Empirical Methods
Time Series Econometrics I & II
Regression with Integrated Processes

Martin Wagner

Department of Economics, University of Klagenfurt

Winter Term 2025

This version: July 30, 2025

Content of this Chapter

- Regression with Integrated Processes
- Cointegrating Regression
- OLS Estimation
- Fully Modified OLS Estimation
- Dynamic OLS Estimation
- IM-OLS Estimation
- Inference
- Fixed-b Inference
- Cointegrating Regression
- Testing the Null of Cointegration

- Consider, for brevity, a simple regression of the form:

$$y_t \quad = \quad \mu + x_t\beta + u_t$$

- The key difference to "Econometrics I" is the assumption that both y_t and x_t are I(1) processes.
- In this respect, two cases have to be distinguished:
    - y_t and x_t are cointegrated, i.e., for some $\beta \neq 0$ in the above equation, the true errors u_t are I(0).
    - y_t and x_t are "related" in a spurious regression, i.e., y_t and x_t are I(1) processes with u_t an I(1) process for all values of β.

Two I(1) processes, y_t and x_t, are cointegrated if there exists a $\beta \neq 0$ and an I(0) process u_t such that:

$$y_t = x_t\beta + u_t$$

Deterministic components (intercept, linear trend, etc.) are *allowed*.

- Unit roots and cointegration imply quite different behavior of regressions and estimators compared to the stationary case.

⇓ Even if y_t and x_t are statistically independent I(1) processes, performing the above regression leads to a seemingly significant $\hat{\beta}$: Nonsense or spurious regression.

⇑ In case of cointegration, the OLS estimator $\hat{\beta}$ is consistent even when x_t is endogenous.

? The limiting distribution of $\hat{\beta}$ depends upon endogeneity and serial correlation structure.

Spurious Regression

- Consider a regression of two independent I(1) processes on each other:

$$
\begin{aligned}
y_t &= y_{t-1} + u_t \\
x_t &= x_{t-1} + v_t,
\end{aligned}
$$

  with u_t and v_t two i.i.d. sequences with finite variances σ_u^2 and σ_v^2 independent of each other.

- Perform OLS estimation of the regression:

$$
\begin{aligned}
y_t &= \mu + x_t\beta + e_t \\
\hat{\beta} &= \frac{\sum_{t=1}^{T}(x_t - \overline{x}_T)y_t}{\sum_{t=1}^{T}(x_t - \overline{x}_T)^2} \\
&= \frac{T^{-2}\sum_{t=1}^{T}(x_t - \overline{x}_T)y_t}{T^{-2}\sum_{t=1}^{T}(x_t - \overline{x}_T)^2}
\end{aligned}
$$

Numerator:

$$\begin{aligned}
T^{-2}\sum_{t=1}^{T}(x_t-\overline{x}_T)y_t &= T^{-2}\sum_{t=1}^{T}x_t y_t - T^{-1}\overline{x}_T\overline{y}_T \\
T^{-2}\sum_{t=1}^{T}x_t y_t &\Rightarrow \sigma_v\sigma_u\int_0^1 W_v(r)W_u(r)dr \\
T^{-1/2}\overline{x}_T &= \frac{1}{T}\sum_{t=1}^{T}\sum_{t=1}^{T}\frac{x_t}{\sqrt{T}} \Rightarrow \sigma_v\int_0^1 W_v(r)dr
\end{aligned}$$

$$T^{-2}\sum_{t=1}^{T}(x_t-\overline{x}_T)y_t \Rightarrow \sigma_v\sigma_u\left(\int_0^1 W_v(r)W_u(r)dr - \int_0^1 W_v(r)dr\int_0^1 W_u(r)dr\right)$$

Spurious Regression

Denominator:

$$T^{-2}\sum_{t=1}^{T}(x_t-\overline{x}_T)^2 \quad \Rightarrow \quad \sigma_v^2\left(\int_0^1 W_v(r)^2dr-\left(\int_0^1 W_v(r)dr\right)^2\right)$$

OLS coefficient estimator:

$$\hat{\beta} \quad \Rightarrow \quad \frac{\sigma_u\left(\int_0^1 W_v(r)W_u(r)dr-\int_0^1 W_v(r)dr\int_0^1 W_u(r)dr\right)}{\sigma_v\left(\int_0^1 W_v(r)^2dr-\left(\int_0^1 W_v(r)dr\right)^2\right)}$$

Spurious Regression

- What about the estimated intercept $\hat{\mu}$?

$$\begin{aligned} \hat{\mu} &= \overline{y}_T - \overline{x}_T\hat{\beta} \\ T^{-1/2}\hat{\mu} &= \frac{\overline{y}_T}{\sqrt{T}} - \frac{\overline{x}_T}{\sqrt{T}}\hat{\beta} \\ &\Rightarrow \sigma_u\left(\int W_u - \frac{\int W_v W_u - \int W_v \int W_u}{\int W_v^2 - \left(\int W_v\right)^2}\int W_v\right) \end{aligned}$$

- Thus, $\hat{\mu} \sim O_p(T^{1/2})$.
- It can also be shown that R^2 converges to a non-zero limit and that the simple t-statistic for β is $O_p(T^{1/2})$.
- What about the OLS residuals:

$$\begin{aligned} \hat{e}_t &= y_t - \hat{\mu} - x_t\hat{\beta} \\ T^{-1/2}\hat{e}_t &= T^{-1/2}y_t - T^{-1/2}\hat{\mu} - T^{-1/2}x_t\hat{\beta}, \end{aligned}$$

i.e., $\hat{e}_t$ behaves like an I(1) process – with dependence upon deterministics and integrated regressor(s).

Testing for Cointegration/Absence of Cointegration

- The above observations lead to a test for the null hypothesis of non-cointegration: Perform a unit root test on $\hat{e}_t$.
- In general, one will need to allow for serial correlation and dependence between u_t and v_t, but without cointegration under the null hypothesis.
- The limiting distribution of the test depends upon:
  - The specification of the deterministic component
  - The number of integrated regressors – but not their dependence structure
- These dependencies reflect the effects of the spurious regression performed to obtain the OLS residuals.

- First, give a suitable definition of a multivariate I(1) process.
- Let $z_t = (y_t, x_t')' \in \mathbb{R}^{1+k}, t \in \mathbb{Z}$ be a stochastic process that fulfills:

$$\frac{1}{\sqrt{T}} z_{[rT]} \quad \Rightarrow \quad \Omega^{1/2} W(r) = B(r),$$

  where $W(r) \in \mathbb{R}^{1+k}$ is vector standard Brownian motion and $\Omega \neq 0$.
- Then $\{z_t\}_{t \in \mathbb{Z}}$ is referred to as an I(1) process.
- A summed up weakly stationary process with non-zero long-run covariance is an I(1) process with this definition.
- A process $\{w_t\}_{t \in \mathbb{Z}}$ is labelled I(0) process if:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[rT]} w_t \quad \Rightarrow \quad \Omega^{1/2} W(r) = B(r), \ \ \Omega \neq 0$$

- Let $\{z_t\}_{t\in\mathbb{Z}}$ be an I(1) process.
- If there exists a vector $\beta \neq 0$ such that $\{\beta' z_t\}_{t\in\mathbb{Z}}$ is I(0), then $\{z_t\}_{t\in\mathbb{Z}}$ is a cointegrated I(1) process.
- The vector β is referred to as cointegrating vector.
- Clearly, cointegrating vectors are not unique and linear combinations of cointegrating vectors are cointegrating vectors.
- Thus, the cointegrating space is a subspace of $\mathbb{R}^{1+k}$.
- The regression-based methods considered at the moment are usually considered for the case of a one-dimensional cointegrating space only (and with some more assumptions for asymptotic analysis in place).

- Going back to the Beveridge-Nelson decomposition (putting a bit more structure on the processes for the moment) we find:

$$\begin{aligned} \Delta y_t &= c(L)\varepsilon_t \\ c(1) &\neq 0 \\ |c(z)| &\neq 0 \; \forall \; |z| \leq 1 \text{ except for possibly } z = 1 \\ \{\varepsilon_t\}_{t\in\mathbb{Z}} & \text{ is } WN(0,\Sigma), \Sigma > 0 \end{aligned}$$

- Consider the BN-decomposition for Δy_t:

$$\begin{aligned} \Delta y_t &= c(L)\varepsilon_t \\ \Delta y_t &= \left(c(1) + \tilde{c}(L)(1-L)\right)\varepsilon_t \\ y_t &= c(1)\sum_{j=0}^{t-1}\varepsilon_{t-j} + \tilde{c}(L)\varepsilon_t + y_0^* \end{aligned}$$

- In the case that $c(1)$ has reduced rank, there are non-zero vectors β in its left kernel:

$$\begin{aligned} \beta' y_t &= \underbrace{\beta' c(1)}_{=0} \sum_{j=0}^{t-1} \varepsilon_{t-j} + \beta' \tilde{c}(L)\varepsilon_t + \beta' y_0^* \\ &= \beta' \tilde{c}(L)\varepsilon_t + \beta' y_0^* \end{aligned}$$

- Under the assumptions listed, the long-run covariance of $\{\Delta y_t\}_{t\in\mathbb{Z}}$ is given by:

$$\Omega_{\Delta y \Delta y} = \sum_{j=-\infty}^{\infty} \mathbb{E}(\Delta y_t \Delta y'_{t-j}) = c(1)\Sigma c(1)'$$

- Thus, reduced rank of Ω is equivalent to cointegration, with:

$$\dim(y_t) - \dim(\text{Coint. space}) = \text{rk}(\Omega) = \text{rk}(c(1))$$

- Now, consider the following spurious regression model:

$$y_t \quad = \quad D_t'\theta_D + x_t'\beta + e_t$$

- Under the null hypothesis of no cointegration $z_t = (y_t, x_t')'$ is a non-cointegrated I(1) process.
- As mentioned above: Appropriate assumptions on the errors driving $\{z_t\}$ are necessary to get FCLT results.
- The absence of cointegration (under H_0) is tested by performing a unit root test on the OLS residuals of the above spurious regression.
- Both ADF- and PP-type tests can be performed.
- The limiting distributions differ from the standard ADF/PP-limiting distributions and depend, in addition to D_t, also on $\dim(x_t)$.

- A residual ADF test is performed using the regression:

$$\Delta\hat{e}_t \quad = \quad \rho\hat{e}_{t-1} + \sum_{j=1}^{p} \phi_j \Delta\hat{e}_{t-j} + v_{t,p}$$

- In general, $p = p(T)$, à la Said and Dickey (1984).
- The test statistic is then the t-statistic for H_0: $\rho = 0$ against H_1: $-1 < \rho < 1$.
- Alternatively, the Phillips-Perron tests can be extended to being used on regression residuals, considering the first order regression:

$$\hat{e}_t = \rho\hat{e}_{t-1} + k_t$$

as starting point.

$$
\begin{aligned}
Z_\rho &= T(\hat{\rho}-1) - \frac{1}{2}(\hat{\omega}_k^2 - \hat{\sigma}_k^2)\left(\frac{1}{T^2}\sum_{t=1}^{T}\hat{e}_{t-1}^2\right)^{-1} \\
Z_t &= \frac{\hat{\sigma}}{\hat{\omega}_k}t_\rho - \frac{1}{2}(\hat{\omega}_k^2 - \hat{\sigma}_k^2)\left(\hat{\omega}_k^2\frac{1}{T^2}\sum_{t=1}^{T}\hat{e}_{t-1}^2\right)^{-1/2},
\end{aligned}
$$

where $\hat{\omega}_k^2$ denotes a consistent estimator of the long-run variance of $\{k_t\}$ and $\hat{\sigma}_k^2$ an estimator of its variance. Under H_0 it holds that:

$$
\begin{aligned}
Z_\rho &\Rightarrow \int_0^1 R(m)dR(m) \\
Z_t &\Rightarrow \int_0^1 R(m)dS(m)
\end{aligned}
$$

- The processes $R(m)$ and $S(m)$ that appear in the limiting distributions are defined as follows:

- First, define $J(r) = (D(r)', W_x(r)')'$, with $D(r)$ the scaled limiting quantities corr. to D_t, $W_x(r)$ a k-dimensional vector of standard Brownian motions and:

$$Q(r) = W_y(r) - \int_0^1 W_y(r)J(r)'dr \left(\int_0^1 J(s)J(s)'ds \right)^{-1} J(r),$$

  with $W_y(r)$ one-dimensional standard Brownian motion independent of $W_x(r)$.

- Then $R(r)$ is defined as:

$$R(r) = Q(r) / \left(\int_0^1 Q(s)^2 ds \right)^{1/2}$$

- Next, define:

$$\begin{pmatrix} f_{yy} & f_{yx} \\ f_{xy} & F_{xx} \end{pmatrix} = \int_0^1 W(r)W(r)'dr,$$

with $W(r) = (W_y(r), W_x(r)')'$.

- Furthermore, define $\kappa = \begin{pmatrix} 1 \\ F_{xx}^{-1} f_{xy} \end{pmatrix}$ and:

$$S(r) = Q(r)/\left(\kappa'\kappa\right)^{1/2}$$

Residual-Based Cointegration Test

- As in unit root testing before, one may want to consider to flip null and alternative hypotheses.
- In unit root testing, this leads to the KPSS test.
- It appears natural to consider extending the KPSS test to residuals from a cointegrating regression.
- However, issues are a bit more delicate here since under cointegration the OLS residuals lead to a residual process that leads to nuisance-parameter-dependent KPSS-type statistics.
- This is because of a nuisance-parameter-dependent limiting distribution of the OLS estimator $\hat{\beta}$ under cointegration.
- These complications – and, in general, problems with inference concerning coefficients in cointegrating relationships – can be overcome by applying suitably modified least squares estimators.
- We will consider three such modifications: Fully Modified, Dynamic and Integrated Modified OLS

Cointegrating Regression

For brevity, consider the case $D_t = 1$

- Consider the regression model:

$$\begin{aligned} y_t &= \mu + x_t'\beta + u_t, \quad t = 1, 2, ..., T, \\ x_t &= x_{t-1} + v_t, \end{aligned}$$

  where y_t is a scalar time series and x_t is a $(k \times 1)$-vector time series.

- Assume that there are no cointegrating relationships among the x_t variables.

- Stack the error processes and define $\eta_t = (u_t, v_t')'$.

- Assume that η_t is an $I(0)$ vector process.

- Assume a FCLT holds for η_t:

$$T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor}\eta_t \Rightarrow B(r) = \Omega^{1/2}W(r),$$

where $W(r)$ is a $(k+1)$-dimensional vector of independent standard Brownian motions with:

$$\Omega = \sum_{j=-\infty}^{\infty}\mathbb{E}(\eta_t\eta_{t-j}') = \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix} > 0$$

- This assumption excludes cointegration amongst the regressors.

- It is convenient to use the Cholesky form of $\Omega^{1/2}$:

$$\Omega^{1/2} = \begin{bmatrix} \omega_{u\cdot v} & \lambda_{uv} \\ \mathbf{0} & \Omega_{vv}^{1/2} \end{bmatrix},$$

  with $\omega_{u\cdot v}^2 = \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$ and $\lambda_{uv} = \Omega_{uv}(\Omega_{vv}^{-1/2})'$.

- Partition $B(r)$ and $W(r)$ as:

$$B(r) = \begin{bmatrix} B_u(r) \\ B_v(r) \end{bmatrix}, \quad W(r) = \begin{bmatrix} W_{u\cdot v}(r) \\ W_v(r) \end{bmatrix},$$

  with the upper blocks scalar and the lower blocks k-dimensional.

- Using this Cholesky decomposition leads to:

$$B(r) = \left[\begin{array}{c} B_u(r) \\ B_v(r) \end{array} \right] = \left[\begin{array}{c} \omega_{u\cdot v} W_{u\cdot v}(r) + \lambda_{uv} W_v(r) \\ \Omega_{vv}^{1/2} W_v(r) \end{array} \right]$$

- Define the one-sided long-run covariance matrix $\Lambda = \sum_{j=1}^{\infty} \mathbb{E}(\eta_{t-j}\eta_t')$, partitioned as:

$$\Lambda = \left[\begin{array}{cc} \Lambda_{uu} & \Lambda_{uv} \\ \Lambda_{vu} & \Lambda_{vv} \end{array} \right]$$

- Note that $\Omega = \Lambda + \Lambda' + \Sigma$, with $\Sigma = \mathbb{E}(\eta_t \eta_t')$, analogously partitioned as:

$$\Sigma = \left[\begin{array}{cc} \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{array} \right]$$

- For OLS, it is well known from, e.g., Phillips and Durlauf (1986) and Stock (1987) that:

$$\begin{pmatrix} T^{1/2}(\hat{\mu}-\mu) \\ T(\hat{\beta}-\beta) \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & \int B_v(r)'dr \\ \int B_v(r)dr & \int B_v(r)B_v(r)'dr \end{pmatrix}^{-1} \times \begin{pmatrix} \int dB_u(r) \\ \int B_v(r)dB_u(r) + \Delta_{vu} \end{pmatrix} = \Theta,$$

  where $\Delta_{vu} = \Sigma_{vu} + \Lambda_{vu}$.

- When u_t and v_t are uncorrelated:
  - $\Delta_{vu} = 0$ and $\lambda_{uv} = 0$
  - $B_v(r)$ and $B_u(r)$ are independent
  - $T(\hat{\beta}-\beta)$ is zero mean Gaussian conditional on $B_v(r)$ and t-, F-statistics have the usual $N(0,1)$ and chi-squared limits provided serial correlation in u_t is handled properly.

- When the regressors are endogenous, the limiting distribution of $T(\hat{\beta} - \beta)$ is more complicated because of correlation between $B_u(r)$ and $B_v(r)$ and the nuisance parameters in the matrix Δ_{vu}.

- Δ_{vu} introduces an asymptotic bias.

- One can no longer condition on $B_v(r)$ to obtain an asymptotic normality result.

- Inference using OLS is “very difficult” in this situation.

- Consider the stochastic process:

$$B_{u\cdot v}(r) = B_u(r) - B_v(r)'\Omega_{vv}^{-1}\Omega_{vu} = \omega_{u\cdot v}W_{u\cdot v}(r)$$

- $B_{u\cdot v}(r)$ is independent of $B_v(r) = \Omega_{vv}^{1/2}W_v(r)$.
- Using $B_{u\cdot v}(r)$, write:

$$\int B_v(r)dB_u(r) + \Delta_{vu} = \int B_v(r)dB_{u\cdot v}(r) + \int B_v(r)dB_v(r)'\Omega_{vv}^{-1}\Omega_{vu} + \Delta_{vu}$$

- Because $B_v(r)$ and $B_{u\cdot v}(r)$ are independent, conditional on $B_v(r)$:

$$\int B_v(r)dB_{u\cdot v}(r) \sim N\left(0, \omega_{u\cdot v}^2 \int B_v(r)B_v(r)'dr\right)$$

- FM-OLS uses two transformations:
  - One removes $\int B_v(r) dB_v(r)' \Omega_{vv}^{-1} \Omega_{uv}'$.
  - The other removes Δ_{vu}.
- The first transformation uses an estimator of $\widehat{\Omega}$:

$$\widehat{\Omega} = T^{-1} \sum_{i=1}^{T} \sum_{j=1}^{T} k(\frac{|i-j|}{M}) \widehat{\eta}_j \widehat{\eta}_i',$$

where $\widehat{\eta}_t = (\widehat{u}_t, \Delta x_t')'$ and $\widehat{u}_t$ are the OLS residuals.

- $k(x)$ is the kernel weighting function and M is the bandwidth.
- Partition $\widehat{\Omega}$ the same way as Ω and define:

$$y_t^+ = y_t - \Delta x_t' \widehat{\Omega}_{vv}^{-1} \widehat{\Omega}_{vu}$$

- The first transformation – using y_t^+ instead of y_t – removes $\int B_v(r)dB_v(r)'\Omega_{vv}^{-1}\Omega_{uv}'$, but changes Δ_{vu} to:

$$\Delta_{vu}^+ = \Delta_{vu} - \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu}$$

- Δ_{vu}^+ is removed by an additive second transformation, using:

$$\hat{\Delta}_{vu}^+ = \widehat{\Delta}_{vu} - \widehat{\Delta}_{vv}\widehat{\Omega}_{vv}^{-1}\widehat{\Omega}_{vu},$$

where:

$$\widehat{\Delta} = T^{-1}\sum_{i=1}^{T}\sum_{j=i}^{T} k(\frac{|i-j|}{M})\widehat{\eta}_j\widehat{\eta}_i'$$

The FM-OLS estimator of Phillips and Hansen (1990) is defined as:

$$\hat{\theta}^{+} = (\tilde{X}'\tilde{X})^{-1}(\tilde{X}'y^{+} - M^{*}),$$

where $\theta = [\mu, \beta']'$, $\tilde{X}$ is the matrix of stacked $[1, x_t']$, y^{+} is the vector of stacked y_t^{+} and:

$$M^{*} = T \begin{pmatrix} 0 \\ \hat{\Delta}_{vu}^{+} \end{pmatrix}$$

Asymptotic Distribution

$$\begin{pmatrix} T^{1/2}(\hat{\mu}^+ - \mu) \\ T(\hat{\beta}^+ - \beta) \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & \int B_v' \\ \int B_v & \int B_v B_v' \end{pmatrix}^{-1} \begin{pmatrix} \int dB_{u\cdot v} \\ \int B_v dB_{u\cdot v} \end{pmatrix}$$

$$= \omega_{u\cdot v} \begin{pmatrix} 1 & \int B_v' \\ \int B_v & \int B_v B_v' \end{pmatrix}^{-1} \begin{pmatrix} \int dW_{u\cdot v} \\ \int B_v dW_{u\cdot v} \end{pmatrix}$$

- Inference requires to scale out $\omega_{u\cdot v}$, using, e.g.,:

$$\widehat{\omega}^2_{u\cdot v} = \widehat{\Omega}_{uu} - \widehat{\Omega}_{uv}\widehat{\Omega}_{vv}^{-1}\widehat{\Omega}_{vu}$$

- Bandwidth and kernel have to be chosen for estimation of $\omega^2_{u\cdot v}$.
- Typically, the bandwidth is chosen using some of the data-dependent rules already mentioned.
- There is, however, also something like a partial fixed$-b$ approach proposed by Jin, Phillips and Sun (2006).
- This approach derives a fixed-b result for $\hat{\omega}^2_{u\cdot v}$ estimated directly from the FM-OLS residuals $\hat{u}_t^+ = y_t^+ - \hat{\mu}^+ - x_t'\hat{\beta}^+$.
- This result, however, relies upon a traditional consistency result for $\hat{\Omega}$ when constructing the correction factors.
- A complete fixed-b analysis of FM-OLS is contained in Vogelsang and Wagner (2014).

- D-OLS performs the necessary corrections by augmenting the regression equation by leads and lags of the first differences of x_t:

$$\begin{aligned} y_t &= \mu + x_t'\beta + \sum_{j=-K}^{K} \Delta x_{t-j}'\gamma_j + u_t^K \\ &= \mu + x_t'\beta + \sum_{j=-K}^{K} v_{t-j}'\gamma_j + u_t^K, \end{aligned}$$

  with $u_t^K = u_t - \sum_{j=-K}^{K} v_{t-j}'\gamma_j$.
- The D-OLS estimator is the OLS estimator in the above equation.

- D-OLS is based on the observation that:

$$u_t^* = u_t - \sum_{j=-\infty}^{\infty} v_{t-j}' \gamma_j,$$

  i.e., the projection error of the projection of $\{u_t\}_{t \in \mathbb{Z}}$ is dynamically uncorrelated with $\{v_t\}_{t \in \mathbb{Z}}$.

- Thus, in an infeasible regression with infinitely many leads and lags regressor endogeneity is "repaired".

- Consistency requires $K \to \infty$ at a suitable rate, $K = o(T^{1/3})$.

- Inference based on $\hat{u}_t^K$ to estimate the conditional long-run variance (also OLS residuals can be used as in FM-OLS).

- There are several choices to be made when using D-OLS: Lead and lag lengths, kernel and bandwidth.
- Lead and lag lengths can be chosen by minimizing information criteria, see Choi and Kurozumi (2012) or Kejrival and Perron (2008).
- Kernel and bandwidth required for estimating the conditional long-run variance estimation are chosen using the data-dependent bandwidth rules mentioned.
- In this respect, $\omega^2_{u \cdot v}$ can be estimated using $\hat{\eta}_t$, as for FM-OLS, or by using the D-OLS residuals $\hat{u}^K_t$.
- Bunzel (2006) provides a fixed-b result based on the D-OLS residuals.

- The IM-OLS estimator is based on two simple tuning-parameter-free transformations.

- Step 1: Partial sum the original equation:

$$y_t = \mu + x_t'\beta + u_t$$

to arrive at:

$$S_t^y = \mu t + S_t^{x\prime}\beta + S_t^u,$$

where $S_t^y = \sum_{j=1}^t y_j$ and S_t^x and S_t^u are defined analogously.

- The benefit of partial summing is that in the expression for $\hat{\theta} - \theta$, a term of the form:

$$\sum_{t=1}^{T} x_t u_t$$

is replaced by a term of the form:

$$\sum_{t=1}^{T} S_t^x S_t^u$$

- The term $T^{-1} \sum_{t=1}^{T} x_t u_t$ is the source of Δ_{vu}.
- The term $T^{-3} \sum_{t=1}^{T} S_t^x S_t^u$ does not lead to such additive components in the limit.

- Partial summing before estimating the model performs the same role for IM-OLS that subtracting M^* plays for FM-OLS.
- This still leaves the problem that correlation between u_t and v_t (x_t) rules out the possibility of conditioning on $B_v(r)$ to obtain a conditional asymptotic normality result.
- Step 2: Add x_t to the partial sum regression:

$$S_t^y = \mu t + S_t^{x\prime}\beta + x_t'\gamma + S_t^u$$

- This step plays the same role for IM-OLS that using y_t^+ instead of y_t plays for FM-OLS.
- The OLS estimator of the parameters of the partial summed and augmented (the IM-OLS) regression is referred to as Integrated Modified OLS estimator.

Asymptotic Distribution IM-OLS

$$\begin{pmatrix} T^{1/2}(\widetilde{\mu} - \mu) \\ T(\widetilde{\beta} - \beta) \\ \widetilde{\gamma} - \Omega_{vv}^{-1}\Omega_{vu} \end{pmatrix} \Rightarrow \omega_{u \cdot v} \left(\Pi \int g(s)g(s)' ds \Pi' \right)^{-1} \Pi \int g(s) W_{u \cdot v}(s) ds$$

$$= \omega_{u \cdot v} (\Pi')^{-1} \left(\int g(s)g(s)' ds \right)^{-1} \int [G(1) - G(s)] dW_{u \cdot v}(s)$$

with:

$$\Pi = \begin{bmatrix} 1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Omega_{vv}^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Omega_{vv}^{1/2} \end{bmatrix}, \quad g(r) = \begin{bmatrix} r \\ \int_0^r W_v(s) ds \\ W_v(r) \end{bmatrix}, \quad G(r) = \int_0^r g(s) ds$$

- Conditional upon $W_v(r)$, the scaled IM-OLS estimator has a mean zero normal asymptotic distribution with covariance matrix:

$$\begin{aligned} V_{IM} &= \omega_{u\cdot v}^2 (\Pi')^{-1} \left(\int g(s)g(s)'ds \right)^{-1} \left(\int [G(1)-G(s)][G(1)-G(s)]'ds \right) \\ &\qquad \times \left(\int g(s)g(s)'ds \right)^{-1} \Pi^{-1} \end{aligned}$$

- Note, furthermore, that $\widetilde{\gamma}$ is asymptotically centered around $\Omega_{vv}^{-1}\Omega_{vu}$, different from zero in case of endogeneity.

- Denote with $k(s) = (s, W_v(s)')'$ and with $\Pi^* = \text{diag}(1, \Omega_{vv}^{1/2})$, then the conditional asymptotic variance of the FM-OLS estimator is given by:

$$V_{FM} = \omega_{u\cdot v}^2 (\Pi^{*\prime})^{-1} \left(\int_0^1 k(s)k(s)'ds \right)^{-1} (\Pi^*)^{-1}$$

- It can be shown that $V_{FM} \leq V_{IM}$, where, for the latter, one has to consider the corresponding sub-block.
- However, such a comparison is misleading: It ignores the impacts of kernel and bandwidth choices on the FM sampling distributions.
- A more relevant comparison is between the conditional variance of the FM-OLS fixed-b limiting distribution and V_{IM}.
- This is not an easy task as, from a fixed-b perspective, the FM limit distribution is a complicated function depending upon kernel and bandwidth.
- Additionally, from a fixed-b perspective, FM-OLS is first-order biased, whereas IM-OLS is not.

- Suppose that a consistent estimator $\check{\omega}^2_{u\cdot v}$ of $\omega^2_{u\cdot v}$ is used and define:

$$\check{V}_{IM} = \check{\omega}^2_{u\cdot v} A_{IM}^{-1} \left(S^{\widetilde{x}\prime} S^{\widetilde{x}} \right)^{-1} (C'C) \left(S^{\widetilde{x}\prime} S^{\widetilde{x}} \right)^{-1} A_{IM}^{-1},$$

where C is the matrix formed by stacking the vector:

$$c_t = S_T^{S^{\widetilde{x}}} - S_{t-1}^{S^{\widetilde{x}}},$$

with $S_t^{S^{\widetilde{x}}} = \sum_{j=1}^{t} S_j^{\widetilde{x}}$ and $A_{IM} = \text{diag}(T^{-1/2}, T^{-1} I_k, I_k)$.

- For $T \to \infty$, it holds that $\check{V}_{IM} \to V_{IM}$.

- One approach to obtain a consistent long-run variance estimator is based on the OLS residuals $\hat{\eta}_t$ of the original equation:

$$\widehat{\omega}^2_{u\cdot v} = \widehat{\Omega}_{uu} - \widehat{\Omega}_{uv}\widehat{\Omega}^{-1}_{vv}\widehat{\Omega}_{vu}$$

- Also, the first differences of the residuals $\widetilde{S}^u_t$ of the IM-OLS regression might be considered:

$$\widetilde{\omega}^2_{u\cdot v} = T^{-1}\sum_{i=2}^{T}\sum_{j=2}^{T} k(\frac{|i-j|}{M})\Delta\widetilde{S}^u_j\Delta\widetilde{S}^u_i$$

- Using the latter, however, leads to conservative tests.

- Under usual assumptions about bandwidths and kernel functions, it holds for $T \to \infty$ that:

$$\widetilde{\omega}^2_{u\cdot v} \Rightarrow \omega^2_{u\cdot v}(1 + d'_\gamma d_\gamma),$$

where d_γ denotes the last k components of:

$$\left(\int g(s)g(s)'ds\right)^{-1} \int [G(1) - G(s)]dw_{u\cdot v}$$

- Denote the consistent variance estimator using $\widehat{u}_t$ as $\widehat{V}_{IM}$ and the variance estimator using $\Delta\widetilde{S}_t^u$ as $\widetilde{V}_{IM}$.

- Test multiple linear hypotheses H_0: $R\theta = r$, $R \in \mathbb{R}^{q\times(1+2k)}$ of full rank.
- A restriction on feasible matrices R is needed because of different convergence rates of coefficients:

$$\lim_{T\to\infty} A_R^{-1} R A_{IM} = R^*,$$

with R^* of full rank q and $A_R \in \mathbb{R}^{q\times q}$.

- The Wald statistics using $\widehat{\omega}^2_{u\cdot v}$ and $\widetilde{\omega}^2_{u\cdot v}$ and their asymptotic null distributions are given by:

$$\widehat{W} = (R\widetilde{\theta} - r)'[R A_{IM} \widehat{V}_{IM} A_{IM} R']^{-1}(R\widetilde{\theta} - r) \Rightarrow \chi^2_q,$$

$$\widetilde{W} = (R\widetilde{\theta} - r)'[R A_{IM} \widetilde{V}_{IM} A_{IM} R']^{-1}(R\widetilde{\theta} - r) \Rightarrow \frac{\chi^2_q}{1 + d'_\gamma d_\gamma}$$

Fixed-b Inference: Simple Example I

- Consider the simple "almost standard" (i.e., HAC) regression:

$$y_t = x_t\beta + u_t,$$

$\frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor} x_t^2 \to rQ$, $Q > 0$ and $z_t = x_t u_t$ s.t. $\frac{1}{T^{1/2}}\sum_{t=1}^{\lfloor rT \rfloor} z_t \Rightarrow \omega W(r)$.

- $\sqrt{T}\left(\hat{\beta} - \beta\right) \to N(0, \omega^2 Q^{-2})$.
- With a consistent estimator $\hat{\omega}^2 \to \omega^2$ it follows that:

$$t_\beta = \frac{\hat{\beta} - \beta_0}{\sqrt{\hat{V}ar(\hat{\beta})}} = \frac{\hat{\beta} - \beta_0}{\hat{\omega}\hat{Q}} \to N(0,1)$$

- Using consistent $\hat{\omega}^2 = \hat{\gamma}_0 + 2\sum_{j=1}^{T-1} k\left(\frac{j}{M}\right)\hat{\gamma}_j$, with $\hat{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T} \hat{z}_t \hat{z}_{t-j}$, "hides" finite sample effects of kernel function $k(\cdot)$ and bandwidth M.

Fixed-b Inference: Simple Example II

- Consider bandwidth proportional to sample size, i.e., $M = bT$.
- Then, under appropriate assumptions, it holds that $\hat{\omega}^2 \Rightarrow \omega^2 P(b,k)$, where $P(b,k)$ is a function of $W(r)$ that depends upon bandwidth b and kernel function $k(\cdot)$.
- This leads to a fixed-b limit distribution of the t-statistic of the form:

$$t_\beta \Rightarrow \frac{W(1)}{P(b,k)}$$

- See, e.g., Kiefer and Vogelsang (2005).
- Critical values can be tabulated for (grid of) values of b and different kernel functions $k(\cdot)$.

- Fixed-b inference is not directly possible using either $\hat{u}_t$ or $\Delta\widetilde{S}_t^u$.
- Fixed-b limits depend upon partial sum processes of residuals:
- The partial sum limit $T^{-1/2}\sum_{t=1}^{[rT]}\hat{\eta}_t$ is a "complicated function" of $\widehat{\Omega}$.
- The partial sum limit of $T^{-1/2}\sum_{t=1}^{[rT]}\Delta\widetilde{S}_t^u$ is proportional to $\omega_{u\cdot v}$, but is correlated, conditional upon $W_v(r)$, with the limit of $\widetilde{\theta}$ (the correlation depends on Π).
- Thus, fixed-b limits of Wald tests based on $\widetilde{\theta}$ and $\widetilde{\omega}_{u\cdot v}^2$ are not pivotal.
- Need: Residual process with partial sum limit proportional to $\omega_{u\cdot v}$ and uncorrelated with $\widetilde{\theta}$.

- Consider the regression:

$$S_t^y = \mu t + S_t^{x\prime}\beta + x_t'\gamma + z_t'\kappa + S_t^{u*},$$

where:

$$z_t = t\sum_{j=1}^{T}\xi_j - \sum_{j=1}^{t-1}\sum_{s=1}^{j}\xi_s, \qquad \xi_t = (t, S_t^{x\prime}, x_t')'$$

- Let $\widetilde{S}_t^{u*}$ denote the OLS residuals from the above further augmented regression and define:

$$\widetilde{\omega}_{u\cdot v}^{2*} = T^{-1}\sum_{i=2}^{T}\sum_{j=2}^{T} k(\frac{|i-j|}{M})\Delta\widetilde{S}_j^{u*}\Delta\widetilde{S}_i^{u*}$$

- This leads to a third estimator of V_{IM}:

$$\widetilde{V}_{IM}^{*} = \widetilde{\omega}_{u\cdot v}^{2*}\left(A_{IM}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}\right)^{-1}(A_{IM}C'CA_{IM})\left(A_{IM}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}\right)^{-1}$$

- The third considered estimator of $\omega^2_{u\cdot v}$ defines a third test statistic:

$$\widetilde{W}^* \quad = (R\widetilde{\theta} - r)'[RA_{IM}\widetilde{V}^*_{IM}A_{IM}R']^{-1}(R\widetilde{\theta} - r)$$

Let $M = bT$, where $b \in (0, 1]$ is held fixed, suppose that R satisfies the formulated asymptotic restriction, then as $T \to \infty$ under the null hypothesis:

$$\widetilde{W}^* \Rightarrow \frac{\chi^2_q}{Q_b(\widetilde{P}^*, \widetilde{P}^*)},$$

where χ^2_q is a chi-square random variable with q degrees of freedom that is independent of $Q_b(\widetilde{P}^*, \widetilde{P}^*)$. When $q = 1$:

$$\widetilde{t}^* = \frac{R\widetilde{\theta} - r}{\sqrt{RA_{IM}\widetilde{V}^*_{IM}A_{IM}R'}} \Rightarrow \frac{Z}{\sqrt{Q_b(\widetilde{P}^*, \widetilde{P}^*)}},$$

where Z is distributed $N(0, 1)$ and is independent of $Q_b(\widetilde{P}^*, \widetilde{P}^*)$.

- The limit process $\widetilde{P}^*$ is given by:

$$\widetilde{P}^*(r) = \int_0^r dw_{u\cdot v}(r) - h(r)' \left(\int_0^1 h(s)h(s)'ds \right)^{-1} \int_0^1 (H(1) - H(s))dW_{u\cdot v}(s),$$

with:

$$h(r) = \left[g(r)', \int_0^r (G(1) - G(s))' \, ds \right]', \qquad H(r) = \int_0^r h(s)ds$$

- The dependence upon $k(\cdot)$ and b is reflected in $Q_b(\cdot,\cdot)$.
- Due to pivotality, critical values can be simulated for given specification of deterministic components, number of integrated regressors, $k(\cdot)$ and b.
- In principle, under assumptions for consistency, $\widetilde{W}^*$ could also be used for standard inference.

- In practice, the two regressions are performed in one step.
- This is achieved by including in the IM-OLS regression not the additional regressors z_t, but z_t orthogonalized with respect to the IM-OLS regressors.
- Fixed-b critical values are available, as well as `MATLAB` code that implements the proposed method.
- The bandwidth required is again, typically, chosen with the data-dependent bandwidth rules already mentioned.

- Having discussed modified OLS estimators allows extending the KPSS test to a test for the null of cointegration.
- This extension was developed by Shin (1994).
- The test statistic $\hat{\eta}_{\text{SHIN}}$ looks exactly as before, with the difference being that now the FM-OLS or D-OLS residuals $\hat{u}_t^+$ are used:[1]

$$\begin{aligned} \text{FM-OLS}: \quad \hat{u}_t^+ &= y_t^+ - \hat{\mu}^+ t - x_t' \hat{\beta}^+ \\ \text{D-OLS}: \quad \hat{u}_t^+ &= y_t - \hat{\mu} t - x_t' \hat{\beta} - \sum_{j=-K}^{K} \Delta x_{t-j}' \hat{\gamma}_j \end{aligned}$$

- The null limiting distribution now depends, in addition to the deterministic components, also on the number of integrated regressors.

[1] For notational brevity, the superscript $^+$ denotes both FM- and D-OLS quantities.

- Denote with $(W_y(r), W_x(r)')' \in \mathbb{R}^{1+k}$ a vector of standard Wiener processes.
- The limit process corresponding to the deterministic component is again denoted by $D(r)$.
- Similarly to before, denote with $\widetilde{W}_x(r)$ the correspondingly detrended version of $W_x(r)$.
- Define the generalized Brownian bridge as:

$$\widehat{W}_y(r) = \quad W_y(r) - \int_0^r D(s)'ds \left(\int_0^1 D(s)D(s)'ds \right)^{-1} \int_0^1 D(s)dW_y(s),$$

which has already been considered in the KPSS discussion for the intercept and intercept and linear trend cases.[2]

---

[2]This also yields the KPSS limiting distribution for more general deterministic components.

- Given the above quantities, define:

$$Q(r) \quad = \quad \widehat{W}_y(r) - \int_0^r \widetilde{W}_x(s)' ds \left(\int_0^1 \widetilde{W}_x(s) \widetilde{W}_x(s)' \right)^{-1} \int_0^1 \widetilde{W}_x(s) d\widehat{W}_y(s)$$

- The null limiting distribution of the Shin statistic is given by:

$$\hat{\eta}_{\mathsf{SHIN}} \quad \Rightarrow \quad \int_0^1 Q(s)^2 ds$$