variables to the structural shocks, while variance decompositions describe the relative importance of the shocks.

4. Stability and Stationarity

Let $x_t \in \mathbb{R}^n$ be a vector of random variables generated by the following Gaussian VAR model:

$$x_t = \mu + \sum_{j=1}^{p} \Pi_j x_{t-j} + \varepsilon_t, \quad t = 1, 2, \ldots, T, \tag{4.1}$$

where for all t

$$\varepsilon_t \sim \text{iid } N_n(0, \Sigma), \tag{4.2}$$

while Σ is positive definite and $x_0, \ldots, x_{1-p}$ is fixed. The parameter p is called the lag length (order) and is assumed to be finite.

Example: $n = p = 1$, i.e. an AR(1) process.

$$x_t = \mu + \Pi_1 x_{t-1} + \varepsilon_t,$$

or using the lag operator

$$(1 - \Pi_1 L) x_t = \mu + \varepsilon_t.$$

If $|\Pi_1| < 1$, then the polynomial $(1 - \Pi_1 z)$ is *invertible* for all $|z| \leq 1$ and the AR(1) process is said to be *stable*. It now follows that

$$\begin{aligned} x_t &= (1 - \Pi_1 L)^{-1} (\mu + \varepsilon_t) \\ &= \left(\textstyle\sum_{j=0}^{\infty} \Pi_1^j L^j \right) (\mu + \varepsilon_t) \\ &= \textstyle\sum_{j=0}^{\infty} \Pi_1^j \mu + \sum_{j=0}^{\infty} \Pi_1^j \varepsilon_{t-j} \\ &= \mu / (1 - \Pi_1) + \textstyle\sum_{j=0}^{\infty} \Pi_1^j \varepsilon_{t-j}. \end{aligned}$$

Hence, an MA(∞) representation of x_t exists since

$$\lim_{j \to \infty} |\Pi_1|^j = 0.$$

It is now easy to compute the mean of x_t. This parameter is given by

$$E[x_t] = \frac{\mu}{1 - \Pi_1}.$$

Similarly, the variance is

$$\begin{aligned} V[x_t] &= E[[\textstyle\sum_{j=0}^{\infty} \Pi_1^j \varepsilon_{t-j}]^2] \\ &= \textstyle\sum_{j=0}^{\infty} \Pi_1^{2j} E[\varepsilon_{t-j}^2] \\ &= \Sigma / (1 - \Pi_1^2). \end{aligned}$$

The autocovariances can be computed similarly. Note first that

$$\begin{aligned} x_t &= \mu/(1-\Pi_1) + \textstyle\sum_{j=0}^{h-1} \Pi_1^j \varepsilon_{t-j} + \sum_{j=h}^{\infty} \Pi_1^j \varepsilon_{t-j} \\ &= \mu/(1-\Pi_1) + \textstyle\sum_{j=0}^{h-1} \Pi_1^j \varepsilon_{t-j} + \sum_{j=0}^{\infty} \Pi_1^{j+h} \varepsilon_{t-h-j}. \end{aligned}$$

Hence, the autocovariances for all $h \geq 1$ are

$$\begin{aligned} C[x_t, x_{t-h}] &= E\left[(\textstyle\sum_{j=0}^{h-1} \Pi_1^j \varepsilon_{t-j} + \sum_{j=0}^{\infty} \Pi_1^{j+h} \varepsilon_{t-h-j})(\sum_{j=0}^{\infty} \Pi_1^j \varepsilon_{t-h-j})\right] \\ &= \textstyle\sum_{j=0}^{\infty} \Pi_1^{2j+h} E[\varepsilon_{t-h-j}^2] \\ &= \Pi_1^h \textstyle\sum_{j=0}^{\infty} \Pi_1^{2j} \Sigma \\ &= \Pi_1^h \Sigma/(1-\Pi_1^2) \\ &= \Pi_1^h V[x_t]. \end{aligned}$$

Finally,

$$\lim_{h\to\infty} C[x_t, x_{t-h}] = 0,$$

since $|\Pi_1| < 1$.

CONCLUSION 1: *When x_t is generated by a stable* $(|\Pi_1| < 1)$ *AR(1) process and ε_t is iid* $(0, \Sigma)$ *(we have not used the assumption of normality), then x_t is also*

1. weakly stationary *since the first and second moments are invariant with respect to time*
2. ergodic *since the dependence between x_t and x_{t-h} (in an absolute sense) declines as the distance h increases.*

Let us now examine the general case. Consider the matrix polynomial

$$\Pi(z) = I_n - \sum_{j=1}^{p} \Pi_j z^j,$$

obtained from

$$\Pi(L)x_t = \mu + \varepsilon_t.$$

QUESTION 3: *Under which condition is $\Pi(z)$ invertible?*

Suppose the inverse exists. Then

$$\Pi(z)^{-1} = \frac{1}{\det[\Pi(z)]}\text{Adj}[\Pi(z)].$$

The adjoint (cofactor) matrix of $\Pi(z)$ always exists. Hence, $\Pi(z)$ is invertible if and only if the determinant is nonzero for all $|z| \leq 1$.

The polynomial $\det[\Pi(z)]$ is of order np since $\Pi(z)$ is $n \times n$ and of order p.

EXAMPLE: Suppose $n = 2$. Then

$$\Pi(z) = \begin{bmatrix} \Pi_{11}(z) & \Pi_{12}(z) \\ \Pi_{21}(z) & \Pi_{22}(z) \end{bmatrix}.$$

Hence, the determinant is given by

$$\det[\Pi(z)] = \Pi_{11}(z)\Pi_{22}(z) - \Pi_{21}(z)\Pi_{12}(z).$$

Now,

$$\Pi_{ij}(z) = \begin{cases} 1 - \sum_{k=1}^{p} \Pi_{ii,k} z^k & \text{if } j = i, \\ -\sum_{k=1}^{p} \Pi_{ij,k} z^k & \text{otherwise.} \end{cases}$$

Hence, for the bivariate case

$$\begin{aligned} \det[\Pi(z)] &= \left(1 - \sum_{k=1}^{p} \Pi_{11,k} z^k\right)\left(1 - \sum_{k=1}^{p} \Pi_{22,k} z^k\right) \\ &\quad - \left(\sum_{k=1}^{p} \Pi_{21,k} z^k\right)\left(\sum_{k=1}^{p} \Pi_{12,k} z^k\right) \\ &= 1 - \sum_{k=1}^{2p} \phi_k z^k \\ &= \prod_{i=1}^{2p} (1 - \lambda_i z). \end{aligned}$$

The parameters ϕ_k are determined directly from $\Pi_{ij,k}$. For instance, $\phi_1 = \Pi_{11,1} + \Pi_{22,1}$, whereas $\phi_2 = \Pi_{11,2} + \Pi_{22,2} + \Pi_{21,1}\Pi_{12,1} - \Pi_{11,1}\Pi_{22,1}$. The third equality above determines the λ_i's from the ϕ_k's. Notice that while ϕ_k is a real number and unique, λ_i is a complex number and typically not unique. That is, we need to use some ordering rule before the λ_i's can be uniquely determined.

CONCLUSION 2: *Let* $\det[\Pi(z)] = \prod_{i=1}^{np}(1 - \lambda_i z)$*, where* $|\lambda_{np}| \geq |\lambda_{np-1}| \geq \ldots \geq |\lambda_1| \geq 0$. *Then* $\Pi(z)$ *is invertible if and only if* $|\lambda_{np}| < 1$.

Let $|\lambda_i|$ denote the modulus of λ_i Suppose, that $\lambda_1 = .5 + .6\imath$, $\lambda_2 = .5 - .6\imath$, where $\imath = \sqrt{-1}$. Then

$$|\lambda_1| = \sqrt{.5^2 + .6^2} = .7810 = |\lambda_2|.$$

An equivalent condition for invertibility is that $\det[\Pi(z)] = 0$ if and only if $|z| > 1$. The z's which imply that the determinant is zero are called roots, and this condition states that all roots must lie outside the unit circle. The λ_i's are eigenvalues of the matrix:

$$\Pi = \begin{bmatrix} \Pi_1 & \Pi_2 & \cdots & \Pi_{p-1} & \Pi_p \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{bmatrix}.$$

This matrix is found when we rewrite the VAR(p) system into a VAR(1) system for $X_t = (x_t, x_{t-1}, \ldots, x_{t-p+1})$, an $np \times 1$ vector. In the AR(1) case, $\Pi = \Pi_1 = \lambda_1$ and the invertibility (stability) condition was found to be $|\Pi_1| < 1$.

The results in Conclusion 1 thus also hold when $np > 1$. That is, if x_t is a stable VAR(p) process, then x_t is weakly stationary and ergodic.

5. Identification and Structural Models

Consider the model

$$B_0 x_t = \gamma + \sum_{j=1}^{p} B_j x_{t-j} + \eta_t, \quad t = 1, 2, \ldots, T, \tag{5.1}$$

where

$$\eta_t \sim \text{iid } N_n(0, \Omega), \tag{5.2}$$

Ω is positive definite, while $x_0, \ldots, x_{1-p}$ are fixed.

If B_0 is invertible, then (5.1) can be written as

$$\begin{aligned} x_t &= B_0^{-1}\gamma + \textstyle\sum_{j=1}^{p} B_0^{-1} B_j x_{t-j} + B_0^{-1}\eta_t \\ &= \mu + \textstyle\sum_{j=1}^{p} \Pi_j x_{t-j} + \varepsilon_t, \end{aligned} \tag{5.3}$$

where

$$\varepsilon_t \sim \text{iid } N_n(0, \Sigma), \tag{5.4}$$

with $\Sigma = B_0^{-1}\Omega(B_0')^{-1}$ being positive definite since Ω is positive definite and B_0 invertible.

Question 4: *Can we identify (uniquely determine) the parameters* $(\gamma, B_0, B_1, \ldots, B_p, \Omega)$ *from* $(\mu, \Pi_1, \ldots, \Pi_p, \Sigma)$*?*

The general answer is, of course, no. This follows directly from the observation that there are n^2 additional parameters in (5.1) relative to (5.3). Accordingly, if the parameters in (5.3) are uniquely determined (from the distribution for x_t), then to achieve identification of the parameters in (5.1) it is necessary (but generally not sufficient) to impose n^2 restrictions (identifying assumptions) on its parameters.

Although uniqueness is often at the heart of what econometricians tend to mean by a structural model, there is no unique definition as to what a structural model is. Let $F(x;\theta)$ be a distribution function for x which depends on a vector of parameters, θ.

DEFINITION 1 (statistics): *A structural model for x is given by a function $F(x;\theta)$ such that θ is uniquely determined from the probability distribution for x.*

QUESTION 5: *Is the VAR model in* (5.3) *a structural model according to this definition?*

The answer is yes. $(\mu, \Pi_1, \dots, \Pi_p, \Sigma)$ is uniquely determined from the first and second moments for x_t. For example, in the case when p is equal to one and the mean of x is zero we have that

$$\begin{aligned}\mu &= 0\\ \Pi_1 &= E[x_t x_{t-1}']E[x_{t-1}x_{t-1}']^{-1}\\ \Sigma &= E[x_t x_t'] + \Pi_1 E[x_{t-1}x_{t-1}']\Pi_1' - E[x_t x_{t-1}']\Pi_1' - \Pi_1 E[x_{t-1}x_t'].\end{aligned}$$

Hence, the VAR parameters are uniquely determined from the population moments of x.

QUESTION 6: *Is the VAR model in* (5.1) *a structural model according to definition 1?*

Since $(\gamma, B_0, B_1, \dots, B_p, \Omega)$ is not uniquely determined from $(\mu, \Pi_1, \dots, \Pi_p, \Sigma)$, the answer must be no. However, once the parameters of (5.1) are uniquely determined, they are indeed structural in this sense.

An alternative notion of what a structure (or structural model) is, comes from David Hendry. My understanding of what he means by a structure in time series econometrics is the following:

DEFINITION 2 ("Hendry"): *A structure is a set of features of the data that remain constant over time, e.g. properties which do not vary across different policy regimes.*

An analogy would be a classroom, where the room is a structure whereas the chairs, tables, students and teachers are not. While this definition has certain appeals from a practical (empirical) point of view, it is of limited theoretical interest since parameters are usually considered constant. In other words, the models in (5.1) and in (5.3) are both structures in Hendry's sense since the parameters are taken to be constant. Moreover, as with Definition 1, there is basically no economics in Hendry's idea of what a structure is.

DEFINITION 3 ("Cowles Commission"): *A structure is a specific set of relationships between (random) variables x and parameters θ, where the latter can be given an economic interpretation.*

The variables x can here include endogenous as well as exogenous variables, while the parameters are taken to be constant (over time or cross sections).

What is important in this definition is that "the parameters" have an economic meaning (that the variables have an economic meaning is implicitly assumed). But this doesn't mean that any transformation of the parameters, e.g. $\phi = f(\theta)$ for a particular function $f(\cdot)$, can be given an economic interpretation. Note also, that this definition does not say that θ satisfies Definition 1. In other words, the structural parameters need not be identified!

DEFINITION 4 ("Sims"): *A structural model is a representation of* (x, θ) *that can be used in decision making, i.e. it generates predictions about the results of different actions.*

This definition indicates that a structural model should be useful for, e.g., policy analysis. In that sense, it is similar to Hendry's idea of a structure. Moreover, Sims's definition suggests that θ is identified, i.e. it satisfies the statistical notion of a structural model. Finally, in order for the results of the actions to be meaningful to an economists, θ must have an economic interpretation. Hence, Sims's definition seems to incorporate all the above notions of what a structural model is. Moreover, it suggests that (5.1) can be a structural model whereas (5.3) cannot.[2]

DEFINITION 5 ("Wold"): *Economic structures (causal relations) are recursive.*

This definition presumes a time series perspective. It states that a variable x_1 can be causal for x_2 if x_1 occurs (is realized) before x_2. However, in practise the sampling frequency of macroeconomic time series is typically (much) lower than the frequency between causal events, thus making use of this definition somewhat doubtful.

The early structural VAR analyses, e.g. Sims (1980), are based on so called *Wold causal chains* with independent innovations (shocks). An economic interpretation is given to the shocks (the actions, e.g. a monetary policy shock) and to the dynamic responses in the endogenous variables (impulse response functions and variance decompositions).

To show that Wold causal chains are exactly identifying, note first that

1. B_0 is lower triangular (recursive) yields $n(n-1)/2$ restrictions.
2. Ω is diagonal (mutually independent innovations) yields $n(n-1)/2$ restrictions.

This gives us a total of $n(n-1)$ identifying restrictions. To exactly identify the parameters of (5.1) we need at least n additional restrictions. These are given by either letting all the diagonal elements of B_0 or of Ω be equal to unity. For impulse responses functions and variance decompositions, these two choices of the n normalizing assumptions are equivalent.

Consider first the case where $\Omega = I_n$. Then $\Sigma = B_0^{-1}(B_0')^{-1}$. Since B_0 is lower triangular, its inverse is also lower triangular. Let P denote the inverse of B_0. With $\Sigma = PP'$ the matrix

[2] This statement assumes that "actions" have an economic meaning.

P is called the Choleski factor of Σ and it can be shown that P is uniquely determined up to an orthogonal transformation N such that N is diagonal with diagonal elements equal to 1 or -1. That is, $P^* = PN$ is also lower triangular and satisfies $\Sigma = P^*P^{*\prime}$. In plain english this means that each structural shock is identified up to its sign!

EXAMPLE: Consider the bivariate case.

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} p_{11} & p_{21} \\ 0 & p_{22} \end{bmatrix} = \begin{bmatrix} p_{11}^2 & p_{11}p_{21} \\ p_{11}p_{21} & p_{22}^2 + p_{21}^2 \end{bmatrix}.$$

Solving for the p_{ij}'s we obtain

$$p_{11} = \sqrt{\sigma_{11}}$$
$$p_{21} = \sigma_{12}/\sqrt{\sigma_{11}}$$
$$p_{22} = \sqrt{\sigma_{22} - (\sigma_{12}^2/\sigma_{11})}.$$

Notice that all p_{ij}'s are real numbers since Σ is assumed to be positive definite. Moreover, we have chosen the orthogonal matrix $N = I_2$.

Consider now the case when we choose to impose the n normalizing assumptions on the diagonal of B_0. In the $n = 2$ case we have that

$$\begin{aligned} \Sigma &= B_0^{-1}\Omega(B_0')^{-1} \\ &= \begin{bmatrix} 1 & 0 \\ -\beta_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \omega_{11} & 0 \\ 0 & \omega_{22} \end{bmatrix} \begin{bmatrix} 1 & -\beta_{21} \\ 0 & 1 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 0 \\ \beta_{21} & 1 \end{bmatrix} \begin{bmatrix} \omega_{11} & 0 \\ 0 & \omega_{22} \end{bmatrix} \begin{bmatrix} 1 & \beta_{21} \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \omega_{11} & \omega_{11}\beta_{21} \\ \omega_{11}\beta_{21} & \omega_{22} + \beta_{21}^2\omega_{11} \end{bmatrix}. \end{aligned}$$

Solving for the structural parameters we obtain

$$\omega_{11} = \sigma_{11},$$
$$\beta_{21} = \sigma_{21}/\sigma_{11},$$
$$\omega_{22} = \sigma_{22} - \sigma_{21}^2/\sigma_{11}.$$

Here we find that $\omega_{ii} > 0$ since Σ is positive definite, while the sign of β_{21} depends on the sign of the covariance between the two residuals in the reduced form VAR.

Alternatively, suppose B_0 and Ω are given by

$$B_0 = \begin{bmatrix} 1 & -\beta_{12} \\ 0 & 1 \end{bmatrix} \qquad \Omega = \begin{bmatrix} \omega_{11} & 0 \\ 0 & \omega_{22} \end{bmatrix}.$$

In this case

$$\begin{aligned} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} &= \begin{bmatrix} 1 & -\beta_{12} \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \omega_{11} & 0 \\ 0 & \omega_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\beta_{12} & 1 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & \beta_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_{11} & 0 \\ 0 & \omega_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \beta_{12} & 1 \end{bmatrix} \\ &= \begin{bmatrix} \omega_{11} + \beta_{12}^2 \omega_{22} & \beta_{12}\omega_{22} \\ \beta_{12}\omega_{22} & \omega_{22} \end{bmatrix}. \end{aligned}$$

Solving for $\beta_{12}, \omega_{11}, \omega_{22}$ we obtain

$$\begin{aligned} \omega_{22} &= \sigma_{22} \\ \beta_{12} &= \sigma_{12}/\sigma_{22} \\ \omega_{11} &= \sigma_{11} - \sigma_{12}^2/\sigma_{22}. \end{aligned}$$

All these choices of B_0, Ω are *observationally equivalent*. The first and the second structural models are equivalent up to a choice of normalization, while the third has very different implications for the behavior of x except when $\sigma_{12} = 0$.

As long as we choose to identify B_0 and Ω from the covariance matrix Σ, all structures will be related to the Choleski decomposition of Σ. For instance, suppose $\Omega = I_n$. Then there exists an infinite number of orthogonal matrices N such that $B_0 = (PN)^{-1}$. In the bivariate case, one such orthogonal matrix is:

$$N = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Since $NN' = I_2$ it follows that $PNN'P' = PP' = \Sigma$. However, the matrix B_0 (for P lower triangular) is now given by

$$B_0 = \frac{1}{\sqrt{2}p_{11}p_{22}} \begin{bmatrix} p_{22} - p_{21} & p_{11} \\ -(p_{22} + p_{21}) & p_{11} \end{bmatrix}.$$

Hence, we no longer have a recursive structure!

EXAMPLE: PIH revisited. Remember that the VAR(1) system for Y_t and C_t can be written as

$$\begin{bmatrix} C_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \mu_Y \\ \mu_Y \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} C_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ w_t \end{bmatrix}, \tag{5.5}$$

where $w_t = u_t + v_t$, while

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_v^2 \end{bmatrix}. \tag{5.6}$$

Similarly, the true structural VAR system is given by

$$\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} C_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \mu_Y \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} C_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix}, \tag{5.7}$$

where

- u_t is a permanent income shock, and
- v_t is a transitory income shock.

QUESTION 7: *Given* (5.5) *and* (5.6), *can we derive* (5.7) *from* $\Sigma = B_0^{-1}\Omega(B_0')^{-1}$ *with* Ω *diagonal and* B_0 *lower triangular with unit diagonal elements?*

Under these conditions we obtain

$$\begin{bmatrix} \sigma_u^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_u^2 \end{bmatrix} = \begin{bmatrix} \omega_{11} & \beta_{21}\omega_{11} \\ \beta_{21}\omega_{11} & \omega_{22} + \beta_{21}^2\omega_{11} \end{bmatrix}.$$

Accordingly,

$$\begin{aligned} \omega_{11} &= \sigma_u^2 \\ \beta_{21} &= 1 \\ \omega_{22} &= \sigma_v^2. \end{aligned}$$

and

$$B_0 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \qquad \Omega = \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}.$$

Premultiplying both sides of (5.5) by B_0 we indeed obtain (5.7).

Now, suppose we change the ordering of the variables while

$$B_0 = \begin{bmatrix} 1 & 0 \\ -\beta_{21} & 1 \end{bmatrix}, \qquad \Omega = \begin{bmatrix} \omega_{11} & 0 \\ 0 & \omega_{22} \end{bmatrix}.$$

Do the resulting "structural shocks" have an economic meaning?

Under the new assumptions we have that

$$\begin{bmatrix} \sigma_u^2 + \sigma_v^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 \end{bmatrix} = \begin{bmatrix} \omega_{11} & \beta_{21}\omega_{11} \\ \beta_{21}\omega_{11} & \omega_{22} + \beta_{21}^2\omega_{11} \end{bmatrix}.$$

Solving these 3 equations for $\beta_{21}, \omega_{11}, \omega_{22}$ we get

$$\begin{aligned} \omega_{11} &= \sigma_u^2 + \sigma_v^2 \\ \beta_{21} &= \sigma_u^2/(\sigma_u^2 + \sigma_v^2) \\ \omega_{22} &= \sigma_u^2 - \sigma_u^4/(\sigma_u^2 + \sigma_v^2). \end{aligned}$$

Notice that $\omega_{ii} > 0$ (as they should be) and that $0 < \beta_{21} < 1$. Moreover, the resulting structural VAR system is now given by

$$\begin{bmatrix} 1 & 0 \\ -\beta_{21} & 1 \end{bmatrix} \begin{bmatrix} Y_t \\ C_t \end{bmatrix} = \begin{bmatrix} \mu_Y \\ (1-\beta_{21})\mu_Y \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 1-\beta_{21} \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ C_{t-1} \end{bmatrix} + \begin{bmatrix} u_t + v_t \\ (1-\beta_{21})u_t - \beta_{21}v_t \end{bmatrix}.$$

Hence, the first "structural shock" is the sum of the permanent and the transitory income shock, while the second is another linear combination of the true structural shocks.

The last case illustrates the Cooley and LeRoy (1985) critique against arbitrary orderings of the variables when the identifying assumptions are based on Wold causal chains. If the identifying assumptions do not rely on a particular economic theory, the resulting "structural shocks" can be pure nonsense shocks. Note, however, that this second example is *not* empirically irrelevant. In fact, models of consumption have a long tradition of using these identifying assumption; see e.g. Davidson, Hendry, Srba, and Yeo (1978). To be fair, the context where it has been used is very different from that of structural VAR's.

If we interpret the second equation in the above structural VAR model as a consumption function we find after a bit of algebra that it can be written as

$$\Delta C_t = (1-\beta_{21})\mu_Y + \beta_{21}\Delta Y_t - \beta_{21}(C_{t-1} - Y_{t-1}) + \psi_t, \qquad (5.8)$$

where $\psi_t = (1-\beta_{21})u_t - \beta_{21}v_t$. We have already noted that $0 < \beta_{21} < 1$ and that income and consumption are CI(1,1) with $(C_t - Y_t)$ being a cointegration relation. The relationship in (5.8) is consistent with a Keynesian consumption function in the empirical modelling tradition of the so called LSE school (Sargan, Hendry, etc.). The cointegration relationship would then have the interpretation of a long run consumption rule (or function). The true values of the parameters suggest that, ceteris paribus, an increase in current income by 1 percent leads to an increase in current consumption by less than 1 percent, while consumption over the long run level in the previous period ($C_{t-1} > Y_{t-1}$) leads to a partial decrease in current consumption.

The assumption which is critical here is that ΔY_t and ψ_t are uncorrelated. In terms of the Wold causal chain this means that income is predetermined, i.e. current income does not depend on current consumption.

To choose between the PIH and the Keynesian consumption function we may turn to examining overidentifying assumptions. In our example, the PIH implies that consumption is a random walk (with drift) and is thus consistent with the Hall (1978) version of this hypothesis. In terms of the VAR in (5.3) this implies 2 restrictions on Π_1. Once we have established that the data is consistent with these restrictions and we choose to use these restrictions in our analysis, the consumption function in (5.8) is no longer an interesting competing theory.

To sum up, we have shown that two sets of identifying assumptions can yield results which makes sense to an economist. The data will not help us choose between these two structures and we can always find an economists who will argue in favor of one of these theories over the other. Still, when we attempt to identify a structural model, economic theory is, in my opinion, the best guide available to us. If competing theories provide restrictions on the parameters of the reduced form VAR, these may help us choose which theory is consistent with the data.

6. Impulse Response Functions and Variance Decompositions

The notion that a set of impulses and a propagation mechanism are useful tools when analysing an economy goes back to Frisch (1933) and Slutzky (1937).

Impulse response analysis addresses the question:

Question 8: *How does x react (over time) to a change in one of the shocks?*

Suppose that our VAR(p) model is stable so that x_t is weakly stationary. The resulting VMA representation of the VAR is then

$$\begin{aligned} x_t &= \Pi(L)^{-1}(\mu + \varepsilon_t) \\ &= \Pi(1)^{-1}\mu + \Pi(L)^{-1}\varepsilon_t \\ &= \delta + C(L)\varepsilon_t, \end{aligned} \tag{6.1}$$

where

$$C(z) = I_n + \sum_{i=1}^{\infty} C_i z^i.$$

The structural shocks, η_t, are related to the VAR innovations, ε_t, according to

$$\eta_t = B_0 \varepsilon_t. \tag{6.2}$$

Hence, we can rewrite the VMA representation as

$$\begin{aligned} x_t &= \delta + C(L) B_0^{-1} B_0 \varepsilon_t \\ &= \delta + R(L) \eta_t, \end{aligned} \tag{6.3}$$

where

$$R(z) = \sum_{i=0}^{\infty} R_i z^i, \tag{6.4}$$

with

$$R_i = \begin{cases} B_0^{-1} & \text{if } i = 0, \\ C_i B_0^{-1} & \text{otherwise.} \end{cases}$$

Example: Let

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} e_1 & e_2 \end{bmatrix} \qquad \Omega^{1/2} = \begin{bmatrix} \sqrt{\omega_{11}} & 0 \\ 0 & \sqrt{\omega_{22}} \end{bmatrix}.$$

Consider the following experiment

$$\eta_t = \begin{cases} \Omega^{1/2} e_j & \text{if } t = t^*, \\ 0 & \text{if } t > t^*. \end{cases} \tag{6.5}$$

Notice that $\Omega^{1/2} e_j = e_j \sqrt{\omega_{jj}}$ measures a one standard deviation increase in the j:th structural shock (while the other shock is zero).

Question 9: *What are the responses in $x_{t^*}, x_{t^*+1}, \ldots$ (relative to $\eta_{t^*+j} = 0$ for all $j \geq 0$) from such a shock?*

Using the structural VMA represenation in (6.3) we find that the impulse response function is:

$$\begin{aligned} \text{resp}(x_{t^*}|\eta_{t^*} = e_j\sqrt{\omega_{jj}}) &= R_0 e_j\sqrt{\omega_{jj}}, \\ \text{resp}(x_{t^*+1}|\eta_{t^*} = e_j\sqrt{\omega_{jj}}) &= R_1 e_j\sqrt{\omega_{jj}}, \\ &\vdots \\ \text{resp}(x_{t^*+i}|\eta_{t^*} = e_j\sqrt{\omega_{jj}}) &= R_i e_j\sqrt{\omega_{jj}}. \end{aligned} \tag{6.6}$$

The vector $R_i e_j$ is the j:th column of R_i. In the limit we have that

$$\lim_{i\to\infty} \text{resp}(x_{t^*+i}|\eta_{t^*} = e_j\sqrt{\omega_{jj}}) = 0,$$

for all $j \in \{1,2\}$ since x_t is ergodic. In other words, for weakly stationary VAR(p) models, the response in x from any shock vanishes in the long run. Hence, we can say that x_t is mean reverting.

QUESTION 10: *Is the experiment in equation* (6.5) *relevant from a statistical point of view?*

Shocks at t^* and $t * +i$ are independent and, moreover, different structural shocks are independent. Thus, the experiment is consistent with the assumptions about η_t.

QUESTION 11: *Is the experiment in equation* (6.5) *relevant from an economics point of view?*

If the shocks can be given a credible economic interpretation, the answer would be yes. However, there is no guarantee that the responses in x will be fully consistent with the economic model which the identification of the shocks is based on. This can occur when the economic model implies overidentifying restrictions on the parameters of the VAR model and these restrictions are not consistent with the data.

An important assumption in structural VAR modelling is that the structural shocks are linear combinations of the residuals in the reduced form VAR model (the so called Wold innovations). To illustrate the relevance of this assumption, consider the following univariate process

$$\eta_t = \alpha\eta_{t-1} + \psi_t - \alpha^{-1}\psi_{t-1}, \tag{6.7}$$

where $|\alpha| < 1$ and $\psi_t \sim \text{iid} \quad N(0, \sigma_\psi^2)$. This looks like an ordinary ARMA(1,1) model, where the AR polynomial is invertible while the MA polynomial is not. Also, the MA coefficient is equal to the inverse of the AR coefficient. The polynomial $(1 - \alpha^{-1}z)/(1 - \alpha z)$ is called a Blaschke factor.

Since the AR polynomial is invertible, it follows that η_t is weakly stationary and that its mean is zero. Moreover, the variance is given by

$$\begin{aligned} E[\eta_t^2] &= E[\alpha^2\eta_{t-1}^2 + \psi_t^2 + \alpha^{-2}\eta_{t-1}^2 + 2\alpha\eta_{t-1}\psi_t - 2\eta_{t-1}\psi_{t-1} - 2\alpha^{-1}\psi_t\psi_{t-1}] \\ &= \alpha^2 E[\eta_{t-1}^2] + \sigma_\psi^2 + \alpha^{-2}\sigma_\psi^2 - 2\sigma_\psi^2. \end{aligned}$$

With $\sigma_\eta^2 = E[\eta_t^2]$ we then obtain

$$\sigma_\eta^2 = \frac{(\alpha^{-2}-1)\sigma_\psi^2}{1-\alpha^2}. \tag{6.8}$$

Similarly, the first autocovariance is given by

$$\begin{aligned} E[\eta_t\eta_{t-1}] &= E[\alpha\eta_{t-1}^2 + \psi_t\eta_{t-1} - \alpha^{-1}\psi_{t-1}\eta_{t-1}] \\ &= \alpha\sigma_\eta^2 - \alpha^{-1}\sigma_\psi^2 \\ &= [\alpha(\alpha^{-2}-1)\sigma_\psi^2 - \alpha^{-1}(1-\alpha^2)\sigma_\psi^2]/(1-\alpha^2) \\ &= 0. \end{aligned}$$

Finally, it can be shown that for all $h \geq 2$

$$E[\eta_t\eta_{t-h}] = \alpha E[\eta_{t-1}\eta_{t-h}] = 0.$$

Accordingly, the parameters α, σ_ψ^2 cannot be uniquely determined from the distribution for η since this random variable is *not* serially correlated. Still, for any pair (α, σ_ψ^2) consistent with the population variance of η, there is a dynamic reaction in η from a shock to ψ. For instance, consider the experiment

$$\psi_t = \begin{cases} \sigma_\psi & \text{if } t = t^*, \\ 0 & \text{if } t > t^*. \end{cases}$$

The response in η_{t^*} is then given by

$$\text{resp}(\eta_{t^*}|\psi_{t^*} = \sigma_\psi) = \sigma_\psi,$$

while for $i \geq 1$ we obtain

$$\text{resp}(\eta_{t^*+i}|\psi_{t^*} = \sigma_\psi) = \alpha^{i-1}(\alpha - \alpha^{-1})\sigma_\psi \neq 0.$$

The impulse response function for x from an experiment where η_t is equal to σ_η at $t = t^*$ and 0 for $t > t^*$ is very different from the impulse response function for x when (for some $\alpha \neq 0$) we consider an experiment based on ψ. This means that the impulse responses are not uniquely determined unless we are willing to either choose a particular value for α, or assume that the structural shocks are linear combinations of the Wold innovations.

QUESTION 12: *Should we worry about Blaschke factors?*

According to Lippi and Reichlin (1993), modern macroeconomic models which are linearized into dynamic systems tend to include noninvertible MA components. While this is certainly a problem from the point of view of estimating a multivariate ARMA model, we should keep in mind that noninvertibility of an MA term does not mean that there exists an AR factor whose coefficient is the inverse of the MA coefficient in question. Still, it emphasizes the point made earlier that sound structural VAR analysis should rest on a firm theoretical basis.

A variance decomposition, or innovation accounting, measures the share of the forecast error variance which is accounted for by a particular shock. Hence, variance decompositions address the question:

QUESTION 13: *How important is a particular shock (relative to all the other shocks) for explaining the fluctuations in x?*

To construct the forecast error variance, from (6.3) we have for all $h \geq 1$ that

$$x_{t+h} = \delta + \sum_{k=0}^{\infty} R_k \eta_{t+h-k}.$$

The optimal prediction of x_{t+h} given all information available at period t is the conditional expectation.[3] Hence,

$$E[x_{t+h}|x_t, x_{t-1}, \dots] = \delta + \sum_{k=h}^{\infty} R_k \eta_{t+h-k}. \tag{6.9}$$

The forecast error is therefore

$$\varphi_{t+h|t} = \sum_{k=0}^{h-1} R_k \eta_{t+h-k}, \tag{6.10}$$

a VMA process of order $(h-1)$. Consequently, the forecast error covariance matrix for x is

$$V_h = E[\varphi_{t+h|t} \varphi'_{t+h|t}] = \sum_{k=0}^{h-1} R_k \Omega R'_k. \tag{6.11}$$

Notice that this covariance matrix is invariant to the choice of identification, i.e. $V_h = \sum_{k=0}^{h-1} C_k \Sigma C'_k$.

For a particular variable $i \in \{1, \dots, n\}$ the h steps ahead forecast error variance is given by the i:th diagonal element of V_h. With e_i being the i:th column of I_n this variance can be written as

$$v_{i,h} = e'_i V_h e_i = \sum_{k=0}^{h-1} e'_i R_k \Omega R'_k e_i. \tag{6.12}$$

[3] By optimal we mean that it has the smallest mean square error among all unbiased predictors.

Let $R_{ij,k}$ denote the (i,j):th element of R_k. It then follows that

$$\begin{aligned} e_i' R_k \Omega R_k' e_i &= \begin{bmatrix} R_{i1,k} & \cdots & R_{in,k} \end{bmatrix} \begin{bmatrix} \omega_{11} & & 0 \\ & \ddots & \\ 0 & & \omega_{nn} \end{bmatrix} \begin{bmatrix} R_{i1,k} \\ \vdots \\ R_{in,k} \end{bmatrix} \\ &= \textstyle\sum_{j=1}^{n} R_{ij,k}^2 \omega_{jj}. \end{aligned}$$

Hence, equation (6.12) can be rewritten as

$$v_{i,h} = \sum_{k=0}^{h-1} \sum_{j=1}^{n} R_{ij,k}^2 \omega_{jj}. \tag{6.13}$$

Multiplying both sides by $1/v_{i,h}$ we thus obtain

$$\begin{aligned} 1 &= \textstyle\sum_{k=0}^{h-1} \sum_{j=1}^{n} R_{ij,k}^2 \omega_{jj} / v_{i,h} \\ &= \textstyle\sum_{j=1}^{n} \left(\sum_{k=0}^{h-1} R_{ij,k}^2 \omega_{jj} / v_{i,h} \right) \\ &= \textstyle\sum_{j=1}^{n} w_{ij,h}. \end{aligned} \tag{6.14}$$

The parameter $w_{ij,h}$ takes values in the unit interval and measures the fraction of the h steps ahead forecast error variance for variable i which is accounted for by shock j.

EXAMPLE: Suppose $n = 2$ with Ω diagonal and B_0 lower triangular with unit diagonal elements. With $R_0 = B_0^{-1}$ the 1 step ahead forecast error variance is

$$\begin{aligned} V_1 &= R_0 \Omega R_0' \\ &= \begin{bmatrix} 1 & 0 \\ \beta_{21} & 1 \end{bmatrix} \begin{bmatrix} \omega_{11} & 0 \\ 0 & \omega_{22} \end{bmatrix} \begin{bmatrix} 1 & \beta_{21} \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \omega_{11} & \beta_{21}\omega_{11} \\ \beta_{21}\omega_{11} & \omega_{22} + \beta_{21}^2 \omega_{11} \end{bmatrix}. \end{aligned}$$

The 1 step ahead forecast error variances are thus

$$\begin{aligned} &\text{variable 1:} \quad \omega_{11} = \sigma_{11} \\ &\text{variable 2:} \quad \omega_{22} + \beta_{21}^2 \omega_{11} = \sigma_{22}. \end{aligned} \tag{6.15}$$

Hence, the 1 step ahead forecast error variance for each variable is invariant with respect to the choice of identification. The variance decompositions, however, are not invariant. For 1 step ahead forecast errors, the share of the total variance for the first variable which is explained by the first (second) shock is unity (zero). For the second variable, the share due to the first shock is $\beta_{21}^2\omega_{11}/\sigma_{22}$ while the share due to the second shock is ω_{22}/σ_{22}. if we instead assume that the B_0 matrix is upper triangular (with unit diagonal elements), for the second variable variable we find that the share of the 1 step ahead forecast error variance due to the second (first) shock is unity (zero). Moreover, for the first variable both shocks may now account for the error variance.

7. Cointegration and Common Trends

It has long been recognized that many macroeconomic time series are trending and thus not well described as weakly stationary. To transform the data into appropriate stationary series various detrending techniques have been considered. Common among these are the linear trend model and the first difference model.

Example: PIH revisited. From equation (2.5) we find that both variables have a linear trend when $\mu_Y \neq 0$. However, removal of this trend does not make the variables stationary since they also include a stochastic trend, $\sum_{i=1}^{t} u_i$. Hence, the linear trend model is not appropriate for rendering nonstationary variables stationary in this model.

By taking first differences, we know from equations (2.8) that these transformations make income and consumption stationary. Still, there does *not* exist a VAR model with finite lag order for the first differences. In fact, if we subtract C_{t-1} from the consumption equation of (3.4) we have that

$$\Delta C_t = \mu_Y + u_t, \tag{7.1}$$

while subtracting Y_{t-1} from the income equation of produces

$$\Delta Y_t = \mu_Y + C_{t-1} - Y_{t-1} + w_t. \tag{7.2}$$

In vector form we thus have that

$$\begin{bmatrix} \Delta C_t \\ \Delta Y_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} C_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ w_t \end{bmatrix}. \tag{7.3}$$

Hence, once the left hand side variables have been transformed into first differences, the levels of lagged income and consumption still appears on the right hand side of the model. Moreover, the matrix of coefficients on the lagged levels has reduced rank (lower rank than dimension). Specifically,

$$\begin{aligned}\Pi &= \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \\ &= \alpha\beta'\end{aligned}$$

where the vectors α, β have rank 1. Finally, the product Πx_{t-1} yields

$$\begin{aligned}\Pi x_{t-1} &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} C_{t-1} \\ Y_{t-1} \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} (C_{t-1} - Y_{t-1}) \, .\end{aligned}$$

Hence, the product produces a vector of weights α on the cointegration relation between consumption and income. VAR models in first differences do *not* take this relation into account and are therefore misspecified.

An alternative way of deriving an appropriate transformation of the variables in the VAR model is to calculate the number of unit roots. If this number is lower than the dimension of the VAR, then a VAR model in first differences will be overdifferenced. In the PIH case we have that

$$\Pi(z) = \begin{bmatrix} 1-z & 0 \\ -z & 1 \end{bmatrix}. \tag{7.4}$$

This matrix polynomial has exactly 1 unit root and no roots inside the unit circle. Hence, the number of variables (2) exceeds the number of unit roots.

To generalize these observations, consider again the VAR(1) model for x_t in (3.5). Subtracting x_{t-1} from both sides we obtain

$$\begin{aligned}\Delta x_t &= \mu + \Pi_1 x_{t-1} - x_{t-1} + \varepsilon_t \\ &= \mu + (\Pi_1 - I_n)\, x_{t-1} + \varepsilon_t \\ &= \mu + \Pi x_{t-1} + \varepsilon_t,\end{aligned} \tag{7.5}$$

where $\Pi = -(I_n + \Pi_1) = -\Pi(1)$ (in terms of the polynomial $\Pi(z) = I_n - \Pi_1 z$).

To ensure that x_t is I(d) with the integer $d \geq 0$, we shall assume that $\det[\Pi(z)] = 0$ if and only if $|z| > 1$ or $z = 1$. In other words, there are neither explosive ($|z| < 1$) nor seasonal ($z = -1$) roots.

1. If rank[Π] $= n$, then there are no unit roots so x_t is weakly stationary.
2. If rank[Π] $= 0$, then $\Pi(z) = (I_n - I_n z)$. Accordingly, $\Delta x_t = \mu + \varepsilon_t$ and x_t is I(1) but not cointegrated.
3. If rank[Π] $= r$ with $r \in \{1, \ldots, n-1\}$ and the number of unit roots is equal to $n - r$, then x_t is CI(1,1) with $\Pi = \alpha\beta'$ and $\beta' x_t$ being I(0).

While the first two cases are not too difficult to understand, the third case is far from obvious. Specifically, what is the importance of the condition that " ... the number of unit roots is equal to $n - r$ ... "?

EXAMPLE: Consider a bivariate VAR(1) model where

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}. \tag{7.6}$$

For this model we have that

$$\Pi(z) = \begin{bmatrix} 1 - 2z & z \\ -z & 1 \end{bmatrix}.$$

Accordingly, $\det[\Pi(z)] = 1 - 2z + z^2 = (1 - z)^2$. Hence, there are 2 unit roots. At the same time,

$$\Pi = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix},$$

has rank 1. Hence, the number of unit roots is greater than $n - r = 1$. In this case, x_t is still integrated but *not* I(1).

If we subtract x_{t-1} from both sides of equation (7.6) we get

$$\begin{aligned}
\begin{bmatrix} \Delta x_{1,t} \\ \Delta x_{2,t} \end{bmatrix} &= \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \left(x_{1,t-1} - x_{2,t-1} \right) + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}.
\end{aligned} \tag{7.7}$$

From this equation it can be seen that $(x_{1,t} - x_{2,t})$ is integrated of an order less than $x_{1,t}$ and $x_{2,t}$. Subtracting $\Delta x_{2,t}$ from $\Delta x_{1,t}$ we obtain

$$\begin{aligned}
\Delta x_{1,t} - \Delta x_{2,t} &= \Delta \left(x_{1,t} - x_{2,t} \right) \\
&= \varepsilon_{1,t} - \varepsilon_{2,t}.
\end{aligned} \tag{7.8}$$

In other words, $(x_{1,t} - x_{2,t})$ is I(1) and we must therefore have that x_t is I(2).

Hence, this example illustrates that the condition " ... the number of unit roots is equal to $(n - r)$ ... " rules out the cases when x_t is I(d) with $d \geq 2$.

In the PIH case, the number of unit roots is exactly equal to $(n - r)$ and thus satisfies the conditions in case (3) above.

Returning to the VAR(1) model in (7.5), we can express the matrix Π as

$$\Pi = \alpha\beta', \tag{7.9}$$

where α, β are $n \times r$ matrices with full column rank. The error correction representation of the model can now be expressed as

$$\Delta x_t = \mu + \alpha\beta' x_{t-1} + \varepsilon_t. \tag{7.10}$$

When Π has reduced rank and the number of unit roots equals the rank reduction $(n - r)$, then x_t is CI(1,1) with $\beta' x_t$ being the r cointegration relations.[4]

Note that the parameters (α, β) are *not* uniquely determined. For any $r \times r$ nonsingular matrix ξ we have that $\beta^{*\prime} x_t = \xi\beta' x_t$ is also I(0). With $\alpha^* = \alpha\xi^{-1}$ it follows that $\Pi = \alpha^*\beta^{*\prime}$. In other words, the cointegration space, $\text{sp}(\beta)$, is uniquely determined from Π, but the basis is not.

In the PIH example, we have that $(C_t - Y_t)$ is I(0), but so is $a(C_t - Y_t)$ for any finite $a \neq 0$.

[4] In the VAR(1) model, the coindition that the number of unit roots is equal to the rank reduction is equivalent to rank$[\alpha'\beta] = r$. In the I(2) example, for instance, we have that $\alpha'\beta = 0$; for parametric conditions for x_t to be I(1) in the VAR(p) model, see Johansen (1991).

QUESTION 14: *How do we invert VAR models with unit roots?*

EXAMPLE: For the PIH, the VAR model with a cointegration constraint is given in (7.3). With $(C_t - Y_t) = -v_t$ the income growth equation can be written

$$\begin{aligned} \Delta Y_t &= \mu_Y + w_t - v_{t-1} \\ &= \mu_Y + w_t + u_{t-1} - (u_{t-1} + v_{t-1}) \\ &= \mu_Y + w_t + u_{t-1} - w_{t-1}. \end{aligned}$$

The MA representation for consumption and income growth is then

$$\begin{bmatrix} \Delta C_t \\ \Delta Y_t \end{bmatrix} = \begin{bmatrix} \mu_Y \\ \mu_Y \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_t \\ w_t \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} u_{t-1} \\ w_{t-1} \end{bmatrix}, \tag{7.11}$$

or

$$\Delta x_t = \delta + (I_2 + C_1 L)\, \varepsilon_t. \tag{7.12}$$

In this case, the inverted error correction model is an MA(1) process for the first differences. As we shall see below, the MA representation for Δx_t is usually of infinite order.

Notice also that $C(z) = I_2 + C_1 z$ is not invertible. Specifically,

$$\det[C(z)] = 1 - z.$$

This is, of course, just the other side of the coin of the fact that there does not exist a finite order VAR model for the first differences.

The result that $C(z)$ has a unit root, means that $C = C(1)$ has reduced rank. In particular,

$$C = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}. \tag{7.13}$$

Notice that C is orthogonal to α and β. Specifically, $\beta' C = 0$ and $C\alpha = 0$. Moreover,

$$\begin{aligned} C(z) - C &= \begin{bmatrix} 1 & 0 \\ z & 1 - z \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ z - 1 & 1 - z \end{bmatrix} \\ &= (1 - z) \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix}. \end{aligned} \tag{7.14}$$

Hence, although $C(z)$ itself does not have a common factor, $(1-z)$, its deviation from C does. We can therefore express the $C(z)$ polynomial as

$$C(z) = C + (1-z)C^*. \tag{7.15}$$

Substituting for $C(z)$ in equation (7.11) we have that

$$\begin{aligned} x_t &= x_{t-1} + \delta + C\varepsilon_t + C^*(\varepsilon_t - \varepsilon_{t-1}) \\ &= [x_{t-2} + \delta + C\varepsilon_{t-1} + C^*(\varepsilon_{t-1} - \varepsilon_{t-2})] + \delta + C\varepsilon_t + C^*(\varepsilon_t - \varepsilon_{t-1}) \\ &= x_{t-2} + \delta 2 + C(\varepsilon_t + \varepsilon_{t-1}) + C^*(\varepsilon_t - \varepsilon_{t-2}) \\ &= x_0 - C^*\varepsilon_0 + \delta t + C\textstyle\sum_{i=1}^{t}\varepsilon_i + C^*\varepsilon_t. \end{aligned} \tag{7.16}$$

We have thus found that the "conditional" MA representation for consumption and income contains (i) an I(1) component $(\delta t + C\sum_{i=1}^{t}\varepsilon_i)$; (ii) an I(0) component $(C^*\varepsilon_t)$; and (iii) initial values $(x_0 - C^*\varepsilon_0)$. The fact that the MA representation includes the third component is reason why I call it a conditional representation.

The I(1) component of the conditional MA representation can also be expressed as

$$\begin{aligned} x_t^{\mathrm{p}} &= \begin{bmatrix} \mu_Y \\ \mu_Y \end{bmatrix} t + \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \textstyle\sum_{i=1}^{t} \begin{bmatrix} u_i \\ w_i \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \left(\mu_Y t + \textstyle\sum_{i=1}^{t} u_i\right) \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_Y \\ \mu_Y \end{bmatrix} t + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \textstyle\sum_{i=1}^{t} \begin{bmatrix} u_i \\ w_i \end{bmatrix} \\ &= \beta_\perp \alpha_\perp' \mu t + \beta_\perp \alpha_\perp' \textstyle\sum_{i=1}^{t} \varepsilon_i. \end{aligned} \tag{7.17}$$

Here, $\alpha_\perp'\alpha = 0$ and $\beta_\perp'\beta = 0$. From equation (7.17) it can be seen that income and consumption have 1 common trend. This trend can be represented by $\alpha_\perp'(\mu t + \sum_{i=1}^{t}\varepsilon_i)$. Moreover, we find that $\beta' x_t^{\mathrm{p}} = 0$ since $\beta' x_t$ is I(0) and cannot include the I(1) component in x_t. Hence, the cointegration vector acts as a detrending model.

To generalize these results to the VAR(1) model, note first that equation (7.10) can be rewritten as

$$x_t = \mu + \left(I_n + \alpha\beta'\right) x_{t-1} + \varepsilon_t. \tag{7.18}$$

Premultiplying this system by β' yields

$$\begin{aligned} \beta' x_t &= \beta'\mu + \beta'\left(I_n + \alpha\beta'\right) x_{t-1} + \beta'\varepsilon_t \\ &= \beta'\mu + \left(I_r + \beta'\alpha\right)\beta' x_{t-1} + \beta'\varepsilon_t, \end{aligned} \tag{7.19}$$

a VAR(1) model for the cointegration relations. Solving this model recursively we obtain

$$\begin{aligned} \beta' x_t &= \textstyle\sum_{i=0}^{\infty} \left(I_r + \beta'\alpha\right)^i \beta'\mu + \sum_{i=0}^{\infty} \left(I_r + \beta'\alpha\right)^i \beta'\varepsilon_{t-i} \\ &= -\left(\beta'\alpha\right)^{-1}\beta'\mu + \textstyle\sum_{i=0}^{\infty} \left(I_r + \beta'\alpha\right)^i \beta'\varepsilon_{t-i}, \end{aligned} \tag{7.20}$$

an MA(∞) representation for the r cointegration relations. Notice that if $\operatorname{rank}(\beta'\alpha) < r$, then the polynomial $(I_r - (I_r + \beta'\alpha)z)$ contains a unit root. This is ruled out by the assumption that the number of unit roots equals $(n - r)$.[5]

Substituting equation (7.20) for $\beta' x_{t-1}$ in (7.10) we have found the MA representation for Δx_t. Specifically, it is given by

$$\begin{aligned} \Delta x_t &= \left(I_n - \alpha(\beta'\alpha)^{-1}\beta'\right)\mu + \varepsilon_t + \textstyle\sum_{i=1}^{\infty} \alpha(I_r + \beta'\alpha)^{(i-1)}\beta'\varepsilon_{t-i} \\ &= \delta + \textstyle\sum_{i=0}^{\infty} C_i \varepsilon_{t-i}, \end{aligned} \tag{7.21}$$

where $C_0 = I_n$.

To show that $C(z)$ has unit roots, note first that

$$\begin{aligned} C &= I_n + \textstyle\sum_{i=1}^{\infty} \alpha(I_r + \beta'\alpha)^{(i-1)}\beta' \\ &= I_n - \alpha(\beta'\alpha)^{-1}\beta'. \end{aligned} \tag{7.22}$$

Second, for any $\alpha_\perp, \beta_\perp \in \mathbb{R}^{n\times(n-r)}$ of rank $(n - r)$ such that $\alpha_\perp'\alpha = 0$ and $\beta_\perp'\beta = 0$ it holds that

$$I_n = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' + \alpha(\beta'\alpha)^{-1}\beta'. \tag{7.23}$$

This can be verified through premultiplication of both sides by $\alpha_\perp'$ or β' or through postmultiplication by α or $\beta_\perp$. The choice of basis for $\alpha_\perp$ and $\beta_\perp$ is irrelevant since $\alpha_\perp^* = \alpha_\perp\zeta$, $\beta_\perp^* = \beta_\perp\xi$ (where ζ and ξ are nonsingular $(n - r) \times (n - r)$ matrices) satisfy

$$\beta_\perp^*(\alpha_\perp^{*\prime}\beta_\perp^*)^{-1}\alpha_\perp^{*\prime} = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'.$$

Using equation (7.23) we thus have that

$$C = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp', \tag{7.24}$$

[5] When $\beta'\alpha$ has full rank r (explosive roots have already been ruled out by assumption), it follows that $\sum_{i=0}^{\infty}(I_r + \beta'\alpha)^i = (I_r - (I_r + \beta'\alpha))^{-1}$, i.e. the matrix $(I_r + \beta'\alpha)$ has all eigenvalues inside the unit circle so that the sum of the exponents from zero to s converges to a finite matrix as s becomes very large.

in the VAR(1) model.[6] Accordingly, we find that

$$\text{rank}[C] = n - r, \tag{7.25}$$

and $C(z)$ thus has r unit units.

To show that $C(z)$ corrected for C has a common unit root, we use brute force. That is

$$\begin{aligned}
C(z) - C &= \textstyle\sum_{i=1}^{\infty} C_i z^i - \sum_{i=1}^{\infty} C_i \\
&= -(1-z)\left[\textstyle\sum_{i=1}^{\infty} C_i\right] - \left[\textstyle\sum_{i=1}^{\infty} C_i\right] z + \textstyle\sum_{i=1}^{\infty} C_i z^i \\
&= -(1-z)\left[\textstyle\sum_{i=1}^{\infty} C_i\right] - \left[\textstyle\sum_{i=1}^{\infty} C_i\right] z + C_1 z + \textstyle\sum_{i=2}^{\infty} C_i z^i \\
&= -(1-z)\left[\textstyle\sum_{i=1}^{\infty} C_i\right] - \left[\textstyle\sum_{i=2}^{\infty} C_i\right] z + \textstyle\sum_{i=2}^{\infty} C_i z^i \\
&= -(1-z)\left[\textstyle\sum_{i=1}^{\infty} C_i\right] - (1-z)\left[\textstyle\sum_{i=2}^{\infty} C_i\right] z - \left[\textstyle\sum_{i=2}^{\infty} C_i\right] z^2 \\
&\quad + \textstyle\sum_{i=2}^{\infty} C_i z^i \\
&= -(1-z)\left[\textstyle\sum_{i=1}^{\infty} C_i\right] - (1-z)\left[\textstyle\sum_{i=2}^{\infty} C_i\right] z - \left[\textstyle\sum_{i=2}^{\infty} C_i\right] z^2 \\
&\quad + C_2 z^2 + \textstyle\sum_{i=3}^{\infty} C_i z^i \\
&= -(1-z)\left[\textstyle\sum_{i=1}^{\infty} C_i\right] - (1-z)\left[\textstyle\sum_{i=2}^{\infty} C_i\right] z - \left[\textstyle\sum_{i=3}^{\infty} C_i\right] z^2 \\
&\quad + \textstyle\sum_{i=3}^{\infty} C_i z^i \\
&= -(1-z)\left[\textstyle\sum_{i=1}^{\infty} C_i\right] - (1-z)\left[\textstyle\sum_{i=2}^{\infty} C_i\right] z - (1-z)\left[\textstyle\sum_{i=3}^{\infty} C_i\right] z^2 \\
&\quad - \left[\textstyle\sum_{i=3}^{\infty} C_i\right] z^3 + \textstyle\sum_{i=3}^{\infty} C_i z^i \\
&= (1-z)\textstyle\sum_{j=0}^{k}\left[-\textstyle\sum_{i=j+1}^{\infty} C_i\right] z^j - \left[\textstyle\sum_{i=k+1}^{\infty} C_i\right] z^k + \textstyle\sum_{i=k+1}^{\infty} C_i z^i .
\end{aligned} \tag{7.26}$$

The last two terms on the right hand side of the last equality vanish as k becomes very large, while the first term converges when the C_i matrices satisfy a summability condition. In that case, we obtain

$$C(z) - C = (1-z)\sum_{j=0}^{\infty} C_j^* z^j, \qquad C_j^* = -\sum_{i=j+1}^{\infty} C_i, \qquad j = 0, 1, \dots . \tag{7.27}$$

The summability condition we require C_i to satisfy is such that the C_j^* matrices are absolutely summable. That is,

$$\begin{aligned}
\textstyle\sum_{j=0}^{\infty} \left| C_j^* \right| &= \textstyle\sum_{j=0}^{\infty} \left| \textstyle\sum_{i=j+1}^{\infty} C_i \right| \\
&= \textstyle\sum_{i=1}^{\infty} i \, |C_i| < \infty .
\end{aligned}$$

Hence, the C_i matrices must be 1-summable. For finite order VAR models, this condition will always be satisfied since its MA representation has exponentially decreasing (in an absolute sense) parameters.

[6] Notice that in the PIH case, $\alpha_\perp' \beta_\perp = 1$.

In our VAR(1) model, the C_j^* matrices are for all $j \geq 0$

$$
\begin{aligned}
C_j^* &= -\sum_{j=i+1}^{\infty} \alpha(I_r + \beta'\alpha)^{(i-1)}\beta' \\
&= -\alpha\left(\sum_{i=j}^{\infty} [I_r + \beta'\alpha]^i\right)\beta' \\
&= -\alpha(I_r + \beta'\alpha)^j \left(\sum_{i=0}^{\infty} [I_r + \beta'\alpha]^i\right)\beta' \\
&= \alpha(I_r + \beta'\alpha)^j (\beta'\alpha)^{-1}\beta'.
\end{aligned}
\tag{7.28}
$$

It is now straightforward to show that these matrices indeed are absolutely summable[7] and thus that the $C(z)$ matrix polynomial can be expressed as

$$
C(z) = C + (1 - z)C^*(z). \tag{7.29}
$$

Substituting for $C(z)$ in equation (7.21) we get

$$
\begin{aligned}
x_t &= x_{t-1} + C\mu + C\varepsilon_t + \sum_{j=0}^{\infty} C_j^*(\varepsilon_{t-j} - \varepsilon_{t-j-1}) \\
&= \tilde{x}_0 + C\mu t + C\sum_{i=1}^{t} \varepsilon_i + \sum_{j=0}^{\infty} C_j^* \varepsilon_{t-j}.
\end{aligned}
\tag{7.30}
$$

Again we find that x_t includes (i) an I(1) component; (ii) an I(0) component; and (iii) initial values, denoted by $\tilde{x}_0$.

The I(1) component is of particular interest in the so called common trends model; see King et al. (1991). Specifically, while this component is made up of n linear combinations of the accumulated Wold innovations, only $(n - r)$ of these combinations are linearly independent. In other words, there are fewer trends than variables. From equation (7.30) we find that

$$
C(\mu t + \sum_{i=1}^{t} \varepsilon_i) = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\left(\alpha_\perp'\mu t + \sum_{i=1}^{t} \alpha_\perp'\varepsilon_i\right). \tag{7.31}
$$

Hence, the reduced form linearly independent $(n - r)$ common trends are given by $(\alpha_\perp'\mu t + \sum_{i=1}^{t} \alpha_\perp'\varepsilon_i)$, while the coefficients on these trends are $\beta_\perp(\alpha_\perp'\beta_\perp)^{-1}$.

Structural common trends models were first suggested by Blanchard and Quah (1989), King et al. (1991), and Shapiro and Watson (1988). The basic idea is to make identifying assumptions about the long run responses in the endogenous variables with respect to the structural shocks. The observation that the number of linearly independent common trends in the reduced form conditional MA representation is smaller than the number of endogenous variables is a central ingredient. This suggests that structural shocks can be decomposed into (i) shocks with permanent effects on x (trend shocks); and (ii) shocks which only have temporary (transitory) effects on x. Moreover, since the I(0) component and the change in the I(1) component in (7.30) are correlated, the structural trend shocks

[7] This follows from the fact that $(I_r + \beta'\alpha)$ has all eigenvalues inside the unit circle.

typically lead to cyclical fluctuations around the trends as well changes in the trends. When x_t contains macroeconomic variables, we can think about this as shocks to growth also having an influence on business cycle fluctuations.

The common trends approach is based on identifying B_0 and Ω using more reduced form parameters than just Σ. In particular, the restrictions implied by cointegration are used for identification through the matrix C.

EXAMPLE: Consider again the PIH. To exactly identify the parameters of a structural VAR model such that it has a common trends interpretation we need to impose 4 identifying assumptions. By letting Ω be the identity matrix we already have 3 of these restrictions. The remaining restriction will be imposed on B_0 such that only one of the structural shocks has a long run effect on x.

Collecting the initial values in $\tilde{x}_0$ the reduced form common trends representation is

$$x_t = \tilde{x}_0 + \delta t + C \sum_{i=1}^{t} \varepsilon_i + C^* \varepsilon_t. \tag{7.32}$$

Again, let $\eta_t = B_0 \varepsilon_t$ be the structural shocks, where $\eta_t = [\varphi_t \;\; \psi_t]'$. The innovation φ_t is a trend shock, while ψ_t is a transitory shock. The structural form of the common trends representation is:

$$x_t = \tilde{x}_0 + \delta t + A \sum_{i=1}^{t} \varphi_t + \Phi \eta_t. \tag{7.33}$$

Here, the 2×1 vector A is defined from $CB_0^{-1} = [A \;\; 0]$, and the 2×2 matrix $\Phi = C^* B_0^{-1}$. Since, $\Omega = I_2$, the parameters of B_0 must also satisfy $B_0^{-1}(B_0')^{-1} = \Sigma$.

QUESTION 15: *What do our 4 identifying assumptions imply for B_0?*

In the PIH case we have for $x_t = [C_t \;\; Y_t]'$ that

$$\begin{aligned} CB_0^{-1} &= \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_{11}^{+} & \beta_{12}^{+} \\ \beta_{21}^{+} & \beta_{22}^{+} \end{bmatrix} \\ &= \begin{bmatrix} \beta_{11}^{+} & \beta_{12}^{+} \\ \beta_{11}^{+} & \beta_{12}^{+} \end{bmatrix}. \end{aligned}$$

For the second column to contain zeros only, the inverse of B_0 must have that $\beta^{+}_{12} = 0$. Since the inverse is lower triangular, it follows that B_0 itself must be lower triangular, i.e.

$$B_0 = \begin{bmatrix} \beta_{11} & 0 \\ \beta_{21} & \beta_{22} \end{bmatrix}.$$

To uniquely determine the remaining 3 elements of B_0 we use the relation $\Sigma = B_0^{-1}(B_0')^{-1}$. This given us

$$\begin{bmatrix} \sigma_u^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_v^2 \end{bmatrix} = \begin{bmatrix} 1/\beta_{11} & 0 \\ -\beta_{21}/(\beta_{11}\beta_{22}) & 1/\beta_{22} \end{bmatrix} \begin{bmatrix} 1/\beta_{11} & -\beta_{21}/(\beta_{11}\beta_{22}) \\ 0 & 1/\beta_{22} \end{bmatrix}$$
$$= \begin{bmatrix} 1/\beta_{11}^2 & -\beta_{21}/(\beta_{11}^2\beta_{22}) \\ -\beta_{21}/(\beta_{11}^2\beta_{22}) & 1/\beta_{22}^2 + \beta_{21}^2/(\beta_{11}^2\beta_{22}^2) \end{bmatrix}.$$

Solving these 3 equations for β_{ij} we obtain:

$$\beta_{11} = 1/\sigma_u \qquad \beta_{21} = -1/\sigma_v \qquad \beta_{22} = 1/\sigma_v. \tag{7.34}$$

These parameters are equivalent to the trend innovation, φ_t, being a permanent income shock, and the temporary innovation, ψ_t, being a transitory income shock. This can be seen from the contemporaneous effects on consumption and income from one standard deviation shocks being

$$B_0^{-1} = \begin{bmatrix} \sigma_u & 0 \\ \sigma_u & \sigma_v \end{bmatrix},$$

where the first column contains the effects on x from the trend shock and the second column the effects from the temporary shock. The long run responses are given by

$$CB_0^{-1} = \begin{bmatrix} \sigma_u & 0 \\ \sigma_u & 0 \end{bmatrix}.$$

The long run is reached after 1 period in this example, and comparing the above results to those in section 2 we find that they are indeed equivalent.

For the n variable case, imposing the necessary n^2 identifying assumptions is somewhat more involved. First, $n(n+1)/2$ restrictions are given by assuming that $\Omega = I_n$. Second, $(n-r)r$ restrictions are obtained from $CB_0^{-1} = [A \;\, 0]$, where A is an $n \times (n-r)$ matrix. These assumptions imply that the first $(n-r)$ structural shocks have a long run effect on at least one of the x variables, while the remaining r shocks have only temporary effects on x. To identify the $(n-r)$ trend shocks $(n-r)(n-r-1)/2$ restrictions need to be imposed on A

(which implies the same number of restrictions on B_0), while the r transitory shocks can be identified from, e.g., restricting $r(r-1)/2$ elements of the final r columns of $\Phi_0 = C_0^* B_0^{-1}$. How to achieve this is discussed in some detail by King et al. (1991), Mellander et al. (1992), and Englund et al. (1994). In addition, the paper by Jacobson et al. (1996) discusses how to relate the structural common trends coefficients of the matrix A to familiar economic theory parameters.

References

Blanchard, O. J., and Quah, D. (1989). "The Dynamic Effects of Aggregate Demand and Supply Disturbances." *American Economic Review*, *79*, 655–673.

Cooley, T. F., and LeRoy, S. F. (1985). "Atheoretical Macroeconometrics: A Critique." *Journal of Monetary Economics*, *16*, 283–308.

Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, S. (1978). "Econometric Modelling of the Aggregate Time-series Relationship between Consumers' Expenditure and Income in the United Kingdom." *Economic Journal*, *88*, 661–692.

Englund, P., Vredin, A., and Warne, A. (1994). "Macroeconomic Shocks in and Open Economy: A Common Trends Representation of Swedish Data, 1871–1990." In V. Bergström and A. Vredin (Eds.), *Measuring and interpreting business cycles* (pp. 125–223). Oxford, England: Clarendon Press.

Frisch, R. (1933). "Propagation Problems and Impulse Problems in Dynamic Economics." In J. Åkerman (Ed.), *Economic essays in honour of gustav cassel* (pp. 171–205). London, England: George Allen.

Hall, R. E. (1978). "Stochastic Implications of the Life Cycle–Permanent Income Hypothesis: Theory and Evidence." *Journal of Political Economy*, *86*, 971–987.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton, N.J.: Princeton University Press.

Jacobson, T., Vredin, A., and Warne, A. (1996). "Common Trends and Hysteresis in Scandinavian Unemployment." *European Economic Review*.

Johansen, S. (1991). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica*, *59*, 1551–1580.

Johansen, S., and Juselius, K. (1990). "Maximum Likelihood Estimation and Inference on Cointegration: With Applications to the Demand for Money." *Oxford Bulletin of Economics and Statistics*, *52*, 169–210.

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.

Lippi, M., and Reichlin, L. (1993). "The Dynamic Effects of Aggregate Demand and Supply Disturbances: Comment." *American Economic Review*, *83*, 644–658.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series.* Berlin, Germany: Springer-Verlag.

Mellander, E., Vredin, A., and Warne, A. (1992). "Stochastic Trends and Economic Fluctuations in a Small Open Economy." *Journal of Applied Econometrics*, *7*, 369–394.

Runkle, D. E. (1987). "Vector Autoregressions and Reality." *Journal of Business and Economic Statistics*, *5*, 437–454.

Shapiro, M. D., and Watson, M. W. (1988). "Sources of Business Cycle Fluctuations." In S. Fischer (Ed.), *Nber macroeconomics annual* (pp. 111–148). Cambridge, Ma: MIT Press.

Sims, C. A. (1980). "Macroeconomics and Reality." *Econometrica*, *48*, 1–48.

Slutzky, E. (1937). "The Summation of Random Causes as the Source of Cyclic Processes." *Econometrica*, *5*, 105–146.

Stock, J. H., and Watson, M. W. (1988). "Variable Trends in Economic Time Series." *Journal of Economic Perspectives*, *2*, 147–174.

Anders Warne, Sveriges Riksbank, 103 37 Stockholm, Sweden
E-mail address: anders.warne@riksbank.se
URL: http://www.riksbank.com/
http://www.farfetched.nu/anders/