Cointegration - general discussion

Definitions:
A time series that requires d differences to get it stationary is said to be "integrated of order d". If the d-th difference has p AutoRegressive and q Moving Average terms, the differenced series is said to be ARMA(p,q) and the original Integrated series to be ARIMA(p,d,q).

Two series X_t and Y_t that are integrated of order d may, through linear combination, produce a series aX_t + bY_t which is stationary (or integrated of order smaller than d) in which case we say that X_t and Y_t are cointegrated and we refer to (a, b) as the cointegrating vector. Granger and Weis discuss this concept and terminology.

An example:

For example, if X_t and Y_t are wages in two similar industries, we may find that both are unit root processes. We may, however, reason that by virtue of the similar skills and easy transfer between the two industries, the difference X_t - Y_t cannot vary too far from 0 and thus, certainly should not be a unit root process. The cointegrating vector is specified by our theory to be (1, -1) or (-1, 1), or (c, -c) all of which are equivalent.

The test for cointegration here consists of simply testing the original series for unit roots, not rejecting the unit root null, then testing the X_t - Y_t series and rejecting the unit root null. We just use the standard D-F tables for all these tests. The reason we can use these D-F tables is that the cointegrating vector was specified by our theory, not estimated from the data.

Numerical examples: \( Y_t = A Y_{t-1} + E_t \)

1. Bivariate, stationary:

\[
\begin{pmatrix}
Y_{1t} \\
Y_{2t}
\end{pmatrix} =
\begin{bmatrix}
1.2 & -0.3 \\
0.4 & 0.5
\end{bmatrix}
\begin{pmatrix}
Y_{1,t-1} \\
Y_{2,t-1}
\end{pmatrix} +
\begin{pmatrix}
e_{1t} \\
e_{2t}
\end{pmatrix},
\quad
E_t \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix} \right)
\]

\[A - \lambda I =
\begin{vmatrix}
1.2 - \lambda & -0.3 \\
0.4 & 0.5 - \lambda
\end{vmatrix}
= \lambda^2 - 1.7\lambda + .6 + .12 = (\lambda - .9)(\lambda - .8)
\]

this is stationary (note \( E_t \) distribution has no impact)
2. Bivariate, nonstationary:

\[
\begin{pmatrix}
Y_{1t} \\
Y_{2t}
\end{pmatrix}
= 
\begin{bmatrix}
1.2 & -0.5 \\
0.2 & 0.5
\end{bmatrix}
\begin{pmatrix}
Y_{1,t-1} \\
Y_{2,t-1}
\end{pmatrix}
+ 
\begin{pmatrix}
e_{1t} \\
e_{2t}
\end{pmatrix},
\]

\[
\mathbf{A} - \lambda \mathbf{I} = 
\begin{bmatrix}
1.2 - \lambda & -0.5 \\
0.2 & 0.5 - \lambda
\end{bmatrix}
= \lambda^2 - 1.7\lambda + .6 + .1 = (\lambda - 1)(\lambda - .7)
\]

This is a unit root process.

Using the spectral decomposition (eigenvalues, vectors) of \( \mathbf{A} \) we have

\[
\begin{bmatrix}
1.2 & -0.5 \\
0.2 & 0.5
\end{bmatrix}
\begin{bmatrix}
1/\sqrt{1.16} & 1/\sqrt{2} \\
.4/\sqrt{1.16} & 1/\sqrt{2}
\end{bmatrix}
= 
\begin{bmatrix}
1/\sqrt{1.16} & 1/\sqrt{2} \\
.4/\sqrt{1.16} & 1/\sqrt{2}
\end{bmatrix}
\begin{bmatrix}
1 & 0 \\
0 & 0.7
\end{bmatrix}
\]

\[
\mathbf{T}^{-1} \mathbf{Y}_t = (\mathbf{T}^{-1} \mathbf{A} \mathbf{T}) \mathbf{T}^{-1} \mathbf{Y}_{t-1} + \mathbf{T}^{-1} \mathbf{E}_t
\]

\[
\mathbf{Z}_t = \Delta \mathbf{Z}_{t-1} + \eta_t
\]

Components of the \( \mathbf{Z}_t \) vector:

\[
\begin{align*}
Z_{1,t} &= Z_{1,t-1} + \eta_{1,t} \text{ "common trend" (unit root)} \\
Z_{2,t} &= 0.7 Z_{2,t-1} + \eta_{2,t} \text{ (stationary root)}
\end{align*}
\]

\[
\begin{align*}
Y_{1,t} &= w_1 Z_{1,t} + w_2 Z_{2,t} = 1/\sqrt{1.16} Z_{1,t} + 1/\sqrt{2} Z_{2,t} \\
Y_{2,t} &= w_3 Z_{1,t} + w_4 Z_{2,t} = .4/\sqrt{1.16} Z_{1,t} + 1/\sqrt{2} Z_{2,t}
\end{align*}
\]

\( \mathbf{Z}_t = \mathbf{T}^{-1} \mathbf{Y}_t \) so that last row of \( \mathbf{T}^{-1} \) is cointegrating vector. Notice that \( \mathbf{A} \) is not symmetric and \( \mathbf{T}^{-1} \neq \mathbf{T}' \).
Engle - Granger method

This is one of the earliest and easiest to understand treatments of cointegration.

\[
Y_{1,t} = w_1 Z_{1,t} + w_2 Z_{2,t}, \quad \text{where } \frac{\sum Z_{1,t}^2}{n^2} \text{ is } O_p(1) \text{ and } \frac{\sum Z_{2,t}^2}{n^2} \text{ is } O_p\left(\frac{1}{n}\right)
\]

so if we regress \(Y_{1,t}\) on \(Y_{2,t}\) our regression coefficient is

\[
\frac{n^2 \sum_{t=1}^{n} Y_{1t} Y_{2t}}{n^2 \sum_{t=1}^{n} Y_{2t}^2} = \frac{n^2 w_1 w_3 \sum Z_{1,t}^2 + O_p\left(\frac{1}{\sqrt{n}}\right)}{n^2 w_3^2 \sum Z_{1,t}^2 + O_p\left(\frac{1}{\sqrt{n}}\right)} = w_1 \frac{w_3}{w_3} + O_p\left(\frac{1}{\sqrt{n}}\right)
\]

and our residual series is thus approximately \(\frac{1}{w_3} \left[ w_3 Y_{1t} - w_1 Y_{2t} \right] = (w_2 - w_1 w_4 / w_3) Z_{2t},\) a stationary series. Thus a simple regression of \(Y_{1,t}\) on \(Y_{2,t}\) gives an estimate of the cointegrating vector and a test for cointegration is just a test that the residuals are stationary. Let the residuals be \(r_t\). Regress \(r_t - r_{t-1}\) on \(r_{t-1}\) (and possibly some lagged differences). Can we compare to our D-F tables? Engle and Granger argue that one cannot do so.

The null hypothesis is that there is no cointegration, thus the bivariate series has 2 unit roots and no linear combination is stationary. We have, in a sense, looked through all possible linear combinations of \(Y_{1,t}\) and \(Y_{2,t}\), finding the one that varies least (least squares) and hence the one that looks most stationary. It is as though we had computed unit root tests for all possible linear combinations then selected the one most likely to reject. We are thus in the area of order statistics. (If you report the minimum heights from samples of 10 men each, the distribution of these minimae will not be the same as the distribution of heights of individual men nor will the distribution of unit root tests from these "best" linear combinations be the same as the distribution you would get for a pre specified linear combination). Engle and Granger provide adjusted critical values. Here is a table comparing their E-G tables to our D-F tables for \(n=100\). E-G used an augmented regression with 4 lagged differences and an intercept to calculate a t statistic \(\tau_\mu\), so keep in mind that part of the discrepancy is due to finite sample effects of the (asymptotically negligible) lagged differences.

<table>
<thead>
<tr>
<th>Prob of smaller (\tau_\mu)</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-G</td>
<td>-3.77</td>
<td>-3.17</td>
<td>-2.84</td>
</tr>
<tr>
<td>D-F</td>
<td>-3.51</td>
<td>-2.89</td>
<td>-2.58</td>
</tr>
</tbody>
</table>
Example:

\( P_t = \text{cash price on delivery date, Texas steers} \)

\( F_t = \text{Futures price} \)

(source: Ken Mathews, NCSU Ag. Econ.)

Data are bimonthly Feb. '76 through Dec. '86 (60 obs.)

1. Test individual series for integration

\[
\nabla P_t = 7.6 - 0.117 P_{t-1} + \sum_{i=1}^{5} \beta_i \nabla P_{t-i} \quad \tau_{\nu}(\text{D-F}) = \frac{-1.117}{0.053} = -2.203
\]

\[
\nabla F_t = 7.7 - 0.120 F_{t-1} + \sum_{i=1}^{5} \beta_i \nabla F_{t-i} \quad \tau_{\nu}(\text{D-F}) = \frac{-1.120}{0.054} = -2.228
\]

each series is integrated, cannot reject at 10%

2. Regress \( F_t \) on \( P_t \):

\[
\nabla R_t = 0.110 - 0.9392 R_{t-1} \quad \tau_{\nu}(\text{E-G}) = \frac{-0.9392}{0.7264} = -7.428
\]

Thus, with a bit of rounding, \( F_t - 1.00 P_t \) is stationary.

The Engle-Granger method requires the specification of one series as the dependent variable in the bivariate regression. Fountis and Dickey (Annals of Stat.) study distributions for the multivariate system.

If \( Y_t = A Y_{t-1} + E_t \) then \( Y_t - Y_{t-1} = -(I - A) Y_{t-1} + E_t \)

We show that if the true series has one unit root then the root of the least squares estimated matrix \( I - \hat{A} \) that is closest to 0 has the same limit distribution, after multiplication by \( n \), as the standard D-F tables and we suggest the use of the eigenvectors of \( I - \hat{A} \) to estimate the cointegrating vector. The only test we can do with this is the null of one unit root versus the alternative of stationarity. Johansen's test, discussed later, extends this in a very nice way. Our result also holds for higher dimension models, but requires the extraction of roots of the estimated characteristic polynomial.

For the Texas steer futures data, the regression gives

\[
\begin{bmatrix}
\nabla P_t \\
\n\nabla F_t
\end{bmatrix} = \begin{bmatrix}
5.3 \\
6.9
\end{bmatrix} + \begin{bmatrix}
-1.77 & 1.69 \\
-1.03 & 0.93
\end{bmatrix} \begin{bmatrix}
P_{t-1} \\
F_{t-1}
\end{bmatrix} + 3 \text{ lagged differences}
\]
where
\[
\begin{bmatrix}
0.54 & -0.82 \\
0.69 & 0.72 \\
-1.77 & 1.69 \\
-1.03 & 0.93 \\
-3.80 & -4.42 \\
-3.63 & -2.84 \\
\end{bmatrix}
\]
indicating that \(-0.69\) \(P_t\) + \(0.72\) \(F_t\) is stationary. This is about .7 times the difference so the two methods agree that \(P_t - F_t\) is stationary, as is any multiple of it.

**Johansen's Method**

This method is similar to that just illustrated but has the advantage of being able to test for any number of unit roots. The method can be described as the application of standard multivariate calculations in the context of a vector autoregression, or VAR. The test statistics are those found in any multivariate text. Johansen's idea, like in univariate unit root tests, is to get the right distribution for these standard calculated statistics. The statistics are standard, their distributions are not.

We start with just a lag one model with mean 0 (no intercept).

\[
\nabla Y_t = \Pi Y_{t-1} + \epsilon_t \quad \text{where } Y_t \text{ is a p-dimensional column vector as is } \epsilon_t
\]

Assume \(E\{\epsilon_t \epsilon_t'\} = \Lambda\).

\[
H_0 : \Pi = -\alpha \beta' = - \left[ \begin{array}{c} \alpha_1 \\ \vdots \\ \alpha_p \end{array} \right]_{p \times p}
\]

\(r = 0 \Rightarrow\) all linear combinations nonstationary
\(r = p \Rightarrow\) all linear combinations stationary
\(0 < r < p \Rightarrow\) cointegration

Note: for any \(\Pi\) there are infinitely many \(\alpha, \beta\) such that \(\Pi = -\alpha \beta'\) [because \(-\alpha \beta' = -\alpha T T^{-1} \beta'\)] so we do not test hypotheses about \(\alpha\) and \(\beta\), only about the rank \(r\).

Now define sums of squares and cross products:

\[
\nabla Y_t' = Y_t' - \Pi Y_{t-1} \\
S_{ij} = \sum_{t=1}^{n} Y_{t-i} Y_{t-1}'
\]

Now write down likelihood (conditional on \(Y_0 = 0\))

\[\mathcal{L} = \frac{1}{(2\pi)^{\frac{p}{2}} |\Lambda|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{n} (\nabla Y_t' - \Pi Y_{t-1}') \Lambda^{-1} (\nabla Y_t' - \Pi Y_{t-1}') \right\}\]
If $\Pi$ is assumed to be full rank ($r = p$) then the likelihood is maximized at the usual estimate - the least squares regression estimate

$$\hat{\Pi} = \sum_{t=1}^{n} (\nabla Y_t \ Y'_{t-1}) \left[ \sum_{t=1}^{n} Y_{t-1} Y'_{t-1} \right]^{-1} = S_{01} S_{11}^{-1}$$

and

$$\hat{\Lambda} = \frac{1}{n} \sum_{t=1}^{n} (\nabla Y_t - \hat{\Pi} Y_{t-1})(\nabla Y_t - \hat{\Pi} Y_{t-1})' = S_{00} - \hat{\Pi} S_{11} \hat{\Pi}'$$

$H_0 : \ r$ stationary linear combinations of $Y_t$ (linearly indep.) and thus (p-r) unit root linear combinations

$H_0 : \ r$ "cointegrating vectors" and (p-r) "common trends"

$H_0 : \ \Pi = -\alpha \beta'$ with $\alpha_{p\times r}$ and $\beta_{p\times r}$

So far we have the unrestricted estimate ($\hat{\Pi}, \hat{\Lambda}$) and can evaluate likelihood there. Principle of likelihood ratio requires that we maximize the likelihood for $\Pi = -\alpha \beta'$ and compare to the unrestricted maximum. That is, we now want to maximize

$$\mathcal{L} = \frac{1}{(2\pi)^{\frac{n}{2}} |\Lambda|^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} \sum_{t=1}^{n} (\nabla Y_t + \alpha \beta' Y_{t-1})' \Lambda^{-1} (\nabla Y_t + \alpha \beta' Y_{t-1}) \right\}$$

**Step 1:**

For any given $\beta$ we can compute $\beta' Y_{t-1}$ and find the corresponding $\alpha$ by regression in the model

$$\nabla Y_t = -\alpha (\beta' Y_{t-1}) + E_t$$

and this is simply

$$\hat{\alpha}(\beta) = -\sum_{t=1}^{n} (\nabla Y_t \ Y'_{t-1} \beta) \left[ \sum_{t=1}^{n} (\beta' Y_{t-1} Y'_{t-1} \beta) \right]^{-1}$$

**Step 2:**

Search over $\beta$ for maximum. To do this, plug $\hat{\alpha}(\beta)$ into the likelihood function which now becomes a function of $\beta$ and $\Lambda$. Now recall (from general regression) that
\[
\exp \left\{ -\frac{1}{2} \text{trace} \left[ X \left( \frac{1}{n} X'X \right)^{-1} X' \right] \right\} = \exp \left\{ -\frac{n}{2} \text{trace} \left[ (X'X)^{-1} X'X \right] \right\} = \exp \left\{ -\frac{n\rho}{2} \right\}
\]
where \( \exp \) is the exponential function and \( p = \text{rank}(X) \). In our case \( X \) has \( t \)th row \( (\nabla Y_t + \alpha \beta' Y_{t-1})' \) and by our usual maximum likelihood arguments we will, for any given \( \beta \), estimate \( \Lambda \) by

\[
\hat{\Lambda}(\beta) = n^{-1} X'X \quad \text{so that} \quad \mathcal{L} = \frac{1}{(2\pi)^{n/2} |\hat{\Lambda}(\beta)|^{1/2}} \exp \left\{ -\frac{n\rho}{2} \right\}
\]

Our goal now is to maximize \( \mathcal{L} \) which we would do by minimizing \(|\hat{\Lambda}(\beta)|\).

**Step 2 a**

Minimize \(|\hat{\Lambda}(\beta)| = |S_{00} - S_{01} \beta (\beta' S_{11} \beta)^{-1} \beta' S_{10}|\)

Recall for a 2x2 matrix we have

\[
\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc = \begin{cases} d(a - b d^{-1}c) \\ a(d - c a^{-1}b) \end{cases}
\]

and similarly for the determinant of a partitioned matrix. Thus

\[
\begin{vmatrix} S_{00} & S_{01} \beta \\ \beta' S_{10} & \beta' S_{11} \beta \end{vmatrix} = \begin{vmatrix} |\hat{\Lambda}(\beta)| & |\beta' S_{11} \beta| \\ |S_{00}| & |\beta' S_{11} \beta - \beta' S_{10} S_{00}^{-1} S_{01} \beta| \end{vmatrix}
\]

so our problem now becomes:

\[
\min_{\beta} \frac{|\hat{\Lambda}(\beta)|}{|S_{00}|} \Rightarrow \min_{\beta} \frac{|\beta' S_{11} \beta - \beta' S_{10} S_{00}^{-1} S_{01} \beta|}{|\beta' S_{11} \beta|} = *
\]

Recall: Cholesky Root

\[
S_{11} \text{ p.s.d. and symmetric} \Rightarrow S_{11} = U'U = \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix}
\]

(U upper triangular) \quad \text{SAS:}

```
PROC IML; U=ROOT(S11);
```

\[
* = \frac{\|\beta' U' (I - (U')^{-1} S_{10} S_{00}^{-1} S_{01} U^{-1}) U \beta\|}{\|\beta' U' U \beta\|}
\]

Note: We have seen that \( \Pi = -\alpha \beta' \) allows a lot of flexibility in choosing the columns of \( \beta \). (Corresponding adjustments in \( \alpha \) will preserve \( \Pi \).) We choose \( \beta' U' U \beta = I \).
Let $Z = U\beta$ 

\[ \beta = U^{-1}Z \]

Fact:

$\zeta_i$ is $i^{th}$ column of $Z$ is eigenvector of symmetric matrix so can get it in SAS.

$Z'(I - (U')^{-1} S_{10} S_{01}^{-1} S_{00} U^{-1})Z = diagonal$ matrix.

1. Cholesky on $S_{11}$

2. $(U')^{-1} S_{10} S_{01}^{-1} S_{00} U^{-1}$

* 3. EIGENVECTORS $\zeta_i$, EIGENVALUES ($\lambda_1 > \lambda_2 > \lambda_3 > \cdots > \lambda_p$)

4. $\hat{\beta} = U^{-1}Z$

5. Get $\alpha$ by regressing $\nabla Y_t$ on $(\hat{\beta}' Y_{t-1}) : \alpha' = (\hat{\beta}' S_{11} \hat{\beta})^{-1}(\hat{\beta}' S_{10})$

* Note: eigenvalues are called "squared canonical correlations" between $\nabla Y_t$ and $Y_{t-1}$.

PROC CANCOR will compute these for you.

Testing

Maximized $\mathcal{L}$ unconditionally
Maximized $\mathcal{L}$ under $H_0$.

Now look at likelihood ratio test.

Summary

1. Choose $\hat{\beta}$ to minimize $|\hat{\Lambda}(\beta)|$

2. $U$ invertible so any $\hat{\beta}$ is expressible as $\hat{\beta} = U^{-1}Z$ for some choice of $Z$.

3. Length of $\hat{\beta}$ vector is arbitrary so we can specify $Z'Z = I$.

4. Pick $Z$ to minimize $|Z'(U')^{-1}(S_{00} - S_{01} S_{11}^{-1} S_{10}) U^{-1}Z|$ and thus $Z$ would be picked as the matrix whose columns are those associated with the smallest (1-$\lambda_i$), that is, the largest "squared canonical correlations" $\lambda_i$.

\[
H_0 : r \leq r_0 \\
H_1 : r > r_0
\]

\[
LRT = \max_{\text{max } H_0(\mathcal{L})} \frac{|\hat{\Lambda}_0|^{n/2}}{|\hat{\Lambda}_1|^{n/2}} = \left[ \prod_{i=1}^{p} \frac{(1-\lambda_i)}{\prod_{i=1}^{r_0} (1-\lambda_i)} \right]^{n/2} = \left[ \prod_{i=r_0+1}^{p} (1 - \lambda_i) \right]^{n/2}
\]
Now in standard likelihood ratio testing we often take the log of the likelihood ratio. The reason for this is that it often leads to a Chi-square limit distribution. There is no hope of that happening in this nonstandard case, but Johansen still follows that tradition, suggesting that we reject when

\[- n \sum_{i=r_0+1}^{p} ln (1 - \lambda_i) \text{ is small, that is we will reject when} \]

\[- n \sum_{i=r_0+1}^{p} ln (1 - \lambda_i) \text{ is large where } \lambda_{r_0+1} > \lambda_{r_0+2} > \lambda_{r_0+3} > \cdots > \lambda_p \]

are the p-\(r_0\) smallest squared canonical correlations. This is Johansen's "trace" test.

To keep things straight .......

1. You use the smallest squared canonical correlations thus making \(1-\lambda_i\) large (nearer 1) and hence making \(-n \sum ln(1-\lambda_i)\) small (i.e. you select the \(\lambda_i\) that best protect \(H_0\))

In a later article, Johansen notes that you may have better power if you opt to test \(H_0 : r=r_0\) versus \(H_1 : r=r_0+1\) and thus use the largest (\(\lambda_{r_0+1}\)) of the smallest \(\lambda\)'s. This is Johansen's "maximal eigenvalue test."

2. Under \(H_0\) or \(H_1\) you have at least \(r_0\) cointegrating vectors and hence at most \(p-r_0\) "common trends." Therefore rejection of the null hypothesis means you have found yet another cointegrating vector.

3. The interpretation of a cointegrating vector is that you have found a linear combination of your vector components that cannot vary too far from 0 (i.e. you have a "law" that cannot be too badly violated). A departure from this relationship would be called an "error" and if we start at any point and forecast into the future with this model, the forecasts will eventually satisfy the relationship. Therefore this kind of model is referred to as an "error correction" model.

**Example:**

\[Z_{1t} = Z_{1,t-1} + e_{1t} \quad \text{We observe} \quad Y_{1t} = Z_{1t} + .9Z_{2t}\]

\[Z_{2t} = .8 Z_{2,t-1} + e_{2t} \quad \text{Y}_{2t} = Z_{1t} - .6Z_{2t}\]

Notice that \(Y_{1t} - Y_{2t} = 1.5Z_{2t}\) is stationary so we are saying that \(Y_{1t}\) can't wander too far from \(Y_{2t}\) and yet both \(Y_i\)'s are nonstationary. They are wandering around, but wandering around together, you might say.

Now in practice we would just observe the \(Y\)'s. Notice that

\[Z_t = \begin{pmatrix} 1 & 0 \\ 0 & .8 \end{pmatrix} Z_{t-1} + E_t \Rightarrow\]
\[
Y_t = \begin{pmatrix} 1 & .9 \\ 1 & .6 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & .8 \end{pmatrix} \begin{pmatrix} 1 & .9 \\ 1 & .6 \end{pmatrix}^{-1} Y_{t-1} + \text{noise}
\]

(exactly)

\[
= \begin{pmatrix} .88 & .12 \\ .08 & .92 \end{pmatrix} Y_{t-1} + \text{noise}
\]

Now suppose \( Y_{1t} = 12 \) and \( Y_{2t} = 2 \). These are not very close to each other and thus are in violation of the equilibrium condition \( Y_{1t} = Y_{2t} \). In the absence of future shocks, does the model indicate that this "error" will "correct" itself?

Error correction:

The next period we forecast \( \hat{Y}_{t+1} = \begin{pmatrix} .88 & .12 \\ .08 & .92 \end{pmatrix} \begin{pmatrix} 12 \\ 2 \end{pmatrix} = \begin{pmatrix} 10.8 \\ 2.8 \end{pmatrix} \) whose components are closer together.

Our two step ahead forecast is \( \begin{pmatrix} .88 & .12 \\ .08 & .92 \end{pmatrix} \begin{pmatrix} 10.8 \\ 2.8 \end{pmatrix} = \begin{pmatrix} 9.84 \\ 3.44 \end{pmatrix} \) and continuing one

finds that \( \begin{pmatrix} .88 & .12 \\ .08 & .92 \end{pmatrix}^{10} \begin{pmatrix} 12 \\ 2 \end{pmatrix} = \begin{pmatrix} 6.644 \\ 5.711 \end{pmatrix} \) and \( \begin{pmatrix} .88 & .12 \\ .08 & .92 \end{pmatrix}^{50} \begin{pmatrix} 12 \\ 2 \end{pmatrix} = \begin{pmatrix} 6.0001 \\ 5.9991 \end{pmatrix} \)

Let us take this example a step further. Modeling the changes in the series as a function of the lagged levels, we have

\[
\nabla Y_t = \begin{pmatrix} .88-1 & .12 \\ .08 & .92-1 \end{pmatrix} Y_{t-1} + \text{noise}
\]

\[
= \begin{pmatrix} - .12 \\ .08 \end{pmatrix}(1 - 1) Y_{t-1} + \text{noise}
\]

so we see that the discrepancy from equilibrium \( (1 - 1)^t Y_{t-1} \), which in our case is 10, is computed then .12 times this is subtracted from \( Y_1 \) and .08 times this is added to \( Y_2 \). The "speed of adjustment" is thus faster in \( Y_1 \) and we end up farther from the original \( Y_1 \) than from the original \( Y_2 \). Also, although the model implies (assuming 0 initial condition) that \( E\{Y_1\} = E\{Y_2\} = 0 \), there is nothing drawing the series back toward their theoretical means (0). Shocks to this series have a permanent effect on the levels of the series, but only a temporary effect on the relationship (equality in this model) between the series components.
**Remaining to do:**

Q1: What are the critical values for the test?
Q2: What if we have more than 1 lag?
Q3: What if we include intercepts, trends, etc.

**Q1.** Usually the likelihood ratio test has a limit Chi-square distribution. For example a regression $F_n^4$ test is such that $4F_n^4 \xrightarrow{d} \chi^2_4$. Also $F_n^1 = t^2 \xrightarrow{d} \chi^2$ is a special case. Now we have seen that the t statistic, $\tau$, has a nonstandard limit distribution expressible as a functional of Brownian Motion $B(t)$ on [0,1]. We found:

$$\tau = \frac{\int_0^1 B(t) \, dB(t)}{\left[ \int_0^1 B^2(t) \, dt \right]^\frac{1}{2}} = \frac{\frac{1}{2}[B^2(1) - 1]}{\left[ \int_0^1 B^2(t) \, dt \right]^\frac{1}{2}}$$

and we might thus expect Johansen's test statistic to converge to a multivariate analogue of this expression. Indeed, Johansen proved that his likelihood ratio (trace) test converges to a variable that can be expressed as a functional of a vector valued Brownian Motion with independent components (channels) as follows (i.e. the error term has variance matrix $\sigma^2$)

$$\text{LRT} \xrightarrow{d} \text{trace} \left\{ \left[ \int_0^1 B(t) \, dB(t) \right]' \left[ \int_0^1 B(t) \, dB'(t) \, dt \right]^{-1} \left[ \int_0^1 B(t) \, dB(t) \right] \right\}$$

For a Brownian motion of dimension $m = 1,2,3,4,5$ (Table 1, page 239 of Johansen) he computes the distribution of the LRT by Monte Carlo. Empirically he notes that these percentiles are quite close to the percentiles of $c \chi^2_f$ where $f=2m^2$ and $c=0.85-.58/f$.

We will not repeat Johansen's development of the limit distribution of the LRT but will simply note that it is what would be expected by one familiar with the usual limit results for F statistics and with the nonstandard limit distributions that arise with unit root processes. As one might expect, the $m=1$ case can be derived from the $\tau$ distribution.

**Q2.** What happens in higher order processes?

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + A_3 Y_{t-3} + \cdots + A_k Y_{t-k} + E_t$$

$$\nabla Y_t = -(I - A_1 - A_2 - \cdots - A_k) Y_{t-1} - (A_2 + A_3 + \cdots + A_k) \nabla Y_{t-1} - (A_3 + A_4 + \cdots + A_k) \nabla Y_{t-2} - \cdots - A_k \nabla Y_{t-k+1} + E_t$$

which has the form

$$\nabla Y_t = -(I - A_1 - A_2 - \cdots - A_k) Y_{t-1} + B_1 \nabla Y_{t-1} + B_2 \nabla Y_{t-2} + \cdots + B_k \nabla Y_{t-k+1} + E_t$$
The characteristic equation is \( |I - A_1 - A_2 - \cdots - A_k| = 0. \) Now if \( m=1 \) is a root of this, we have \( |I - A_1 - A_2 - \cdots - A_k| = 0. \) The number of unit roots in the system is the number of 0 roots for the matrix \((I - A_1 - A_2 - \cdots - A_k)\) and the rank of this matrix is the number of cointegrating vectors. I am assuming a vector autoregression in which each component has at most one unit root (i.e. differencing makes each component stationary) here. While this parameterization most closely resembles the usual unit root testing framework, Johansen chooses an equivalent way to parameterize the model, placing the lagged levels at lag \( k \) instead of lag 1. His model is written:

\[
\nabla Y_t = - (I - A_1) \nabla Y_{t-1} - \cdots - (I - A_1 - A_2 - \cdots - A_{k-1}) \nabla Y_{t-k+1} - (I - A_1 - A_2 - \cdots - A_k) Y_{t-k} + E_t
\]

and he is checking the rank of \( \Pi = (I - A_1 - A_2 - \cdots - A_k) \).

Recall: Regression can be done in 3 steps. Suppose we are regressing \( Y \) on \( X_1, X_2, \ldots, X_{k-1}, X_k \) and we are interested in the coefficient (matrix) of \( X_k \). We can get that coefficient matrix in 3 steps:

- **Step 1:** Regress \( Y \) on \( X_1, X_2, \ldots, X_{k-1} \) --- residuals \( R_Y \)
- **Step 2:** Regress \( X_k \) on \( X_1, X_2, \ldots, X_{k-1} \) --- residuals \( R_k \)
- **Step 3:** Regress \( R_Y \) on \( R_k \) -- gives same coefficient matrix as in full regression.

so Johansen does the following in higher order models:

- **Step 1:** Regress \( \nabla Y_t \) on \( \nabla Y_{t-1}, \nabla Y_{t-2}, \ldots, \nabla Y_{t-k+1} \) --- residuals \( R_{\nabla Y} \)
- **Step 2:** Regress \( Y_{t-k} \) on \( \nabla Y_{t-1}, \nabla Y_{t-2}, \ldots, \nabla Y_{t-k+1} \) --- residuals \( R_k \)
- **Step 3:** Squared canonical correlations between \( R_{\nabla Y} \) on \( R_k \)

The idea is very nice. Johansen maximizes the likelihood for any \( \Pi_k \) with respect to the other parameters by performing steps 1 and 2. Having done this, he's back to a lag 1 type of problem.

By analogy, if in ordinary unit root testing the null hypothesis were true, you could estimate the autoregressive coefficients consistently by regressing the first difference on the lagged first differences. Having these estimates, \( \hat{\alpha}_1, \ldots, \hat{\alpha}_p \), you could compute a "filtered" version of \( Y \), namely \( \hat{Z}_t = Y_t - \hat{\alpha}_1 Y_{t-1} - \cdots - \hat{\alpha}_p Y_{t-p} \) and under the null hypothesis, \( Z_t = Y_t - \alpha_1 Y_{t-1} - \cdots - \alpha_p Y_{t-p} \) is a random walk so you could regress \( \hat{Z}_t \) on \( \hat{Z}_{t-1} \) and, in large samples, compare the results to our D-F unit root tables both for the coefficient and t statistic. Note that all of Johansen's statistics are multivariate analogues of \( \tau \) - there is nothing like the "normalized bias test" \( n(\hat{\rho} - 1) \).
Example:

\[ \mathbf{Y}_t = A_1 \mathbf{Y}_{t-1} + A_2 \mathbf{Y}_{t-2} + A_3 \mathbf{Y}_{t-3} + \mathbf{E}_t, \quad \mathbf{Y}'_t = (Y_{1t}, Y_{2t}, Y_{3t}, Y_{4t}) \]

100 observations

Step 1: Regress \( \nabla \mathbf{Y}_t \) on \( \nabla \mathbf{Y}_{t-1}, \nabla \mathbf{Y}_{t-2} \)
Step 2: Regress \( \mathbf{Y}_{t-3} \) on \( \nabla \mathbf{Y}_{t-1}, \nabla \mathbf{Y}_{t-2} \)
Step 3: Squared canonical correlations 0.010, 0.020, 0.08, 0.460

Test \( H_0 : r = 0 \) vs. \( H_1 : r > 0 \).
\( r_0 = 0, \ p=4, \) so use 4-0 = 4 smallest canonical correlations.
LRT = -100 \[ \ln(0.99) + \ln(0.98) + \ln(0.92) + \ln(0.54) \] = 72.98
Johansen, Table 1, gives critical value for \( m=4 \) (\( H_0 \) implies \( m=4 \) common trends)
Critical value is 41.2. Reject \( H_0 \). There is at least 1 cointegrating vector.

Test \( H_0 : r = 1 \) vs. \( H_1 : r > 1 \).
\( r_0 = 1, \ p=4, \) so use 4-1 = 3 smallest canonical correlations.
LRT = -100 \[ \ln(0.99) + \ln(0.98) + \ln(0.92) \] = 11.36
(look in Johansen's table under \( m=3 \) common trends)
Critical value is 23.8. Do not reject \( H_0 \). There are no more cointegrating vectors.

Q3 Johansen later wrote a paper addressing the intercept case. The trend issue is a bit tricky as is true in the univariate case. Even the intercept case has some interesting features in terms of how intercepts in the model for the observed series translate into the underlying canonical series (common trends).

An additional note: We are used to taking \(-2 \ln(L)\) in maximum likelihood estimation and likelihood ratio testing. We do this because, under certain regularity conditions, it produces test statistics with standard limiting distributions, (Chi-square usually). In this case, we do not have the required regularity conditions for the test of \( r \) (number of cointegrating vectors) otherwise, Johansen would not have needed to do any new tabulations. Because this is the case, Johansen could have opted to use other functions of the eigenvalues such as \( n \sum \lambda_i \) in place of his \( -n \sum \ln(1 - \lambda_i) \) statistic. We show (for the case of a univariate series - just one \( \lambda \)) that both of these statistics converge to the same distribution and it is the distribution \( r^2 \). First note that for the series \( \hat{Y}_t = Y_{t-1} + e_t \) we have \( \nabla \hat{Y}_t = \hat{Y}_t - Y_{t-1} = e_t \) so our regression MSE is \( \sum e_t^2 / n = S_{00}, \) our standard error is \( \sqrt{MSE/\sum Y_{t-1}^2} = \sqrt{S_{00}/(n S_{11})} \) and \( \hat{\rho} - 1 = S_{01}/S_{11} \) so
Now the "Cholesky root" of the univariate variable $S_{11}$ is, of course, just $\sqrt{S_{11}}$ and Johansen looks at the matrix $\sqrt{S_{11}} \left( 1 - \frac{S_{01}}{\sqrt{S_{11}}} \frac{1}{\sqrt{S_{00}}} \frac{S_{01}}{S_{11}} \right) \sqrt{S_{11}} = \sqrt{S_{11}}(1 - \frac{\tau^2}{n}) \sqrt{S_{11}}$ calling its eigenvalues $1-\lambda_i$. We see immediately that $n \lambda_i$ is just $\tau^2$ so we already know its distribution (the multivariate cases are analogous).

This also shows that $\lambda_i = O_p(1/n)$ so if we expand Johansen's statistic, using the Taylor series for $\ln(1 + x)$ expanded around $x = 0$, we then have

$$-n \ln(1 - \lambda_i) = -n(\ln(1) + \frac{1}{1+0}( -\lambda_i ) + O_p(1/n^2) )$$

from which we see that $-n \ln(1 - \lambda_i) = n\lambda_i + O_p(1/n)$ proving, as claimed, that these two statistics have the same limit distribution (of course there are some extra details needed for the multivariate case). Notice that for a single $\lambda$ these two statistics are monotone transforms of each other so even in finite samples, provided we had the right distributions, they would give exactly equivalent tests. For the more interesting multivariate case, they are the same only in the limit.

Demo: Amazon.com High and Low stock prices

```
DATA AMAZON;
  INPUT DATE OPEN HIGH LOW CLOSE VOLUME;
  TITLE "DATA ON AMAZON.COM STOCK";
  /** DATA FROM INTERNET, YAHOO SITE ****/ ;
  
  CLOSE_LEVEL=CLOSE;  VOL_LEVEL=VOLUME;
  OPEN=LOG(OPN);  CLOSE=LOG(CLOSE);  HIGH=LOG(HIGH);
  LOW=LOG(LW);  VOLUME=LOG(VOLUME);
  TITLE2 "DATA IN LOGARITHMS";
  HHAT=LOW+.076;
  SPREAD=HIGH-LOW;
  FORMAT DATE DATE7.;
  CARDS;
  14389 117.25   121.125 111.375  111.5625 7755100
  14388 128.625  129.375 116      117.5    7126900
  (more data)
```
RUN;
PROC SORT;
  BY DATE;
RUN;
PROC ARIMA DATA=AMAZON;
  I VAR=SPREAD STATIONARITY = (ADF=(2));
  E P=3;
RUN;

The spread = log(high)-log(low) seems stationary:

### Augmented Dickey-Fuller Unit Root Tests

<table>
<thead>
<tr>
<th>Type</th>
<th>Lags</th>
<th>Tau</th>
<th>Pr &lt; Tau</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero Mean</td>
<td>2</td>
<td>-3.00</td>
<td>0.0028</td>
</tr>
<tr>
<td>Single Mean</td>
<td>2</td>
<td>-7.65</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Trend</td>
<td>2</td>
<td>-8.05</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

### Conditional Least Squares Estimation

| Parameter | Estimate     | Standard Error | t Value | Pr > |t| |
|-----------|--------------|----------------|---------|------|-----|
| MU        | 0.07652      | 0.0043870      | 17.44   | <.0001|
| AR1,1     | 0.38917      | 0.04370        | 8.91    | <.0001|
| AR1,2     | 0.04592      | 0.04702        | 0.98    | 0.3293|
| AR1,3     | 0.18888      | 0.04378        | 4.31    | <.0001|

Constant Estimate 0.028775
Variance Estimate 0.00141
Std Error Estimate 0.03755
The fit seems fine so we expect to find log(high)-log(low) = cointegration vector:

<table>
<thead>
<tr>
<th>Lag Square</th>
<th>ChiSq</th>
<th>DF</th>
<th>Autocorrelations</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>4.63</td>
<td>3</td>
<td>-0.011 -0.018 -0.009 -0.047 0.072 -0.033</td>
</tr>
<tr>
<td>12</td>
<td>9.43</td>
<td>9</td>
<td>0.037 0.041 0.035 0.025 0.029 0.058</td>
</tr>
<tr>
<td>18</td>
<td>12.71</td>
<td>15</td>
<td>-0.011 -0.046 0.025 -0.014 0.016 0.052</td>
</tr>
<tr>
<td>24</td>
<td>21.14</td>
<td>21</td>
<td>0.017 0.023 -0.067 -0.074 0.059 0.038</td>
</tr>
<tr>
<td>30</td>
<td>25.13</td>
<td>27</td>
<td>0.025 0.049 0.014 -0.012 -0.006 0.063</td>
</tr>
<tr>
<td>36</td>
<td>28.86</td>
<td>33</td>
<td>0.038 0.049 -0.023 0.016 -0.045</td>
</tr>
<tr>
<td>42</td>
<td>33.05</td>
<td>39</td>
<td>0.049 0.055 0.023 0.010 0.039 0.003</td>
</tr>
<tr>
<td>48</td>
<td>36.51</td>
<td>45</td>
<td>0.030 -0.035 -0.050 -0.038 0.006 -0.004</td>
</tr>
</tbody>
</table>

Model for variable SPREAD
Estimated Mean 0.076524

Autoregressive Factors
Factor 1: 1 - 0.38917 B**(1) - 0.04592 B**(2) - 0.18888 B**(3)

Try Johansen's method (PROC VARMAX). Lots of output produced.
Individual series should be nonstationary:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Rho</th>
<th>Pr &lt; Rho</th>
<th>Tau</th>
<th>Pr &lt; Tau</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH</td>
<td>Zero Mean</td>
<td>0.84</td>
<td>0.8848</td>
<td>1.86</td>
<td>0.9854</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-1.02</td>
<td>0.8844</td>
<td>-0.83</td>
<td>0.8082</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-17.67</td>
<td>0.1083</td>
<td>-2.86</td>
<td>0.1780</td>
</tr>
<tr>
<td>LOW</td>
<td>Zero Mean</td>
<td>0.84</td>
<td>0.8846</td>
<td>1.72</td>
<td>0.9795</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-1.13</td>
<td>0.8735</td>
<td>-0.87</td>
<td>0.7987</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-22.19</td>
<td>0.0426</td>
<td>-3.22</td>
<td>0.0823</td>
</tr>
</tbody>
</table>

There follow cross covariances, cross-correlations etc.

Next: VAR representation (3 lags, 2x2 coefficient matrices) and partial autoregressive matrices - i.e. fit 1 then 2 then 3 vector lags and report last coefficient matrix.

<table>
<thead>
<tr>
<th>Lag</th>
<th>Variable</th>
<th>HIGH</th>
<th>LOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HIGH</td>
<td>0.93267</td>
<td>0.14810</td>
</tr>
<tr>
<td></td>
<td>LOW</td>
<td>0.56697</td>
<td>0.51960</td>
</tr>
<tr>
<td>2</td>
<td>HIGH</td>
<td>-0.09905</td>
<td>0.02425</td>
</tr>
<tr>
<td></td>
<td>LOW</td>
<td>-0.13872</td>
<td>0.05341</td>
</tr>
<tr>
<td>3</td>
<td>HIGH</td>
<td>0.17072</td>
<td>-0.18205</td>
</tr>
<tr>
<td></td>
<td>LOW</td>
<td>0.00254</td>
<td>-0.01333</td>
</tr>
</tbody>
</table>

Schematic Representation of Partial Autoregression

<table>
<thead>
<tr>
<th>Lag</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH</td>
<td>+.</td>
<td>..</td>
<td>..</td>
<td>..</td>
<td>..</td>
</tr>
<tr>
<td>LOW</td>
<td>++</td>
<td>..</td>
<td>..</td>
<td>..</td>
<td>..</td>
</tr>
</tbody>
</table>

+ is > 2*std error, - is < -2*std error, . is between
Next, the partial canonical correlations and Johansen trace tests

<table>
<thead>
<tr>
<th>Lag</th>
<th>Correlation1</th>
<th>Correlation2</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.99521</td>
<td>0.42734</td>
<td>4</td>
<td>595.91</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>2</td>
<td>0.11409</td>
<td>0.07557</td>
<td>4</td>
<td>9.49</td>
<td>0.0499</td>
</tr>
<tr>
<td>3</td>
<td>0.17292</td>
<td>0.01049</td>
<td>4</td>
<td>15.19</td>
<td>0.0043</td>
</tr>
<tr>
<td>4</td>
<td>0.09158</td>
<td>0.03170</td>
<td>4</td>
<td>4.74</td>
<td>0.3148</td>
</tr>
<tr>
<td>5</td>
<td>0.08729</td>
<td>0.01140</td>
<td>4</td>
<td>3.91</td>
<td>0.4189</td>
</tr>
</tbody>
</table>

**Cointegration Rank Test Using Trace**

<table>
<thead>
<tr>
<th>Rank=r</th>
<th>Rank&gt;r</th>
<th>Critical Value</th>
<th>Drift in Eigenvalue</th>
<th>Trace Value</th>
<th>Drift in ECM</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.1204</td>
<td>65.4985</td>
<td>15.34</td>
<td>Constant</td>
<td>Linear</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.0013</td>
<td>0.6559</td>
<td>3.84</td>
<td>Constant</td>
<td>Constant</td>
</tr>
</tbody>
</table>

**Cointegration Rank Test Using Trace Under Restriction**

<table>
<thead>
<tr>
<th>Rank=r</th>
<th>Rank&gt;r</th>
<th>Critical Value</th>
<th>Drift in Eigenvalue</th>
<th>Trace Value</th>
<th>Drift in ECM</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.1204</td>
<td>71.1499</td>
<td>19.99</td>
<td>Constant</td>
<td>Constant</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.0123</td>
<td>6.2589</td>
<td>9.13</td>
<td>Constant</td>
<td>Linear</td>
</tr>
</tbody>
</table>

**Hypothesis of the Restriction**

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Drift</th>
<th>Drift in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis in ECM</td>
<td>Process</td>
<td></td>
</tr>
<tr>
<td>H0</td>
<td>Constant</td>
<td>Constant</td>
</tr>
<tr>
<td>H1</td>
<td>Constant</td>
<td>Linear</td>
</tr>
</tbody>
</table>
The restriction is that the cointegrating vector annihilates the intercept term hence the drift is constant in the "error correction mechanism" as well as in the process itself in analogy to $Y_t = 0 + \rho Y_{t-1} + \epsilon_t$. Without the restriction, an intercept appears in the ECM but a linear trend appears in the process in analogy to $Y_t = 2 + \rho Y_{t-1} + \epsilon_t$, which has either a mean $2/(1-\rho)$ or drift 2 per time period depending on whether $\rho$ is less than 1 or equal to 1. Either way, we decide there is more than 0 (Trace > Critical Value) but not more than 1 (Trace < Critical Value).

Now we may be interested in testing whether this restriction holds. To do this we have to take a stand on the cointegrating rank. A rough plot suggests drift in the underlying processes

**Plot of HIGH*DATE.** Symbol used is 'H'.

**Plot of LOW*DATE.** Symbol used is 'L'.

**NOTE:** 927 obs hidden.
The tests for ranks 0 and 1 are given. We are interested in the rank 1 test as we believe the cointegrating rank is indeed 1 and the two tests disagree on whether or not the restriction (that the cointegrating vector annihilates the intercept) holds. For the rank 1 model, we reject the restriction (as the graph above seems to suggest).

### Hypothesis Test of the Restriction

<table>
<thead>
<tr>
<th>Rank</th>
<th>Restricted Eigenvalue</th>
<th>Restricted Eigenvalue</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1203</td>
<td>0.1204</td>
<td>2</td>
<td>5.65</td>
<td>0.0593</td>
</tr>
<tr>
<td>1</td>
<td>0.0013</td>
<td>0.0123</td>
<td>1</td>
<td>5.60</td>
<td>0.0179</td>
</tr>
</tbody>
</table>

There follow estimates of the $\alpha$ and $\beta$ matrices without the restriction and under the restriction. From the above, we are interested in the results for the unrestricted model and we think the rank is 1 so we are also interested only in the first columns of $\alpha$ and $\beta$ from which the 2x2 impact matrix $\pi = \alpha\beta'$ is computed.

### Long-Run Parameter Beta Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH</td>
<td>1.00000</td>
<td>1.00000</td>
</tr>
<tr>
<td>LOW</td>
<td>-1.01036</td>
<td>-0.24344</td>
</tr>
</tbody>
</table>

### Adjustment Coefficient Alpha Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH</td>
<td>-0.06411</td>
<td>-0.00209</td>
</tr>
<tr>
<td>LOW</td>
<td>0.35013</td>
<td>-0.00174</td>
</tr>
</tbody>
</table>

In addition a 3x3 matrix $\beta$ and a 2x3 matrix $\alpha$ which relate the 2-vector $(H_t - H_{t-1}, L_t - L_{t-1})'$ to the 3-vector $(H_{t-1}, L_{t-1}, 1)'$ when the restriction holds. We are not interested in that one as we have rejected the restriction. So without the restriction, our impact matrix $\pi = \alpha\beta'$ becomes

\[
\begin{pmatrix}
-0.06 \\
0.35
\end{pmatrix}
\begin{pmatrix}
1.00 \\
-1.01
\end{pmatrix}
= \begin{pmatrix}
-0.064 & 0.065 \\
0.350 & -0.354
\end{pmatrix}.
\]
Estimates of the intercepts are given as well:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH</td>
<td>0.00857</td>
</tr>
<tr>
<td>LOW</td>
<td>-0.01019</td>
</tr>
</tbody>
</table>

Parameter Alpha * Beta' Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>HIGH</th>
<th>LOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH</td>
<td>-0.06411</td>
<td>0.06478</td>
</tr>
<tr>
<td>LOW</td>
<td>0.35013</td>
<td>-0.35376</td>
</tr>
</tbody>
</table>

Combining these items we have the following:

\[
\begin{pmatrix}
H_t - H_{t-1} \\
L_t - L_{t-1}
\end{pmatrix} = \begin{pmatrix}
0.00857 \\
-0.01019
\end{pmatrix} + \begin{pmatrix}
-0.06 \\
0.35
\end{pmatrix} \begin{pmatrix}
1.00 & -1.01
\end{pmatrix} \begin{pmatrix}
H_{t-1} \\
L_{t-1}
\end{pmatrix} +
\]

+ lagged difference terms + e

where the 2x2 estimate of the covariance matrix \( \Sigma \) for e also appears on the printout as well as the 2x2 estimated coefficient matrices for the lagged differences. A schematic representation of the vector autoregressive model is shown as well.
The next result is a bit surprising. Even though we showed that 
\[(1.00 \quad -1.00)(H_{t-1} \quad L_{t-1})'\]
is stationary (if we choose 1 and -1 beforehand and thus allowed the use of ordinary unit root tests), even though 
\[(1.00 \quad -1.00)\] and 
\[(1.00 \quad -1.01)\] are almost the same (perhaps the same from a practical if not statistical perspective) and even though we have shown that there is 1 cointegrating vector, we will REJECT the hypothesis that the cointegrating vector takes the form of a difference. Since we rejected here, the disagreement cannot be attributed to low power of the test. Here is the relevant piece of output:

\[
\begin{array}{ccc}
\text{Variable} & 1 & 2 \\
\hline
\text{HIGH} & 1.00000 & 0.00000 \\
\text{LOW} & -1.00000 & 0.00000 \\
1 & 0.00000 & 1.00000 \\
\end{array}
\]

\[
\begin{array}{ccc}
\text{Variable} & 1 & 2 \\
\hline
\text{HIGH} & 1.00000 & \\
\text{LOW} & -1.00000 & \\
1 & -0.01706 & \\
\end{array}
\]
Adjustment Coefficient
Alpha with Respect to
Hypothesis on Beta

Variable 1

HIGH 0.05334
LOW 0.13957

Test for Restricted Long-Run Coefficient Beta

<table>
<thead>
<tr>
<th>Index</th>
<th>Restricted Eigenvalue</th>
<th>Eigenvalue</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1204</td>
<td>0.1038</td>
<td>1</td>
<td>9.44</td>
<td>0.0021</td>
</tr>
</tbody>
</table>

Much more output is produced, including a large array of residual diagnostics, information criteria, impulse response functions, etc.