Macroeconometrics, Spring 2022
Bent E. Sørensen
February 10, 2022

1 Cointegration.

There are by now many surveys and coverage in textbooks, although most of these are not too clear and multivariate models are best covered in Johansen's (1995) book.

Cointegration and unit-root theory is the innovation in theoretical econometrics that created the most interest among economists in the lates 1980s and 1990s. The definition in the simple case of 2 time series x_t and y_t, that are both integrated of order one (this is abbreviated I(1), and means that the process contains a unit root), is the following:

Definition:

x_t and y_t are said to be cointegrated if there exists a parameter α such that

$$u_t \;=\; y_t - \alpha x_t$$

is a stationary process.

This turns out to be a pathbreaking way of looking at time series. Why? because it seems that lots of lots of economic series behaves that way and because this is often predicted by theory. The first thing to notice is of course that economic series behave like I(1) processes, i.e. they seem to "drift all over the place"; but the second thing to notice is that they seem to drift in such a way that the they do not drift away from each other. If you formulate this statistically you come up with the cointegration model.

The famous paper by Davidson, Hendry, Srba and Yeo (1978) (DHSY), argued heuristically for models that imposed the "long run" condition that the series modeled should not be allowed to drift arbitrarily far from each other. In the 60s and 70s people estimated (in particular) consumption functions in either levels:

$$c_t = \beta_0 + \beta_1 y_t + lags + u_t.$$

and found very larger R^2s. But typically the estimated coefficients were not not very stable when the sample changed so other economists found that the differenced relation

$$\Delta c_t = \alpha_0 + \alpha_1 \Delta y_t + lags + e_t\,,$$

was more stable even if the R^2s were way smaller. DHSY suggest that that the level regression should be interpreted as long-run relation with the right-hand side as a targer for c_t, so that

the best relation would be "error correction model" (ECM)

$$\Delta c_t = \alpha_0 + \gamma * (c_{t-1} - \beta_1 y_{t-1}) + \kappa_1 \Delta c_{t-1} + \kappa_2 \Delta y_{t-1} + lags + u_t \,,$$

where we would expect γ to be negative, with the interpretation that is c were about its long-run target, consumption growth would tend to be below average to a "correct" the "error." DHSY found that the γ coefficient indeed was negative and that the ECM was quite stable compared to purely levels or differences equations.
But, as I keep preaching, you cannot put apples and bananas in the same regressions so how can you have levels and differences together? Granger suggested co-integration theory as a statistically coherent way of dealing with this. As it turns out, aggregate consumption and income both behave like random walks. Therefore the levels regression return a stationary error $c_t - \beta_1 y_t$ that you can include in the ECM and all terms are stationary. ("All apples, no bananas.")

One reason unit roots and cointegration is so important is the following. Consider the regression

$$y_t = \alpha_0 + \alpha_1 x_t + u_t \ . \tag{1}$$

A: Assume that x_t is a random walk and that y_t is an *independent* random walk (so that x_t is independent of y_s for all s). Then the true value of α_1 is of course 0, but the limiting distribution of $\hat{\alpha_1}$ is such that $\hat{\alpha_1}$ converges to a function of Brownian motions which is often far from zero and more often that not a standard t-test will reject that the coefficient is 0. This is called a **spurious regression**, and was first noted by Monte Carlo studies by Granger and Newbold (1974). One can furthermore show that the t-statistic *diverges* at rate $\sqrt{T}$, so in large samples you would almost always reject the true value of 0. The problem does reveal itself in typical OLS regression output though—if you find a very high R^2 and a very low Durbin-Watson value, you usually face a regression where unit roots should be taken into account. (Sadly, no-one prints Durbin-Watson statistics any more, but if the R^2 is too good to be true, the situation is usually not good and not true.) I So consider any reported time series regression with R^2 coefficients above (say) .95 with extreme suspicion. But you should always test if you time series data has unit roots and, of course, if you test and find unit roots, then you can get standard consistent estimators by running the regression

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta x_t + e_t \ , \tag{2}$$

since you now regress a stationary variable on a stationary variable, and the classical statistical theory applies.

B: Assume that x_t is a random walk and the y_t is another random walk, such that (1) holds for non-zero α_1, but with the error term u_t following a unit root distribution. In this case you still get inconsistent estimates and you need to estimate the relation (2). This may also be called a spurious regression, even though there actually is a relation between x_t and y_t—I don't think there is quite an established way of referring to this situation (which is often "forgotten" in the discussion of case A and Case C).

C: Now assume that (1) holds with a stationary error term. This is exactly the case where x and y are cointegrated. In this case, $\hat{\alpha}_1$ is not only consistent, but it converges to the true value at rate T. We say that the OLS estimater is **superconsistent**. In the case where x_t is a simple random walk and u_t is serially uncorrelated you will find that $T*(\hat{\alpha}_1 - \alpha_1)$ is asymptotically distributed as a $\int_0^1 B_2 dB_1 / (\int_0^1 B_2^2 dt)$, where B_1 and B_2 are independent Brownian motions. This limiting distribution has mean zero, but more importantly the standard t-test is asymptotically normally distributed. In the situation where there is serial correlation or x_t may be correlated with u_s for some s, you do not get a symmetric asymptotic distribution of $T*(\hat{\alpha}_1 - \alpha_1)$, and even if the estimate converges at the very fast rate T, this may translate into non-negligible bias in typical macro data samples. In this situation the t-statistic is no longer asymptotically normal. However, the distribution typically depends on only the number of lags and is tabulated or built into programs like Stata.
Note that in the situation where x_t and y_t are cointegrated the regression (2) is still consistent (although you would introduce a unit root in the MA representation for the error term). So the differenced regression is always consistent and one could argue that it would be "safe" to always estimate this relation. The loss of efficiency is, however, very big and most macro time series are so short that the gain in efficiency from running the cointegrating regression can be critical for getting decent results.

Consider the error correction model again: Assume we estimate the co-integrating coefficient (or vector) $\hat{\beta}_1$ and then consider the term $x_t = c_{t-1} - \hat{\beta}_1 y_{t-1}$, where I use the lable x_t. For the purpose of estimating the coefficients of the error correction model, we can consider x_t an observed, pre-determined, variable and estimate

$$\Delta c_t = \alpha_0 + \gamma x_t + \kappa_1 \Delta c_{t-1} + \kappa_2 \Delta y_{t-1} + lags + u_t \,,$$

using least squares using the usual asymptotic t-stats.. Why is that OK, when we have a parameter pre-estimated? The answer is that because $\hat{\beta}_1$ converges at rate T which is much faster than the $\sqrt{T}$ convergence of the parameters in the error correction equation, the estimation

error for this variable is ignorable asymptotically.

For many purposes the above definition of cointegration is too narrow. It is true that economic series tend to move together but in order to obtain a linear combination of the series, that is stationary one may have to include more variables. The general definition of co-integration (for the I(1) case) is therefore given as restrictions on a VAR model. This theory is conceptually straightforward and well developed by Johansen, but it treats all variables symmetrically and sometimes it is much easier to have intuition for cases where and "x" variable impacts on a "y" variable.

Definition *A vector of I(1) variables y_t is said to be cointegrated if there exist at vector β_i such that $\beta_i' y_t$ is trend stationary. If there exist r such linearly independent vectors β_i, i = 1,...,r, then y_t is said to be cointegrated with cointegrating rank r. The matrix $\beta = (\beta_1, ...\beta_r)$ is called the cointegrating matrix.*

Note that $\beta' y_t$ is an r-dimensional vector of trend-stationary variables.
Also note that this definition is symmetric in the variables, i.e. there is no designated left-hand side variable. This is usually an advantage for the statistical testing but of course it makes it harder for the economic intuition.
Note the β_i vectors are individually identified only up to scale since $\beta_i y_t$ stationary implies that $c\beta_i y_t$ is stationary. This of course implies that one can normalize one of the coefficients to one - but only in the case where one is willing to impose the a priori restriction that this coefficient is not zero. As far as the identification of the matrix β is concerned it is also clear that if β is a cointegrating matrix then $\beta F'$ is also a cointegrating matrix for any non-singular matrix F.
Finally note that the treatment of constants and drift terms are suppressed here. One has to consider those for any practical applications of cointegration methods.

1.1 Cointegration in the autoregressive representation

The general VAR(k) model can be written as

$$\Delta y_t \;=\; \Pi y_{t-1} \;+\; \sum_{j=1}^{k-1} \Gamma_j \Delta y_{t-j} \;+\; e_t \;,$$

as considered earlier.
If Π is equal to zero this means that there is no cointegration. This is the model that is implicit in the Box-Jenkins method. The variables may be I(1); but that can easily be "cured"

by taking differences (in order to achieve the usual asymptotic distribution theory).
If Π has full rank then all y_t must be stationary since the left hand side and the other right hand side variables are stationary (since we limit ourselves to variables that are either I(0) or I(1)).
The most interesting case is when Π has less than full rank but is not equal to zero. This is the case of cointegration. In this case Π can be written as $\Pi = \alpha\beta'$ (yes, this β corresponds to the cointegration matrix introduced above), where α and β are $n \times r$ matrices. Note that α and β are only identified up to non-singular transformations since $\Pi = \alpha\beta' = \alpha F^{-1}(\beta F')'$ for any non-singular F. This lack of identification can sometimes render results from multivariate cointegration analysis impossible to interpret and finding a proper way of normalizing β (and thereby α) is often the hardest part of the work. α can be interpreted as a "speed of adjustment towards equilibrium".

1.2 Cointegration in the moving average representation

The multivariate Wold-representation states that the stationary series Δy_t can be written as

$$(1-L)y_t \;=\; \Psi(L)e_t \;,$$

which, by the Beveridge-Nelson decomposition, can be written as

$$(1)\quad y_t \;=\; \Psi(1)S_t \;+\; \Psi^*(L)e_t \;,$$

where S_t is the n-dimensional random walk $S_t = \sum_{s=1}^{t} e_s$ and $\Psi^*(L) = (1-L)^{-1}(\Psi(L)-\Psi(1))$. Now $\beta' y_t$ is stationary in the case of cointegration, so that

$$\beta' y_t \;=\; \beta'\Psi(1)S_t \;+\; \beta'\Psi^*(L)e_t \;,$$

is stationary, which implies that $\beta'\Psi(1)S_t$ is equal to 0. This gives another characterization of cointegration that may be useful for testing.

One can show that the representation (1) can be reformulated in the case of cointegration as

$$y_t \;=\; \tilde{\Psi}S_t^* \;+\; \Psi^*(L)e_t \;,$$

where S_t^* is the (n-r)-dimensional random walk. This is called a common trend representation in Stock and Watson 1988, and this representation can also be used as the basis for cointegration tests (some of which are included in coint package for GAUSS).

1.3 Testing for cointegration

1.4 The Engle-Granger test

The most well known test, suggested by Engle and Granger (1987) (sometimes known as the EG test) is to run a static regression (after first having verified that y_t and x_t both are I(1))

$$y_t \;=\; \theta' x_t \;+\; e_t \;,$$

where x_t is one- or higher-dimensional. The asymptotic distribution of θ is not standard, but the test suggested by Engle and Granger was to estimate $\hat{\theta}$ by OLS and the test for unit roots in

$$\hat{e}_t \;=\; y_t \;-\; \hat{\theta}' x_t \;.$$

Note, that since the unit root tests test the null-hypothesis of a unit root, most cointegration tests test the **Null of no cointegration**. Unfortunately the limiting distribution of for example the t-test, does not have the limiting distribution tabulated by Dickey and Fuller. The limiting distribution does, however, resemble the Dickey-Fuller distribution even though you need a separate table for each dimension of the regressor. Typically, you will allow for dynamics in the residual and perform the equivalent of the ADF test (using the slightly different critical values in this case). Such a procedure is usually called a Cointegration ADF test, abbreviated CADF-test. Engle and Granger (1987) compared different tests and recommended the CADF test. They supplied critical values based on Monte Carlo simulations for the case of just one regressor. Engle and Yoo (1987) extend those tables to the case of more than one regressor, and MacKinnon (1991) has the most complete tables available so far. You can find the critical values for this residual based test in Hamilton Table B.9.

New tests for unit roots in residuals from a potentially cointegrating relation (like the Phillips-Perron tests) have been suggested since the publication of Engle and Granger (1987) and critical values have been simulated for some of those (see Phillips and Ouliaris (1990) for critical values for the PP test - these values are built into the COINT package), but it seems that the CADF test stands up pretty well. Again, you have to be careful if the series contains trends. If the x_t series contain a trend (or may contain a trend) then you should be careful to include a trend in the cointegrating regression, otherwise the asymptotic critical values will be different. In the case of a one-dimensional x_t, that include a deterministic trend, a regression of y_t on x_t that does not include the trend will give you an asymptotically normal coefficient (this is not too surprising since a deterministic trend always will dominate a unit root trend). Bruce Hansen's article in Journal of Econometrics (Hansen (1992)) treats this topic in more detail. Also note that Campbell and Perron (1991) refer to the case where there is a deterministic

trend in y_t and x_t, but not in $y_t - \theta' x_t$ as "deterministic cointegration." If you include a trend in the CADF test, as I suggested, you use (e.g.) Table IIc in Phillips and Ouliaris (1990). Hamilton p 596-7 suggests (based on Hansen (1992)) a slightly more efficient test where you do not include a time trend in the regression—I find the biggest drawback of this strategy that this test is not invariant to whether there is a trend in the data or not.

1.4.1 Estimation of the parameters in case of cointegration

NOTE: Be aware that the issue of efficient *estimation* of parameters in cointegrating relationships is quite a different issue from the issue of *testing* for cointegration.

Engle and Granger (1987) suggested the following simple two-step estimator (which is not efficient and therefore *not* recommended). First estimate the static cointegrating relationship $y_t = \theta' x_t + e_t$, then define $z_t = y_t - \hat{\theta}' x_t$, and finally estimate the **error correction model**

$$\Delta y_t = A_1(L)\Delta y_{t-1} + A_2(L)\Delta x_{t-1} + \gamma z_{t-1} + e_t \ .$$

The fact that $\hat{\theta}$ is super-consistent implies that the parameters of the lag polynomials have the same (asymptotically normal) distribution as they would have if θ had been known.

1.5 The Johansen ML estimator

The best way of testing for unit roots is by using the system ML estimator of Johansen (1988,1991) is a test for cointegration restrictions in a VAR representation. "Johansen" estimation is treated in much detail in the book by Johansen (1995). This estimator also gives you asymptotically efficient estimates of the cointegrating vectors (the β's) and of the adjustment parameters (the α's).

"Johansen's method" is the maximum likelihood estimator of the so-called reduced rank model. We start with the AR(k) model

$$\Delta y_t \ = \ \mu \ + \ \Gamma_1 \Delta y_{t-1} \ + \ ... \ + \ \Gamma_{k-1} \Delta y_{t-k+1} \ + \ \Pi y_{t-k} \ + \ e_t \ ,$$

which under the assumption of cointegration of order k can be written as

$$\Delta y_t \ = \ \mu \ + \ \Gamma_1 \Delta y_{t-1} \ + \ ... \ + \ \Gamma_{k-1} \Delta y_{t-k+1} \ + \ \alpha\beta' y_{t-k} \ + \ e_t \ , \tag{3}$$

where α and β both have dimension $p \times k$. The number of parameters in the unrestricted model is $p + kp^2 + p(p+1)/2$. Let $Z_{0t} = \Delta y_t$, $Z_{1t} = (\Delta y'_{t-1}, ..., \Delta y'_{t-k+1}, 1)'$ and $Z_{kt} = y_{t-k}$. Define the moment matrices as

$$M_{ij} \ = \ T^{-1} \sum_{t=1}^{T} Z_{it} Z'_{jt} \qquad (i, j = 0, 1, k) \ ,$$

We first regress Z_{it} , $i = 0, k$ on Z_{1t} and get residuals R_{it} , $i = 0, k$. You should think of this as "purging" Δy_t and y_{t-k} of the short-run parameters which can be considered "noise" in the cointegrating relation. We are also purging those variables of their mean, and if you want to include exogenous variables (e.g. dummy variables) they should also be included in the Z_{1t} vector. Denote the residual sum of squares from regressing Z_0 and Z_k on Z_1 as S_{ij} ; $i, j = 0, k$, in other words

$$S_{ij} = \frac{1}{T}\sum_{t=1}^{T} R_{it}R'_{jt} \ .$$

The maximum likelihood estimator of α and β is a function of these residuals. Johansen (1988,1991) shows that $\hat{\beta}$ can be found from choosing the eigenvectors $(\hat{v_1}, ..., \hat{v_r})$, where $\hat{V} = (\hat{v}_1, ..., \hat{v}_p)$ are the eigenvectors of the equation

$$(*) \quad |\lambda S_{kk} - S_{k0}S_{00}^{-1}S_{0k}| \ = \ 0 \ ,$$

normalized such that $\hat{V}'S_{kk}\hat{V} = I$, and ordered in the order of the corresponding eigenvalues such that $\hat{\lambda}_1 > ... > \hat{\lambda}_p > 0$. Make sure you get the intuition here: Cointegration of order r implies that $\lambda_1 \neq 0, ..., \lambda_r \neq 0$, while $\lambda_{r+1} = ... = \lambda_p = 0$. (Since the estimated eigenvalues are continuous random variables they are different from zero (and from each other) with probability 1.) And it is intuitively clear now that you want the eigenvectors corresponding to the non-zero eigenvalues to be the estimators of the cointegrating vectors.

In order to find those eigenvalues pre- and post-multiply the equation above by $S_{kk}^{-1/2}$ (you can use the Cholesky factorization in e.g. GAUSS to get $S_{kk}^{-1/2}$, but the inverse of any matrix X that satisfies $XX' = S_{kk}$ will do) and get the equivalent problem

$$(**) \quad |\lambda - S_{kk}^{-1/2}S_{k0}S_{00}^{-1}S_{0k}S_{kk}^{'-1/2}| \ = \ 0 \ .$$

Note that this is a standard eigenvalue problem that programs like GAUSS can solve directly. The eigenvalues will be the ones that you are looking for. The eigenvectors (u_i, say) that GAUSS gives you will be normalized such that $u_i'u_i = 1$ so you will use $(\hat{v_1}, ..., \hat{v_r}) = S_{kk}^{-1/2}u_1, ..., S_{kk}^{-1/2}u_r$.

In order to give some interpretation of this equation remember that the least squares $\hat{\Pi}$ can be obtained by regressing R_{0t} on R_{kt}, by the Frisch-Waugh theorem. So the least squares estimate of Π is

$$\hat{\Pi} \ = \ S_{kk}^{-1}S_{k0} \ .$$

Now note that

$$S_{kk}^{-1/2}S_{k0}S_{00}^{-1}S_{0k}S_{kk}^{-1/2} \ = \ S_{kk}^{1/2}S_{kk}^{-1}S_{k0}S_{00}^{-1/2}S_{00}^{-1/2}S_{0k}S_{kk}^{-1}S_{kk}^{1/2} \ = \ (S_{kk}^{1/2}\hat{\Pi}S_{00}^{-1/2})(S_{kk}^{1/2}\hat{\Pi}S_{00}^{-1/2})' \ .$$

The intuitively natural approach would be to consider the eigenvalues of $\hat{\Pi}\hat{\Pi}\hat{\Pi}'$ and you can see that this is actually what the Maximum Likelihood algorithm does apart from the fact that $\hat{\Pi}$ has been normalized by pre-multiplying by $S_{kk}^{-1/2}$ and post-multiplying by $S_{00}^{-1/2}$.

The maximized likelihood function is

$$L_{\max}^{-2/T}(r) \;=\; |S_{00}| \prod_{i=1}^{r} (1 - \hat{\lambda}_i) \; .$$

Notice that this is a function of the estimated eigenvalues where all the eigenvalues except the largest r eigenvectors are set equal to zero. So for example the test for one cointegrating vector against no cointegrating vectors consist of testing whether the largest eigenvalue is significantly different from zero. Johansen further finds

$$\hat{\alpha} \;=\; S_{0k}\hat{\beta} \; ,$$

$$\{\hat{\Gamma}_1, ..., \hat{\Gamma}_{k-1}, \hat{\mu}\} \;=\; (M_{01} \;-\; \hat{\alpha}\hat{\beta}' M_{k1}) M_{11}^{-1} \; ,$$

and

$$\hat{\Lambda} \;=\; S_{00} - \hat{\alpha}\hat{\alpha}' \; .$$

The likelihood ratio test statistic H for the hypothesis that $\Pi = \alpha\beta'$ is of rank r against the unrestricted model where Π has full rank p is

$$H \;=\; -2\ln(Q) \;=\; -T \sum_{i=r+1}^{p} \ln(1 - \hat{\lambda}_i) \; .$$

Note that the Null hypothesis here is that there are $(p - r)$ unit roots. This corresponds to the simple residual based test previously, where we have $p = 2$ (if the X variable is one dimensional), and we test for 1 cointegrating relation, the null is then that there are 2 unit roots. This test statistic is often referred to as the "trace"-statistic, see e.g. Johansen and Juselius (1992). Note that this statistic is expected to be close to zero if there is at most r (linearly independent) cointegrating vectors. Another test that is often used it the "$\lambda - max$" test which looks at $-T\ln(1 - \hat{\lambda}_{r+1})$—the idea being that if the $(r + 1)$th eigenvalue can be accepted to be zero, then all the smaller eigenvalues can also. This test is a test of $r + 1$ cointegrating vectors against r cointegrating vectors.

The asymptotic distribution of the likelihood ratio test is a functional of multivariate Brownian motion (Johansen (1991)), and is tabulated for values of p up to 11 in Osterwald-Lenum (1992) and reproduced in Hamilton – Table B. 10. The case that allows for a a deterministic trend in the variables is the one that you will "normally" use – this is denoted "Case 3" in Table B. 10.

Often you do not really want to test whether there is (say) 3 cointegrating vectors against no cointegrating vectors, rather you want to make a *decision* on to what is the *number of cointegrating vectors*. In the situation where you directly want to test $r+1$ cointegrating vectors against r cointegrating vectors you should of course use the "$\lambda - max$" test, but this test will not give you a consistent way of deciding the cointegration rank. A consistent (in the sense that you with probability 1 will not underestimate the number of cointegrating vectors) way to do this, using the trace test, is to *start by testing for zero* cointegrating vectors. (I.e. if your system is 4 dimensional, you compare the test statistic $-T\sum_{i=1}^{4}\ln(1-\hat{\lambda}_i)$ to the row labelled 4 in Hamilton Table B.10). If you reject zero cointegrating vectors, you then test for (at most) 1 cointegrating vectors. (In the 4-dimensional case, you compare the test statistic $-T\sum_{i=2}^{4}\ln(1-\hat{\lambda}_i)$ to the row labelled 3 in Hamilton Table B.10). If this is not rejected you stop and decide that $r=1$ - if you reject this you move on until you can not longer reject and stop there. See Johansen (1992b) for further details.

Even though there is a constant in the error correction representation (eqn. (**??**)), this may not translate into a deterministic trend in y_t. Note that this is not the same as what Campbell and Perron (1992) refer to as "deterministic cointegration", namely the case where there is trend in y_t but no trend in $\beta' y_t$. Johansen (1991) derives the likelihood ratio test (which we will denote H^*) for reduced rank in the case where there is a constant in the ECM but no trend in y_t, see Johansen (1991) or Johansen (1995) for the full explanation. Johansen (1992b) discusses how to obtain a consistent test for the number of stochastic trends and for trend in y_t at the same time. See Johansen (1991) for the derivation of the maximum likelihood estimator when there may or may not be trend. It turns out to be very convenient to program the Maximum Likelihood estimator in this case: all you have to do is to move the vector of ones in to Z_{kt} and delete if from Z_{1t}. (The Johansen (1991) article also has the most readable proof of the Granger representation theorem in my opinion).

There are two drawbacks of the Johansen method. One is that it takes a little getting used to interpreting the results and formulating hypotheses in this setting. In the VAR system all variables are treated symmetrically, as opposed to the standard univariate models that usually have a clear interpretation in terms of exogenous and endogenous variables. The other drawback of the VAR system is one has to model all the variables at the same time, which will be a problem if the relation for some variable is flawed. This may give bias in the whole system and one may have been better of conditioning on that variable. Further, the multidimensional VAR model uses many degrees of freedom

1.6 Hypothesis testing in the Johansen model

This is just a very short introduction - Bannerjee et al. p. 276-277 have a slightly longer (but also much too short for the purpose of really learning it) introduction and Hamilton p. 648-650 has a discussion that is also very brief. The best place to look for more results is the Johansen and Juselius (1992) paper, which explains the method and uses them on an actual economic model, or you may look in Johansen (1995), chapter 7. Here is just an example. Assume that your system consists of 3 variables (y_{1t}, y_{2t}, y_{3t}) and that you have found that there are 2 cointegrating vectors. In other words you have an estimate of β where

$$\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \end{pmatrix} .$$

Assume then that you have a long-run hypothesis of the form $y_{1t} - y_{3t} = 0$ (e.g. the PPP-model has this form). This would be the case if for example $\beta_{11} = -\beta_{13}$ and $\beta_{12} = 0$. Note that the model only identifies the parameters up to scale. But even worse the β matrix is only identified up to rotations so it is not obvious how one can do a Wald test. However a Likelihood Ratio test will give you a standard χ^2-test (if you condition on cointegration of order 2). Under the alternative you have to parameterize β as (H, ψ), where ψ is a 3×1 vector of parameters, and

$$H = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} .$$

Johansen and Juselius explain how one can estimate models with restrictions of this and similar forms (and also how to impose linear restrictions on α). (It is actually surprisingly easy to modify the program that does the "standard" Johansen ML-estimator to impose these restrictions). I will not go further into this here, but notice that the hypoteses have to formulated in the form of linear relationships rather than as explicit null-restrictions. I need to explain how to find the degrees of freedom (df) for such a hypothesis. Theoretically we can normalize the β vector to

$$\begin{pmatrix} 1 & 0 \\ \beta_{21} & 1 \\ \beta_{31} & \beta_{32} \end{pmatrix} ,$$

so there are 3 df in the β matrix. The reason I use the word "theoretically" is that you may have to re-order the rows since we have not imposed the restriction that any particular β_{ij} is equal to zero. (You may have a model that implies such a restriction, but the general ML estimator does not do that and Johansen himself is quite critical against imposing such a restriction *a*

priori). In the restricted model there are 2 df (since you can normalize one of the coefficients (but again you may not know which one *a priori*) to be 1. So the "PPP" hypothesis that one of the cointegration vectors is $(1, 0, -1)$ can then be tested against the $\chi^2(1)$ distribution. The null hypothesis is estimated by modifying the ML-algorithm for the unrestricted model - I will not go into the details, see Johansen and Juselius (1992), but it is not exceedingly hard to do.

1.6.1 "Common Trends"

Following Johansen (1988,1991) one can choose a set of vectors $\beta_\perp$ such that the matrix $\{\beta, \beta_\perp\}$ has full rank and $\beta'\beta_\perp = 0$. In other words the $p \times (p-r)$ matrix $\beta_\perp$ is orthogonal to the matrix β and the columns of $\beta_\perp$ are orthogonal to the columns of β. The vectors $\beta'_\perp y_t$ constitute the noncointegrated part of the time series y_t. We call $\beta_\perp$ the *common trends loading matrix* and will refer to the space spanned by $\beta'_\perp y_t$ as the *unit root space* of the series y_t. See also Stock and Watson (1988).

1.7 Multicointegration

"Multicointegration" denotes the case where there is cointegration between processes of different order of integration, f.ex. where z_t is I(2), x_t is I(1), and there exist γ such that $z_t - \gamma x_t$ is I(0). We will not have time to go into the theory of multicointegration, but you should now it exists and where to look in the literature. Some models imply a relation between I(2) and I(1) variables, and by extending the ideas of cointegration for I(1) variables above, these models can be analyzed. See Granger and Lee (1990) for a simple introduction, and Johansen (1992c) and Johansen (1992d) for a comprehensive treatment. Also see Stock and Watson (1992.

1.8 Asymptotically efficient single equation methods

The simple two-step estimator of Engle and Granger is not asymptotically efficient; but recently several asymptotically efficient single equation methods have been proposed. Phillips (1991b) suggests a regression in the spectral domain, Phillips and Loretan (1991) suggests a non-linear error-correction estimation, Phillips and Hansen (1990) suggests an instrumental regression with a correction a la Phillips-Perron, Saikkonen (1991) suggests a simple trick of including leads as well as lags in the lag-polynomials of the error correction model in order to achieve asymptotic efficiency, Saikkonen (1992) suggests a simple GLS type estimator, whereas Park's (1991) CCR estimator transforms the data so that OLS afterwards gives asymptotically efficient estimators, and finally Engle and Yoo (1991) suggest a 3 step estimator starting from the static Engle-Granger estimation. From all of those estimators it is possible to obtain simple t-values

for the short term adjustment parameters.
The routines for implementing these single equation procedures are all available in the COINT package for GAUSS, except Engle and Yoo (1991). NOTE, however, that it is not obvious what you are estimating if the system contains more than one cointegrating relation. This makes a very strong case for using the Johansen test in the case of a higher dimensional system, where you rarely can rule out the possibility of more than one conintegrating vector *a priori*.