Ch. 22 Unit Root in Vector Time Series

(June 12, 2017)

1 Multivariate Wiener Processes and Multivariate FCLT

Section 2.1 of Chapter 21 described univariate standard Brownian motion $W(r)$ as a scalar continuous-time process ($W : r \in [0, 1] \rightarrow \mathbb{R}^1$). The variable $W(r)$ has a $N(0, r)$ distribution across realization, and for any given realization, $W(r)$ is continuous function of the date r with independent increments. If a set of k such independent processes, denoted $W_1(r), W_2(r), ..., W_k(r)$, are collected in a ($k \times 1$) vector $\mathbf{w}(r)$, the results is $k-$*dimentional standard Brownian motion*.

Definition.

A k-dimensional standard Brownian motion $\mathbf{w}(\cdot)$ is a continuous-time process associating each date $r \in [0, 1]$ with the $(k \times 1)$ vector $\mathbf{w}(r)$ satisfying the following:

(a). $\mathbf{w}(0) = \mathbf{0}$;

(b). For any dates $0 \leq r_1 < r_2 < ... < r_k \leq 1$, the changes $[\mathbf{w}(r_2) - \mathbf{w}(r_1)]$, $[\mathbf{w}(r_3) - \mathbf{w}(r_2)], ..., [\mathbf{w}(r_k) - \mathbf{w}(r_{k-1})]$ are independent multivariate Gaussian with $[\mathbf{w}(s) - \mathbf{w}(v)] \sim \mathbf{N}(\mathbf{0}, (s - v)\mathbf{I}_k)$;

(c). For any given realization, $\mathbf{w}(r)$ is continuous in r with probability 1. ■

Analogous to the univariate case, we can define a multivariate random walk as follows.

Definition.
Let the $k \times 1$ random vector $\mathbf{y_t}$ follow $\mathbf{y}_t = \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$, $t = 1, 2, ...,$ where $\mathbf{y}_0 = \mathbf{0}$ and $\boldsymbol{\varepsilon}_t$ is a sequence of *i.i.d.* random vector such that $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ and $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \boldsymbol{\Omega}$, a finite positive definite matrix. Then $\mathbf{y_t}$ is a multivariate (k-dimensional) random walk.■

We form the re-scaled partial sums as

$$\mathbf{w}_T(r) \equiv \boldsymbol{\Omega}^{-1/2} T^{-1/2} \sum_{t=1}^{[Tr]^*} \boldsymbol{\varepsilon}_t.$$

The components of $\mathbf{w}_T(r)$ are the individual partial sums

$$W_{Tj}(r) = T^{-1/2} \sum_{t=1}^{[Tr]^*} \tilde{\varepsilon}_{tj}, \quad j = 1, 2, ..., k,$$

where $\tilde{\varepsilon}_{tj}$ is the jth element of $\boldsymbol{\Omega}^{-1/2}\boldsymbol{\varepsilon}_t$.

The Function Central Limit Theorem (FCLT) provides conditions under which $\mathbf{w}_T(r)$ converges to the multivariate standard Wiener process $\mathbf{w}(r)$. The simplest multivariate FCLT is the multivariate Donsker's theorem.

Theorem. (Multivariate Donsker)
Let $\boldsymbol{\varepsilon}_t$ be a sequence of *i.i.d.* random vector such that $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ and $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \boldsymbol{\Omega}$, a finite positive definite matrix. Then $\mathbf{w}_T(\cdot) \Longrightarrow \mathbf{w}(\cdot)$. ■

Quite general multivariate FCLTs are available. For example, we may applied FCLT to serially dependent vector processes using a generalization of (70) and Theorem 12 of Chapter 21.

Theorem. (FCLT when $\mathbf{u}_t$ is a vector $MA(\infty)$ process):
Let

$$\mathbf{u}_t = \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s \boldsymbol{\varepsilon}_{t-s},$$

then

$$\mathbf{w}_T(\cdot) \Longrightarrow \mathbf{w}(\cdot),$$

where $\mathbf{w}_T(r) \equiv \mathbf{\Psi}(1)^{-1}\mathbf{\Omega}^{-1/2}T^{-1/2}\sum_{t=1}^{[Tr]^*}\mathbf{u_t}$, $\boldsymbol{\varepsilon}_t$ is a k dimensional $i.i.d.$ random vector with variance covariance $\mathbf{\Omega}$, and if $\psi_{ij}^{(s)}$ denote the row i, column j element of $\mathbf{\Psi}_s$,

$$\sum_{s=0}^{\infty} s \cdot |\psi_{ij}^{(s)}| < \infty$$

for each $i, j = 1, 2, ..., k$.

Proof.

Using multivariate Beveridge-Nelson decomposition and from that to derive the long run variance matrix of $\mathbf{u}_t$ to be $\frac{1}{T}E[\sum \mathbf{u}_t]^2 = \mathbf{\Psi}^2(1)\mathbf{\Omega}$. ■

2 Vector Autoregression Containing Unit Roots

Let $\mathbf{y}_t$ be an $(k \times 1)$ vector autoregressive process $(VAR(p))$, i.e.

$$[\mathbf{I}_k - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2 - ... - \mathbf{\Phi}_p L^p]\mathbf{y}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t. \tag{22-1}$$

The scalar algebra in (33) of Chapter 21 works perfectly well for matrices, establishing that for any value $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$,..., $\mathbf{\Phi}_p$, the following polynomials are equivalent:

$$\begin{aligned}&[\mathbf{I}_k - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2 - ... - \mathbf{\Phi}_p L^p]\\ &= (\mathbf{I}_k - \boldsymbol{\rho} L) - (\boldsymbol{\xi}_1 L + \boldsymbol{\xi}_2 L^2 + ... + \boldsymbol{\xi}_{p-1} L^{p-1})(1-L),\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{\rho} &\equiv \mathbf{\Phi}_1 + \mathbf{\Phi}_2 + ... + \mathbf{\Phi}_p \\ \boldsymbol{\xi}_s &\equiv -[\mathbf{\Phi}_{s+1} + \mathbf{\Phi}_{s+2} + ... + \mathbf{\Phi}_p] \quad for\ s = 1, 2, ..., p-1.\end{aligned} \tag{22-2}$$

It follows that any $VAR(p)$ process (22-1) can always be written in the form

$$(\mathbf{I}_k - \boldsymbol{\rho} L)\mathbf{y}_t - (\boldsymbol{\xi}_1 L + \boldsymbol{\xi}_2 L^2 + ... + \boldsymbol{\xi}_{p-1} L^{p-1})(1-L)\mathbf{y}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t$$

or

$$\mathbf{y}_t = \boldsymbol{\xi}_1 \triangle \mathbf{y}_{t-1} + \boldsymbol{\xi}_2 \triangle \mathbf{y}_{t-2} + ... + \boldsymbol{\xi}_{p-1}\mathbf{y}_{t-p+1} + \mathbf{c} + \boldsymbol{\rho}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{22-3}$$

There are tow meanings of a VAR process contains unit roots.

(a). First, if the first difference of $\mathbf{y}_t$ follows a $VAR(p-1)$ process:

$$\triangle \mathbf{y}_t = \boldsymbol{\xi}_1 \triangle \mathbf{y}_{t-1} + \boldsymbol{\xi}_2 \triangle \mathbf{y}_{t-2} + ... + \boldsymbol{\xi}_{p-1}\mathbf{y}_{t-p+1} + \mathbf{c} + \boldsymbol{\varepsilon}_t,$$

requiring from (22-3) that

$$\boldsymbol{\rho} = \mathbf{I}_k$$

or from (22-2) that

$$\mathbf{\Phi}_1 + \mathbf{\Phi}_2 + ... + \mathbf{\Phi}_p = \mathbf{I}_k. \tag{22-4}$$

(b). Second, recalling from (8) of Chapter 18 that a $VAR(p)$ such as in (22-1) will be said to contain at least one unit root ($z = 1$) if the following determinant is zero:

$$|\mathbf{I}_k - \mathbf{\Phi}_1 - \mathbf{\Phi}_2 - ... - \mathbf{\Phi}_p| = 0. \tag{22-5}$$

□

Note that (22-4) implies (22-5) but (22-5) does not imply (22-4). Vector autoregression for which (22-5) holds but (22-4) does not will be considered in Chapter 23.

3 Spurious Regression

3.1 Asymptotics for Spurious Regression

Consider a regression of the form

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t, \tag{22-6}$$

for which elements of y_t and $\mathbf{x}_t$ might be nonstationary. If there does not exist some population value for $\boldsymbol{\beta}$ for which the disturbance $u_t = y_t - \mathbf{x}_t'\boldsymbol{\beta}$ is $I(0)$, then OLS is quite likely to produce *spurious results*. In an extreme condition that Y_t and $\mathbf{x}_t$ are independent random walks, as we shall see, the OLS estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_T$ is not consistent for $\boldsymbol{\beta} = \mathbf{0}$ but instead converge to a particular random variable. Because there is truly no relation between Y_t and $\mathbf{x}_t$, and because $\hat{\boldsymbol{\beta}}_T$ is **incapable** of revealing this, we call this a case of "spurious regression".

This spurious-regression phenomenon was first considered by Yule (1926), and the dangers of spurious regression were forcefully brought to the economists by the Monte Carlo studies of Granger and Newbold (1974) and latter explained theoretically by Phillips (1986).

$\boxed{\textbf{Theorem.}}$ (Spurious Regression, two independent random walks):
Let X_t and Y_t be independent random walks, $X_t = X_{t-1} + \eta_t$ and $Y_t = Y_{t-1} + \zeta_t$, and η_t is independent of ζ_t. Consider the regression equation for Y_t in terms of X_t, formally as $Y_t = X_t\beta + u_t$, where $\beta = 0$ and $u_t = Y_t$, reflecting the lack of any relations between Y_t and X_t. Then the OLS estimator of β, $\hat{\beta}_T \xrightarrow{L} (\sigma_2/\sigma_1)\left[\int_0^1 W_1(r)^2 dr\right]^{-1} \int_0^1 W_1(r)W_2(r)dr$, where $\sigma_1^2 = E(\eta_t^2)$ and $\sigma_2^2 = E(\zeta_t^2)$.

Proof.
To proceed, we write

$$W_{1T}(r_{t-1}) = T^{-1/2}\sum_{s=1}^{t-1}\eta_s/\sigma_1 = T^{-1/2}X_{t-1}/\sigma_1,$$

$$W_{2T}(r_{t-1}) = T^{-1/2}\sum_{s=1}^{t-1}\zeta_s/\sigma_2 = T^{-1/2}Y_{t-1}/\sigma_2$$

or

$$T^{-1/2}X_{t-1} = \sigma_1 W_{1T}(r_{t-1}) \tag{22-7}$$

and

$$T^{-1/2}Y_{t-1} = \sigma_2 W_{2T}(r_{t-1}), \tag{22-8}$$

where $\sigma_1^2 \equiv \lim_{T\to\infty} Var(T^{-1/2}\sum_{t=1}^{T}\eta_t)$ and $\sigma_2^2 \equiv \lim_{T\to\infty} Var(T^{-1/2}\sum_{t=1}^{T}\zeta_t)$, and $r_{t-1} = (t-1)/T$ as before.

From Donsker's theorem and the continuous mapping theorem we have that $T^{-2}\sum_{t=1}^{T} X_{t-1}^2 \Rightarrow \sigma_1^2 \int_0^1 W_1(r)dr$ and also $T^{-2}\sum_{t=1}^{T} Y_{t-1}^2 \Rightarrow \sigma_2^2 \int_0^1 W_2(r)dr$. The multivariate version of Donsker's theorem states that[1]

$$\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1/2} T^{-1/2} \sum_{t=1}^{[Tr]^*} \begin{bmatrix} \eta_t \\ \zeta \end{bmatrix} \Rightarrow \begin{bmatrix} W_1(r) \\ W_2(r) \end{bmatrix}$$

or

$$\begin{bmatrix} T^{-1/2}X_T(r) \\ T^{-1/2}Y_T(r) \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma_1 W_1(r) \\ \sigma_2 W_2(r) \end{bmatrix}.$$

From (22-7) and (22-8) we have

$$\begin{aligned} T^{-1}\cdot T^{-1}\sum_{t=1}^{T} X_{t-1}Y_{t-1} &= T^{-1}\sum_{t=1}^{T} \sigma_1 W_{1T}(r_{t-1})\sigma_2 W_{2T}(r_{t-1}) \\ &= \sigma_1\sigma_2 T^{-1}\sum_{t=1}^{T} W_{1T}(r_{t-1})W_{2T}(r_{t-1}) \\ &= \sigma_1\sigma_2 \sum_{t=1}^{T}\int_{(t-1)/T}^{t/T} W_{1T}(r)W_{2T}(r)dr \\ &= \sigma_1\sigma_2 \int_0^1 W_{1T}(r)W_{2T}(r)dr \\ &\Rightarrow \sigma_1\sigma_2 \int_0^1 W_1(r)W_2(r)dr, \end{aligned}$$

where we have use the fact that $W_{1T}(r)$ and $W_{2T}(r)$ is constant for $(t-1)/T \le r < t/T$ and the continuous mapping theorem to the mapping

$$(x, y) \mapsto \int_0^1 x(a)y(a)da.$$

[1]Here, in fact we do not have to require that η_t and ζ_t to be uncorrelated to get the same results.

Hence for convenience treating $\hat{\beta}_{T-1}$ instead of $\hat{\beta}_T$ we have

$$
\begin{aligned}
\hat{\beta}_{T-1} - 0 &= \left(T^{-2}\sum_{t=1}^{T} X_{t-1}^2\right)^{-1}\left(T^{-2}\sum_{t=1}^{T} X_{t-1}Y_{t-1}\right) \\
&= \left(\sigma_1^2\int_0^1 W_1^2(r)dr\right)^{-1}\sigma_1\sigma_2\int_0^1 W_1(r)W_2(r)dr \\
&= (\sigma_2/\sigma_1)\left(\int_0^1 W_1^2(r)dr\right)^{-1}\int_0^1 W_1(r)W_2(r)dr, \qquad (22\text{-}9)
\end{aligned}
$$

which is nondegenerate random variable. $\hat{\beta}_T$ is then not consistent for $\beta = 0$, so the regression is "spurious". ■

The spurious regression problem become clear upon inspection of (22-9). The true value of the derivative of Y_t with respect to X_t is zero because the errors generating X_t and Y_t series in the regression are independent. Yet $\hat{\boldsymbol{\beta}}_T$ fails to converge in probability to zero and instead has a non-degenerate distribution.

Using similar techniques, Phillips (1986) show that $T^{-1/2}t_{\hat{\boldsymbol{\beta}}_T}$ has a non-degenerate distribution, or in other words that the t-statistic for $\hat{\boldsymbol{\beta}}_T$ has a divergent distribution. Hence as $T \to \infty$, the probability of a significant t-value arising in a regression such as (22-8) approach one, leading to spurious inference about the existence of a relationship between X_t and Y_t.

The spurious regression problem not only arise from independent random walks, it even appears among non-cointegrated generally $I(1)$ process.

Theorem. (Spurious Regression, not cointegrated $I(1)$ process, Hamilton's Parametric Method):

Consider an $(k \times 1)$ vector $\mathbf{y}_t$ whose first difference is described by

$$(1-L)\mathbf{y}_t = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t = \sum_{s=0}^{\infty}\boldsymbol{\Psi}_s\boldsymbol{\varepsilon}_{t-s},$$

for $\boldsymbol{\varepsilon}_t$ an $i.i.d.$ vector with mean zero, variance $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \boldsymbol{\Omega} = \mathbf{PP}'$, and finite fourth moment and where $\{s\cdot\boldsymbol{\Psi}_s\}_{s=0}^{\infty}$ is absolutely summable. Let $g = (k-1)$ and $\boldsymbol{\Lambda} = \boldsymbol{\Psi}(1)\mathbf{P}$. Partition $\mathbf{y}_t$ as $\mathbf{y}_t = (Y_{1t}, \mathbf{y}_{2t}')'$, and partition $\boldsymbol{\Lambda\Lambda}'$ as

$$\boldsymbol{\Lambda\Lambda'} = \begin{bmatrix} \Sigma_{11} & \boldsymbol{\Sigma}_{21}' \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

where Σ_{11} is (1×1) and $\boldsymbol{\Sigma}_{22}$ is $(g \times g)$.

Suppose that $\mathbf{\Lambda\Lambda'}$ is nonsingular, and define

$$(\sigma_1^*)^2 \equiv (\Sigma_{11} - \mathbf{\Sigma}_{21}'\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}).$$

Let $\mathbf{L}_{22}$ denote the Cholesky factor of $\mathbf{\Sigma}_{22}^{-1}$ and consider the consequence of an *OLS* regression of the first variable on the others and a constant,

$$Y_{1t} = \hat{\alpha}_T + \mathbf{y}_{2t}'\hat{\boldsymbol{\beta}}_T + \hat{u}_t, \tag{22-10}$$

and ant null hypothesis of the form $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{q}$, where $\mathbf{R}$ is a known $(r \times g)$ matrix representing r separate hypothesis involving $\boldsymbol{\beta}$ and $\mathbf{q}$ is a known $r \times 1$ vector. Then the following hold.

(a). The *OLS* estimate $\hat{\alpha}_T$ and $\hat{\boldsymbol{\beta}}_T$ are characterized by

$$\begin{bmatrix} T^{-1/2}\hat{\alpha}_T \\ \hat{\boldsymbol{\beta}}_T - \mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma_1^* h_1 \\ \sigma_1^*\mathbf{L}_{22}\mathbf{h}_{22} \end{bmatrix},$$

where

$$\begin{bmatrix} h_1 \\ \mathbf{h}_2 \end{bmatrix} \equiv \begin{bmatrix} 1 & \int_0^1 [\mathbf{w}_2(r)]'dr \\ \int_0^1 \mathbf{w}_2(r)dr & \int_0^1 [\mathbf{w}_2(r)][\mathbf{w}_2(r)]'dr \end{bmatrix}^{-1} \times \begin{bmatrix} \int_0^1 W_1(r)dr \\ \int_0^1 \mathbf{w}_2(r)W_1(r)dr \end{bmatrix}$$

and $W_1(r)$ denotes scalar standard Brownian motion, $\mathbf{w}_2(r)$ denotes g-dimensional standard Brownian motion with $\mathbf{w}_2(r)$ independent of $W_1(r)$.

(b). The sum of squared errors *SSE* from the *OLS* estimation of (22-10) satisfies

$$T^{-2} \cdot SSE \xrightarrow{L} (\sigma_1^*)^2 \cdot H,$$

where

$$\begin{aligned} H \;\equiv\; & \int_0^1 [W_1(r)]^2 dr - \left\{ \left[\int_0^1 W_1(r)dr \quad \int_0^1 [W_1(r)][\mathbf{w}_2(r)]'dr \right] \right. \\ & \left. \times \begin{bmatrix} 1 & \int_0^1 [\mathbf{w}_2(r)]'dr \\ \int_0^1 \mathbf{w}_2(r)dr & \int_0^1 [\mathbf{w}_2(r)][\mathbf{w}_2(r)]'dr \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 W_1(r)dr \\ \int_0^1 \mathbf{w}_2(r)W_1(r)dr \end{bmatrix} \right\}. \end{aligned}$$

(c). The *OLS F* test satisfies

$$\begin{aligned} T^{-1}F_T \xrightarrow{L} & (\sigma_1^* \mathbf{R}^* \mathbf{h}_2 - \mathbf{q}^*)' \\ & \times \left\{ \sigma_1^* H [\mathbf{0}\ \mathbf{R}^*] \times \begin{bmatrix} 1 & \int_0^1 [\mathbf{w}_2(r)]' dr \\ \int_0^1 \mathbf{w}_2(r) dr & \int_0^1 [\mathbf{w}_2(r)][\mathbf{w}_2(r)]' dr \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}' \\ \mathbf{R}'^* \end{bmatrix} \right\} \\ & \times (\sigma_1^* \mathbf{R}^* \mathbf{h}_2 - \mathbf{q}^*) \div r, \end{aligned}$$

where

$$\begin{aligned} \mathbf{R}^* &\equiv \mathbf{R}\mathbf{L}_{22}, \\ \mathbf{q}^* &\equiv \mathbf{q} - \mathbf{R}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \end{aligned}$$ ■

Result (a) implies that neither estimator is consistent. The estimator of the constant, $\hat{\alpha}_T$ actually diverge, and must divided by $T^{1/2}$ to obtain a random variable with a well-specified distribution. The estimator $\hat{\alpha}_T$ itself is likely to get farther and farther from the true value of zero as the sample T increase. Thing does not get better when we look at $\hat{\boldsymbol{\beta}}_T$. Different arbitrary large sample will have randomly differing estimators $\hat{\boldsymbol{\beta}}_T$. Those usual happenings that $\hat{\boldsymbol{\beta}}_T \xrightarrow{p} \mathbf{0}$ and must multiplied by some increasing function of T in order to obtain a nondegenerate asymptotic distribution does not occur.

Result (b) implies that the usual *OLS* estimator of the variance of u_t

$$s_T^2 = (T-k)^{-1} SSE_T,$$

again diverge as $T \to \infty$. To obtain an estimator that does not grow with the sample size, the sums of squared errors has to be divided by T^2 rather than T. In this respect, the residual $\hat{u}_t$ from a spurious regression behave like a unit root process; if ξ_t is a scalar $I(1)$ series, then $T^{-1}\sum \xi_t^2$ diverge and $T^{-2}\sum \xi_t^2$ converges.

Result (c) means that any *OLS* t or F test based on the spurious regression also diverge; the *OLS F* statistics must be divided by T to obtains a variable that does not grow with the sample size. Since an F test of a single restriction is the square of the corresponding t test, any t statistics would have to be divided by $T^{1/2}$ to obtain a convergent variable. Thus, as the sample size become large, it becomes increasingly that the absolute value of an *OLS* t test will exceed any arbitrary finite value (such as the usual critical value of $t = 2$). For example, in the regression of (22-10), it appears that Y_{1t} and $\mathbf{y}_{2t}$ are significantly related whereas in reality they are completely independent.

Should we be totally pessimistic on the regression of unit root process from above results ? There is, in fact, one case of major importance where the correlation properties of Y_{1t} and $\mathbf{y}_{2t}$ do interfere with these qualitative results. The conditions in this Theorem require that $\mathbf{\Lambda\Lambda'}$ is nonsingular. From the fact that rank $(\mathbf{\Lambda\Lambda'})$ =rank $(\mathbf{\Lambda})$, $\mathbf{\Lambda} = \mathbf{\Psi}(1)\mathbf{P}$, and $\mathbf{P}$ is nonsingular we require that $\mathbf{\Psi}(1)$ is nonsingular or that the determinant $|\mathbf{\Psi}(1)| \neq 0$. If we allow $\mathbf{\Psi}(1)$ to be singular, then the asymptotic theory of this theorem no longer holds as stated. The condition that $\mathbf{\Psi}(1)$ is singular is a necessary conditions for Y_{1t} and $\mathbf{y}_{2t}$ to be cointegrated in the sense of Engle and Granger (1987). See Chapter 23 for details.

Example.

The following is the code to generate the spurious regression. Let $Y_t = Y_{t-1} + u_t$ and $X_t = X_{t-1} + v_t$, where[2]

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \overset{i.i.d}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}\right).$$

Consider the sequence of an OLS regression of Y_t on X_t and a constant

$$Y_t = \alpha + \beta X_t + \varepsilon_t, \; t = 1, 2, ..., T.$$

It can see that the *OLS* estimates of $\hat{\alpha}$ and $\hat{\beta}$ is inconsistent and the t-ratio to test the null hypothesis that $\alpha = 0$ and $\beta = 0$, $t_{\hat{\alpha}}$ and $t_{\hat{\beta}}$, is increasing with sample. We always incorrectly reject the null hypothesis.

(a). Plot inconsistency of $\hat{\alpha}$ and $\hat{\beta}$.

(b). Plot the t ratio, $t_{\hat{\alpha}}$ and $t_{\hat{\beta}}$.

[2] It is important to note that here as long as this two $I(1)$ are not cointegrated, even u_t and v_t are correlated, the spurious regression phenomenon still exists.

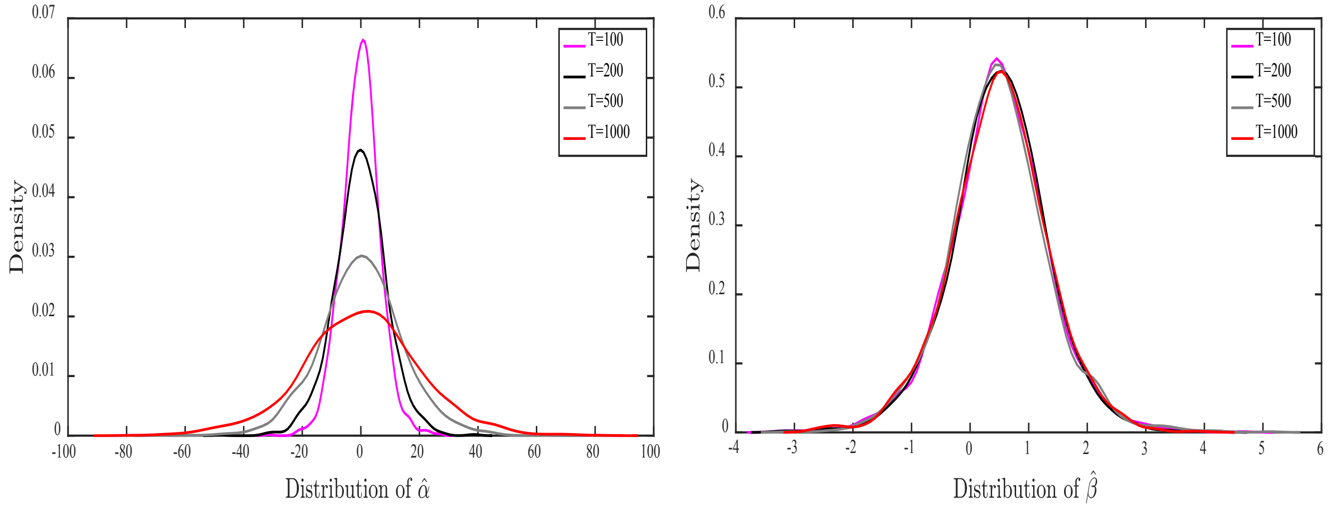

Figure (22-1a). Inconsistency of $\hat{\alpha}$ and $\hat{\beta}$ in a spurious regression.

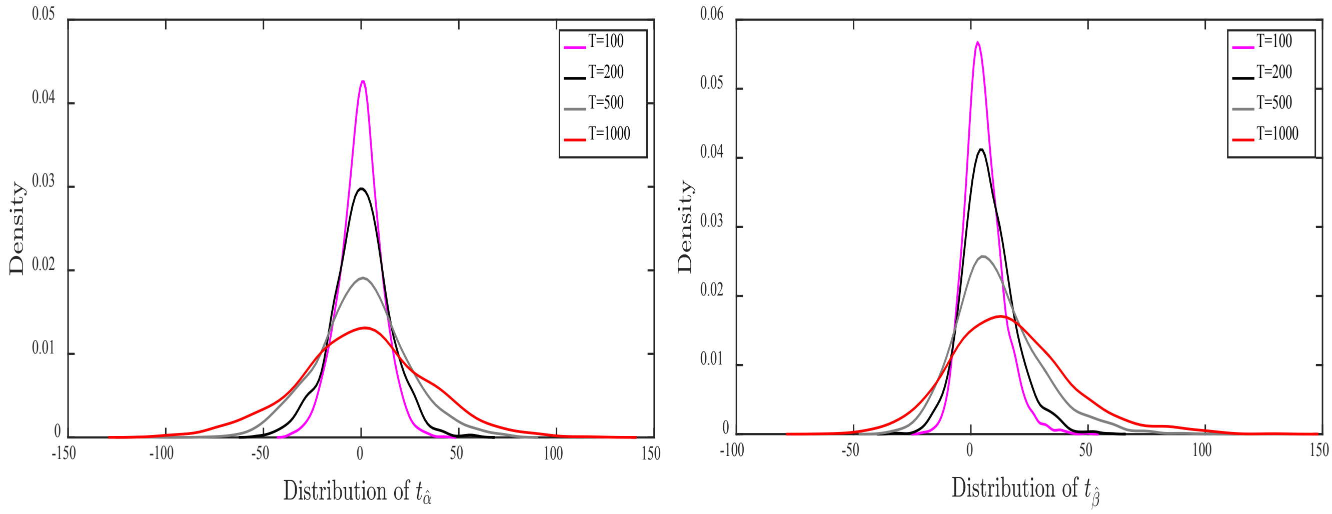

Figure (22-1b). The two t-ratios to test the null hypothesis that $\alpha = 0$ and $\beta = 0$, $t_{\hat{\alpha}}$ and $t_{\hat{\beta}}$, are increasing with sample size.

■

3.2 Cures For Spurious Regression

Many researchers recommend routinely differencing apparently nonstationary variables before estimating regression (for example, Gordon (1984)):

$$\triangle Y_{1t} = a + \triangle \mathbf{y}'_{2t}\mathbf{b} + v_t,$$

which is believe to avoid the spurious regression problem as well as the nonstandard distributions for certain hypotheses associated with the levels regression (22-10). While this is the ideal cure for the problem discussed in this section, there are two different situations in which it might be inappropriate.

(a). First, if a economic theory specify a linear relation between Y_{1t} and $\mathbf{y}_{2t}$ in level as in (22-10), then the parameters has its own economical interpretation, for example, $\partial C_t/\partial Y_t = \beta$ is the marginal propensity to consume which must be positive under normal condition. However, a regression in differenced data, the parameters has different economic implication, e.g. $\partial \triangle C_t/\partial \triangle Y_t = b$, which may be positive or negative even though $\partial C_t/\partial Y_t = \beta$ must be positive. Thus, differenceing the data before regression avoids the econometrics's problem but incurs additionally the economic interpretation problem.

(b). Second, if both Y_{1t} and $\mathbf{y}_{2t}$ are $I(1)$ process, there is an interesting class of models for which the dynamic relation between Y_{1t} and $\mathbf{y}_{2t}$ will be misspecified if the researchers simply differences both Y_{1t} and $\mathbf{y}_{2t}$. The class of models, known as *cointegrated process*, is discussed in the following chapters.

Shei-Mountain (3886m, 'Snow-Mountain').
The second highest peak in Taiwan.

End of this Chapter