

Department of Economics
University of California at Berkeley

M. Jansson

Multivariate Nonstationary Time Series

When estimating multivariate models with nonstationary data, we face a nontrivial model selection problem. On the one hand, all of the results derived in the univariate unit root model have multivariate analogues. That is, there are multivariate unit root models in which one needs to use first differences of the data when doing inference. On the other hand, there is an important class of models, cointegration models, where the use of data in first differences introduces noninvertibility/over-differencing problems.

Multivariate Unit Root Models

All of the results derived in the univariate unit root model have multivariate analogues. Some generalizations are more obvious than others, however. By analogy with the univariate case without deterministic components, suppose z_t is a zero-mean observed k -dimensional time series generated by

$$z_t = z_{t-1} + v_t, \quad v_t = \Psi(L) \varepsilon_t \quad (t = 1, \dots, T),$$

where $z_0 = 0$, $\Psi(L) = \sum_{i=0}^{\infty} \Psi_i L^i$, and $\varepsilon_t \sim i.i.d. (0, \Sigma)$. This model generalizes the univariate ‘case 2’ model. When $\Psi(L) = I$, we have a multivariate ‘case 1’ model.

As in the univariate case, the limiting distributions of estimators and test statistics can be characterized in terms of functionals of (multivariate) Brownian motions. A continuous time k -dimensional process $\{B(r) : 0 \leq r \leq 1\}$ is a *Brownian motion with variance* Σ , denoted $B \sim BM(\Sigma)$, if

- $B(0) = 0$
- For $0 \leq r_1 < r_2 \leq 1$, $B(r_2) - B(r_1)$ is independent of $\{B(r) : 0 \leq r \leq r_1\}$ and is distributed $\mathcal{N}[0, \Sigma \cdot (r_2 - r_1)]$
- Any realization of $B(\cdot)$ is continuous

If $B \sim BM(I)$, B is a k -dimensional Wiener process. Clearly, $B \sim BM(\Sigma)$ if $B(\cdot) = \Sigma^{1/2}W(\cdot)$ for some Wiener process W (and vice versa), where $\Sigma^{1/2}$ is a $k \times k$ matrix such that $\Sigma^{1/2}\Sigma^{1/2'} = \Sigma$.

VAR(1) estimation, case 1. When $\Psi(L) = I$, the model is

$$z_t = z_{t-1} + \varepsilon_t \quad (t = 1, \dots, T),$$

where $z_0 = 0$ and $\varepsilon_t \sim i.i.d. (0, \Sigma)$.

Suppose the following equation is estimated by OLS:

$$z_t = \hat{A}z_{t-1} + \hat{\varepsilon}_t.$$

The OLS estimator $\hat{A} = \left(\sum_{t=1}^T z_t z_{t-1}' \right) \left(\sum_{t=1}^T z_{t-1} z_{t-1}' \right)^{-1}$ satisfies

$$T(\hat{A} - I) = \left(\frac{1}{T} \sum_{t=1}^T z_{t-1} \varepsilon_t' \right)' \left(\frac{1}{T^2} \sum_{t=1}^T z_{t-1} z_{t-1}' \right)^{-1}.$$

Donsker's theorem generalizes in an obvious way to the multivariate case:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T \rfloor} \varepsilon_t \rightarrow_d \Sigma^{1/2} W(\cdot),$$

where W is a k -dimensional Wiener process. Upon defining $Z_T(r) = z_{\lfloor Tr \rfloor} / \sqrt{T}$, we therefore have:

$$\frac{1}{T^2} \sum_{t=1}^T z_{t-1} z_{t-1}' = \int_0^1 Z_T(r) Z_T(r)' dr \rightarrow_d \Sigma^{1/2} \left[\int_0^1 W(r) W(r)' dr \right] \Sigma^{1/2}.$$

In the univariate case, the derivation of the limiting distribution of the OLS estimator was completed by showing that

$$\frac{1}{T} \sum_{t=1}^T y_{t-1} \varepsilon_t \rightarrow_d \frac{1}{2} \sigma^2 [W(1)^2 - 1].$$

As it turns, the following multivariate analogue of that result does *not* hold:

$$\frac{1}{T} \sum_{t=1}^T z_{t-1} \varepsilon_t' \rightarrow_d \frac{1}{2} \Sigma^{1/2} [W(1) W(1)' - I] \Sigma^{1/2}.$$

By analogy with the univariate case, we have:

$$z_t z_t' = (z_{t-1} + \varepsilon_t)(z_{t-1} + \varepsilon_t)' = z_{t-1} z_{t-1}' + z_{t-1} \varepsilon_t' + \varepsilon_t z_{t-1}' + \varepsilon_t \varepsilon_t'$$

and

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T z_{t-1} \varepsilon_t' + \frac{1}{T} \sum_{t=1}^T \varepsilon_t z_{t-1}' &= \frac{1}{T} z_T z_T' - \frac{1}{T} \sum_{t=1}^T \varepsilon_t \varepsilon_t' \\ &= Z_T(1) Z_T(1)' - \frac{1}{T} \sum_{t=1}^T \varepsilon_t \varepsilon_t' \\ &\rightarrow_d \Sigma^{1/2} [W(1) W(1)'] \Sigma^{1/2} - \Sigma \\ &= \Sigma^{1/2} [W(1) W(1)' - I] \Sigma^{1/2}. \end{aligned}$$

Unlike the univariate case, $T^{-1} \sum_{t=1}^T z_{t-1} \varepsilon_t' = \left(T^{-1} \sum_{t=1}^T \varepsilon_t z_{t-1}' \right)' \neq T^{-1} \sum_{t=1}^T \varepsilon_t z_{t-1}'$ in general, so

$$\frac{1}{T} \sum_{t=1}^T z_{t-1} \varepsilon'_t \neq \frac{1}{2} \left(\frac{1}{T} \sum_{t=1}^T z_{t-1} \varepsilon'_t + \frac{1}{T} \sum_{t=1}^T \varepsilon_t z'_{t-1} \right)$$

and a different method of proof must be employed in the multivariate case.

A result that *does* generalize to multivariate settings can be obtained by making use of the fact that

$$\frac{1}{2} \left[W(1)^2 - 1 \right] = \int_0^1 W(r) dW(r),$$

where the integral on the right hand side is an Itô integral, the definition of which can be found in stochastic calculus textbooks such as “Brownian Motion and Stochastic Calculus” by Karatzas and Shreve. The univariate result

$$\frac{1}{T} \sum_{t=1}^T y_{t-1} \varepsilon_t \rightarrow_d \frac{1}{2} \sigma^2 \left[W(1)^2 - 1 \right]$$

can therefore be rephrased as follows:

$$\frac{1}{T} \sum_{t=1}^T y_{t-1} \varepsilon_t \rightarrow_d \sigma^2 \int_0^1 W(r) dW(r).$$

Chan and Wei (1988, *Annals of Statistics*, 16, 367-401) have shown that the result in the preceding display has the following multivariate analogue:

$$\frac{1}{T} \sum_{t=1}^T z_{t-1} \varepsilon'_t \rightarrow_d \Sigma^{1/2} \left[\int_0^1 W(r) dW(r)' \right] \Sigma^{1/2'}.$$

As a consequence,

$$\begin{aligned} T(\hat{A} - I) &= \left(\frac{1}{T} \sum_{t=1}^T z_{t-1} \varepsilon'_t \right)' \left(\frac{1}{T^2} \sum_{t=1}^T z_{t-1} z'_{t-1} \right)^{-1} \\ &\rightarrow_d \left(\Sigma^{1/2} \left[\int_0^1 W(r) dW(r)' \right] \Sigma^{1/2'} \right)' \left(\Sigma^{1/2} \left[\int_0^1 W(r) W(r)' dr \right] \Sigma^{1/2'} \right)^{-1} \\ &= \Sigma^{1/2} \left[\int_0^1 W(r) dW(r)' \right]' \left[\int_0^1 W(r) W(r)' dr \right]^{-1} \Sigma^{-1/2}. \end{aligned}$$

Remark. A Wald test of $A = I$ can be based on

$$F = \text{vec}(\hat{A} - I)' \left(\left(\sum_{t=1}^T z_{t-1} z'_{t-1} \right)^{-1} \otimes \hat{\Sigma} \right)^{-1} \text{vec}(\hat{A} - I),$$

where $\hat{\Sigma} = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t'$. Its null distribution is a multivariate analogue of the Dickey-Fuller distribution:

$$F \rightarrow_d tr \left(\left[\int_0^1 W(r) dW(r)' \right]' \left[\int_0^1 W(r) W(r)' dr \right]^{-1} \left[\int_0^1 W(r) dW(r)' \right] \right).$$

In particular, the limiting null distribution of F is non-standard, but does not depend on any nuisance parameters.

Similarly, a multivariate version of the Dickey-Fuller coefficient test can be based on $tr \left[T \left(\hat{A} - I \right) \right]$:

$$tr \left[T \left(\hat{A} - I \right) \right] \rightarrow_d tr \left(\left[\int_0^1 W(r) W(r)' dr \right]^{-1} \left[\int_0^1 W(r) dW(r)' \right] \right). \quad \blacksquare$$

Stochastic integrals. Let W be a scalar Wiener process and let F be a (possibly random and/or vector valued) function. It turns out to be nontrivial to define an integral of F with respect to W . The usual (Riemann-Stieltjes) construction proceeds as follows:

- Pick any $0 \leq t_0 < t_1 < \dots < t_n \leq 1$.
- Pick any $\{s_i : 1 \leq i \leq n\}$ with $t_{i-1} \leq s_i \leq t_i \forall i$.
- Compute the sum $\sum_{i=1}^n F(s_i) [W(t_i) - W(t_{i-1})]$.
- Take limits as $n \rightarrow \infty$ and $\max_{1 \leq i \leq n} |t_i - t_{i-1}| \rightarrow 0$.

This construction fails when W is a Wiener process. Roughly speaking, the Itô construction is:

- Pick a particular partition $0 \leq t_0 < t_1 < \dots < t_n \leq 1$.
- Compute the sum $\sum_{i=1}^n F(t_{i-1}) [W(t_i) - W(t_{i-1})]$.
- Take limits as $n \rightarrow \infty$ ($\Rightarrow \max_{1 \leq i \leq n} |t_i - t_{i-1}| \rightarrow 0$).

The resulting (well defined) limit is a random variable denoted

$$\int_0^1 F(r) dW(r).$$

The properties of this random variable depend on the properties of F . For our purposes, the following results suffice:

1. Suppose F is non-random and continuous. Then

$$\int_0^1 F(r) dW(r) \sim \mathcal{N} \left[0, \int_0^1 F(r) F(r)' dr \right].$$

Moreover,

$$\int_0^1 F(r) dW(r) = F(r) W(r) \Big|_{r=0}^1 - \int_0^1 F'(r) W(r) dr$$

whenever F is differentiable. For example (set $F(r) = 1$),

$$\int_0^1 dW(r) = W(1)$$

and (set $F(r) = 1 - r$)

$$\int_0^1 (1 - r) dW(r) = \int_0^1 W(r) dr \sim \mathcal{N}\left(0, \frac{1}{3}\right).$$

2. Suppose F is continuous, random and independent of W . Then

$$\int_0^1 F(r) dW(r) \sim \mathcal{N}\left[0, \int_0^1 F(r) F(r)' dr\right]$$

conditional on $\{F(r) : 0 \leq r \leq 1\}$. For example, if W_x and W_y are independent Wiener processes of dimension m and 1, respectively,

$$\int_0^1 W_x(r) dW_y(r) \sim \mathcal{N}\left[0, \int_0^1 W_x(r) W_x(r)' dr\right]$$

conditional on $\{W_x(r) : 0 \leq r \leq 1\}$.

3. Suppose $F = F_1 + F_2$, where F_1 is of bounded variation and F_2 is a Brownian motion. Then

$$\int_0^1 F(r) dW(r) = F(r) W(r) \Big|_{r=0}^1 - \int_0^1 dF(r) W(r) - \int_0^1 dE[F_2(r) W(r)].$$

For example (take $F(r) = W(r) \Rightarrow E[F(r) W(r)] = r$),

$$\begin{aligned} \int_0^1 W(r) dW(r) &= W(r)^2 \Big|_{r=0}^1 - \int_0^1 dW(r) W(r) - \int_0^1 dr \\ &= W(1)^2 - \int_0^1 W(r) dW(r) - 1, \end{aligned}$$

implying

$$\frac{1}{2} [W(1)^2 - 1] = \int_0^1 W(r) dW(r).$$

VAR(1) estimation, case 2. Suppose

$$z_t = z_{t-1} + v_t, \quad v_t = \Psi(L) \varepsilon_t \quad (t = 1, \dots, T),$$

where $z_0 = 0$, $\Psi(L) = \sum_{i=0}^{\infty} \Psi_i L^i$ with $|\Psi(1)| \neq 0$, $\sum_{i=1}^{\infty} i \|\Psi_i\| < \infty$, and $\varepsilon_t \sim i.i.d. (0, \Sigma)$.

The full rank assumption $|\Psi(1)| \neq 0$ is stronger than the assumption $\Psi(1) \neq 0$. The summability condition on $\{\Psi_i : i \geq 1\}$ is expressed in terms of the *Euclidean norm* $\|\cdot\|$ defined as

$$\|M\| = \left(\sum_{i=1}^m \sum_{j=1}^n m_{ij}^2 \right)^{1/2} = \text{tr}(M'M)^{1/2}$$

for any $m \times n$ matrix M with elements $\{m_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$.

The BN decomposition of $\Psi(L)$ is

$$\Psi(L) = \Psi(1) + (1-L)\tilde{\Psi}(L),$$

where $\tilde{\Psi}(L) = \sum_{i=0}^{\infty} \tilde{\Psi}_i L^i$ with $\tilde{\Psi}_i = -\sum_{j=i+1}^{\infty} \Psi_j$. The summability condition $\sum_{i=1}^{\infty} i \|\Psi_i\| < \infty$ implies $\sum_{i=0}^{\infty} \|\tilde{\Psi}_i\| < \infty$. Using these algebraic results and the multivariate version of Donsker's theorem, it can be shown that

$$T^{-1/2} \sum_{t=1}^{\lfloor T \cdot \rfloor} v_t \rightarrow_d \Psi(1) \Sigma^{1/2} W(\cdot) = \Omega^{1/2} W(\cdot),$$

where $\Omega = \Psi(1) \Sigma \Psi(1)'$ in the long-run covariance matrix of v_t .

Upon defining $Z_T(r) = z_{\lfloor Tr \rfloor} / \sqrt{T}$, we therefore have:

$$\frac{1}{T^2} \sum_{t=1}^T z_{t-1} z'_{t-1} = \int_0^1 Z_T(r) Z_T(r)' dr \rightarrow_d \Omega^{1/2} \left[\int_0^1 W(r) W(r)' dr \right] \Omega^{1/2'}$$

As in case 1, we have to be careful when generalizing the univariate result

$$\frac{1}{T} \sum_{t=1}^T y_{t-1} v_t \rightarrow_d \frac{1}{2} \omega^2 [W(1)^2 - 1] + \frac{1}{2} [\omega^2 - \gamma_v(0)],$$

where $\omega^2 = \sigma^2 \psi(1)^2$ and $\gamma_v(0) = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2$. Using the relations

$$\frac{1}{2} [W(1)^2 - 1] = \int_0^1 W(r) dW(r)$$

and

$$\frac{1}{2} [\omega^2 - \gamma_v(0)] = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E(y_{t-1} v_t) = \sum_{i=1}^{\infty} \gamma_v(-i), \quad \gamma_v(-i) = E(v_{t-i} v_t),$$

the univariate result is:

$$\frac{1}{T} \sum_{t=1}^T y_{t-1} v_t \rightarrow_d \omega^2 \int_0^1 W(r) dW(r) + \lambda, \quad \lambda = \sum_{i=1}^{\infty} \gamma_v(-i).$$

The following multivariate analogue of that result is due to Phillips (1988, *Econometric Theory*, 4, 528-533):

$$\frac{1}{T} \sum_{t=1}^T z_{t-1} v_t' \rightarrow_d \Omega^{1/2} \left[\int_0^1 W(r) dW(r)' \right] \Omega^{1/2'} + \Lambda,$$

where

$$\Lambda = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E(z_{t-1} v_t') = \sum_{i=1}^{\infty} \Gamma_v(-i), \quad \Gamma_v(-i) = E(v_{t-i} v_t')$$

As a consequence,

$$\begin{aligned} T(\hat{A} - I) &= \left(\frac{1}{T} \sum_{t=1}^T z_{t-1} v_t' \right)' \left(\frac{1}{T^2} \sum_{t=1}^T z_{t-1} z_{t-1}' \right)^{-1} \\ &\rightarrow_d \left(\Omega^{1/2} \left[\int_0^1 W(r) dW(r)' \right] \Omega^{1/2'} + \Lambda \right)' \left(\Omega^{1/2} \left[\int_0^1 W(r) W(r)' dr \right] \Omega^{1/2'} \right)^{-1}. \end{aligned}$$

Remark. Let

$$\hat{A}^+ = \hat{A} - T \cdot \hat{\Lambda}' \left(\sum_{t=1}^T z_{t-1} z_{t-1}' \right)^{-1},$$

where $\hat{\Lambda}$ is a consistent estimator of Λ such as

$$\hat{\Lambda} = \sum_{i=1}^b \left(1 - \frac{i}{b} \right) \hat{\Gamma}_v(-i), \quad \hat{\Gamma}_v(-i) = \frac{1}{T} \sum_{t=i+1}^T \hat{v}_{t-i} \hat{v}_t'$$

which is consistent if $b \rightarrow \infty$ but $b/\sqrt{T} \rightarrow 0$ as $T \rightarrow \infty$. We have:

$$\begin{aligned} T(\hat{A}^+ - I) &= T(\hat{A} - I) - \hat{\Lambda}' \left(T^{-2} \sum_{t=1}^T z_{t-1} z_{t-1}' \right)^{-1} \\ &\rightarrow_d \Omega^{1/2} \left[\int_0^1 W(r) dW(r)' \right]' \left[\int_0^1 W(r) W(r)' dr \right]^{-1} \Omega^{-1/2}. \end{aligned}$$

A Wald test of $A = I$ can be based on

$$F^+ = \text{vec}(\hat{A}^+ - I)' \left[\left(\sum_{t=1}^T z_{t-1} z_{t-1}' \right)^{-1} \otimes \hat{\Omega} \right]^{-1} \text{vec}(\hat{A}^+ - I),$$

where $\hat{\Omega}$ is a consistent estimator of Ω such as the Newey-West estimator

$$\hat{\Omega} = \hat{\Gamma}_v(0) + \sum_{i=1}^b \left(1 - \frac{i}{b} \right) \left(\hat{\Gamma}_v(i) + \hat{\Gamma}_v(i)' \right), \quad \hat{\Gamma}_v(i) = T^{-1} \sum_{t=i+1}^T \hat{v}_t \hat{v}_{t-i}'$$

Its null distribution can be characterized as follows:

$$F^+ \rightarrow_d \text{tr} \left(\left[\int_0^1 W(r) dW(r)' \right]' \left[\int_0^1 W(r) W(r)' dr \right]^{-1} \left[\int_0^1 W(r) dW(r)' \right] \right). \quad \blacksquare$$

As does its univariate counterpart, the multivariate unit root model gives rise to nonstandard distributional results. In the univariate unit root model, standard theory applies if the first difference of the data is used. A similar result holds in the multivariate unit root model. It turns out, however, that there is an important class of multivariate nonstationary time series model for which this is not the case. These models, cointegration models, have $\Psi(1) \neq 0$ but $|\Psi(1)| = 0$ and will be studied later. Before doing so, we illustrate another problem that can arise when multivariate models are estimated in levels.

Spurious regressions. Suppose $\{y_t : 1 \leq t \leq T\}$ and $\{x_t : 1 \leq t \leq T\}$ are independent random walks of dimension 1 and m , respectively:

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \varepsilon_t^y \\ \varepsilon_t^x \end{pmatrix} \sim i.i.d. \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{yy} & 0 \\ 0 & \Sigma_{xx} \end{pmatrix} \right].$$

Moreover, suppose the following equation is estimated by OLS:

$$y_t = \hat{\beta}' x_t + \hat{u}_t \quad (t = 1, \dots, T)$$

Since $\{y_t : 1 \leq t \leq T\}$ and $\{x_t : 1 \leq t \leq T\}$ are independent, intuition would suggest that $\hat{\beta}$ should be insignificantly different from zero in moderate samples and that $\hat{\beta} \rightarrow_p 0$ and $R^2 \rightarrow_p 0$. In an influential Monte Carlo study, Granger and Newbold (1974, *Journal of Econometrics*, 2, 111-120) found that regressions of this type tend to produce a moderately large R^2 and a large regression F -statistic.

The OLS estimator is

$$\hat{\beta} = \left(\sum_{t=1}^T x_t x_t' \right)^{-1} \left(\sum_{t=1}^T x_t y_t \right).$$

The regression F -statistic is

$$\begin{aligned} F &= m^{-1} \cdot \hat{\beta}' \left[\hat{\sigma}^2 \left(\sum_{t=1}^T x_t x_t' \right)^{-1} \right]^{-1} \hat{\beta} \\ &= m^{-1} \cdot \hat{\sigma}^{-2} \left(\sum_{t=1}^T x_t y_t' \right)' \left(\sum_{t=1}^T x_t x_t' \right)^{-1} \left(\sum_{t=1}^T x_t y_t' \right) \\ &= \frac{T-m}{m} \cdot \frac{R^2}{1-R^2}, \end{aligned}$$

where $\hat{\sigma}^2 = (T-m)^{-1} \sum_{t=1}^T \hat{u}_t^2$ and

$$R^2 = \frac{\sum_{t=1}^T y_t^2 - \sum_{t=1}^T \hat{u}_t^2}{\sum_{t=1}^T y_t^2}.$$

Let

$$Y_T(r) = T^{-1/2} y_{[Tr]}.$$

and

$$X_T(r) = T^{-1/2} x_{[Tr]}$$

By the multivariate FCLT,

$$\begin{pmatrix} Y_T(\cdot) \\ X_T(\cdot) \end{pmatrix} \rightarrow_d \begin{pmatrix} \sigma_{yy}^{1/2} W_y(\cdot) \\ \Sigma_{xx}^{1/2} W_x(\cdot) \end{pmatrix},$$

where W_y and W_x are independent Wiener processes of dimension 1 and m , respectively. By CMT,

$$\begin{aligned} \hat{\beta} &= \left(\frac{1}{T^2} \sum_{t=1}^T x_t x_t' \right)^{-1} \left(\frac{1}{T^2} \sum_{t=1}^T x_t y_t \right) \\ &= \left[\int_0^1 X_T(r) X_T(r)' dr + \frac{1}{T} X_T(1) X_T(1)' \right]^{-1} \left[\int_0^1 X_T(r) Y_T(r) dr + \frac{1}{T} X_T(1) Y_T(1)' \right] \\ &\rightarrow_d \sigma_{yy}^{1/2} \Sigma_{xx}^{-1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right]^{-1} \left[\int_0^1 W_x(r) W_y(r) dr \right]. \end{aligned}$$

In other words, $\hat{\beta}$ has a non-degenerate limiting distribution (with mean zero). The limiting distribution of R^2 is also non-degenerate,

$$\begin{aligned} R^2 &= \frac{\left(\sum_{t=1}^T y_t x_t' \right) \left(\sum_{t=1}^T x_t x_t' \right)^{-1} \left(\sum_{t=1}^T x_t y_t \right)}{\sum_{t=1}^T y_t^2} \\ &= \frac{\left[\int_0^1 Y_T(r) X_T(r)' dr \right] \left[\int_0^1 X_T(r) X_T(r)' dr \right]^{-1} \left[\int_0^1 X_T(r) Y_T(r) dr \right]}{\int_0^1 Y_T(r)^2 dr} + o_p(1) \\ &\rightarrow_d \frac{\left[\int_0^1 W_y(r) W_x(r)' dr \right] \left[\int_0^1 W_x(r) W_x(r)' dr \right]^{-1} \left[\int_0^1 W_x(r) W_y(r) dr \right]}{\int_0^1 W_y(r)^2 dr}, \end{aligned}$$

as is the limiting distribution of $T^{-1}F$:

$$\frac{1}{T}F = \frac{T-m}{T \cdot m} \cdot \frac{R^2}{1-R^2}$$

$$\xrightarrow{d} \frac{1}{m} \cdot \frac{\left[\int_0^1 W_y(r) W_x(r)' dr \right] \left[\int_0^1 W_x(r) W_x(r)' dr \right]^{-1} \left[\int_0^1 W_x(r) W_y(r)' dr \right]}{\int_0^1 W_y(r)^2 dr - \left[\int_0^1 W_y(r) W_x(r)' dr \right] \left[\int_0^1 W_x(r) W_x(r)' dr \right]^{-1} \left[\int_0^1 W_x(r) W_y(r)' dr \right]}.$$

The F -statistic diverges even though $\{y_t\}$ and $\{x_t\}$ are completely independent. This phenomenon, the *spurious regression problem*, illustrates what might happen if we run regressions in levels when the data is nonstationary. The qualitative conclusions remain the same if we include deterministic terms in the regression, allow Δy_t and Δx_t to be serially correlated and/or imperfectly correlated with each other. The “autocorrelation robust” regression statistic

$$G = \hat{\beta}' \left[\hat{\omega}^2 \left(\sum_{t=1}^T x_t x_t' \right)^{-1} \right]^{-1} \hat{\beta},$$

where $\hat{\omega}^2$ is the Newey-West estimator computed from $\{\hat{u}_t\}$, also diverges, albeit at a slower rate than F .

The spurious regression example resembles the multivariate unit root example. An obvious solution to the spurious regression problem is to run regressions in first differences.

Cointegration

A bivariate example. Suppose y_t and x_t are $I(1)$ (i.e., each series has a unit root). Then y_t and x_t are *cointegrated* if $y_t - \beta x_t$ is $I(0)$ for some β . Therefore, Δy_t and Δx_t are stationary, as is $y_t - \beta x_t$. Cointegration is an implication of many economic theories. For instance, a simple version of the permanent income hypothesis suggests that consumption and income should be cointegrated if they each have a unit root.

A bivariate cointegration model is the following:

$$\begin{aligned} y_t &= \beta x_t + u_t^y \\ \Delta x_t &= u_t^x, \end{aligned}$$

where $\beta \neq 0$ and

$$\begin{pmatrix} u_t^y \\ u_t^x \end{pmatrix} \sim i.i.d. \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{yy} & 0 \\ 0 & \sigma_{xx} \end{pmatrix} \right].$$

Since $\Delta x_t = u_t^x \sim WN(0, \sigma_{xx})$, $x_t \sim ARIMA(0, 1, 0)$. In particular, $x_t \sim I(1)$. Moreover,

$$\Delta y_t = \beta \Delta x_t + \Delta u_t^y = \beta u_t^x + u_t^y - u_{t-1}^y$$

has a non-invertible $MA(1)$ representation, so $y_t \sim ARIMA(0, 1, 1)$. In particular, $y_t \sim I(1)$. The linear combination $y_t - \beta x_t = u_t^y \sim WN(0, \sigma_{yy})$ is $I(0)$, so y_t and x_t are cointegrated.

The vector moving average (VMA) representation of $(\Delta y_t, \Delta x_t)'$ is

$$\begin{aligned} \begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} &= \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u_t^y \\ u_t^x \end{pmatrix} + \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} u_{t-1}^y \\ u_{t-1}^x \end{pmatrix} \\ &= \Psi(L) \begin{pmatrix} u_t^y \\ u_t^x \end{pmatrix}. \end{aligned}$$

Since

$$\Psi(1) = \begin{pmatrix} 0 & \beta \\ 0 & 1 \end{pmatrix}$$

is singular, $\Psi(L)$ is non-invertible and $(\Delta y_t, \Delta x_t)'$ does not have a VAR representation.

The VAR representation of $(y_t, x_t)'$ is

$$\begin{pmatrix} 1 & -\beta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_t^y \\ u_t^x \end{pmatrix}$$

\Downarrow

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0 & \beta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} e_t^y \\ e_t^x \end{pmatrix}, \quad \begin{pmatrix} e_t^y \\ e_t^x \end{pmatrix} = \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u_t^y \\ u_t^x \end{pmatrix}.$$

The eigenvalues of

$$\begin{pmatrix} 0 & \beta \\ 0 & 1 \end{pmatrix}$$

are zero and one, so $(y_t, x_t)'$ is nonstationary.

An alternative representation is

$$\begin{aligned} \begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} &= \begin{pmatrix} -1 & \beta \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} e_t^y \\ e_t^x \end{pmatrix} \\ &= \begin{pmatrix} -1 \\ 0 \end{pmatrix} (1, -\beta) \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} e_t^y \\ e_t^x \end{pmatrix}. \end{aligned}$$

The multivariate case. Let z_t be a $k \times 1$ vector of integrated time series (i.e. each element is $I(1)$). If there exists a full rank $k \times r$ matrix γ such that $\gamma' z_t$ is $I(0)$, but $\tilde{\gamma}' z_t$ is non-stationary whenever $\tilde{\gamma}$ is a full rank $k \times (r+1)$ matrix, then z_t is *cointegrated with cointegration rank r* . The cointegration rank r is unique. On the other hand, the matrix γ appearing in the definition of cointegration is not unique,

because $(\gamma M)' z_t = M' \gamma' z_t$ is $I(0)$ whenever M is a nonsingular $r \times r$ matrix and $\gamma' z_t$ is $I(0)$.

Representations. Suppose z_t is cointegrated with cointegration rank r .

The *triangular representation* of z_t is

$$\begin{aligned} y_t &= \beta' x_t + u_t^y \\ \Delta x_t &= u_t^x, \end{aligned}$$

where $(u_t^y, u_t^x)'$ is stationary and the rows of

$$z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix} \begin{array}{c} \updownarrow r \\ \updownarrow k - r \end{array}$$

have been permuted (if necessary) in such a way that we can set

$$\gamma = \begin{pmatrix} I_r \\ -\beta \end{pmatrix}$$

in the definition of cointegration. The triangular representation is useful for estimation. In particular, the parameter γ is identified.

The *moving average representation* of Δz_t is

$$\Delta z_t = \Psi(L) \varepsilon_t,$$

where $\varepsilon_t \sim WN(\Sigma)$ and $\Psi(1)$ has rank $k - r$. Notice that the rank deficiency of $\Psi(1)$ equals the cointegration rank of z_t . Indeed, it is easy to show that

$$\gamma' \Psi(1) = 0.$$

The moving average representation is useful for trend extraction.

If z_t is generated by a $VAR(p)$,

$$z_t = A_1 z_{t-1} + \dots + A_p z_{t-p} + \varepsilon_t,$$

then the *vector error correction representation* of z_t is

$$\Delta z_t = \delta \gamma' z_{t-1} + \Phi_1 \Delta z_{t-1} + \dots + \Phi_{p-1} \Delta z_{t-p} + \varepsilon_t,$$

where δ is a full rank $k \times r$ matrix, as is γ . In particular, the $k \times k$ matrix $\delta \gamma'$ has (reduced) rank r . The vector error correction representation is useful for estimation. Since $\delta \gamma' = \delta (M')^{-1} M' \gamma'$ for any nonsingular $r \times r$ matrix M , γ is not identified unless further restrictions (such as $\gamma = (I_r, -\beta)'$) are imposed.

Estimation. When z_t is cointegrated, $\Psi(L)$ in the moving average representation of Δz_t is noninvertible and we cannot simply estimate a VAR in first differences. On the other hand, distributional results are

non-standard if an unrestricted VAR in levels is estimated. In contrast, it turns out that estimators obtained from models imposing the true cointegration rank are well behaved. The most interesting problem is that of estimating γ , the parameter characterizing the cointegrating relations. Indeed, inference on the parameters characterizing the transient dynamics of the system (i.e. the serial correlation pattern of $(u_t^y, u_t^x)'$ in the triangular representation and $\delta, \Phi_1, \dots, \Phi_{p-1}$ in the vector error correction representation) is completely standard and will not be discussed here. For details, see Watson's (1994) handbook chapter.

Estimates of the parameter γ can be obtained from the triangular model or the vector error correction model. The latter model can be estimated by reduced rank regression (which is maximum likelihood under normality) once identifying assumptions on γ have been made. This estimation method, the Johansen procedure, is described in Hamilton's chapter 20. An alternative approach, and the approach we will take, is to base estimation on the triangular model. In that model, identification is achieved by specifying y_t and x_t in such a way that $z_t = (y_t, x_t)'$ is cointegrated with $\gamma = (I_r, -\beta')$, where β is an unknown parameter.

For simplicity, we focus on the case where y_t is a scalar time series. Moreover, we ignore deterministic terms, since the inclusion of these does not change the results of primary interest. We will study the model

$$\begin{aligned} y_t &= \beta' x_t + u_t^y \\ \Delta x_t &= u_t^x, \end{aligned}$$

where y_t is a scalar, x_t is an m -vector ($m = k - 1$) with $x_0 = 0$ and

$$\begin{pmatrix} u_t^y \\ u_t^x \end{pmatrix} = \Psi(L) \varepsilon_t.$$

Three cases will be considered:

1. $\Psi(L) = I_k$ and

$$\varepsilon_t = \begin{pmatrix} \varepsilon_t^y \\ \varepsilon_t^x \end{pmatrix} \sim i.i.d. \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{yy} & 0 \\ 0 & \Sigma_{xx} \end{pmatrix} \right].$$

2. $\Psi(L) = I_k$ and

$$\varepsilon_t = \begin{pmatrix} \varepsilon_t^y \\ \varepsilon_t^x \end{pmatrix} \sim i.i.d. \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{yy} & \sigma'_{xy} \\ \sigma_{xy} & \Sigma_{xx} \end{pmatrix} \right].$$

3. $\Psi(L) = \sum_{i=0}^{\infty} \Psi_i L^i$ and $\varepsilon_t \sim i.i.d. (0, \Sigma)$.

In case 1 there is no serial correlation and the regressor x_t is strictly exogenous in the sense that $\{u_t^y\}$ is independent of $\{x_t\}$. In case 1, the model reduces to the neoclassical normal regression model studied in e.g. Econ 240A. It turns out that many results from case 1 carry over to cases 2, where "endogeneity" is introduced, and case 3, where serial correlation is accommodated.

CASE 1: No serial correlation, exogenous regressors

The model is:

$$\begin{aligned} y_t &= \beta' x_t + \varepsilon_t^y \\ \Delta x_t &= \varepsilon_t^x, \end{aligned}$$

where y_t is a scalar, x_t is an m -vector ($m = k - 1$) with $x_0 = 0$ and

$$\varepsilon_t = \begin{pmatrix} \varepsilon_t^y \\ \varepsilon_t^x \end{pmatrix} \sim i.i.d. \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{yy} & 0 \\ 0 & \Sigma_{xx} \end{pmatrix} \right].$$

The OLS estimator

$$\hat{\beta} = \left(\sum_{t=1}^T x_t x_t' \right)^{-1} \left(\sum_{t=1}^T x_t y_t \right)$$

is also the MLE. Its distributional properties are well known:

$$\hat{\beta} \sim \mathcal{N} \left[\beta, \sigma_{yy} \left(\sum_{t=1}^T x_t x_t' \right)^{-1} \right]$$

conditional on $\{x_t\}$. The unconditional distribution of $\hat{\beta}$ is a (covariance) mixture of normals with mean β .

Inference on β is straightforward. The hypothesis $H_0 : R\beta = q$, where R is a full rank $r \times m$ matrix, can be tested by means of

$$F = \frac{1}{r} (R\hat{\beta} - q)' \left[\hat{\sigma}^2 R \left(\sum_{t=1}^T x_t x_t' \right)^{-1} R' \right]^{-1} (R\hat{\beta} - q),$$

where $\hat{\sigma}^2 = (T - m)^{-1} \sum_{t=1}^T \hat{\varepsilon}_t^y$, $\hat{\varepsilon}_t^y = y_t - \hat{\beta}' x_t$. Under H_0 , $F \sim F(r, T - m)$ conditional on $\{x_t\}$. This conditional distribution does not depend on $\{x_t\}$ and is therefore the unconditional distribution as well.

Asymptotics. As usual, we need to rely on asymptotic results in more realistic models than the case 1 model. For comparison, we therefore characterize the asymptotic properties of $\hat{\beta}$ and F in case 1. We have:

$$T(\hat{\beta} - \beta) \sim \mathcal{N} \left[0, \sigma_{yy} \left(\frac{1}{T^2} \sum_{t=1}^T x_t x_t' \right)^{-1} \right]$$

conditional on $\{x_t\}$. Moreover,

$$\frac{1}{T^2} \sum_{t=1}^T x_t x_t' \rightarrow_d \Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2'}$$

where W_x is an m -dimensional Wiener process. Heuristically, we might therefore expect the limiting distribution of $T(\hat{\beta} - \beta)$ to be covariance mixture of normals with mean zero and “variance”

$$\sigma_{yy} \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2'} \right)^{-1}.$$

More formally, using the results from the multivariate unit root model, we have:

$$\begin{aligned}
T(\hat{\beta} - \beta) &= \left(\frac{1}{T^2} \sum_{t=1}^T x_t x_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^T x_t \varepsilon_t^y \right) \\
&= \left(\frac{1}{T^2} \sum_{t=1}^T x_t x_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^T x_{t-1} \varepsilon_t^y + \frac{1}{T} \sum_{t=1}^T \varepsilon_t^x \varepsilon_t^y \right) \\
&\rightarrow_d \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2'} \right)^{-1} \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_y(r) \right] \sigma_{yy}^{1/2} \right),
\end{aligned}$$

where W_x and W_y are independent Wiener processes of dimension m and 1, respectively.

Since $T(\hat{\beta} - \beta) = O_p(1)$, $\hat{\beta}$ is superconsistent. As in the unit root model, this result is driven by the fact that the regressor (x_t) and the error term (ε_t^y) are I(1) and I(0), respectively. By the properties of the Itô integral,

$$\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_y(r) \right] \sigma_{yy}^{1/2} \sim \mathcal{N} \left(0, \sigma_{yy} \Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2'} \right)$$

conditional on $\{W_x(r) : 0 \leq r \leq 1\}$. Therefore, the limiting distribution of $T(\hat{\beta} - \beta)$ is a covariance mixture of normals:

$$\begin{aligned}
&\left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2'} \right)^{-1} \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_y(r) \right] \sigma_{yy}^{1/2} \right) \\
&\sim \mathcal{N} \left[0, \sigma_{yy} \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2'} \right)^{-1} \right]
\end{aligned}$$

conditional on $\{W_x(r) : 0 \leq r \leq 1\}$.

Under $H_0 : R\beta = q$, $F \sim F(r, T - m)$ and

$$r \cdot F = (R\hat{\beta} - q)' \left[\hat{\sigma}^2 R \left(\sum_{t=1}^T x_t x_t' \right)^{-1} R' \right]^{-1} (R\hat{\beta} - q) \rightarrow_d \chi^2(r),$$

so asymptotic inference on β can be based on the χ^2 distribution.

CASE 2: No serial correlation, endogenous regressors

The model is:

$$\begin{aligned}
y_t &= \beta' x_t + \varepsilon_t^y \\
\Delta x_t &= \varepsilon_t^x,
\end{aligned}$$

where y_t is a scalar, x_t is an m -vector ($m = k - 1$) with $x_0 = 0$ and

$$\varepsilon_t = \begin{pmatrix} \varepsilon_t^y \\ \varepsilon_t^x \end{pmatrix} \sim i.i.d. \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{yy} & \sigma'_{xy} \\ \sigma_{xy} & \Sigma_{xx} \end{pmatrix} \right].$$

Estimation. Adding and subtracting $\sigma'_{xy} \Sigma_{xx}^{-1} \Delta x_t = \sigma'_{xy} \Sigma_{xx}^{-1} \varepsilon_t^x$ on the right hand side of the equation

$$y_t = \beta' x_t + \varepsilon_t^y,$$

we have:

$$y_t = \beta' x_t + \delta' \Delta x_t + \varepsilon_t^{y,x},$$

where $\delta = \Sigma_{xx}^{-1} \sigma_{xy}$ and $\varepsilon_t^{y,x} = \varepsilon_t^y - \sigma'_{xy} \Sigma_{xx}^{-1} \varepsilon_t^x \sim i.i.d. \mathcal{N}(0, \sigma_{yy.x} = \sigma_{yy} - \sigma'_{xy} \Sigma_{xx}^{-1} \sigma_{xy})$. Since $\{\varepsilon_t^{y,x}\}$ is independent of $\{x_t\}$, we are back in the neoclassical normal regression model and an unbiased estimator of β can be obtained by estimating the following equation by OLS:

$$y_t = \tilde{\beta}' x_t + \tilde{\delta}' \Delta x_t + \tilde{\varepsilon}_t^{y,x} \quad (t = 1, \dots, T).$$

Conditional on $\{x_t\}$,

$$\tilde{\beta} \sim \mathcal{N} \left[\beta, \sigma_{yy} \left(\sum_{t=1}^T x_t^\Delta x_t^{\Delta'} \right)^{-1} \right],$$

where $x_t^\Delta = x_t - \left(\sum_{s=1}^T x_s \Delta x_s' \right) \left(\sum_{s=1}^T \Delta x_s \Delta x_s' \right)^{-1} \Delta x_t$. The unconditional distribution of $\tilde{\beta}$ is a (covariance) mixture of normals with mean β .

The hypothesis $H_0 : R\beta = q$, where R is a full rank $r \times m$ matrix, can be tested by means of

$$F = \frac{1}{r} \left(R\tilde{\beta} - q \right)' \left[\hat{\sigma}^2 R \left(\sum_{t=1}^T x_t^\Delta x_t^{\Delta'} \right)^{-1} R' \right]^{-1} \left(R\tilde{\beta} - q \right),$$

where $\hat{\sigma}^2 = (T - 2m)^{-1} \sum_{t=1}^T (\tilde{\varepsilon}_t^{y,x})^2$. Under H_0 , $F \sim F(r, T - 2m)$ conditional on $\{x_t\}$. This conditional distribution does not depend on $\{x_t\}$ and is therefore the unconditional distribution as well.

The estimator $\tilde{\beta}$ is the *dynamic OLS* (DOLS) estimator of β . It can be shown that $\tilde{\beta}$ is the MLE of β . In finite samples, $\tilde{\beta}$ is different from the (static) OLS estimator $\hat{\beta} = \left(\sum_{t=1}^T x_t x_t' \right)^{-1} \left(\sum_{t=1}^T x_t y_t \right)$ from

$$y_t = \hat{\beta}' x_t + \hat{\varepsilon}_t^y \quad (t = 1, \dots, T).$$

Asymptotics. By the properties, of OLS,

$$T \left(\tilde{\beta} - \beta \right) = \left(\frac{1}{T^2} \sum_{t=1}^T x_t^\Delta x_t^{\Delta'} \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^T x_t^\Delta \varepsilon_t^{y,x} \right).$$

Using the results

$$\begin{aligned}\frac{1}{T} \sum_{t=1}^T \Delta x_t \Delta x_t' &\rightarrow_p E(\Delta x_t \Delta x_t') = \Sigma_{xx}, \\ \frac{1}{T} \sum_{t=1}^T x_t \Delta x_t' &= O_p(1), \\ \frac{1}{T} \sum_{t=1}^T \Delta x_t \varepsilon_t^{y,x} &\rightarrow_p E(\Delta x_t \varepsilon_t^{y,x}) = 0,\end{aligned}$$

we have:

$$\frac{1}{T^2} \sum_{t=1}^T x_t \Delta x_t \Delta x_t' = \frac{1}{T^2} \sum_{t=1}^T x_t x_t' + o_p(1) \rightarrow_d \Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2}$$

and

$$\frac{1}{T} \sum_{t=1}^T x_t \Delta \varepsilon_t^{y,x} = \frac{1}{T} \sum_{t=1}^T x_t \varepsilon_t^{y,x} + o_p(1) \rightarrow_d \Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_y(r) \right] \sigma_{yy,x}^{1/2},$$

where W_x and W_y are independent Wiener processes of dimension m and 1, respectively. As a consequence,

$$T(\tilde{\beta} - \beta) \rightarrow_d \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2} \right)^{-1} \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_y(r) \right] \sigma_{yy,x}^{1/2} \right),$$

a covariance mixture of normals:

$$\begin{aligned}&\left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2} \right)^{-1} \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_y(r) \right] \sigma_{yy,x}^{1/2} \right) \\ &\sim \mathcal{N} \left[0, \sigma_{yy,x} \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2} \right)^{-1} \right]\end{aligned}$$

conditional on $\{W_x(r) : 0 \leq r \leq 1\}$.

Under $H_0 : R\beta = q$, $F \sim F(r, T - m)$ and

$$r \cdot F = (R\tilde{\beta} - q)' \left[\tilde{\sigma}^2 R \left(\sum_{t=1}^T x_t x_t' \right)^{-1} R' \right]^{-1} (R\tilde{\beta} - q) \rightarrow_d \chi^2(r),$$

so asymptotic inference on β can be based on the χ^2 distribution.

The static OLS estimator satisfies

$$\begin{aligned} T(\hat{\beta} - \beta) &= \left(\frac{1}{T^2} \sum_{t=1}^T x_t x_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^T x_t \varepsilon_t^y \right) \\ &= \left(\frac{1}{T^2} \sum_{t=1}^T x_t x_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^T x_t \varepsilon_t^{y \cdot x} + \frac{1}{T} \sum_{t=1}^T x_t \varepsilon_t^{x'} \delta \right). \end{aligned}$$

Since

$$\begin{aligned} T^{-1} \sum_{t=1}^T x_t \varepsilon_t^{x'} &= T^{-1} \sum_{t=1}^T x_{t-1} \varepsilon_t^{x'} + T^{-1} \sum_{t=1}^T \varepsilon_t^x \varepsilon_t^{x'} \\ &\rightarrow_d \Sigma_{xx}^{1/2} \int_0^1 W_x(r) dW_x(r)' \Sigma_{xx}^{1/2'} + \Sigma_{xx}, \end{aligned}$$

we have:

$$\begin{aligned} T(\hat{\beta} - \beta) &\rightarrow_d \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2'} \right)^{-1} \\ &\quad \cdot \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_y(r) \right] \sigma_{yy \cdot x}^{1/2} + \Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_x(r)' \right] \Sigma_{xx}^{-1/2} \sigma_{xy} + \sigma_{xy} \right) \end{aligned}$$

and

$$T(\hat{\beta} - \tilde{\beta}) \rightarrow_d \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2'} \right)^{-1} \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_x(r)' \right] \Sigma_{xx}^{-1/2} \sigma_{xy} + \sigma_{xy} \right).$$

As is $\tilde{\beta}$, $\hat{\beta}$ is superconsistent. On the other hand, the limiting distributions of $\tilde{\beta}$ (the MLE) and $\hat{\beta}$ (the QMLE assuming $\sigma_{xy} = 0$) are different unless $\sigma_{xy} = 0$. The right hand side in the preceding display can be interpreted as a “simultaneous equations bias” term in the limiting representation of $T(\hat{\beta} - \beta)$. The distribution of the test statistic

$$\left(R\hat{\beta} - q \right)' \left[\hat{\sigma}^2 R \left(\sum_{t=1}^T x_t x_t' \right)^{-1} R' \right]^{-1} \left(R\hat{\beta} - q \right), \quad \hat{\sigma}^2 = \frac{1}{T-m} \sum_{t=1}^T \hat{\varepsilon}_t^{y2},$$

is not asymptotically pivotal under $H_0 : R\beta = q$.

Efficiency. From a practical point of view, $\tilde{\beta}$ is a more attractive estimator than $\hat{\beta}$ because it is straightforward to use $\tilde{\beta}$ to conduct inference. More importantly, it turns out that there is a well defined sense in which $\tilde{\beta}$ is asymptotically efficient relative to $\hat{\beta}$. Heuristically, it is pretty obvious that $\tilde{\beta}$ is superior to

$\hat{\beta}$ since the mixture normal limiting representations of $T(\tilde{\beta} - \beta)$ and $T(\hat{\beta} - \beta)$ have the same variance, but only the former has zero mean. A more satisfactory statement will be developed next.

Suppose $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimators of a (scalar) parameter $\theta \in \Theta \subseteq \mathbb{R}$ such that

$$\sqrt{n}(\hat{\theta}_1 - \theta) \rightarrow_d \mathcal{N}[0, \sigma_1^2(\theta)],$$

and

$$\sqrt{n}(\hat{\theta}_2 - \theta) \rightarrow_d \mathcal{N}[0, \sigma_2^2(\theta)].$$

In such cases, $\hat{\theta}_1$ is said to be asymptotically efficient relative to $\hat{\theta}_2$ if $\sigma_1^2(\theta) \leq \sigma_2^2(\theta)$ for all $\theta \in \Theta$. The applicability of this definition of asymptotic relative efficiency is limited by the fact that asymptotic normality with mean zero is required (the \sqrt{n} -consistency requirement is also restrictive, but is easy to relax). A more general definition is the following. A δ_n -consistent estimator $\hat{\theta}_1$ (i.e. $\delta_n(\hat{\theta}_1 - \theta)$ has a non-degenerate limiting distribution) is said to be asymptotically efficient relative to $\hat{\theta}_2$ if

$$\lim_{n \rightarrow \infty} \Pr[-c \leq \delta_n(\hat{\theta}_1 - \theta) \leq c] \geq \lim_{n \rightarrow \infty} \Pr[-c \leq \delta_n(\hat{\theta}_2 - \theta) \leq c]$$

for all $c \geq 0$ and all $\theta \in \Theta$. According to this definition, $\hat{\theta}_1$ is asymptotically efficient relative to $\hat{\theta}_2$ if

$$\sqrt{n}(\hat{\theta}_1 - \theta) \rightarrow_d \mathcal{N}[0, \sigma_1^2(\theta)],$$

and

$$\sqrt{n}(\hat{\theta}_2 - \theta) \rightarrow_d \mathcal{N}[\mu_2(\theta), \sigma_2^2(\theta)]$$

with $\sigma_1^2(\theta) \leq \sigma_2^2(\theta)$ for all $\theta \in \Theta$.

Vector-valued estimators can also be accommodated. Suppose $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimators of a parameter vector $\theta \in \Theta \subseteq \mathbb{R}^k$ such that

$$\sqrt{n}(\hat{\theta}_1 - \theta) \rightarrow_d \mathcal{N}[0, \Sigma_1(\theta)],$$

and

$$\sqrt{n}(\hat{\theta}_2 - \theta) \rightarrow_d \mathcal{N}[0, \Sigma_2(\theta)].$$

In such cases, $\hat{\theta}_1$ is said to be asymptotically efficient relative to $\hat{\theta}_2$ if $\Sigma_2(\theta) - \Sigma_1(\theta)$ is positive semidefinite for all $\theta \in \Theta$. More generally, a δ_n -consistent estimator $\hat{\theta}_1$ (i.e. $\delta_n(\hat{\theta}_1 - \theta)$ has a non-degenerate limiting distribution, where $\{\delta_n\}$ is a sequence of $k \times k$ matrices) is said to be asymptotically efficient relative to $\hat{\theta}_2$ if

$$\lim_{n \rightarrow \infty} \Pr[\delta_n(\hat{\theta}_1 - \theta) \in C] \geq \lim_{n \rightarrow \infty} \Pr[\delta_n(\hat{\theta}_2 - \theta) \in C]$$

for all convex sets $C \subseteq \mathbb{R}^k$, symmetric about the origin (i.e. $x \in C$ implies $-x \in C$) and all $\theta \in \Theta$. In particular, $\hat{\theta}_1$ is asymptotically efficient relative to $\hat{\theta}_2$ if

$$\sqrt{n} \left(\hat{\theta}_1 - \theta \right) \rightarrow_d \mathcal{N} [0, \Sigma_1(\theta)],$$

and

$$\sqrt{n} \left(\hat{\theta}_2 - \theta \right) \rightarrow_d \mathcal{N} [\mu_2(\theta), \Sigma_2(\theta)],$$

where $\Sigma_2(\theta) - \Sigma_1(\theta)$ is positive semidefinite for all $\theta \in \Theta$.

Let \tilde{B} and \hat{B} denote the random variables appearing in the limiting representations of $T(\tilde{\beta} - \beta)$ and $T(\hat{\beta} - \beta)$, respectively:

$$\tilde{B} = \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2'} \right)^{-1} \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_y(r) \right] \sigma_{yy.x}^{1/2} \right).$$

and

$$\begin{aligned} \hat{B} &= \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2'} \right)^{-1} \\ &\cdot \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_y(r) \right] \sigma_{yy.x}^{1/2} + \Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_x(r)' \right] \Sigma_{xx}^{-1/2} \sigma_{xy} + \sigma_{xy} \right) \end{aligned}$$

Conditional on $\{W_x(r) : 0 \leq r \leq 1\}$, \tilde{B} and \hat{B} are both normally distributed with

$$E(\tilde{B}|W_x) = 0,$$

$$E(\hat{B}|W_x) = \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2'} \right)^{-1} \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_x(r)' \right] \Sigma_{xx}^{-1/2} \sigma_{xy} + \sigma_{xy} \right),$$

and

$$V(\tilde{B}|W_x) = V(\hat{B}|W_x) = \sigma_{yy.x} \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2'} \right)^{-1}.$$

As a consequence,

$$\Pr(\tilde{B} \in C|W_x) \geq \Pr(\hat{B} \in C|W_x)$$

for any convex sets $C \subseteq \mathbb{R}^m$, symmetric about the origin. In particular,

$$\begin{aligned}
\lim_{T \rightarrow \infty} \Pr \left[T \left(\tilde{\beta} - \beta \right) \in C \right] &= \Pr \left(\tilde{B} \in C \right) \\
&= E \left[\Pr \left(\tilde{B} \in C | W_x \right) \right] \\
&\geq E \left[\Pr \left(\hat{B} \in C | W_x \right) \right] \\
&= \lim_{T \rightarrow \infty} \Pr \left[T \left(\hat{\beta} - \beta \right) \in C \right]
\end{aligned}$$

for any β and any convex sets C , symmetric about the origin. Therefore, $\tilde{\beta}$ is asymptotically efficient relative to $\hat{\beta}$.

It can be shown that the maximum likelihood estimator $\tilde{\beta}$ asymptotically efficient relative to any “regular” estimator and the limiting representation of $T \left(\tilde{\beta} - \beta \right)$ therefore serves as an efficiency bound in the present estimation problem. A precise statement can be found in Phillips (1991, *Econometrica*, 59, 283-306).

Remark. Any estimator $\tilde{\beta}$ satisfying

$$T \left(\tilde{\beta} - \beta \right) \rightarrow_d \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Sigma_{xx}^{1/2'} \right)^{-1} \left(\Sigma_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_y(r) \right] \sigma_{yy.x}^{1/2} \right)$$

is asymptotically efficient. Since

$$T \left(\tilde{\beta} - \beta \right) = \left(\frac{1}{T^2} \sum_{t=1}^T x_t x_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^T x_t \varepsilon_t^{y.x} \right) + o_p(1),$$

the infeasible estimator

$$\tilde{\beta}^+(\delta) = \left(\frac{1}{T^2} \sum_{t=1}^T x_t x_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^T x_t y_t^+(\delta) \right), \quad y_t^+(\delta) = y_t - \delta' \Delta x_t,$$

is asymptotically efficient. A feasible efficient estimator is $\tilde{\beta}^+(\hat{\delta})$, where $\hat{\delta}$ is any consistent estimator of $\delta = \Sigma_{xx}^{-1} \sigma_{xy}$ such as $\hat{\delta} = \hat{\Sigma}_{xx}^{-1} \hat{\sigma}_{xy}$, where $\hat{\Sigma}_{xx} = T^{-1} \sum_{t=1}^T \Delta x_t \Delta x_t'$ and $\hat{\sigma}_{xy} = T^{-1} \sum_{t=1}^T \Delta x_t \hat{\varepsilon}_t^y$. ■

CASE 3: Serial correlation, endogenous regressors

The model is:

$$\begin{aligned}
y_t &= \beta' x_t + u_t^y \\
\Delta x_t &= u_t^x,
\end{aligned}$$

where y_t is a scalar, x_t is an m -vector ($m = k - 1$) with $x_0 = 0$ and

$$\begin{pmatrix} u_t^y \\ u_t^x \end{pmatrix} = \begin{pmatrix} \Psi_y(L) \\ \Psi_x(L) \end{pmatrix} \varepsilon_t = \Psi(L) \varepsilon_t,$$

where $\Psi(L) = \sum_{i=0}^{\infty} \Psi_i L^i$ with $|\Psi(1)| \neq 0$, $\sum_{i=1}^{\infty} i \|\Psi_i\| < \infty$, and $\varepsilon_t \sim i.i.d. (0, \Sigma)$.

To obtain a well behaved estimator of β , we must eliminate “simultaneous equation bias”. Doing so is more complicated when u_t^y and u_t^x are both mutually and serially correlated. Moreover, we must take serial correlation into account when doing inference.

We can eliminate “simultaneous equation bias” by including “omitted” variables to the right hand side of the regression equation in such a way that the error term becomes asymptotically uncorrelated (in an appropriate sense) with the regressor x_t . It can be shown that there exists a two-sided lag polynomial $\delta(L) = \sum_{i=-\infty}^{\infty} \delta_i L^i$ of (possibly) infinite order such that

$$u_t^{y,x} = u_t^y - \delta(L)' u_t^x = u_t^y - \delta(L)' \Delta x_t,$$

is uncorrelated with $\{u_s^x : s \in \mathbb{Z}\}$. If $\delta(L)' \Delta x_t$ was observed, we could simply regress $y_t - \delta(L)' \Delta x_t$ on x_t using OLS, yielding the estimator

$$\begin{aligned} \ddot{\beta} &= \left(\sum_{t=1}^T x_t x_t' \right)^{-1} \left(\sum_{t=1}^T x_t [y_t - \delta(L)' \Delta x_t] \right) \\ &= \beta + \left(\sum_{t=1}^T x_t x_t' \right)^{-1} \left(\sum_{t=1}^T x_t u_t^{y,x} \right), \end{aligned}$$

whose asymptotic behavior can be characterized as follows:

$$\begin{aligned} T(\ddot{\beta} - \beta) &= \left(\frac{1}{T^2} \sum_{t=1}^T x_t x_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^T x_t u_t^{y,x} \right) \\ &\rightarrow_d \left(\Omega_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Omega_{xx}^{1/2'} \right)^{-1} \left(\Omega_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_y(r) \right] \omega_{yy,x}^{1/2} \right), \end{aligned}$$

where

$$\Omega = \begin{pmatrix} \omega_{yy} & \omega'_{xy} \\ \omega_{xy} & \Omega_{xx} \end{pmatrix} = \Psi(1) \Sigma \Psi(1)'$$

is the long-run covariance matrix of u_t and $\omega_{yy,x} = \omega_{yy} - \omega'_{xy} \Omega_{xx}^{-1} \omega_{xy}$ is the long-run variance of u_t^y conditional on u_t^x . Even when $u_t^{y,x}$ is serially correlated, it turns out that OLS is asymptotically equivalent to GLS. In this sense, the infeasible estimator $\ddot{\beta}$ is asymptotically efficient and the task is to find a feasible estimator which is asymptotically equivalent to $\ddot{\beta}$.

In practice, when $\delta(L)' \Delta x_t$ is unobserved we need to estimate (some function of) $\delta(L)$ in order to get a well behaved, feasible estimator of β . The simplest approach is to approximate $\delta(L)$ by a finite-order two-sided polynomial and estimate the following equation by OLS:

$$y_t = \tilde{\beta}' x_t + \sum_{i=-p}^p \tilde{\delta}'_i \Delta x_{t-i} + \tilde{u}_t^{y,x} \quad (t = p+1, \dots, T-p).$$

The estimator $\tilde{\beta}$ is the DOLS estimator of β . If $\delta(L) = \sum_{i=-p}^p \delta_i L^i$, it is easy to show that

$$\begin{aligned} T(\tilde{\beta} - \beta) &= \left(\frac{1}{T^2} \sum_{t=p+1}^{T-p} x_t x_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=p+1}^{T-p} x_t u_t^{y,x} \right) + o_p(1) \\ &\rightarrow_d \left(\Omega_{xx}^{1/2} \left[\int_0^1 W_x(r) W_x(r)' dr \right] \Omega_{xx}^{1/2} \right)^{-1} \left(\Omega_{xx}^{1/2} \left[\int_0^1 W_x(r) dW_y(r) \right] \omega_{yy,x}^{1/2} \right). \end{aligned}$$

Even if $\delta(L)$ is not of finite order, the DOLS estimator is asymptotically equivalent to $\tilde{\beta}$ provided p satisfies $p \rightarrow \infty$ and $p/\sqrt[3]{T} \rightarrow 0$ as $T \rightarrow \infty$ and certain additional regularity conditions hold (for details, see Saikkonen (1991, *Econometric Theory*, 7, 1-21)).

The hypothesis $H_0 : R\beta = q$ can be tested using

$$G = \left(R\tilde{\beta} - q \right)' \left[\tilde{\omega}_{yy,x} R \left(\sum_{t=1}^T x_t x_t' \right)^{-1} R' \right]^{-1} \left(R\tilde{\beta} - q \right),$$

where $\tilde{\omega}_{yy,x}$ is a consistent estimator of $\omega_{yy,x}$ such as the Newey-West estimator constructed using $\{\tilde{u}_t^{y,x}\}$. Under H_0 , $G \rightarrow_d \chi^2(q)$.

An asymptotically equivalent estimator can be obtained by regressing y_t on x_t and $\Delta x_{t-p}, \dots, \Delta x_{t+p}$ using feasible GLS. This estimator is known as the *DGLS estimator* of β . Two alternative estimators, the *fully modified OLS (FM-OLS) estimator* and the *canonical cointegration regression (CCR) estimator*, are constructed (essentially) by regressing $y_t - \hat{\delta}(1)' \Delta x_t$ on x_t and correcting for bias, where $\hat{\delta}(1)$ is a consistent estimator of $\delta(1) = \Omega_{xx}^{-1} \omega_{xy}$. In addition, β can be estimated efficiently using band spectrum regression (a frequency domain method) or reduced rank regression of a vector autoregressive model.