Unit roots in vector time series

A. Vector autoregressions with unit roots

Scalar autoregression

True model:

$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \cdots + \phi_p \Delta y_{t-p} + \varepsilon_t$

Estimated model:

$\Delta y_t = c + \eta y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \cdots$
$+ \phi_p \Delta y_{t-p} + \varepsilon_t$

Results:

$\sqrt{T}\,(\hat{\phi}_j - \phi_{j0})$ is asymptotically normal (same distribution as if imposed $\eta = 0$)

$T(\hat{\eta} - \eta_0)$ is nonstandard

If instead estimate in the form

$$y_t = c + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots$$
$$+\rho_{p+1} y_{t-p-1} + \varepsilon_t$$

then

$\sqrt{T}(\hat{\rho}_j - \rho_{j0})$ is asymptotically normal (same distribution as if imposed $\rho_1 + \rho_2 + \cdots + \rho_{p+1} = 1$)

vector $\mathbf{y}_t = (n \times 1)$

True model:

$$\Delta \mathbf{y}_t = \boldsymbol{\phi}_1 \Delta \mathbf{y}_{t-1} + \boldsymbol{\phi}_2 \Delta \mathbf{y}_{t-2} + \cdots$$
$$+ \boldsymbol{\phi}_p \Delta \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

Estimated model:

$$\Delta \mathbf{y}_t = \mathbf{c} + \boldsymbol{\eta} \mathbf{y}_{t-1} + \boldsymbol{\phi}_1 \Delta \mathbf{y}_{t-1} + \boldsymbol{\phi}_2 \Delta \mathbf{y}_{t-2} + \cdots$$
$$+ \boldsymbol{\phi}_p \Delta \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

$\boldsymbol{\eta} = (n \times n)$ matrix (true $\boldsymbol{\eta} = \mathbf{0}$)

Results:

$\sqrt{T}$ times elements $\hat{\phi}_{ij}$ asymptotically normal (same distribution as if imposed $\boldsymbol{\eta} = \mathbf{0}$)

T times elements $\hat{\eta}_{ij}$ nonstandard

If we instead estimate VAR in levels:

$$\mathbf{y}_t = \mathbf{c} + \boldsymbol{\rho}_1 \mathbf{y}_{t-1} + \boldsymbol{\rho}_2 \mathbf{y}_{t-1} + \cdots + \boldsymbol{\rho}_{p+1} \mathbf{y}_{t-p-1} + \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\rho}_1 = \mathbf{I}_n + \boldsymbol{\eta} + \boldsymbol{\Phi}_1$$

$$\boldsymbol{\rho}_2 = \boldsymbol{\Phi}_2 - \boldsymbol{\Phi}_1$$

$$\vdots$$

$$\boldsymbol{\rho}_p = \boldsymbol{\Phi}_p - \boldsymbol{\Phi}_{p-1}$$

$$\boldsymbol{\rho}_{p+1} = -\boldsymbol{\Phi}_p$$

Examples of statistics with standard distributions

Any individual element of $\hat{\boldsymbol{\rho}}$ converges at $\sqrt{T}$ to Normal

Any t test on individual element is asymptotically $N(0,1)$

$H_0 : \boldsymbol{\rho}_{p+1} = \mathbf{0}$ (i.e., p lags sufficient) identical to test of $H_0 : \Phi_p = \mathbf{0}$ and is asymptotically valid

Examples of statistics with nonstandard distributions

(1) Hypothesis involving sum of coefficients:

$$\hat{\boldsymbol{\rho}}_1 + \hat{\boldsymbol{\rho}}_2 + \cdots \hat{\boldsymbol{\rho}}_{p+1} = \mathbf{I}_n + \hat{\boldsymbol{\eta}}$$

(2) Hypothesis that variable 2 does not Granger-cause variable 1

$$H_0 : \rho_1^{(2,1)} = \rho_2^{(2,1)} = \cdots = \rho_{p+1}^{(2,1)} = 0$$

$$\Leftrightarrow \eta^{(2,1)} = \phi_1^{(2,1)} = \phi_2^{(2,1)} = \cdots = \phi_p^{(2,1)} = 0$$

Conclusion: if true model should be a VAR in differences (impose $\eta = \mathbf{0}$) and you instead estimate as VAR in levels, it is not a capital crime.

Problems will be

- loss of efficiency if don't impose a true restriction
- certain hypothesis tests have nonstandard distribution

Unit roots in vector time series

A. Vector autoregressions with unit roots

B. Spurious regressions

What about spurious regression?
Example:

$$\Delta y_{1t} = \varepsilon_{1t}$$
$$\Delta y_{2t} = \varepsilon_{2t}$$
$$\boldsymbol{\varepsilon}_t \sim \text{i.i.d.}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right)$$

$\Rightarrow y_{1t}, y_{2t}$ completely unrelated

Suppose we regress

$$y_{1t} = \alpha + \gamma y_{2t} + u_t$$
$$\begin{bmatrix} \hat{\alpha} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \sum 1 & \sum y_{2t} \\ \sum y_{2t} & \sum y_{2t}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_{1t} \\ \sum y_{2t} y_{1t} \end{bmatrix}$$

true $\alpha = \gamma = 0$

$$\Upsilon_T = \begin{bmatrix} T^{-1/2} & 0 \\ 0 & 1 \end{bmatrix}$$

will analyze $\Upsilon_T \hat{\boldsymbol{\beta}}$

Note radical departure from usual

$$\Upsilon_T = \begin{bmatrix} T^{1/2} & 0 \\ 0 & T^{1/2} \end{bmatrix}$$

$\hat{\gamma}$ does not converge to anything!

$\hat{\alpha}$ diverges to $\pm\infty$!

$$\begin{bmatrix} T^{-1/2} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} T^{-1/2} & 0 \\ 0 & 1 \end{bmatrix} \times$$

$$\begin{bmatrix} T & \sum y_{2t} \\ \sum y_{2t} & \sum y_{2t}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-3/2} & 0 \\ 0 & T^{-2} \end{bmatrix}^{-1} \times$$

$$\begin{bmatrix} T^{-3/2} & 0 \\ 0 & T^{-2} \end{bmatrix} \begin{bmatrix} \sum y_{1t} \\ \sum y_{2t} y_{1t} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & T^{-3/2} \sum y_{2t} \\ T^{-3/2} \sum y_{2t} & T^{-2} \sum y_{2t}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-3/2} \sum y_{1t} \\ T^{-2} \sum y_{2t} y_{1t} \end{bmatrix}$$

$$\xrightarrow{L} \begin{bmatrix} 1 & \sigma_2 \int_0^1 W_2(r)dr \\ \sigma_2 \int_0^1 W_2(r)dr & \sigma_2^2 \int_0^1 [W_2(r)]^2 dr \end{bmatrix}^{-1} \times$$

$$\begin{bmatrix} \sigma_1 \int_0^1 W_1(r)dr \\ \sigma_1 \sigma_2 \int_0^1 W_1(r) W_2(r)dr \end{bmatrix}$$

Can further show OLS $\hat{\sigma}_{\hat{\gamma}} \xrightarrow{p} 0$

$\Rightarrow$ OLS t test of $H_0 : \gamma = 0$ diverges to $\pm\infty$ even though H_0 is true!

If regress one trend on another trend with no time trend in regression: $R^2 \to 1$.
If regress one random walk on another random walk with no lags in regression: $R^2 \to 1$.
Recommendation: never do a spurious regression.

Cure for spurious regression: include lags of y_1 and y_2:
$y_{1t} = c + \beta_1 y_{2t} + \beta_2 y_{2,t-1} + \beta_3 y_{1,t-1} + \varepsilon_t$
This is equivalent to estimating
$\Delta y_{1t} = c + \gamma_1 \Delta y_{2t} + \gamma_2 y_{2,t-1} + \gamma_3 y_{1,t-1} + \varepsilon_t$

$\sqrt{T}\,\hat{\gamma}_1$ will be asymptotically normal (and same distribution as if imposed $\gamma_2 = \gamma_3 = 0$)
$T\hat{\gamma}_2$ and $T\hat{\gamma}_3$ will be nonstandard

$$y_{1t} = c + \beta_1 y_{2t} + \beta_2 y_{2,t-1} + \beta_3 y_{1,t-1} + \varepsilon_t$$
$$\Delta y_{1t} = c + \gamma_1 \Delta y_{2t} + \gamma_2 y_{2,t-1} + \gamma_3 y_{1,t-1} + \varepsilon_t$$
$$\hat{\beta}_1 = \hat{\gamma}_1$$
$$\sqrt{T}\hat{\beta}_1 \xrightarrow{L} N(0, V_1)$$
$$\hat{\beta}_2 = \hat{\gamma}_2 - \hat{\gamma}_1$$
$$\sqrt{T}\hat{\beta}_2 = \sqrt{T}(\hat{\gamma}_2 - \hat{\gamma}_1)$$
$$\xrightarrow{L} -\sqrt{T}\hat{\gamma}_1 \xrightarrow{L} N(0, V_1)$$

Conclusion: the capital crime would be not to include lags in regression

$y_{1t} = \alpha + \gamma y_{2t} + u_t$

True u_t (for $\alpha = \gamma = 0$) is infinitely serially correlated (random walk)

Unit roots in vector time series

A. Vector autoregressions with unit roots
B. Spurious regression
C. Cointegration– single equation methods

$$\Delta \mathbf{y}_t = \mathbf{c} + \boldsymbol{\eta}\mathbf{y}_{t-1} + \boldsymbol{\phi}_1 \Delta \mathbf{y}_{t-1} + \boldsymbol{\phi}_2 \Delta \mathbf{y}_{t-2} + \cdots + \boldsymbol{\phi}_p \Delta \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

When $\boldsymbol{\eta} = \mathbf{0}$ this becomes a VAR in differences

When $\boldsymbol{\eta}$ is unrestricted this becomes a VAR in levels

Intermediate case:

$0 < \text{rank}(\boldsymbol{\eta}) < n$

Example:

$\boldsymbol{\eta} = \mathbf{ab}'$ for $\mathbf{a}$ and $\mathbf{b}$ $(n \times 1)$ vectors

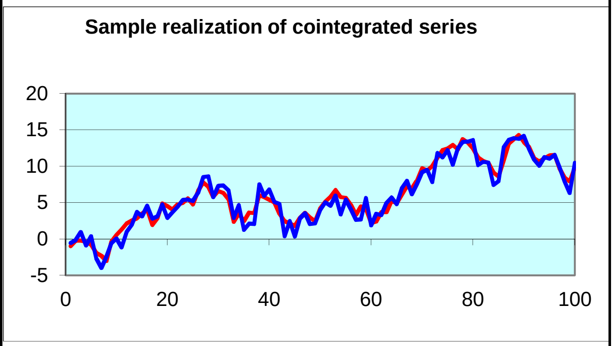
$$\Delta \mathbf{y}_t = \mathbf{c} + \boldsymbol{\eta}\mathbf{y}_{t-1} + \boldsymbol{\phi}_1 \Delta \mathbf{y}_{t-1} + \boldsymbol{\phi}_2 \Delta \mathbf{y}_{t-2} + \cdots + \boldsymbol{\phi}_p \Delta \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

$$\Delta \mathbf{y}_t = \mathbf{c} + \mathbf{ab}'\mathbf{y}_{t-1} + \boldsymbol{\phi}_1 \Delta \mathbf{y}_{t-1} + \boldsymbol{\phi}_2 \Delta \mathbf{y}_{t-2} + \cdots + \boldsymbol{\phi}_p \Delta \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

level $\mathbf{y}_{t-1}$ only matters through the linear combination $\mathbf{b}'\mathbf{y}_{t-1}$

Example:

$$y_{1t} = y_{1,t-1} + \varepsilon_{1t}$$
$$y_{2t} = y_{1,t-1} + \varepsilon_{2t}$$



$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

$$\Delta \mathbf{y}_t = \mathbf{ab}'\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

Definition: a vector $\mathbf{y}_t$ is said to be cointegrated if

(1) each element $y_{jt} \sim I(1)$

(2) $\exists \mathbf{b} \neq \mathbf{0} : \mathbf{b}'\mathbf{y}_t \sim I(0)$

One way to estimate $\mathbf{b}$:

If $b_1 \neq 0$, then OLS estimation of

$$y_{1t} = c + \gamma_2 y_{2t} + \gamma_3 y_{3t} + \cdots + \gamma_n y_{nt} + u_t$$

gives superconsistent estimates of normalized value of $\mathbf{b}$

$T(\hat{\gamma}_j - \gamma_{j0}) \xrightarrow{L}$ nonstandard

Example:

$$y_{1t} = y_{1,t-1} + \varepsilon_{1t}$$
$$y_{2t} = \beta y_{1t} + \varepsilon_{2t}$$
$$\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right)$$
$$\mathbf{y}_1 = (y_{11}, \ldots, y_{1T})'$$
$$\mathbf{y}_2 | \mathbf{y}_1 \sim N(\beta \mathbf{y}_1, \sigma_2^2 \mathbf{I}_T)$$

Regression $y_{2t} = \beta y_{1t} + \varepsilon_{2t}$

satisfies Gaussian regression model

OLS t test of $H_0 : \beta = \beta_0$ has exact small sample Student t distribution with $T - 1$ degrees of freedom

$$\hat{\beta} = \beta_0 + \frac{\sum y_{1t}\varepsilon_{2t}}{\sum y_{1t}^2}$$
$$T(\hat{\beta} - \beta_0) = \frac{T^{-1} \sum y_{1t}\varepsilon_{2t}}{T^{-2} \sum y_{1t}^2}$$
$$\xrightarrow{L} \frac{\sigma_1 \sigma_2 \int_0^1 W_1(r) dW_2(r) dr}{\sigma_1^2 \int_0^1 [W_1(r)]^2 dr}$$

If instead

$$\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$$

Then regressor y_{1t} is correlated with residual

$$y_{2t} = \beta y_{1t} + \varepsilon_{2t}$$

But still

$$T^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{p} 0$$

$$T(\hat{\beta} - \beta_0) \xrightarrow{p} \text{nonstandard}$$

In general, for

$y_{1t} = c + \gamma_2 y_{2t} + \gamma_3 y_{3t} + \cdots + \gamma_n y_{nt} + u_t$

$y_{it} \sim I(1)$ for $i = 1, \ldots, n$

$u_t \sim I(0)$ (so cointegrated)

u_t correlated with y_{it}

then OLS estimates are superconsistent

$T(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{L}$ nonstandard

However, if $\mathbf{y}_t$ is not cointegrated, OLS gives spurious regression.

How to tell the difference:

if spurious, $u_t \sim I(1)$

if cointegrated, $u_t \sim I(0)$

Estimate $\hat{\gamma}$ by OLS, do Dickey-Fuller test on residuals $\hat{u}_t$

Compare with Table B.9, Case 2.

Using Table B.9:

- Case 1: Estimated “cointegrating regression” contains no constant term
- Case 2: Estimated “cointegrating regression” contains a constant term and right-hand variables not trended
- Case 3: Estimated “cointegrating regression” contains a constant term and right-hand variables are trended

Unit roots in vector time series

A. Vector autoregressions with unit roots
B. Spurious regression
C. Cointegration– single equation methods
D. Cointegration– full information maximum likelihood

Consider an $(n \times 1)$ vector $\mathbf{y}_t$ characterized by $0 < h < n$ different cointegrating relations:

$$\Delta\mathbf{y}_t = \mathbf{c} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \boldsymbol{\zeta}_1\Delta\mathbf{y}_{t-1} + \boldsymbol{\zeta}_2\Delta\mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

$\boldsymbol{\alpha}$ $(n \times h)$

$\boldsymbol{\beta}$ $(n \times h)$

$\rho(\boldsymbol{\alpha}) = \rho(\boldsymbol{\beta}) = h$

$\boldsymbol{\varepsilon}_t \sim$ i.i.d. $N(\mathbf{0}, \boldsymbol{\Omega})$

Step 1: Do OLS regressions of each of the n elements of $\Delta\mathbf{y}_t$ and n elements of $\mathbf{y}_{t-1}$ on a constant and $p-1$ lags of $\Delta\mathbf{y}_{t-j}$

$$
\begin{aligned}
\Delta \mathbf{y}_t &= \hat{\boldsymbol{\pi}}_0 + \hat{\Pi}_1 \Delta \mathbf{y}_{t-1} + \hat{\Pi}_2 \Delta \mathbf{y}_{t-2} + \\
&\quad \cdots + \hat{\Pi}_{p-1} \Delta \mathbf{y}_{t-p+1} + \hat{\mathbf{u}}_t \\
\mathbf{y}_{t-1} &= \hat{\boldsymbol{\theta}} + \hat{\Xi}_1 \Delta \mathbf{y}_{t-1} + \hat{\Xi}_2 \Delta \mathbf{y}_{t-2} + \\
&\quad \cdots + \hat{\Xi}_{p-1} \Delta \mathbf{y}_{t-p+1} + \hat{\mathbf{v}}_t
\end{aligned}
$$

Step 2: Form the variance-covariance matrices of these residuals

$$
\begin{aligned}
\hat{\Sigma}_{\mathrm{VV}} &= T^{-1} \sum_{t=1}^{T} \hat{\mathbf{v}}_t \hat{\mathbf{v}}_t' \\
\hat{\Sigma}_{\mathrm{UU}} &= T^{-1} \sum_{t=1}^{T} \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' \\
\hat{\Sigma}_{\mathrm{UV}} &= T^{-1} \sum_{t=1}^{T} \hat{\mathbf{u}}_t \hat{\mathbf{v}}_t' \\
\hat{\Sigma}_{\mathrm{VU}} &= \hat{\Sigma}_{\mathrm{UV}}'
\end{aligned}
$$

Step 3: Calculate the eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_n$ (ordered $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_n$) and eigenvectors $\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_n$ of the matrix

$$\hat{\Sigma}_{\mathrm{VV}}^{-1} \hat{\Sigma}_{\mathrm{VU}} \hat{\Sigma}_{\mathrm{UU}}^{-1} \hat{\Sigma}_{\mathrm{UV}}$$

Then the MLE of the space of cointegrating vectors (the space spanned by columns of β) is spanned by $\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_h$

And the maximum value achieved for the log likelihood under the restriction that there are exactly h cointegrating relations is given by

$$\mathcal{L}^* = -(Tn/2)\log(2\pi) - (Tn/2)$$
$$-(T/2)\log|\hat{\Sigma}_{UU}| - (T/2)\sum_{i=1}^{h}\log(1-\hat{\lambda}_i)$$

Implication: a likelihood ratio test of the null hypothesis of h cointegrating relations against the alternative that there are n cointegrating relations (i.e., *every* linear combination of $\mathbf{y}_t$ is stationary under H_A) is given by

$$2(\mathcal{L}_A^* - \mathcal{L}_0^*) = -T\sum_{i=h+1}^{n}\log(1-\hat{\lambda}_i)$$

Critical values from Table B.10.

Test of null of h cointegrating relations against alternative of $h+1$ relations:

$2(\mathcal{L}_A^* - \mathcal{L}_0^*) = -T\log(1 - \hat{\lambda}_{h+1})$

Compare with critical values in Table B.11

Unit roots in vector time series

A. Vector autoregressions with unit roots
B. Spurious regression
C. Cointegration– single equation methods
D. Cointegration– full information maximum likelihood
E. Testing hypotheses about the cointegrating vector

Suppose we have a maintained hypothesis that there are exactly h cointegrating relations, and want to test the hypothesis that only linear combinations of $\mathbf{D}'\mathbf{y}_t$ are involved in cointegration, where $\mathbf{D}'$ is known $(q \times n)$ matrix and $q < n$

Example: $n = 3$ and variable 1 is hypothesized not to appear in cointegrating relation:

$$\mathbf{D}' = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Let $\hat{\Sigma}_{\mathbf{UU}}$ $\hat{\Sigma}_{\mathbf{VV}}$ and $\hat{\Sigma}_{\mathbf{VU}}$ be same unrestricted matrices defined above and let

$$\tilde{\Sigma}_{\mathbf{VV}} = \mathbf{D}'\hat{\Sigma}_{\mathbf{VV}}\mathbf{D}$$

$$\tilde{\Sigma}_{\mathbf{UV}} = \hat{\Sigma}_{\mathbf{UV}}\mathbf{D}$$

Let $\tilde{\lambda}_i$ denote the ith largest eigenvalue of

$$\tilde{\Sigma}_{\mathbf{VV}}^{-1}\tilde{\Sigma}_{\mathbf{VU}}\tilde{\Sigma}_{\mathbf{UU}}^{-1}\tilde{\Sigma}_{\mathbf{UV}}$$

Then the likelihood ratio test of the null hypothesis is given by

$$2(\mathcal{L}_A^* - \mathcal{L}_0^*) = -T\sum_{i=1}^{h} \log(1-\hat{\lambda}_i) + T\sum_{i=1}^{h} \log(1-\tilde{\lambda}_i)$$

which asymptotically has a $\chi^2(hn - hq)$ distribution

Unit roots in vector time series

A. Vector autoregressions with unit roots
B. Spurious regression
C. Cointegration– single equation methods
D. Cointegration– full information maximum likelihood
E. Testing hypotheses about the cointegrating vector
F. Summary and overview

What to do with all these issues?

- should we use differences or growth rates?
- should we use error-correction representation or not?

If we always just estimate in levels:

- benefit– get consistent estimate of truth no matter what
- costs– some hypothesis tests invalid, inefficient estimates

Recommendation:

- Make best effort to identify nature of unit roots and cointegration and estimate as appropriate
- Compare with levels estimate
- If same answer great, if different answer, try to reconcile