Estimation and Interpretation of Dynamic Relationships

Oscar Jorda
Version 1 - 2/24/00

1 Introduction

The object of this brief report is to review some of the basic time series techniques available for the study of dynamic systems. It is often of interest to obtain a collection of "stylized facts" from the data that any theoretical model of the economy should match. However, in practice, practitioners quickly become far more ambitious than their econometrics permit. First, the most common dynamic descriptors involve the calculation of dynamic second order moment conditions that are often compatible with a large number of structural models of the economy. Second, the structural interpretation of the statistics computed require that the practitioner make a number of untestable identification assumptions. The profession seems to have agreed upon a rule that rewards techniques that produce the most unambiguous results with the smallest number of assumptions. VARs and the study of their impulse response functions and variance decompositions certainly seem to fit the bill. On the other hand, cointegration tests and tests for serial correlation features provide a collection of strong statistical properties of the data that typically impose strong restrictions on the form of most theoretical models.

2 Building Blocks - The Wold Decomposition Theorem

The Wold decomposition theorem constitutes one of the basic tenets of linear time series modelling. It is based on the ability to linearly forecast any stationary process by means of uncomplicated and conventional linear time series properties. While this result is convenient, it makes no statement regarding the true form of the dynamic data generating process other than that implied by the first and second order moment conditions.

Theorem 1 *Any mean zero, covariance stationary process,* $\left\{ \underset{n\times 1}{\mathbf{y}_t} \right\}$ *can be represented in the form*

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\psi}(\mathbf{L})\boldsymbol{\varepsilon}_t$$

where $\psi(L)$ *is a matrix polynomial in the lag operator with*

- $\psi(0) = I_n$
- $\sum_{j=1}^{\infty} j|\psi_j| < \infty$ *(one summability condition that ensures stationarity)*
- $\underset{n\times 1}{\boldsymbol{\varepsilon}_t}$ *are one-step ahead* ***linear forecast errors*** *in* y_t *given information on lagged values of* y_t

- $\boldsymbol{\mu}_t$ *is linearly deterministic. Usually we consider a constant or a constant and a time trend.*

Remark 1 *The* $\boldsymbol{\varepsilon}_t$ *are linear forecast errors for* $\mathbf{y}_t$. *However, this does not imply* $E(\boldsymbol{\varepsilon}_t|\mathbf{y}_{t-1},...) = 0$. *Linear projections find the best linear combination of past* $\mathbf{y}_t$ *to fit* $\mathbf{y}_t$ *while the conditional expectation is the best guess of* $\mathbf{y}_t$ *using linear* ***and*** *nonlinear combinations of past* y_t.

Remark 2 *The shocks* ε_t *therefore, need not be the "true" shocks to the system.*

Remark 3 *The Wold representation is* ***a*** *representation of the time series which captures its second moment properties but may not be the representation of the time series* $\mathbf{y}_t$. *Representations based on non-forecast errors are perfectly possible.*

Remark 4 *Two time series with the same Wold representation are the same time series since the Wold representation is unique.*

The last two remarks are probably the most important in highlighting the link existing between the properties of the data and the data generating process, which I will equate to the structural economic model. The Wold decomposition is powerful because of its generality and because it enables us to investigate the dynamic properties of the time series on the basis of its second conditional dynamic moments. However, it is important to understand that the Wold decomposition can be, at most, a good approximation of the true data generation process. The last remark justifies the empirical approach used to test most theoretical models of the economy: If a theoretical model were to produce all moment conditions to match those implied by the Wold decomposition theorem, we could conclude that the observed data was generated by our model of the economy. What is harder to evaluate is the success of our model when this match is not exact. In addition, one can construct economic models that do not have a Wold decomposition, which adds to the difficulty of this procedure.

3 Vector Autoregressions

A convenient way to construct linear projections of $\mathbf{y}_t$ on the basis of past information is the vector autoregression (VAR). Although it is not the most parsimonious representation possible, it is much more helpful in terms of estimation and forecasting. VARMA models are more parsimonious representations and in fact many systems have Wold representations that are well summarized by small order VARMA models. However, estimation of generic VMA components requires recursive non-linear optimization techniques on MLE or QMLE expressions. In addition, the computation of forecasts is slightly more cumbersome. Since any invertible VMA has a VAR representation (often with infinite lags, however), it is not surprising that they have become the "standard" way to summarize the dynamics of a system.

3.1 Reduced Form VARs

$$\mathbf{y}_t = \mathbf{c} + \boldsymbol{\phi}_1 \mathbf{y}_{t-1} + ... + \boldsymbol{\phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

3.2 VMA(∞) Representation

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\psi}_1 \boldsymbol{\varepsilon}_{t-1} + ...$$

Any reduced form VAR of a covariance stationary process has an infinite VMA representation (which will coincide with the Wold representation if the VAR is correctly specified). The VMA representation is convenient here because it is much easier to interpret in terms of the statistics that are often of interest. In particular, the matrix of coefficients ψ_s has the interpretation,

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\varepsilon}_t'} = \psi_s$$

The function that describes this partial derivative for the j^{th} component of the vector $\mathbf{y}_t$ with respect to a shock on the i^{th} component at time t, for $s = 1, 2, ...$ is what we call the impulse response function. However, note that under the reduced form equivalent representation of the VAR

$$E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \Omega$$

which is usually a non-diagonal matrix. The implication of this result is that the linear forecast errors are contemporaneously correlated. The interpretation of the impulse response function is appealing from an economic point of view: It measures the response of the j^{th} component of $\mathbf{y}_t$ to an unanticipated disturbance in the i^{th} component. The "strait-jacket" of the Lucas critique makes this type of statement appealing since the unanticipated aspect of the impulse reassures us that the agents in the economy will not reoptimize and thus change the structural parameters.

Unfortunately, reduced form impulse response functions have the disadvantage that they do not allow us to separate the effect of the shocks that are correlated. For instance if ε_{yt} and ε_{rt} represent a technology shock and a monetary shock respectively, it does not make much sense to study the response to a unit shock to ε_{rt} since by the nature of their correlation, we know that part of the response is motivated by movements in ε_{yt}. In addition, we usually attach a causal interpretation to the response of the system to a shock. Therefore, while technology shocks maybe highly correlated with monetary shocks, we still pursue the understanding of the response of the economy to a monetary shock in isolation.

3.3 Identification of the "Structural Shocks"

The impulse response function that we are interested in, is that which answers the question,

$$\frac{\partial \widehat{E}(y_{i,t+s}|y_{j,t}, \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ...)}{\partial y_{j,t}} = \frac{\partial \widehat{E}(y_{i,t+s}|y_{j,t}, \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ...)}{\partial \varepsilon_{j,t}} \frac{\partial \varepsilon_{j,t}}{\partial u_{j,t}} = \widehat{E}(y_{i,t+s}|y_{j,t}, \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ...) - \widehat{E}(y_{i,t+s}|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ...)$$

which is different from the one we can directly obtain from a reduced form VAR. The variance-covariance matrix of the reduced form shocks, $E(\varepsilon_t \varepsilon_t') = \Omega$ is a real, symmetric, positive definite matrix which can always be rewritten as

$$\Omega = ADA'$$

where D is a diagonal matrix. Alternatively, we can also find the following decomposition

$$\Omega = PI_nP' = AD^{1/2}D^{1/2}A'$$

Given this transformation, we can define a vector of orthogonal residuals (i.e. contemporaneously uncorrelated), often labeled as "structural" shocks, as

$$\mathbf{u}_t \equiv A^{-1}\boldsymbol{\varepsilon}_t \text{ such that } E(\mathbf{u}_t\mathbf{u}_t') = E(A^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'(A^{-1})') = A^{-1}(ADA')(A^{-1})' = D$$

or

$$\mathbf{v}_t \equiv P^{-1}\boldsymbol{\varepsilon}_t \text{ such that } E(\mathbf{v}_t\mathbf{v}_t') = I$$

Therefore, the vector of orthogonal residuals, $\mathbf{u}_t$ (or $\mathbf{v}_t$ for orthogonal residuals with variance 1) can be expressed as a linear combination of the reduced form residuals. Going back to our original question,

$$\begin{aligned}
\frac{\partial \widehat{E}(y_{i,t+s}|y_{j,t},\mathbf{y}_{t-1},\mathbf{y}_{t-2},...)}{\partial y_{j,t}} &= \frac{\partial \widehat{E}(y_{i,t+s}|y_{j,t},\mathbf{y}_{t-1},\mathbf{y}_{t-2},...)}{\partial \varepsilon_{j,t}}\frac{\partial \varepsilon_{j,t}}{\partial u_{j,t}} = \psi_s a_j \text{ or} \\
\frac{\partial \widehat{E}(y_{i,t+s}|y_{j,t},\mathbf{y}_{t-1},\mathbf{y}_{t-2},...)}{\partial \varepsilon_{j,t}}\frac{\partial \varepsilon_{j,t}}{\partial v_{j,t}} &= \psi_s a_j d_{jj}^{1/2} = \psi_s p_j
\end{aligned}$$

where a_j is the j^{th} column of the matrix A and where the response is to a shock of size one. If we choose the normalization provided by $\mathbf{v}_t$ then we have the response to a shock of size one standard deviation.

To summarize, in order to compute a more "structural looking" impulse response function, we need to find a transformation matrix (A or P) that orthogonalizes the reduced form residuals. The next section discusses some of the more traditional choices of transformation matrix.

3.4 Common Identification Assumptions

Reduced form VARs can be seen as general representations of structural models of the form

$$B_0\mathbf{y}_t = k + B_1\mathbf{y}_{t-1} + ... + B_p\mathbf{y}_{t-p} + \mathbf{u}_t \qquad E(\mathbf{u}_t\mathbf{u}_t') = D \tag{1}$$

Since B_0 is invertible (otherwise one of the variables in the system is a perfect linear combination of the others) then pre-multiplying by B_0^{-1}, we have the reduced form,

$$\mathbf{y}_t = \mathbf{c} + \boldsymbol{\phi}_1\mathbf{y}_{t-1} + ... + \boldsymbol{\phi}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t \tag{2}$$

where

- $\mathbf{c} = B_0^{-1}k$
- $\boldsymbol{\phi}_j = B_0^{-1}B_j$ for $j = 1, ..., p$
- $\boldsymbol{\varepsilon}_t = B_0^{-1}\mathbf{u}_t$
- $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = E(B_0^{-1}\mathbf{u}_t\mathbf{u}_t'(B_0^{-1})') = \Omega$

Both, (1) and (2) have the respective Wold representations,

$$\mathbf{y}_t = \widetilde{\mu} + \widetilde{\psi}_0 \mathbf{u}_t + \widetilde{\psi}_1 \mathbf{u}_{t-1} + ...$$

and

$$\mathbf{y}_t = \mu + \boldsymbol{\varepsilon}_t + \psi_1 \boldsymbol{\varepsilon}_{t-1} + ...$$

with $\widetilde{\psi}_0 = B_0^{-1}$ and $\widetilde{\psi}_j = B_0^{-1}\psi_j$.

3.4.1 Sims Orthogonalization Assumption - Specification of B_0 (or $\widetilde{\psi}_0$)

The object of orthogonalizing the reduced form residuals therefore consists on finding a transformation matrix A (or in terms of the structural model, B_0) that expresses the structural residuals in terms of linear combinations of the reduced form residuals. As the reader may have anticipated, these transformation matrices are not unique and therefore depend on the assumptions the practitioner is willing to impose on the nature of the contemporaneous correlation of the dependent variables.

The most popular orthogonalization scheme is that originally proposed by Sims (1980). The orthogonalization operated by assuming that the transformation matrix A has a lower triangular structure, that is,

$$A = \begin{bmatrix} 1 & 0 & 0 & ... & 0 \\ a_{21} & 1 & 0 & ... & 0 \\ a_{31} & a_{32} & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ a_{n1} & ... & ... & ... & 1 \end{bmatrix} \tag{3}$$

The implication of this assumption is that, *given the ordering of the variables within the vector* $\mathbf{y}_t$, the contemporaneous correlations are given a hierarchical structure (sometime referred to as a Wold causal ordering), such that,

- y_{1t} can be explained by $\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ...$
- y_{2t} can be explained by y_{1t}, and $\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ...$
- y_{3t} can be explained by y_{1t}, y_{2t} and $\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ...$
- ...
- y_{nt} can be explained by $y_{1t}, y_{2t}, ..., y_{n-1,t}$ and $\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ...$

The Choleski decomposition of the variance-covariance matrix of the reduced for residuals, $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \Omega$ yields a unique, lower triangular matrix with the structure described by the matrix (3).

Remark 5 *The uniqueness of the Choleski decomposition is with respect to the ordering within the vector* $\mathbf{y}_t$*. However, for each reordering of* $\mathbf{y}_t$*, there is a different unique Choleski decomposition.*

Remark 6 *The advantage of imposing this type of constraint on the nature of the contemporaneous correlations is that the reduced form VAR can be estimated without imposing any constraints. Then, one can compute the variance-covariance matrix of the reduced form residuals from which the Choleski decomposition can then be calculated. The matrix A so calculated can then be used to compute the so identified "structural" impulse response functions.*

3.4.2 Blanchard and Quah Orthogonalization Restrictions on $\psi(1)$

So far we have been working under the assumption that $\mathbf{y}_t$ was covariance-stationary. This assumption was convenient to derive the main results but does not detract from the generality of the derivations. When $\mathbf{y}_t$ has unit roots, there are alternative identification restrictions that can be used. Let $\boldsymbol{\Delta}\mathbf{y}_t$ be $I(0)$, then, it has a Wold decomposition given by

$$\Delta\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\psi}_1\boldsymbol{\varepsilon}_{t-1} + ...$$

or, using the Beveridge-Nelson decomposition,

$$\Delta\mathbf{y}_t = \underset{trends}{\boldsymbol{\psi}(1)\boldsymbol{\varepsilon}_t} + \underset{cycles}{\Delta\boldsymbol{\psi}^*(L)\boldsymbol{\varepsilon}_t}$$

where,

$$\psi_i^* = -\sum_{j>i}^{\infty}\psi_j; \qquad \psi_0^* = I_n - \psi(1)$$

Blanchard and Quah (1989) assume that $\boldsymbol{\psi}(1)$ is lower triangular which implies that the long-run response of y_{1t} to shocks on other variables is zero. Blanchard and Quah assume that, in a VAR with GNP in the first equation, "demand shocks" (i.e., shocks coming from the rest of the variables in the system) have no long-run effects.

Although cointegration will be discussed below, it is instructive to point out that if $\boldsymbol{\psi}(1)$ is lower triangular, it is a full rank matrix and therefore, there is no cointegration in the system. Thus, in order to use an assumption of this nature, it is advisable to test for cointegration and modify the type of assumptions made on $\boldsymbol{\psi}(1)$.

3.4.3 Other Identification Assumptions and Remarks

The type of restrictions placed on A by Sims' approach, although convenient, may not be sustainable on theoretical grounds, by the specific model under consideration. The construction of "structural" residuals from the reduced form residuals only requires that the structural matrix B_0 (or A) have enough restrictions to allow for the system to have just identified "structural" residuals. The estimation of this VAR can no longer be based on the reduced form VAR and requires methods of estimation that incorporate the constraints (for example Limited Information Maximum Likelihood, LIML). Identification of the "structural" residuals requires that both rank and order conditions be satisfied, namely:

- **Order Condition:** B_0 and D can have no more unknown parameters than Ω. Since Ω is symmetric, this means $n(n+1)/2$ distinct values. Given that D is diagonal, it has n parameters and therefore B_0 can not have more than $n(n-1)/2$ true parameters.

- **Rank Condition:** See Hamilton (1994) pg. 333 and Gianinni (1992) on how to check for rank conditions.

In the spirit of Sims' approach, other researchers impose somewhat less restrictive identification assumptions that rely on partitioning the matrix A into blocks.

To reiterate, the motivation for studying impulse response functions stems from a strong desire to avoid the Lucas critique. However, in practice, there are many situations in which the interesting question lies, not so much as to how does the economy react to unanticipated shocks as it is to analyze the impact of systematic behavior. For example, in monetary VARs, the unanticipated component of monetary policy tends to be relatively insignificant in explaining other variables in the system. This seems natural since we would not expect the Federal Reserve to react randomly to the economy's disturbances

4 Variance Decompositions

Given the Wold representation, the s period ahead forecast error can be expressed as

$$\mathbf{y}_{t+s}-\widehat{E}(\mathbf{y}_{t+s}|\mathbf{y}_t,...) = \boldsymbol{\varepsilon}_{t+s}+\boldsymbol{\psi}_1\boldsymbol{\varepsilon}_{t+s-1}+...+\boldsymbol{\psi}_{s-1}\boldsymbol{\varepsilon}_{t+1}$$

with mean square error given by the expression,

$$\begin{aligned} MSE(\widehat{y}_{t+s|t}) &= E\left[(\mathbf{y}_{t+s}-\widehat{y}_{t+s|t})(\mathbf{y}_{t+s}-\widehat{y}_{t+s|t})'\right] = \\ &\quad \Omega+\psi_1\Omega\psi_1'+...+\psi_{s-1}\Omega\psi_{s-1}' \end{aligned}$$

Recall that from our previous discussion on orthogonalizing reduced form errors, we had that the structural errors could be expressed as linear combinations of the reduced form errors. In particular, $\varepsilon_t = Au_t$. Therefore,

$$MSE(\widehat{y}_{t+s|t}) = D\left[AA'+\psi_1 AA'\psi_1'+...+\psi_{s-1}AA'\psi_{s-1}'\right]$$

Consequently, the contribution of the j^{th}orthogonalized innovation to the $MSE$of the $s-$*period* ahead forecast is,

$$d_{jj}\left[a_j a_j'+\psi_1 a_j a_j'\psi_1'+...+\psi_{s-1}a_j a_j'\psi_{s-1}'\right]$$

As $s\to\infty$, a covariance stationary process will have $MSE(\widehat{y}_{t+s|t})\to\Gamma_0$, the unconditional variance.

The typical use of the variance decomposition is to make statements regarding the "percentage of the variance explained by innovations in variable j." Just as when we discussed impulse response functions, this statement is only sensible if we can attach some structural meaning to the innovation under consideration. Identification of the structural innovations requires the same type of assumptions on the transformation matrix A as we made above.

5 Assumption Free Measures of Dynamic Covariation

5.1 Den Haan's Measure of Dynamic Covariation

To avoid having to identify structural innovations, Den Haan (1999) proposes the following statistic. Consider a reduced form VAR of the form we have been studying,

$$\mathbf{y}_t = \mathbf{c}+\boldsymbol{\phi}_1\mathbf{y}_{t-1}+...+\boldsymbol{\phi}_p\mathbf{y}_{t-p}+\boldsymbol{\varepsilon}_t$$

Let $\widehat{e}_j(t+k|t) \equiv y_{j,t+k} - \widehat{E}(y_{j,t+k}|\mathbf{y}_t, \mathbf{y}_{t-1}, ...)$. Define $COV_{ij}(K) \equiv E(\widehat{e}_j(t+k|t)\widehat{e}_i(t+k|t)')$. The plot of this covariance for different values of k is a plot of the dynamic covariance between $y_i$and y_j. To put this measure in the context of the models we have been considering, note that,

$$\mathbf{y}_{t+k} - \widehat{E}(\mathbf{y}_{t+k}|\mathbf{y}_t, \mathbf{y}_{t-1}, ...) = \boldsymbol{\varepsilon}_{t+k} + \psi_1 \boldsymbol{\varepsilon}_{t+k-1} + ... + \psi_{k-1}\boldsymbol{\varepsilon}_{t+1}$$

Therefore,

$$COV(K) = \Omega + \psi_1 \Omega \psi_1' + ... + \psi_{k-1}\Omega\psi_{k-1}'$$

or more specifically,

$$COV_{ij}(K) = \sigma_{ij}(1 + \psi_{1,i}\psi_{1,j}' + ... + \psi_{k-1,i}\psi_{k-1,j}')$$

Therefore, Den Haan's measure amounts to calculating the covariance for the K^{th} periods ahead forecast errors.

5.2 Vector Autocorrelations

Vector autocorrelations (or more conventionally known as cross-correlograms in the Box-Jenkins tradition) consist in calculating the following dynamic correlations,

$$\rho_{ij}(s) = \frac{E\left[(y_{i,t} - \mu_j)(y_{j,t+s} - \mu_i)\right]}{\sigma_i \sigma_j}$$

When $i = j$, this amounts to calculating the autocorrelation function. Note that, in principle $\rho_{ji}(s) \neq \rho_{ij}(s)$. The right relation is $\rho_{ij}(s) = \rho_{ji}(-s)$. Vector autocorrelations provide measures of dynamic correlation between two variables.

Remark 7 *Although the vector autocorrelation is not defined for non-stationary data, note that if y_i and y_j are cointegrated, their cross correlation never dies out in pretty much the same fashion as when we look at the autocorrelation function of a unit root process.*

From a practical point of view, matching the sample vector autocorrelations of the data with those obtained from a model may be the best we can do. Aside from constituting another collection of stylized facts, it is hard to attach any structural meaning unless one is willing to assume that the only source of the covariation is contained within the two variables under study.

5.3 Vector Partial Autocorrelations

In univariate time series analysis we are familiar with the partial autocorrelation function, which is defined as the last coefficient in a linear projection of y_i on its most m recent values. In other words, it is the correlation between y_{it} and y_{it-m} which is unexplained by the intervening $m-1$ lags.

$$\rho_{im} = Corr[y_{it} - E^*(y_{it}|y_{it-1}, ..., y_{it-m+1}), y_{it-m})$$

where E^* denotes the linear projection operator. Therefore, it seems natural to extend the concept of the partial autocorrelation to a multivariate context. Define the partial vector autocorrelation as,

$$\lambda_{ij}(m) = Corr\left[y_{it} - E^*(y_{it}|\mathbf{y}_{t-1}, ...\mathbf{y}_{t-m}), y_{jt-m}\right]$$

The partial autocorrelation defined in this manner has a much closer interpretation to a structural impulse response function (from a causal point of view) since we are analyzing the dynamic covariation of the variables y_i and y_j conditional on all the intervening information captured by the lags of all the variables in the system. However, it maintains its assumption-free nature in that its construction does not impose any a priori restrictions other than those reflecting the choice of variables in the system.

Computation of vector partial autocorrelations is straight forward. A natural estimate of the m^{th} cross partial autocorrelation is the last coefficient in an OLS regression of $\widetilde{y}_{it}$ on the m most recent lags of $\widetilde{\mathbf{y}}_t$

$$\widetilde{y}_{it} = A_1^m \widetilde{y}_{t-1} + ... + A_m^m \widetilde{y}_{t-m} + e_{it}$$

where the $\widetilde{\mathbf{y}}_t$ have been standardized to have mean zero and variance one. The sample estimate of the vector A_m^m is the estimate of the m partial auto and cross autocorrelations. Furthermore $V(\widehat{a}_{j,m}^m) \simeq 1/T$ for $m > p$ (the lag length of the VAR).

6 Common Trends and Common Cycles

The literature on unit roots and cointegration has lent a host of new econometric results with which to analyze the data in search of relevant economic properties. These approaches allow us to uncover a different collection of dynamic patterns in the data.

Let $\Delta \mathbf{y}_t \sim I(0)$, then it has a vector Wold decomposition given by,

$$\Delta \mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\psi}(\mathbf{L})\boldsymbol{\varepsilon}_t$$

Assuming $\mu = 0$ for convenience and using the Beveridge-Nelson decomposition, we have

$$\Delta \mathbf{y}_t = \psi(1)\boldsymbol{\varepsilon}_t + \Delta\psi^*(L)\boldsymbol{\varepsilon}_t$$

where

$$\psi_i^* = -\sum_{j>i}^{\infty} \psi_j; \qquad \psi_0^* = I_n - \psi(1)$$

or

$$\mathbf{y}_t = \underset{trends}{\psi(1)\Delta\boldsymbol{\varepsilon}_t} + \underset{cycles}{\psi^*(L)\boldsymbol{\varepsilon}_t}$$

The Beveridge-Nelson (BN) decomposition can be also cast in terms of what is called the Beveridge-Nelson-Stock-Watson (BNSW) decomposition, namely

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\gamma}\boldsymbol{\tau}_t + \mathbf{c}_t \\ \boldsymbol{\tau}_t &= \boldsymbol{\tau}_{t-1} + \boldsymbol{\delta}'\boldsymbol{\varepsilon}_t \end{aligned}$$

with

$$\boldsymbol{\tau}_t = \boldsymbol{\delta}' \sum_{s=0}^{\infty} \varepsilon_{t-s}; \qquad c_t = \psi^*(L)\boldsymbol{\varepsilon}_t$$

Cointegration asks the question: Are there linear combinations of $\mathbf{y}_t$that are stationary? In other words, can one find vectors collected in the matrix A, such that, $A'\mathbf{y}_t \sim I(0)$? From the BN decomposition, we know the trends are collected in the term $\psi(1)\Delta\boldsymbol{\varepsilon}_t$ so cointegration consists in finding a matrix $\underset{n\times k}{A}$ such that $A'\psi(1) = 0$. The implication of this statement is that the matrix $\psi(1)$ has therefore less than full rank.

The most popular cointegration test (Johansen's test) is based precisely on testing the rank of the matrix $\psi(1)$ and finding the linear combinations that satisfy the condition $A'\psi(1) = 0$. Due to the nature of non-stationary data and the superconsistency properties that the parameter estimates of the cointegration relations exhibit, cointegration tests are popular in macroeconomics. Intuitively, they also yield fairly strong statements: Two nonstationary series, which by virtue of their nature have a forecast variance that goes to infinity as the forecast horizon goes further into the future, are nevertheless related in such a way that even in the most distant of future, they will never be apart from each other by more than they are today.

In principle, there is nothing that precludes us from asking similar questions regarding the short-run properties of the data, or what we labeled as cycles in the BN decomposition. While the cointegration matrix A is such that $A'\mathbf{y}_t = A'c_t$ we can also ask if there are linear combinations of the elements in $\mathbf{y}_t$, say, F,such that $F'\mathbf{y}_t = F'\gamma\tau_t$. This condition implies that the matrix F is such that,

$$F'c_t = 0 = F'\psi^*(L)$$

This property is sometimes labeled *codependence* or $\mathbf{y}_t$ is said to have *common serial correlation features*. One may have assumed that the linear combinations that make the series in the vector $\mathbf{y}_t$ to be cointegrated would also make the vector exhibit codependence. However, note that if $F'\psi^*(L) = 0$, then, in particular, $F'\psi_0^* = 0$. However, note that $\psi_0^* = I - \psi(1)$. Therefore, $F'(I - \psi(1)) = 0$ implies that $F'\psi(1) \neq 0$. Thus, codependence relations are orthogonal to cointegration relations.

6.1 Canonical Correlation Analysis

Canonical correlations are particularly useful in testing for the commonality of trends and cycles but have other applications as well. They are seldom discussed in regular Time Series courses but they are quite useful. Let $\underset{n\times 1}{y_t}$ and $\underset{k\times 1}{x_t}$ be two vectors of random variables (usually expressed in deviations from the mean for simplicity). Assume $n < k$. Canonical correlation analysis consists in finding the linear combinations of the elements of y_t that are *maximally correlated* with linear combinations of x_t.

Define $\underset{m\times 1}{\eta_t} \equiv \underset{m\times n}{K}{}' y_t$ and $\underset{m\times 1}{\xi_t} \equiv \underset{m\times k}{A}{}' x_t$ where $m = \min(n, k)$. The choice of K and A is made such that:

1. The individual elements of η_t and ξ_t have unit variance and are mutually uncorrelated,

$$\left.\begin{array}{l} E(\eta_{it}, \eta_{jt}) \\ E(\xi_{it}, \xi_{jt}) \end{array}\right\} = \left\{\begin{array}{l} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{array}\right.$$

   or in matrix form, $K'\Sigma_{yy}K = I_m$ and $A'\Sigma_{xx}A = I_m$.

2. The i^{th} element of η_t is uncorrelated with the j^{th} element of ξ_t for $i \neq j$. That is, $E(\eta_{it}, \xi_{jt}) = 0$ for $i \neq j$.

3. $E(\eta_t, \xi_t) = K'\Sigma_{xy}A = R$, where

$$R = \begin{bmatrix} r_1 & 0 & ... & 0 \\ 0 & r_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & ... & 0 & r_m \end{bmatrix}$$

4. The elements of R are usually ordered such that $1 \geq r_1 \geq ... \geq r_m \geq 0$. The r_i are denominated as "*i^{th} canonical correlation between y_t and x_t*."

The i^{th} sample canonical correlation can be then calculated as the i^{th} largest eigenvalue of the matrix

$$\widehat{\Sigma}_{yy}^{-1}\widehat{\Sigma}_{yx}\widehat{\Sigma}_{xx}^{-1}\widehat{\Sigma}_{xy}$$

The mechanics of finding the canonical correlations therefore involve rather simple matrix algebra and can be found in Reinsel (1994) or Hamilton (1994) for example. More useful is the test statistic for the null hypothesis that the s smallest canonical correlations are zero. This test, developed in Tiao and Tsay (1985) has the following structure,

$$-(T-k)\sum_{i=1}^{s}\log(1-r_i^2)$$

and is distributed as a χ^2.

6.2 Testing for Cointegration and Codependence

We begin first with the test for codependence. We are interested in finding linear combinations of the vector $\mathbf{y}_t$ which display no serial correlation. Suppose that this vector is non-stationary and displays cointegration (for generality's sake). Let $\mathbf{W}_t = \{\mathbf{\Delta y}_{t-1}, ..., \mathbf{\Delta y}_{t-p-1}, \mathbf{z}_{t-1}\}$, that is, $\mathbf{W}_t$ collects all the relevant lags and error correction terms (if there is no cointegration this just means that the terms $\mathbf{z}_{t-1} = 0$). Then a test for zero canonical correlations between $\mathbf{y}_t$ and $\mathbf{W}_t$ is a test for codependence: A zero canonical correlation in this context is a linear combination of the elements in $\mathbf{y}_t$ that is uncorrelated with all the lags and cointegration terms in the VAR, the definition of codependence.

Johansen's cointegration test works in a similar fashion with two caveats: (1) we need to construct a set of auxiliary regressions first, and (2) the distribution of the test for zero canonical correlations will be non-standard since it involves vectors of non-stationary variables.

6.2.1 Digression

Any n-variate VAR

$$\mathbf{y}_t = \mathbf{c} + \boldsymbol{\phi}_1\mathbf{y}_{t-1} + ... + \boldsymbol{\phi}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

can be rewritten as,

$$\mathbf{\Delta y}_t = \boldsymbol{\xi}_1\mathbf{\Delta y}_{t-1} + ... + \boldsymbol{\xi}_{p-1}\mathbf{\Delta y}_{t-p-1} + \boldsymbol{\alpha} + \boldsymbol{\xi}_0\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- If $rank(\xi_0) = n$ then there are n cointegrating relationships which is another way of saying that any linear combination of the elements $\mathbf{y}_t$ is stationary. Therefore, $\mathbf{y}_t$ is itself stationary. This rank condition is equivalent to saying that $rank(\psi(1)) = 0$. There is no reason to specify the VAR in differences.

- If $rank(\xi_0) = 0$ then there is no cointegration and the VAR in differences imposes the appropriate restriction that there are n different trends in the system. Note that in this case, $rank(\psi(1)) = n$.

- If $rank(\xi_0) = h$, $0 < h < n$, then we say that there are h cointegrating relations or $n - h$ trends. In this case $rank(\psi(1)) = n - h$

The Johansen test is a rank test on ξ_0. It involves the following steps:

1. Calculate the auxiliary regressions,

$$\begin{aligned} \Delta y_t &= \beta_0 + \beta_1 \Delta y_{t-1} + ... + \beta_{p-1} \Delta y_{t-p-1} + u_t \\ y_{t-1} &= \gamma_0 + \gamma_1 \Delta y_{t-1} + ... + \gamma_{p-1} \Delta y_{t-p-1} + v_t \end{aligned}$$

2. Test the number of zero canonical correlations between $\widehat{u}_t$ and $\widehat{v}_t$. For example, consider the null hypothesis that there are h cointegrating relations versus the null that there are n (i.e., the system is stationary), then an appropriate test statistic is,

$$-T \sum_{i=h+1}^{n} \log(1 - r_i)$$

   which has a non-standard distribution that was calculated by Johansen. The r_i are the canonical correlations between $\widehat{u}_t$ and $\widehat{v}_t$.