

Testing for Unit Roots and Cointegration

Bo Sjö

Aug 2008

This guide explains the use of the DF, ADF and DW tests of unit roots. For co-integration the Engle and Granger's two step procedure, the CRDW test, the error correction test, the dynamic equation approach and finally Johansen's multivariate VAR approach. The paper also gives a brief practical guide to the formulation and estimation of the Johansen test procedure, including the I(2) analysis.

Contents

| | | |
|----------|--|-----------|
| 1 | Unit Root Tests: Determining the order of integration. | 2 |
| 1.1 | Why test for integration in the first place? | 2 |
| 1.2 | The DW Test. | 3 |
| 1.3 | Augmented Dickey-Fuller test | 3 |
| 1.3.1 | Restricted ADF Model: $\Delta x_t = \pi x_{t-1} + \varepsilon_t$ | 4 |
| 1.3.2 | ADF model: $\Delta x_t = \alpha + \pi x_{t-1} + \varepsilon_t$ | 5 |
| 1.4 | General ADF model: $\Delta x_t = \alpha + \beta t + \pi x_{t-1} + \varepsilon_t$ | 5 |
| 1.5 | Alternatives to the Dickey-Fuller Test | 6 |
| 2 | Dealing with I(2) variables, seasonals and outliers | 6 |
| 3 | Testing the significance of the constant and/or the trend. | 7 |
| 3.1 | Testing the constant in $\Delta x_t = \alpha + \pi x_{t-1} + \varepsilon_t$ | 7 |
| 3.2 | Testing the significance of α in $\Delta x_t = \alpha + \beta(t - \bar{t}) + \pi x_{t-1} + \varepsilon_t$ | 8 |
| 3.3 | The significance of β in $\Delta x_t = \alpha + \beta(t - \bar{t}) + \pi x_{t-1} + \varepsilon_t$ | 8 |
| 4 | Testing Cointegration | 9 |
| 4.1 | The cointegrating regression Durbin-Watson statistic | 9 |
| 4.2 | Engle and Granger's two-step procedure | 9 |
| 4.3 | Tables for cointegration tests in order of appearance: | 11 |
| 4.4 | MacKinnon critical values. | 12 |
| 4.5 | Test for cointegration in an ADL model | 12 |
| 4.6 | Testing cointegration in a single equation ECM | 12 |
| 5 | The significance of a deterministic time trend in unit root tests. | 13 |
| 6 | The Johansen test of Cointegration | 13 |
| 6.1 | Testing for different trends in the VECM | 18 |
| 6.2 | Testing for I(2) in the VECM model | 19 |
| 6.3 | Testing the vector | 20 |
| 7 | References | 21 |

1 Unit Root Tests: Determining the order of integration.

This guide presents some simple tests for unit roots and cointegration. There is a large number of tests in these areas. This guide deals with the some basic tests; the Dickey-Fuller test and Johansen's test.¹ The reason for starting with the Dickey-Fuller test is that there is no uniformly better test. This should not be understood as a motive for not performing other types of unit root tests. tests. Comparing different results from different test methods is a good way of testing the sensitivity of your conclusions. Once you understand how this test work, and its limitations, you will understand the use of other tests. The advantage with the DF test is that, once you master it you will understand the meaning and purpose of other tests. the Johansen test is central for testing cointegration. However, it relies to a large extent on asymptotic properties. If applied to small samples the results might be difficult to understand. Again, understanding the limitations of a test often leads to a better modelling decisions in the end.

1.1 Why test for integration in the first place?

Testing for the order of integration is standard in applied econometric work. The way a test is performed in applied work depends on the motive behind the test. We can find two motives behind unit root tests. The first is knowing the order of integration is crucial for setting up an econometric model and do inference. The second motive is that economic theory suggests that certain variables should be integrated, a random walk or a martingale process. In this situation, it is motivated to perform very detailed tests, and take great care in finding exact critical values. The unit root test is motivated by theory, it will be one test in combination with other tests.²

The second, and the most common, motive is to investigate the properties of the prior to the construction of an econometric model. In this case, unit root tests are mainly a descriptive tool performed to classify series as stationary and non-stationary. Since integrated variables lead to non-standard distributions and perhaps spurious regression results, the recommendation is the following; If a data series appear to be non-stationary, assume as the maintained hypothesis, that it is non-stationary and integrated. Reject this hypothesis only, and only if, there is clear evidence for rejection. Once you have been able to classify your variables as integrated, stationary or perhaps deterministic trend stationary, you are in position to sort out long-run and short-run effects in your model, and to set up a model where statistical inference will be meaningful.

How you want to set up a test for a unit root depends on the purpose of the study. In a finite sample it is in practice impossible to construct a test that, in a finite sample, can distinguish between a unit root, and a root close to unity, say 0.999. The latter root is stationary, the former is not. In the long run these processes will have very different behavior. The intuition behind this is that the distribution of a unit root process must take place in continuous time, where it will difficult to distinguish between a value of unit and a value very close to unit, since there is always another decimal to consider. No test of integration, no matter how complex or famous, is a substitute for the use of graphics in combination with common sense.

¹See Sjöo (1998) for some alternative test methods.

²In this context se also the section concering aggregation in Sjöo(1998).

The most common test for testing $I(1)$ versus $I(0)$ is the Dickey-Fuller test. This test has as the null that the series is $I(1)$, which in general might be hard to reject. An alternative is the KPSS test which has the null of $I(0)$. Also the Johansen test has $I(0)$ as the null when used to test the order of integration. The reason that we focus on DF test is that it is simple and there is no uniformly better alternative.

1.2 The DW Test.

A simple, and unreliable, test for $I(1)$ is the DW-test. To understand the test, recall that the DW-value is calculated as $DW = 2(1 - \hat{\rho})$, where $\rho = \hat{\rho}$ is the estimated first order autocorrelation. Thus, if x_t is a random walk ρ will equal unity and the DW value is zero. Under the null that x_t is a random walk, the DW statistic calculated from the first order autocorrelation of the series $x_t = x_{t-1} + \varepsilon_t$, will approach unity. As $T \rightarrow \infty$, the DW value approaches 0 under the null of a random walk. A DW value significantly different from zero rejects the hypothesis that x_t is a random walk and $I(1)$, in favor of the alternative that x_t is not $I(1)$, and perhaps $I(0)$. The test is limited by the assumption that x_t is a random walk variable. The test is not good for integrated variables in general. The critical value at the 5% level for the maintained hypothesis of $I(1)$ versus $I(0)$ is 0.17. A higher value rejects $I(1)$.

1.3 Augmented Dickey-Fuller test

A basic test for the order of integration is the Dickey-Fuller test. Assume that x_t is random walk process, $x_t = x_{t-1} + \varepsilon_t$, then the regression model becomes $x_{t-1} = \rho x_{t-1} + \varepsilon_t$. Subtract x_{t-1} from both sides of the equation,

$$\Delta x_t = \pi x_{t-1} + \varepsilon_t, \tag{1}$$

where $\pi = (1 - \rho)$. In this model we know, under the null, that $\hat{\pi}$ is biased downwards.³ Because of this, the significance of π is tested as a one-sided "t-test". The problem is that the test statistics associated with $\hat{\pi}$ is non-standard. If x_t is a stationary variable $\hat{\pi}$ would asymptotically follow a normal distribution, and standard tests would be possible. It can be shown that if x_t is a random walk, that the distribution of $\hat{\pi}$ is skewed under the null. Dickey and Fuller simulated the correct test statistics for $H_0 : \hat{\pi} = 0$, under the assumption of a random walk process. Instead of using standard t-tables, to perform the "t-test", we use the non-standard Dickey-Fuller distribution instead. (Notice that the Dickey-Fuller distribution changes depending on how the test equation is set up.)

The Dickey-Fuller test can be set up in three ways, depending on what we want the alternative hypothesis to be. The null is always that x_t is a random walk without drift. The alternative can be that x_t is stationary $I(0)$, or that x_t is driven by a deterministic trend (t), alternatively that x_t is driven by a deterministic quadratic trend (t^2). For each model the empirical distribution of $t_{\hat{\beta}}$ is different, and tabulated in separate tables. The three tests are presented

³See Hendry (1995) page 110 for a discussion. Notice that Dickey originally developed F-tests for testing the joint significance of π and constants. These tests does not offer any advantage over the t-tests presented here, due to the downward bias that motivates one sided testing.

below. The empirical distribution is always simulated under the assumption a random walk with white noise residuals, $\varepsilon \sim iid(0, \sigma^2)$. In general, this is not the case, Δx_t is likely to be have an ARMA representation. In this situation the autoregressive structure can be dealt with by augmenting the regression equation with lagged Δx_t variables, such that $\hat{\varepsilon}_t$ in the regression model becomes white noise and the Dickey-Fuller distributions are valid. If Δx_t contains a moving average process the situation is more complex. The augmentation is now at best viewed as an approximation. A solution is offered by Phillips and Perron (1988), who finds a non-parametric way of adjusting the estimated variance so that the tabulated distribution is valid.

If x_t contains seasonal deterministic factors they can be added to the model as seasonal dummies without affecting the asymptotic distribution of the test statistics.

Impulse dummies that remove extreme values can be added to the models without affecting the asymptotic distribution of the test statistics.

Shifts in a deterministic trend components of x_t affects the distribution, however. Step dummies that capture changes in deterministic growth cannot be included in the model without changing the distribution.⁴⁵ Furthermore, the test is sensitive to large negative MA processes in the series. An alternative, might be that the variable is near integrated, with an MA(1) process close to minus one. In the following we look at the three ways of setting up an Augmented Dickey-Fuller (ADF) test.

1.3.1 Restricted ADF Model: $\Delta x_t = \pi x_{t-1} + \varepsilon_t$

The Dickey-Fuller "t-statistics" for the significance of $\hat{\pi}$ is based on the estimated Dickey-Fuller (DF) model

$$\Delta x_t = \pi x_{t-1} + \varepsilon_t. \tag{2}$$

Alternatively, in the case of autocorrelation in the observed series, estimate the Augmented Dickey-Fuller (ADF) model:

$$\Delta x_t = \pi x_{t-1} + \sum_{i=1}^k \gamma_i \Delta x_{t-i} + \varepsilon_t. \tag{3}$$

The null hypothesis is that $x_t = x_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim NID(0, \sigma^2)$. The augmentation ($k > 0$) will not affect the asymptotic distribution of the test statistic. The test can also be performed on variables in first differences as a test for $I(2)$ -ness.⁶ Under the null $\hat{\pi}$ will be negatively biased in a limited sample, thus unless x_t is explosive, only a one sided test is necessary for determining $H_0 : \pi = 0$ ($x_t \sim I(1)$) against $H_a : \pi < 0$ ($x_t \sim I(0)$). This a very restricted

⁴The alternative hypothesis of segmented deterministic trends is discussed by Perron (1989), see also Enders (1994). A test procedure is offered by Banerjee, Lumsdaine and Stock (1992).

⁵Notice that if you on a priori grounds identify segmented trends (shifts in the constant) and introduces these into the model, the outcome is that the empirical distribution of t_β is driven from the Dickey-Fuller distribution towards the normal distribution. Thus, if you continue testing for unit root using the Dickey-Fuller test you will not over-reject the null of having an integrated variable, and you can avoid spurious regression results in a following regression model.

⁶These comments also applies to the ADF models with constants, trend, and seasonals.

model, there is alternative hypothesis for explaining an observed non-stationary process. A significant positive value implies an explosive process, which can be a very difficult alternative hypothesis to handle.

The critical values are tabulated in Fuller (1976), p. 373, Table 8.5.2 upper part. Also in Benerjee et al (1993) p. 103, Table 4. upper part. The critical value at 5% and 25 observations is -1.95.

1.3.2 ADF model: $\Delta x_t = \alpha + \pi x_{t-1} + \varepsilon_t$

The Dickey-Fuller "t-statistics" for the significance of ρ is based on the estimated model

$$\Delta x_t = \alpha + \pi x_{t-1} + \varepsilon_t \tag{4}$$

Alternatively, in the case of autocorrelation in the observed series, estimate the augmented Dickey-Fuller model:

$$\Delta x_t = \alpha + \pi x_{t-1} + \sum_{i=1}^k \gamma_i \Delta x_{t-i} + \varepsilon_t. \tag{5}$$

The null hypothesis is that $x_t = x_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim NID(0, \sigma^2)$. Under the null $\hat{\pi}$ will be negatively biased in a limited sample, thus only a one sided test is necessary for determining $H_0 : \pi = 0$ ($x_t \sim I(1)$) against $H_a : \pi < 0$. ($x_t \sim I(0)$). This model is less restricted, because it allows a deterministic trend as $x_t = \alpha t + \pi x_{t-1} + \varepsilon_t$. The critical values are tabulated in Fuller (1976), p. 373, Table 8.5.2 middle part. Also in Benerjee et al (1993) p. 103, Table 4. middle part. The critical value at 5% and 25 observations is -3.00.

1.4 General ADF model: $\Delta x_t = \alpha + \beta t + \pi x_{t-1} + \varepsilon_t$

The Dickey-Fuller "t-statistics" for the significance of ρ is based on the estimated model

$$\Delta x_t = \alpha + \beta t + \pi x_{t-1} + \varepsilon_t \tag{6}$$

Alternatively, in the case of autocorrelation in the observed series, estimate the augmented Dickey-Fuller model:

$$\Delta x_t = \alpha + \beta t + \pi x_{t-1} + \sum_{i=1}^k \gamma_i \Delta x_{t-i} + \varepsilon_t. \tag{7}$$

The null hypothesis is that $x_t = x_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim NID(0, \sigma^2)$. Under the null $\hat{\pi}$ will be negatively biased in a limited sample, thus only a one sided test is necessary for determining $H_0 : \pi = 0$ [$x_t \sim I(1)$] against $H_a : \pi < 0$ [$x_t \sim I(0)$]. This model is less restricted, because it allows a deterministic trend as $x_t = \alpha t + \beta t^2 + \pi x_{t-1} + \varepsilon_t$. The critical values are tabulated in Fuller (1976), p. 373, Table 8.5.2 lower part. Also in Benerjee et. al. (1993) p. 103, Table 4. lower part. The critical value at 5% and 25 observations is -3.60.

It is a good strategy to start with the model containing both a constant and a trend, because this model is the least restricted. If a unit root is rejected here, due to a significant $\hat{\pi}$, there is no need to continue testing. If $\hat{\pi} = 0$ cannot be rejected, the improve efficiency in a model without a time trend might be better. But, there is also the augmentation to consider.

1.5 Alternatives to the Dickey-Fuller Test

The Dickey-Fuller tests are simulated on the assumption that the alternative is a random walk, with or without drift terms, and that the residual process is white noise. The test is quite sensitive to the presence of a negative MA(1) process (-1). The KPSS test has as the null that the variable is stationary, $I(1)$. The DF-test has as the null that the variable is integrated. The KPSS test is perhaps better, if there is a priori knowledge suggesting $I(0)$ as a reasonable maintained hypothesis. The Perron test has $I(1)$ as the maintained hypothesis, like the ADF-test, but allows for segmented deterministic trends in the alternative. The alternative in the ADF-test allows only for deterministic trends, linear or quadratic, over the sample period. Other tests in the literature, like the Phillips -Perron test, tries to find ways of dealing with deviations for having white noise in the estimated model. These tests aims at either adjusting the estimated tests statistic so that it "fits in" with the simulated Dickey-Fuller values better that by augmenting the Dickey-Fuller model with lagged dependent variables, or they try to adjust the test statistics and present new (simulated) critical values. A careful researcher, who wants to make a strong point about the order of integration should perform many tests before deciding on the order of integration.

Again, in a finite sample it is impossible to distinguish between an integrated variable and a "near-integrated variable", that is a variable close to a unit root process. For a finite sample it is always possible to fit a unit root process, or a stationary process such that it will not be possible to distinguish these two from each other. So, the question to be asked is what is the best way to describe the non-stationary process that you are about to model, and perform statistical inference on? What is the best approximation to the true underlying distributions of the variables and of the test statistics? In general, the answer is assume a stochastic trend.

2 Dealing with I(2) variables, seasonals and outliers

If the data contains seasonal variation, seasonality can either be removed through X12 without affecting the unit root property, or seasonal dummies can be included. Dummies of impulse type has no affect on the asymptotic distribution of the test statistic. The same holds for outliers.

A variable might contain unit roots as various seasonal frequencies. These tests are not discussed here.

$I(2)$ processes can be tested with the models above, x_t is replaced with Δx_t .

$$\Delta^2 x_t = \alpha + \pi \Delta x_{t-1} + \sum_{i=1}^k \gamma_i \Delta^2 x_{t-i} + \varepsilon_t. \tag{8}$$

When testing for $I(2)$ a trend term is not a plausible alternative. The two interesting models here are the ones with and without a constant term. Furthermore, lag length in the augmentation can also be assumed to be shorter.

The "Pantula Principle" is important here, namely that since higher order integration dominates lower order integration, all tests of integration of order d are always performed under the assumption that the variable is not integrated of order $d + 1$.

TESTING FOR UNIT ROOTS AND COINTEGRATION 7

In practice this means that if $I(2)$ against the alternative of $I(1)$ is tested and not rejected, it makes no sense of testing for $I(1)$ against $I(0)$, because this test might be totally misleading. In fact, it is not uncommon to find that $I(1)$ is rejected, when an $I(2)$ variable is tested for $I(1)$ against $I(0)$, leading to the wrong assumption of stationarity. In this situation a graph of the variable will tell you that something is wrong.

Remark 1 *In this memo we do not discuss seasonal unit roots. This does not mean that they do not exist, only that they might not be very common.*

Remark 2 *There are a number of alternative tests for unit roots. They are all good at something, but without a priori information there is no uniformly better unit root test, see Sjöö (2000) for a survey.*

Remark 3 *How much time should you spend on unit root tests? It depends on the nature of the problem, and whether there are some interesting economic results that depends on whether data is $I(0)$ or $I(1)$. In general unit root tests are performed to confirm that the data series are likely $I(1)$ variables, so one has to be aware of spurious regression and long-run cointegrating relationships. In that case $I(1)$ or possibly $I(2)$ is only rejected if rejection can be done clearly, otherwise the null of integration is maintained.*

Remark 4 *Remember that there is number of alternative hypothesis if $I(1)$ is rejected, Sjöö (2000) provides a list.*

Remark 5 *The most critical factor in the ADF test is to find the correct augmentation, lag length of Δx_{t-k} . In general the test performs well if the true value of k is known. In practice it has to be estimated, and the outcome of the test might change depending on the choice of k . It might therefore be necessary to show the reader how sensitive the conclusions are to different choices of k . A "rule-of-thumb" is to, counting from lag $k = 1$, to include all significant lags plus one additional lag.*

3 Testing the significance of the constant and/or the trend.

3.1 Testing the constant in $\Delta x_t = \alpha + \pi x_{t-1} + \varepsilon_t$

In the Dickey-Fuller regression, not only the slope parameter of x_{t-1} has a difficult distribution, the constant and the parameter associated with the trend term also follow non-standard distributions. As long as we cannot reject the null that x_t is a random walk the distribution of other parameters in the model are also affected. The parameters of the augmentation is not affected. Under the null, Δx_t is stationary, it follows that x_{t-1} etc. are also stationary and the test statistics as asymptotically standard. It is possible to use the t-statistics to determine the significance of the lags. The "t-statistics" for the significance of α in the following estimated model is non-standard as long as $\pi = 0$ is not rejected,

$$\Delta x_t = \alpha + \pi x_{t-1} + \varepsilon_t. \quad (9)$$

Alternatively, in the case of autocorrelation in the observed series, estimate the augmented Dickey-Fuller model:

$$\Delta x_t = \alpha + \pi x_{t-1} + \sum_{i=1}^k \gamma_i \Delta x_{t-i} + \varepsilon_t. \quad (10)$$

The null hypothesis is that $x_t = x_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim NID(0, \sigma^2)$. Under this null, $\hat{\alpha}$ has a non-standard, but symmetric distribution. The critical values are tabulated in Dickey and Fuller (1979), p. 1062, Table I. The 5% critical value for the significance of α in a sample of 25 observation is 2.97. Remember that once the null hypothesis of a unit root is rejected have α and β asymptotic standard normal distributions. This is also true for the models below. (Extra: Haldrup (1992) shows that the mean correction for the trend in the Dickey-Fuller tables above is not correct. The significance test for α should be done in a model where the trend is not corrected for its mean. The table gives the empirical distribution of the "t-test" for α , in a model where the trend is not mean corrected. The null hypothesis is that $x_t = x_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim NID(0, \sigma^2)$. Under the null $\hat{\alpha}$ has a non-standard, but symmetric distribution. The critical values are tabulated in Haldrup (1992, p. 19, Table 1. The 5% critical value for the significance of α in a sample of 25 observation is 3.33.)

3.2 Testing the significance of α in $\Delta x_t = \alpha + \beta(t - \bar{t}) + \pi x_{t-1} + \varepsilon_t$.

The "t-statistics" for the significance of α in the following estimated model is also non-standard,

$$\Delta x_t = \alpha + \beta(t - \bar{t}) + \pi x_{t-1} + \varepsilon_t, \quad (11)$$

where $(t - \bar{t})$ indicate that the time trend is adjusted for its mean.

Alternatively, in the case of autocorrelation in the observed series, estimate the augmented Dickey-Fuller model:

$$\Delta x_t = \alpha + \beta(t - \bar{t}) + \pi x_{t-1} + \sum_{i=1}^k \gamma_i \Delta x_{t-i} + \varepsilon_t. \quad (12)$$

The null hypothesis is that $x_t = x_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim NID(0, \sigma^2)$. Under the null, $\hat{\alpha}$ has a non-standard, but symmetric distribution. The critical values are tabulated in Dickey and Fuller (1979), p. 1062, Table II. The 5% critical value for the significance of α in a sample of 25 observation is 3.59.

3.3 The significance of β in $\Delta x_t = \alpha + \beta(t - \bar{t}) + \pi x_{t-1} + \varepsilon_t$.

The "t-statistics" for the significance of β in the following model is non-standard, under the null of x_t is a random walk,

$$\Delta x_t = \alpha + \beta(t - \bar{t}) + \pi x_{t-1} + \varepsilon_t, \quad (13)$$

where $(t - \bar{t})$ indicate that the time trend is adjusted for its mean.

Alternatively, in the case of autocorrelation in the observed series, estimate the augmented Dickey-Fuller model:

$$\Delta x_t = \alpha + \beta(t - \bar{t}) + \pi x_{t-1} + \sum_{i=1}^k \gamma_i \Delta x_{t-i} + \varepsilon_t. \quad (14)$$

The null hypothesis is that $x_t = x_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim NID(0, \sigma^2)$. Under the null $\bar{\beta}$ has a non-standard, but symmetric distribution. The critical values are tabulated in Dickey and Fuller (1979), p. 1062, Table III. Thus, the 5% critical value for the significance of β in a sample of 25 observation is 3.25.

4 Testing Cointegration

Once variable have been classified as integrated of order $I(0)$, $I(1)$, $I(2)$ etc. is possible to set up models that lead to stationary relations among the variables, and where standard inference is possible. The necessary criteria for stationarity among non-stationary variables is called cointegration. Testing for cointegration is necessary step to check if your modelling empirically meaningful relationships. If variables have different trends processes, they cannot stay in fixed long-run relation to each other, implying that you cannot model the long-run, and there is usually no valid base for inference based on standard distributions. If you do not find cointegration it is necessary to continue to work with variables in differences instead.

4.1 The cointegrating regression Durbin-Watson statistic

The DW test statistic can be used as a quick test of cointegration. Estimate the assumed cointegrating equation

$$x_{1,t} = \beta_1 + \beta_2 x_{2,t} + \dots + \beta_p x_{p,t} + u_t \quad (15)$$

and calculate the Durbin-Watson test statistics for first order autocorrelation. Under the null hypothesis that $x_{1,t}$ is a random walk and that $\beta_2 = \dots = \beta_p = 0$, so there is no cointegration, and \hat{u}_t becomes a random walk with theoretical first order autocorrelation equal to unity. Under the null of no cointegration, the DW value will not be significantly different from zero. Therefore, a Cointegrating Regression Durbin-Watson (CRDW) test statistic different from zero implies cointegration. The critical values for the CRDW test must be simulated for different sample sizes and number of regressor in the model, under the null that the variables are independent random walks. This test suffers from two major problems, it is extremely sensitive to the assumption of $x_{1,t}$ being a true random walk and, the critical values of the test statistic are not consistent as the number of regressors (p) increases over the sample size. The practical use of this test is therefore extremely limited.

4.2 Engle and Granger's two-step procedure

There are several tests of cointegration. The Johansen test is the most fundamental test. Engle and Granger (1987) formulated one of the first test of cointegration (or common stochastic trends). This test has the advantage that it is intuitive, easy to perform and once you master it you will also realize it

TESTING FOR UNIT ROOTS AND COINTEGRATION 10

limitations and why there are other tests. The intuition behind the test motivates its role as the first cointegration test to learn. Start by estimating the so-called co-integrating regression (the first step),

$$x_{1,t} = \beta_1 + \beta_2 x_{2,t} + \dots + \beta_p x_{p,t} + u_t \quad (16)$$

where p is the number of variables in the equation. In this regression we assume that all variables are $I(1)$ and might cointegrate to form a stationary relationship, and thus a stationary residual term $\hat{u}_t = x_{1,t} - \beta_1 - \beta_2 x_{2,t} - \dots - \beta_p x_{p,t}$ (In the tabulated critical values $p = n$). This equation represents the assumed economically meaningful (or understandable) steady state or equilibrium relationship among the variables. If the variables are cointegrating, they will share a common trend and form a stationary relationship in the long run. Furthermore, under cointegration, due to the properties of super convergence, the estimated parameters can be viewed as correct estimates of the long-run steady state parameters, and the residual (lagged once) can be used as an error correction term in an error correction model. (Observe that the estimated standard errors from this model are generally useless when the variables are integrated. Thus, no inference using standard distribution is possible. Do not print the standard errors or the t-statistics from this model)

The second step, in Engle and Granger's two-step procedure, is to test for a unit root in the residual process of the cointegrating regression above. For this purpose set up a ADF test like,

$$\Delta \hat{u}_t = \alpha + \pi \hat{u}_{t-1} + \sum_{i=1}^k \gamma_i \Delta \hat{u}_{t-i} + v_t, \quad (17)$$

where the constant term α (in most cases) can be left out to improve the efficiency of the estimate. Under the null of no cointegration, the estimated residual is $I(1)$ because $x_{1,t}$ is $I(1)$, and all parameters are zero in the long run. Finding the lag length so the residual process becomes with the noise is extremely important. The empirical t-distribution is not identical to the Dickey-Fuller, though the tests are similar. The reason is that the unit root test is now applied to a derived variable, the estimated residual from a regression. Thus, new critical values must be tabulated through simulation.

The maintained hypothesis is no cointegration. Thus, finding a significant π implies co-integration. The alternative hypothesis is that the equation is a cointegrating equation, meaning that the integrated variable $x_{1,t}$ cointegrates at least with one of the variables on the right hand side.

If the dependent variable is integrated with $d > 0$, and at least one regressor is also integrated of the same order, cointegration leads to a stationary $I(0)$ residual. But, the test does not tell us if $x_{1,t}$ is cointegrating with all, some or only one of the variables on the right hand side. Lack of cointegration means that the residual has the same stochastic trend as the dependent variable. The integrated properties of the dependent variable will if there is no cointegration pass through the equation to the residual. The test statistics for $H_0 : \pi = 0$ (no co-integration) against $H_a : \pi < 0$ (co-integration), changes with the number of variables in the co-integrating equation, and in a limited sample also with the number of lags in the augmentation ($k > 0$).

Asymptotically, the test is independent of which variable occurs on the left hand side of the cointegrating regression. By choosing one variable on the left

hand side the cointegrating vector are said to be normalized around that variable, implicitly you are assuming that the normalization corresponds to some long-run economic meaningful relationship. But, this is not always correct in limited samples, there are evidence that normalization matters (Ng and Perron 1995). If the variables in the cointegrating vectors have large differences in variances, some might be near integrated (having a large negative MA(1) component), such factors might affect the outcome of the cointegration test. The most important thing, however, is that the normalization make an economic sense, since the economic interpretation is what counts in the end. If testing in different ways give different conclusions you are perhaps free to argue for or against cointegration.

There are three main problems with the two-step procedure. First, since the tests involves an ADF test in the second step, all problems of ADF tests are valid here as well, especially choosing the number of lags in the augmentation is a critical factor. Second, the test is based on the assumption of one cointegrating vector, captured by the cointegrating regression. Thus, care must be taking when applying the test to models with more that two variables. If two variables cointegrate adding a third integrated variable to the model will not change the outcome of the test. If the third variable do not belong in the cointegrating vector, OLS estimation will simply put its parameter to zero, leaving the error process unchanged. Logical chains of bi-variate testing is often necessary (or sufficient) to get around this problem.

Third, the test assumes a common factor in the dynamics of the system. To see why this is so, rewrite the simplest two-variable version of the test as,

$$\Delta u_t = \pi u_{t-1} + v_t. = \tag{18}$$

$$[x_{1,t} - \beta_2 x_{2,t}] = [x_{1,t-1} - \beta_2 x_{2,t-1}] + v_t. \tag{19}$$

If this common factor restriction does not hold, the test can be expected to perform badly.

The advantage of the procedure is that it is easy, and therefore relatively costless to apply compared with other approaches. Especially for two variables it can work quite well, but remember that the common factor restriction is a severe restriction since all short-run dynamics is forced to the residual process. In this respect the dynamic test advocated by Hendry and others might be expected to perform better.

Remark 6 *There is also a three step procedure for testing cointegration, building on the two steps presented here. It is not that much better though, instead look for alternatives below, preferable the Johansen approach.*

4.3 Tables for cointegration tests in order of appearance:

- Dickey-Fuller test of co-integration ADF(0), from Engle and Yoo (1987). Table 2, p. 157.
- Augmented Dickey-Fuller test of co-integration ADF(4), from Engle and Yoo (1987), Table 3, p. 158.
- CRDW, ADF(1) and ADF(4), from Banerjee et al (1993), Table 7.1 p. 209.

4.4 MacKinnon critical values.

Mackinnon (1991, p.272-275) presents response surface estimates of the critical values of Dickey-Fuller test statistics [ADF(0)] for unit roots and co-integration for all sample sizes. Thus $N=1$ corresponds to unit root test, $N>1$ to co-integration test. For any number of observation (T) the critical value is given by,

$$\beta_{\infty} + \beta_1/T + \beta_2/T^2 \tag{20}$$

Select the appropriate number of variables, the critical value and put in the β -values in the formula to get the critical value for the sample size T .

4.5 Test for cointegration in an ADL model

In the cointegrating regression discussed above, the static estimation implies that all short-run dynamics is pushed into the residual process. An alternative is to estimate the autoregressive distributed lag model,

$$A(L)y_t = B(L)x_t + u_t, \tag{21}$$

and choose the lag length so that the residuals are white noise. The model is in levels, but can easily be rewritten as an error correction model. $A(1)^{-1}B(1)$ represents the long-run steady state solution of the model, and is therefore the error correction mechanism. To have cointegration the polynomial $A(L)$ should not contain any unit root. Otherwise the model will not converge to a long-run steady state solution. This unit root test has its own empirical distribution. PCGIVE calculates the statistics, under the long-run solution of the model, and indicates significance. A significant value rejects no cointegration. The critical values are given in Banerjee, Dolado, and Mestre (1992). This test is not recommended, since you cannot directly discuss the economic meaning of your long-run (cointegrating) solution. The single equation ECM is better alternative, perhaps in combination with Engle and Granger's two step procedure.

4.6 Testing cointegration in a single equation ECM

A quite pragmatic approach is to use a priori information to estimate an error correction mechanisms first and then include it in an equation of variables in first differences. Theory might suggest a long-run equilibrium relationship, this can be estimated either in a static or through an ADL model, and the error term ($\hat{\varepsilon}_{t-1}$) can be used as an error correction mechanism, in the equation

$$\Delta x_{1,t} = \alpha + \sum_{i=1}^k \gamma_{1,i} \Delta x_{1,t-i} + \dots + \sum_{i=1}^k \gamma_{n,i} \Delta x_{n,t-1} + \lambda \hat{\varepsilon}_{t-1} + u_t,$$

assuming n variables in the ECM, the lag length in the model is chosen such that u_t is NID(0, σ^2) residual term. In this equation the error correction term is the only potentially I(1) variable, assuming that there are no I(2) variables. The t-statistics of λ will under these assumptions be closer to the normal distribution than the Dickey-Fuller. Thus, the normal distribution, perhaps at 1% risk level might be an approximation. Furthermore, since the remaining variables are all I(0) their distribution are not affected by ($\hat{\varepsilon}_{t-1}$) being I(1), so no harm in terms of spurious results is done by including the ECM term.⁷ This pragmatic

⁷See Banerjee et.al. (1993) for the details.

approach can be motivated by a strong a priori assumption about the long-run relationship, and that it is better to have some representation of the long-run adjustment in the model than no long-run adjustment, in particular since it causes no harm. As more observations become available, the future will tell us if we are on the right track.

5 The significance of a deterministic time trend in unit root tests.

In the random walk model with a drift (or for any integrated variable) a significant constant (=time trend) implies that the asymptotic distribution of the variable is normal. Thus, if we find a significant drift term there is no risk of spurious regression result in the long run, since the variable has a standard normal distribution in the long-run. The deterministic trend will asymptotically dominate the stochastic trend as $t \rightarrow \infty$. In a limited sample, however, it is not sufficient for the deterministic trend to be significant, it must also be sufficiently large to dominate the stochastic trend to motivate the use of standard test statistics.

If x_t is non-stationary, $E\{\Delta \log x_t\} \neq 0$, indicates a constant deterministic growth over time. To test if the drift term is sufficiently large to dominate over the stochastic trend in a limited sample, regress $\Delta x_t = \alpha + v_t$. Calculate the standard error of v_t , use this estimated standard error to standardize the estimated $\hat{\alpha}$. The standardized $\hat{\alpha}$ can then be used as an estimate μ the constant term in a Dickey-Fuller test where the DGP for x_t is $x_t = \mu + x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim NID(0,1)$. For high values of μ in combination with high values of T , will the "t-value" of ρ be close to the standard normal. As an example, for a standardized value of $\hat{\alpha}$ around unity, and a sample size of 400, the empirical distribution of $t_{\rho=0}$ at the 5% risk level is 2.06, which is quite close to the standard normal value of 1.96. In this situation the normal distribution, or the ordinary, t-statistics, would be an acceptable approximation for testing the significance of ρ . Critical values from Banerjee et al p. 170-171, Table 6.3.

6 The Johansen test of Cointegration

The superior test for cointegration is Johansen's test. This is a test which has al desirable statistical properties. The weakness of the test is that it relies on asymptotic properties, and is therefore sensitive to specification errors in limited samples. In the end some judgement in combination with economic and statistical model building is unavoidable. Start with a VAR representation of the variables, in this case what we think is the economic system we would like to investigate. We have a p -dimensional process, integrated of order d , $\{\mathbf{x}\}_t \sim I(d)$, with the VAR representation,

$$\mathbf{A}_k(\mathbf{L})\mathbf{x}_t = \boldsymbol{\mu}_0 + \boldsymbol{\Psi}\mathbf{D}_t + \boldsymbol{\varepsilon}_t. \tag{22}$$

The empirical VAR is formulated with lags and dummy variables so that the residuals become a white noise process. The demands for a well-specified model is higher than for an ARIMA model. Here we do test for all components in the residual process. The reason being that the critical values are determined

conditionally on a normal distribution of the residual process. Typically, we will assume that the system is integrated of order one. If there are signs of $I(2)$ variables, we will transform them to $I(1)$ before setting up the VAR. By using the difference operator $\Delta = 1 - L$, or $L = 1 - \Delta$, the VAR in levels can be transformed to a vector error correction model (VECM),

$$\Delta \mathbf{x}_t = \Gamma_1 \Delta \mathbf{x}_{t-1} + \dots + \Gamma_{k-1} \Delta \mathbf{x}_{t-k+1} + \Pi \mathbf{x}_{t-1} + \boldsymbol{\mu}_0 + \Psi \mathbf{D}_t + \boldsymbol{\varepsilon}_t. \quad (23)$$

where the Γ_i : s and Π are matrixes of variables. The lag length in the VAR is k lags on each variable. After transforming the model, using $L = 1 - \Delta$, we 'lose' on lag at the end, leading to $k - 1$ lags in the VECM. In a more compact for the VECM becomes

$$\Delta \mathbf{x}_t = \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \Pi \mathbf{x}_{t-1} + \boldsymbol{\mu}_0 + \Psi \mathbf{D}_t + \boldsymbol{\varepsilon}_t. \quad (24)$$

The number of cointegrating vectors, are identical to the number of stationary relationships in the Π -matrix. If there is no cointegration, all rows in Π must be filled with zeros. If there are stationary combinations, or stationary variables, in Π then some parameters in the matrix will be non-zero. There is a simple mathematical technique for answering the problem raised here. The rank of Π matrix determines the number independent rows in Π , and therefore also the number of cointegrating vectors. The rank of Π is given by the number of significant eigenvalues found in $\hat{\Pi}$. Each significant eigenvalue represent a stationary relation. Under the null hypothesis of $\{\mathbf{x}\}_t \sim I(d)$, with $d > 1$, the test statistic for determining the significance of the eigenvalues is non-standard, and must be simulated.

- If Π has reduced rank there are co-integrating relations among the \mathbf{x} : s . Thus, $\text{rank}(\Pi) = 0$, implies that all x ' s are non-stationary. There is no combination of variables that leads to stationarity. The conclusion is that modelling should be done in first differences instead, since there are no stable linear combination of the variables in levels. If $\text{rank}(\Pi) = p$, so Π has full rank, then all variables in \mathbf{x}_t must be stationary.

If Π has reduced rank, $0 < r < p$, the cointegrating vectors are given as, $\Pi = \boldsymbol{\alpha} \boldsymbol{\beta}'$ where β_i represents the i : th co-integration vector, and α_j represents the effect of each co-integrating vector on the $\Delta x_{p,t}$ variables in the model. Once the rank of Π is determined and imposed on the model, the model will consist of stationary variables or expression, and estimated parameters follows standard distributions.

- The reduced rank test of Π determines only the number of co-integrating vectors (r) and the number of common trends ($p - r$). The econometrician must through normalization, and tests of the α and β matrices identify the cointegrating vector(s) so they represent meaningful economic relationships. Notice that the inclusion of a stationary $I(0)$ non-white noise variable in \mathbf{x}_t automatically leads to a reduced rank of Π . The corresponding "co-integrating" vector will then consist of one variable only, namely the stationary variable. A test of the $I(1)$ hypothesis in the VAR model is strongly recommended.. Also finding

a cointegrating vector when $p > 2$ does not mean that all p variables are needed to form a stationary vector. Some of the β 's might be zero. It is necessary to test if some variables can be excluded from a vector and still have cointegration.⁸

Remark 7 *The test builds on a VAR with Gaussian errors (meaning in practice normal distribution). It is therefore necessary to test the estimated residual process extremely carefully. These test must also accompany any published work. It is advisable to avoid too long lag structures, since this eats up degrees of freedom and make the underlying dynamic structure difficult to understand, say in terms of some underlying stochastic differential data generating process. There is, generally, a trade off between the lag structure and choice of dummy variables for picking up outliers. A lag length of 2, and a suitable choice of dummy variables for creating white noise residuals, is often a good choice.*

Remark 8 *The critical values are only valid asymptotically. (Simulations are made for a Brownian motion with 400 observations) Unfortunately, there is no easy way of simulating the small sample distributions of the test statistics. (Interesting work is being done by applying methods of both strapping). Thus, the test statistics should be used as a approximation. The accepted hypothesis concerning number of co-integrating vectors and common trends must allow for a meaningful economic interpretation.*

Remark 9 *Originally, Johansen derived two tests, the λ -max (or maximum eigenvalue test) and the trace test. The Max test is constructed as*

$$\lambda_{\max}[\mathcal{H}_1(r-1) | \mathcal{H}_1(r)] = -T \log(1 - \hat{\lambda}_r), \quad (25)$$

for $r = 0, 1, 2, \dots, p-2, p-1$. The null is that there exists r cointegrating vectors against the alternative of $r+1$ vectors. The trace test is

$$\lambda_{\text{trace}}[\mathcal{H}_1(r) | \mathcal{H}_0] = -T \sum_{i=r+1}^p \log(1 - \hat{\lambda}_i) \quad (26)$$

where the null hypothesis is $\lambda_i = 0$, so only the first r eigenvalues are non-zero. It has been found that the trace test is the better test, since it appears to be more robust to skewness and excess kurtosis.. Therefore, make your decision on the basis of the trace test. Furthermore, the trace test can be adjusted for degrees of freedom, which can be of importance in small samples. Reimers (1992) suggests replacing T in the trace statistics by $T - nk$.

Remark 10 *There are several tabulations of the test statistics. The most accurate is found in Johansen (1995) for the trace statistics. The tabulated values in Osterwald-Lenum (1992) are also quite accurate.⁹*

⁸Notice that PCFIML ver. 8 imposes a normalization structure automatically. To get other normalizations the model must therefore be re-estimated with a different ordering of the variables.

⁹Other tabulations include, Johansen (1988), Johansen and Juselius(1990), Johansen (1991), but they are not as accurate as the ones in Johansen (1995).

Remark 11 *The tests for determining the number of co-integrating vectors are nested. The test should therefore be performed by starting from the hypothesis of zero cointegrating vectors. Thus $H_{0,1} : 0$ co-integrating vectors is tested against the alternative $H_{a,1} : \text{at least one co-integrating vector}$. If $H_{0,1}$ is rejected the next test is $H_{0,2} : 1$ cointegrating vector against $H_{a,2} : \text{at least two co-integrating vectors}$, etc.*

Remark 12 *The test statistics determines the number of co-integrating vectors. If there are $I(2)$ processes in the data, one or several of the cointegrating vectors found in test can be picking up co-integration from $I(2)$ to $I(1)$ relations, which then co-integrates with other $I(1)$ relations in the model. Tests for $I(2)$ -ness is therefore, strongly recommended in the cases where $I(2)$ -ness is present. The $I(2)$ test is explained below.*

Remark 13 *The test only determines the number of cointegrating vectors, not what they look like or how they should be understood. You must therefore (1) test if all variables are $I(1)$, or perhaps $I(2)$. If you include one stationary variable in the VAR, the result is at least one significant eigenvalue, representing the $I(0)$ variable. (2) Knowing that all variables in the vector are $I(1)$, or perhaps $I(2)$, does not mean that they are needed there to form a stationary co-integrating vector. You should therefore test which variables form the cointegrating vector or vectors. If a variable is not in the vector it will have a parameter of zero.*

The test statistics varies depending on the inclusion of constants and trends in the model. The standard model can be said to include an unrestricted constant but no specific trend term. For a p -dimensional vector of variables (x_t) the estimated model is,

$$\Delta \mathbf{x}_t = \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \alpha \beta' \mathbf{x}_{t-1} + \mu_0 + \Psi \mathbf{D}_t + \varepsilon_t, \quad (27)$$

where μ_0 is an unrestricted constant term, \mathbf{D}_t is a vector impulse dummies and/or seasonal dummies (seasonal factors will be centralized around zero), selected such that the residual term is $\varepsilon_t \sim N_p(0, \Sigma)$.¹⁰ This equation with an unrestricted constant allows for a constant term in the cointegrating vector, to account for different levels of the \mathbf{x}' s, and constant terms in $\Delta \mathbf{x}_t$, representing linear deterministic trends in the \mathbf{x}_t 's.

The critical values of the *trace* test are tabulated in Johansen (1995) Table 15.3. (Corresponds to Table 1. Case 1, in Osterwald-Lenum (1992)). For a process $p = 4$, the estimated eigenvalues are ordered in size from the largest to the smallest. Assume that the associated estimated *trace* statistics of the eigenvalues are (49.14, 19.06, 8.89, 2.35). The test is performed as follows.

First, at 5 % risk level, the value of 49.14 is higher than the critical value 47.21. Thus we reject the assumption of zero co-integrating vectors, in favour of the hypothesis that there is at least one co-integrating vector in the system.

Second, the test statistics of the second largest eigenvalue, is 19.06 which is smaller than the corresponding critical value of 29.38 at 5%. Thus, we cannot

¹⁰For a model with quarterly data the vector of seasonal dummies could be (-0.25,-0.25,-0.25,-0.75), because the seasonal dummies has a mean of zero they will not affect the estimation of the constants, which pick up any unrestricted deterministic trends in x_i .

reject the null hypothesis that there is only one co-integrating vector. Here the test stops, the remaining values are now uninteresting, even if they should indicate significance. Due to the nested properties of the test procedure, the remaining test statistics are not reliable once we found a non-significant eigenvalue.

The model above represent a standard approach. It can be restricted or be made more general by allowing for linear deterministic trends in the β -vectors and in $\Delta \mathbf{x}_t$. The latter corresponds to quadratic trends in the levels of the variables, an assumption which is often totally unrealistic in economics. In the following we use the notation in CATS in RATS and label the alternatives as Model 1, 2, 3, 4 and 5, respectively. Among these model, only Model 2, 3 and 4 are really interesting in empirical work.

Model 1, is the most restrictive model. Restricting the constant term to zero, both in the first difference equations and in the cointegrating vectors leads to

$$\Delta \mathbf{x}_t = \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \alpha \beta' \mathbf{x}_{t-1} + \Psi \mathbf{D}_t + \varepsilon_t. \quad (28)$$

In general, this model is uninteresting because it assumes that all variables in the cointegrating vectors have the same mean. The test statistics for Model 1 is tabulated in Table 15.1, Johansen (1995) and in the CATS Manual. Corresponds to Osterwald-Lenum (1992) Table 0. Case 0*.

A less restricted model (Model 2) sets $\mu_0 = 0$, but allows for constant in the cointegrating vectors in the estimated model,

$$\Delta \mathbf{x}_t = \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \alpha [\beta', \beta_0] [\mathbf{x}_{t-1}, 1] + \Psi \mathbf{D}_t + \varepsilon_t. \quad (29)$$

The test statistics for Model 2 is tabulated in Table 15.2, Johansen (1995) and in the CATS Manual. Corresponds to Table 1. Case 0*, in Osterwald-Lenum (1992).

The third alternative is the model discussed in the introduction (Model 3), where μ_0 is left unrestricted in the equation, thereby including both deterministic trend in the \mathbf{x} 's and constants in the cointegrating vectors,

$$\Delta \mathbf{x}_t = \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \alpha \beta' \mathbf{x}_{t-1} + \mu_0 + \Psi \mathbf{D}_t + \varepsilon_t. \quad (30)$$

The test statistics for Model 3 is tabulated in Table 15.3 Johansen (1995) and in the CATS Manual. Corresponds to Table 1. Case 1, in Osterwald-Lenum (1992).

The fourth alternative (Model 4), is to allow for constants and deterministic trends in the cointegrating vectors, leading to

$$\begin{aligned} \Delta \mathbf{x}_t = & \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \alpha [\beta', \beta_1, \beta_0]' [\mathbf{x}_{t-1}, t, 1] \\ & + \mu_0 + \Psi \mathbf{D}_t + \varepsilon_t \end{aligned} \quad (31)$$

The test statistics for Model 4 is tabulated in Table 15.4, Johansen (1995) and in the CATS Manual. Corresponds to Table 2*. Case 1*, in Osterwald-Lenum (1992). In practice this is the model of last resort. If no meaningful cointegration vectors are found using Models 2 or 3, a trend component in the vectors might do the trick. Having a trend in the cointegrating vectors can be understood as a type of growth in target problem, sometimes motivated with productivity growth, technological development etc. In other situations we conclude that there is some growth in the data which the model cannot account for.

Finally, we can allow for a quadratic trend in the \mathbf{x}_t 's. This leads to model 5, the least restricted of all models,

$$\Delta \mathbf{x}_t = \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \alpha [\beta', \beta_1, \beta_0]' [\mathbf{x}_{t-1}, t, 1] \mathbf{x}_{t-1} + \mu_0 + \mu_1 t + \Psi \mathbf{D}_t + \varepsilon_t. \tag{32}$$

This model is quite unrealistic and should not be considered in applied work. The reason why this is an unrealistic model is the difficulty in motivating quadratic trends in a multivariate model. For instance, from an economic point of view it is totally unrealistic to assume that technological or productivity growth is an increasingly expanding process. In the following section we consider how to choose between these models.

Of these five models, Model 3 with the unrestricted constant is the basic model, and the one to chose for most applications.

6.1 Testing for different trends in the VECM

In some situations estimating the "standard" model is not sufficient, we want to say more about the model, and how variables and cointegrating relations evolve over time. To select the most appropriate model, Johansen suggests estimating all models and perform the reduced rank test across the model. In practice, only Model 2, 3 and 4 are of interest. As an example assume that the model contains three variables, $p = 3$. For each model we get 3 trace-test statistics, and 3 max-test statistics. Concentrating on the trace statistics, we have 3x3 trace test statistics if all three models are estimated. For this example, label the estimated trace statistics as $tr_{i,j}$ where i refers to Model i and, j refers to the test of the number of significant eigenvalues j . Organize all 3x3 tests statistics as below,

| $H_0 : r$ | $p - r$ | Model 2 | Model 3 | Model 4 |
|-----------|---------|------------------------|------------------------|------------------------|
| 0 | 3 | $tr_{2,0}$ | $\rightarrow tr_{3,0}$ | $\rightarrow tr_{4,0}$ |
| 1 | 2 | $\rightarrow tr_{2,1}$ | $\rightarrow tr_{3,1}$ | $\rightarrow tr_{4,1}$ |
| 2 | 1 | $\rightarrow tr_{2,2}$ | $\rightarrow tr_{3,2}$ | $\rightarrow tr_{4,2}$ |
| 3 | 0 | $\rightarrow tr_{2,3}$ | $\rightarrow tr_{3,3}$ | $\rightarrow tr_{4,3}$ |

r is the number of cointegrating vectors, it follows that $p - r$ is the number of unit roots (common trends) in the system. The test is performed by starting in row one, testing Model 2 for $r = 0$, if the test statistic ($tr_{2,0}$) exceeds its critical value continue to test $r = 0$ for Model 3. If $tr_{3,0}$ exceeds it critical value continue with $tr_{4,0}$. If all three model reject the assumption of zero cointegrating vectors, in favour of at least one cointegrating vector, continue with row 2, where H_0

is $r = 1$. Suppose that it is first when we use $tr_{4,2}$ that we cannot reject any more. This leads to the conclusion that there are two cointegrating vectors in the system, and that Model 4 is the better model for describing the system. Once we cannot reject a hypothesis in this sequence there is no reason to go further. In fact looking at test statistics beyond this point might be misleading, in the same way as testing for $I(2)$ v.s $I(1)$ should always come before testing for $I(1)$ v.s $I(0)$.

6.2 Testing for I(2) in the VECM model

The presence of $I(2)$ variables and $I(2)$ relations are difficult to handle from a statistical point of view and difficult to understand in an economic interesting way. Thus, $I(2)$ relations should be tested for, and if they are found be eliminated through data transformation. There are two ways of testing for $I(2)$ -ness, the first and most useful way is to start with the assumption that all variables are at most $I(1)$, estimate the VAR and test for the number of cointegrating vectors. The $I(2)$ test can there after be applied as a misspecification test of the $I(1)$ model, meaning that the model is maintained if $I(2)$ -ness is rejected.

The alternative, is to test for both $I(2)$ and cointegration in the same step following Parulo (1996). (This part will be added here).

Starting with a VAR model, that can be written as

$$\Delta \mathbf{x}_t = \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \Pi \mathbf{x}_{t-1} + \boldsymbol{\mu}_0 + \Psi \mathbf{D}_t + \boldsymbol{\varepsilon}_t. \quad (33)$$

If there are $I(2)$ variables rewrite the system in second difference form,

$$\Delta^2 x_t = \Gamma \Delta x_{t-1} + \Pi x_{t-2} + \sum_{i=1}^{k-2} \Phi_i \Delta^2 x_{t-i} + \boldsymbol{\mu}_0 + \Psi \mathbf{D}_t + \boldsymbol{\varepsilon}_t, \quad (34)$$

where $\Gamma = I - \sum_{i=1}^{k-1} \Gamma_i$, and $\Phi_i = -\sum_{j=i+1}^{k-1} \Gamma_j$, for $i = 1, 2, \dots, k-2$. The test is performed in two steps, first for the number of cointegrating vectors, the rank of Π , as before. Second conditionally on the chosen rank of Π test the rank of Γ to find the number of $I(2)$ relations. We have that $\Pi = \alpha\beta'$, of dimension $p \times r$. Given α and β define the orthogonal of these as α_{\perp} and β_{\perp} , one can then define a new matrix such that $\alpha_{\perp} \Gamma \beta_{\perp} = \varphi\eta$ with dimension $(p-r) \times s_1$. For an estimated rank (s_1) of this matrix the number of $I(2)$ variables are $s_2 = p - r - s_1$. The value of s_1 is determined from the $I(2)$ program in the following way. The program produces a matrix of test values organized such that each row corresponds to a given value of r . The matrix of test statistics for $p = 5$ is

| $p - r$ | r | | | | | | | <i>Trace</i> | <i>Crit.val(r)</i> |
|---------|-----|----------|----------|----------|----------|----------|--|--------------|--------------------------------|
| p | 0 | H_{00} | H_{01} | H_{02} | H_{03} | H_{04} | | tr_1 | |
| $p - 1$ | 1 | | H_{10} | H_{11} | H_{12} | H_{13} | | tr_2 | |
| $p - 2$ | 2 | | | H_{20} | H_{21} | H_{22} | | tr_3 | |
| | 3 | | | | H_{30} | H_{31} | | tr_4 | |
| | 4 | | | | | H_{40} | | tr_5 | |
| | | | | | | | | | <i>Crit.val(s₁)</i> |

where H_{ij} is a test statistics for the joint hypothesis that $r = i$ and $s_1 = j$. Thus, first test for r , in the usual way. That is compare tr_i with the relevant

critical values. For a given r then read the corresponding row and test against the critical values against the horizontal axis. The program does not print the critical values they are of course your choice depending on the model. For determining s_1 use a more restricted model. It is unlikely that there is a quadratic trend in the $I(2)$ model. Perhaps use critical values for a model no intercept term.

How to look for $I(2)$ in general. First look at graphs of the data series. Second compare the graph of the cointegration vectors $\beta'x_t$ against graphs of the conditional cointegration vectors $\beta'R_t$, where R_t represent the vector looks like after the x'_t s are corrected for the changes in the variables Δx_t . That is $x_t | \Delta x_{t-i}$. If $\beta'x_t$ looks non-stationary, but $\beta'R_t$ looks stationary that a sign of $I(2)$ -ness. A third way is to look at the characteristic polynomial for the estimated values. Some programs will rewrite the estimated in as a system of first order equations (the companion form) and calculate the eigenvalues of this system. If it seems like there is one or more unit roots in the system than indicated by $p - r$ then that is a sign of $I(2)$ -ness.¹¹

However, is it likely to have $I(2)$ variables? Is really inflation $I(1)$ for all times, or is the average rate of inflation simply shifting between different periods as a consequence of changes in the monetary regime? As an example shifting from fixed exchange rates to flexible, and back. There is no clear answer to the question, besides that since $I(2)$ is a little bit mind bending, look for changes in policy that explains shift dummies instead.

6.3 Testing the vector

The Johansen procedure test for cointegration will only identify the number of stationary vector among your variables. It is therefore necessary to test the vector carefully in order to identify them. From the cointegration test you get $\hat{\Pi} = \alpha\beta'$. This parameter vector produces stationary relations. It is therefore possible to test which $\alpha : s$ and $\beta : s$ that are significant. These test helps to identify the cointegrating vector in terms of economic relations, which variables are affected (driven) by the vectors, which $\beta : s$ are significant, are they different from specific pre-determined values. You might want to test if some variables are different from unity as an example.

Test for stationarity, exclusion are always necessary. Test of exogeneity provide further information. These tests should be done for different choices of r cointegrating vectors. It might be that your cointegrating vector is only made up be one (or more) stationary variable. Notice that the inclusion of a constant might make some variables stationary in the context of your model. The tests of exogeneity is harder to analyze.. It tells you basically if it will be possible to reestimate the system and keep some variable exogenous.. The test is also a type Granger causality test, showing which variables that are predicted by the vector in a reduced system, but nothing more. In a structural model the alpha coefficient might be different. Once the basic tests are performed, you can also impose different values on the coefficients to exactly identify the vector as an economic interesting relation.

¹¹The alternative understanding is of course that r is chosen to be to high. Too few unit roots are imposed on the system.

7 References

- Banerjee, Anindya, Juan Dolado, John W. Galbraith, and David F. Hendry (1993) *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press, Oxford.
- Dickey, D.A. and W.A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, 74, p. 427-431.
- Engle, R.F. and B.S. Yoo, (1987) "Forecasting and Testing in Co-integrated Systems", *Journal of Econometrics*, 35, 143-159.
- Fuller, W.A. (1976) *Introduction to Statistical Time Series*, John Wiley, New York.
- Johansen, Sören (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.
- MacKinnon, J.G. (1991) "Critical Values for Co-Integration Tests", in R.F. Engle and C.W. J. Granger (eds) *Long-Run Economic Relationships*, Oxford University Press, p. 267-276
- Osterwald-Lenum, M. (1992) "A Note with Fractiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics: Four Cases", *Oxford Bulletin of Economics and Statistics*, 54, 461-472.
- Reimers, H-E (1992) Comparisons of tests for multivariate cointegration, *Statistical Papers* **33**, 335-359.
- Sjö, Boo (2000) Lectures in Modern Economic Time Series Analysis, memo. (Revised, 1997,1998, 2000)

TESTING FOR UNIT ROOTS AND COINTEGRATION 22

I: For an estimated DF equation without constat

II: For an estimated DF equation with a constant

III: For an estimated DF equation with a constant and a time trend. **Table**

1. Critical values for Dickey-Fuller's Unit Root Test¹²

| n | 0.01 | 0.025 | 0.05 | 0.10 | 0.90 | 0.95 | 0.975 | 0.99 |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|
| I | | | | | | | | |
| 25 | -2.66 | -2.26 | -1.95 | -1.60 | 0.92 | 1.33 | 1.70 | 2.16 |
| 50 | -2.62 | -2.25 | -1.95 | -1.61 | 0.91 | 1.31 | 1.66 | 2.08 |
| 100 | -2.60 | -2.24 | -1.95 | -1.61 | 0.90 | 1.29 | 1.64 | 2.03 |
| 250 | -2.58 | -2.23 | -1.95 | -1.62 | 0.89 | 1.29 | 1.63 | 2.01 |
| 500 | -2.58 | -2.23 | -1.95 | -1.62 | 0.89 | 1.28 | 1.62 | 2.00 |
| ∞ | -2.58 | -2.23 | -1.95 | -1.62 | 0.89 | 1.28 | 1.62 | 2.00 |
| II | | | | | | | | |
| 25 | -3.75 | -3.33 | -3.00 | -2.63 | -0.37 | 0.00 | 0.34 | 0.72 |
| 50 | -3.58 | -3.22 | -2.93 | -2.60 | -0.40 | -0.03 | 0.29 | 0.66 |
| 100 | -3.51 | -3.17 | -2.89 | -2.58 | -0.42 | -0.05 | 0.26 | 0.63 |
| 250 | -3.46 | -3.14 | -2.88 | -2.57 | -0.42 | -0.07 | 0.24 | 0.61 |
| 500 | -3.44 | -3.13 | -2.87 | -2.57 | -0.43 | -0.07 | 0.23 | 0.60 |
| ∞ | -3.43 | -3.12 | -2.86 | -2.57 | -0.44 | -0.07 | 0.23 | 0.60 |
| III | | | | | | | | |
| 25 | -4.38 | -3.95 | -3.60 | -3.24 | -1.14 | -0.80 | -0.50 | -0.15 |
| 50 | -4.15 | -3.80 | -3.50 | -3.18 | -1.19 | -0.87 | -0.58 | -0.24 |
| 100 | -4.04 | -3.73 | -3.45 | -3.15 | -1.22 | -0.90 | -0.62 | -0.28 |
| 250 | -3.99 | -3.69 | -3.43 | -3.13 | -1.23 | -0.92 | -0.64 | -0.31 |
| 500 | -3.98 | -3.68 | -3.42 | -3.13 | -1.24 | -0.93 | -0.65 | -0.32 |
| ∞ | -3.96 | -3.66 | -3.41 | -3.12 | -1.25 | -0.94 | -0.66 | -0.33 |

¹²The DGP is a random walk, $x_t = x_{t-1} + \varepsilon_t$.

I: For an estimated DF equation without constat

II: For an estimated DF equation with a constant

III: For an estimated DF equation with a constant and a time trend.

TESTING FOR UNIT ROOTS AND COINTEGRATION 23

T: number of observations

ADF(1), ADF(4): critical values for the second step using 1 or 4 lags in the augmentation. **Table 2. Critical Values for 2-step Cointegration Tests**¹³

| n | T | CRDW | ADF(1) | ADF(4) |
|----------|-----|------|--------|--------|
| 2 | 50 | 0.72 | -3.43 | -3.29 |
| | 100 | 0.38 | -3.38 | -3.17 |
| | 200 | 0.20 | -3.37 | -3.25 |
| 3 | 50 | 0.89 | -3.82 | -3.75 |
| | 100 | 0.48 | -3.76 | -3.62 |
| | 200 | 0.25 | -3.74 | -3.78 |
| 4 | 50 | 1.05 | -4.18 | -3.98 |
| | 100 | 0.58 | -4.12 | -4.02 |
| | 200 | 0.30 | -4.11 | -4.13 |
| 5 | 50 | 1.19 | -4.51 | -4.15 |
| | 100 | 0.68 | -4.48 | -4.36 |
| | 200 | 0.35 | -4.42 | -4.43 |

¹³n: number of variables in cointegrating regression model.

T: number of observations

ADF(1), ADF(4): critical values for the second step using 1 or 4 lags in the augmentation.

Table 3.1 Critical Values For Johansen's Trace Test VECM.1

| p-r | 90% | 95% | 99% |
|-----|--------|--------|--------|
| 1 | 2.98 | 4.14 | 7.01 |
| 2 | 10.35 | 12.21 | 16.16 |
| 3 | 21.58 | 24.08 | 29.19 |
| 4 | 36.58 | 39.71 | 45.99 |
| 5 | 57.54 | 59.23 | 66.70 |
| 6 | 78.30 | 82.61 | 91.11 |
| 7 | 104.93 | 109.93 | 119.58 |
| 8 | 135.16 | 140.74 | 151.70 |
| 9 | 169.30 | 175.47 | 187.82 |
| 10 | 207.21 | 214.07 | 226.94 |
| 11 | 248.77 | 256.23 | 270.47 |
| 12 | 293.83 | 301.95 | 318.14 |

Table 3.2. Critical Values for Johansen's Trace Test VECM.2

| p-3 | 90% | 95% | 99% |
|-----|--------|--------|--------|
| 1 | 7.50 | 9.13 | 12.73 |
| 2 | 17.79 | 19.99 | 24.73 |
| 3 | 31.88 | 34.79 | 40.84 |
| 4 | 49.91 | 53.42 | 60.42 |
| 5 | 71.66 | 75.74 | 83.93 |
| 6 | 97.17 | 101.84 | 111.38 |
| 7 | 126.71 | 132.00 | 142.34 |
| 8 | 159.74 | 165.73 | 177.41 |
| 9 | 196.66 | 203.34 | 216.08 |
| 10 | 237.35 | 244.56 | 258.31 |
| 11 | 281.63 | 289.70 | 304.89 |
| 12 | 333.26 | 338.09 | 354.32 |

Table 3.3. Critical Values for Johansen's Trace Test VECM 3.

| p-r | 90% | 95% | 99% |
|-----|--------|--------|--------|
| 1 | 2.71 | 3.84 | 6.63 |
| 2 | 13.31 | 15.34 | 19.69 |
| 3 | 26.70 | 29.38 | 34.87 |
| 4 | 43.84 | 47.21 | 53.91 |
| 5 | 64.74 | 68.68 | 76.37 |
| 6 | 89.37 | 93.92 | 102.94 |
| 7 | 117.73 | 123.04 | 133.04 |
| 8 | 149.99 | 155.75 | 166.95 |
| 9 | 185.83 | 192.30 | 204.64 |
| 10 | 228.55 | 232.60 | 246.17 |
| 11 | 272.03 | 276.37 | 291.58 |
| 12 | 319.21 | 323.93 | 339.64 |

Table 3.4. Critical Values for Johansen's Trace Test VECM.4.

| p-r | 90% | 95% | 99% |
|-----|--------|--------|--------|
| 1 | 10.56 | 12.39 | 16.39 |
| 2 | 22.95 | 25.47 | 30.65 |
| 3 | 39.08 | 42.20 | 48.59 |
| 4 | 58.96 | 62.61 | 70.22 |
| 5 | 82.68 | 86.96 | 95.37 |
| 6 | 110.00 | 114.96 | 124.61 |
| 7 | 141.31 | 146.75 | 157.53 |
| 8 | 176.13 | 182.45 | 194.12 |
| 9 | 214.72 | 221.56 | 234.65 |
| 10 | 257.08 | 264.23 | 278.80 |
| 11 | 306.47 | 311.12 | 326.73 |
| 12 | 356.39 | 361.07 | 377.54 |

Table 3.5 Critical Values for Johansen's Trace Test VECM.5

| p-r | 90% | 95% | 99% |
|-----|---------|---------|---------|
| 1 | 2.71 | 3.84 | 6.35 |
| 2 | 15.94 | 18.15 | 22.78 |
| 3 | 31.57 | 34.55 | 40.61 |
| 4 | 50.67 | 54.11 | 61.27 |
| 5 | 73.62 | 77.79 | 86.11 |
| 6 | 99.97 | 104.76 | 114.22 |
| 7 | 130.39 | 135.66 | 146.05 |
| 8 | 130.39* | 135.66* | 146.05* |
| 9 | 201.82 | 208.52 | 221.46 |
| 10 | 243.28 | 250.53 | 264.52 |
| 11 | 291.51 | 296.02 | 311.41 |
| 12 | 340.54 | 345.27 | 361.68 |