Moreover, we consider the following:

$$\frac{1}{\sqrt{T}\sigma} y_{t'} = \frac{1}{\sqrt{T}\sigma}(\epsilon_1 + \epsilon_2 + \cdots + \epsilon_{t'}) \longrightarrow N(0, r') \equiv W(r').$$

For $t' > t$, we have the following:

$$\begin{aligned}\frac{1}{\sqrt{T}\sigma} y_{t'} &= \frac{1}{\sqrt{T}\sigma}(\epsilon_1 + \cdots + \epsilon_t + \epsilon_{t+1} + \cdots \epsilon_{t'}) \longrightarrow N(0, r') \equiv W(r') \\ &= \frac{1}{\sqrt{T}\sigma} y_t + \frac{1}{\sqrt{T}\sigma}(\epsilon_{t+1} + \cdots \epsilon_{t'}).\end{aligned}$$

Therefore, we have

$$\frac{1}{\sqrt{T}\sigma} y_{t'} - \frac{1}{\sqrt{T}\sigma} y_t = \frac{1}{\sqrt{T}\sigma}(\epsilon_{t+1} + \cdots \epsilon_{t'}) \longrightarrow N(0, r' - r) \equiv W(r') - W(r).$$

That is, $W(r)$ is independent of $W(r') - W(r)$ for $r' > r$.

Moreover, note as follows:

$$\frac{1}{T\sqrt{T}\sigma}\sum_{t=1}^{T} y_t = \frac{1}{T}\sum_{t=1}^{T}\Big(\frac{y_t}{\sqrt{T}\sigma}\Big) \longrightarrow \int_0^1 W(r)\mathrm{d}r$$

where $\dfrac{1}{T}$ and $\displaystyle\sum_{t=1}^{T}$ are replaced by dr and $\displaystyle\int_0^1$ as T goes to infinity.

We devide the time interval $(0,\ 1)$ into T time intervals $\Big(\dfrac{t}{T},\ \dfrac{t+1}{T}\Big)$.

That is, time interval $(1,\ T)$ is transformed into (0, 1).

(*) We know that $\dfrac{y_t}{\sqrt{T}\sigma} \longrightarrow W(r)$ as $\dfrac{t}{T} \longrightarrow r$.

Summary: <u>Properties of $W(r)$ for $0 < r < 1$</u>:

1. $W(r) \equiv N(0, r) \quad \Longrightarrow \quad W(r)$ is a random variable.

2. $W(1) \equiv N(0, 1)$

3. $W(1)^2 \equiv \chi^2(1) \quad \Longrightarrow$ Remember that $Z^2 \sim \chi^2(1)$ when $Z \sim N(0, 1)$.

4. $W(r)$ is independent of $W(r') - W(r)$ for $r < r'$.

5. $W(r_4) - W(r_3)$ is independent of $W(r_2) - W(r_1)$ for $0 \leq r_1 < r_2 < r_3 < r_4 \leq 1$.

   $\Longrightarrow$ The interval between r_4 and r_3 is not overlapped with the interval between r_2 and r_1.

- **True Model $y_t = y_{t-1} + \epsilon_t$ vs Estimated Model $y_t = \phi y_{t-1} + \epsilon_t$:** Under $\phi = 1$, we estimate ϕ in the regression model:

$$y_t = \phi y_{t-1} + \epsilon_t$$

OLS of ϕ is:

$$\hat{\phi} = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2} = \phi + \frac{\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

As mentioned aove, the numerator is related to:

$$\frac{1}{\sigma^2 T}\sum_{t=1}^{T} y_{t-1}\epsilon_t = \frac{1}{2}\left(\frac{y_T}{\sigma\sqrt{T}}\right)^2 - \frac{1}{2\sigma^2}\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 \longrightarrow \frac{1}{2}W(1)^2 - \frac{1}{2}$$

which is rewritten by using the Brownian motion $W(1)$.

The denominator is:

$$\frac{1}{\sigma^2 T^2}\sum_{t=1}^{T} y_{t-1}^2 \approx \frac{1}{\sigma^2 T^2}\sum_{t=1}^{T} y_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t}{\sigma\sqrt{T}}\right)^2 \longrightarrow \int_0^1 W(r)^2 \mathrm{d}r$$

where $\dfrac{1}{T} \longrightarrow \mathrm{d}r$ and $\dfrac{y_t}{\sigma\sqrt{T}} \longrightarrow W(r)$ for $\dfrac{t}{T} \longrightarrow r$.

Thus, under $\phi = 1$, $T(\hat{\phi} - \phi)$ is asymptotically distributed as follows:

$$T(\hat{\phi}-\phi) = T(\hat{\phi}-1) = \frac{\dfrac{1}{\sigma^2 T}\displaystyle\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\dfrac{1}{\sigma^2 T^2}\displaystyle\sum_{t=1}^{T} y_{t-1}^2} \longrightarrow \frac{\dfrac{1}{2}\Big(W(1)^2 - 1\Big)}{\displaystyle\int_0^1 W(r)^2 \mathrm{d}r}$$

- t value:

In the regression model: $y_t - y_{t-1} \equiv \Delta y_t = \rho y_{t-1} + \epsilon_t$, OLSE of $\rho = \phi - 1$ is given by $\hat{\rho} = \hat{\phi} - 1$.

t value is $\dfrac{\hat{\rho}}{s_\rho} = \dfrac{\hat{\phi}-1}{s_\phi}$, where s_ρ and s_ϕ denote the standard errors of $\hat{\rho}$ and $\hat{\phi}$.
Note that $s_\rho = s_\phi$ because of $\mathrm{V}(\hat{\rho}) = \mathrm{V}(\hat{\phi}-1) = \mathrm{V}(\hat{\phi})$.

The standard error of $\hat{\phi}$, denoted by s_ϕ, is given by: $s_\phi^2 = \dfrac{s^2}{\sum_{t=1}^{T} y_{t-1}^2}$, where $s^2 = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T}(y_t - \hat{\phi} y_{t-1})^2$, called the standard error of regression.

$$
\begin{aligned}
s^2 &= \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{\phi} y_{t-1})^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}(\epsilon_t - (\hat{\phi}-1)y_{t-1})^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 - 2\frac{1}{T}\frac{1}{T}(\hat{\phi}-1)\sum_{t=1}^{T} y_{t-1}\epsilon_t + \frac{1}{T}(\hat{\phi}-1)^2\sum_{t=1}^{T} y_{t-1}^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 - 2\frac{\sigma^2}{T}\Big[T(\hat{\phi}-1)\Big]\Big[\frac{1}{T\sigma^2}\sum_{t=1}^{T} y_{t-1}\epsilon_t\Big] + \frac{\sigma^2}{T}\Big[T(\hat{\phi}-1)\Big]^2\Big[\frac{1}{T^2\sigma^2}\sum_{t=1}^{T} y_{t-1}^2\Big] \\
&\longrightarrow \sigma^2.
\end{aligned}
$$

The random variables in $\Big[\cdot\Big]$ converge in distribution..

Note that in the right hand side of the fourth line the second and third terms go to zero, because we know the followings:

$$\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 \longrightarrow \sigma^2$$

$$T(\hat{\phi}-1) \longrightarrow \frac{\frac{1}{2}\Big(W(1)^2-1\Big)}{\int_0^1 W(r)^2 \mathrm{d}r}$$

$$\frac{1}{T\sigma^2}\sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow \frac{1}{2}W(1)^2-\frac{1}{2}$$

$$\frac{1}{T^2\sigma^2}\sum_{t=1}^{T} y_t^2 \longrightarrow \int_0^1 W(r)^2 \mathrm{d}r$$

Therefore, from $s_\phi^2 = \dfrac{1}{T^2\sigma^2}\dfrac{s^2}{\frac{1}{T^2\sigma^2}\sum_{t=1}^{T} y_{t-1}^2}$, we obtain $T^2 s_\phi^2 \longrightarrow \Big(\int_0^1 W(r)^2 \mathrm{d}r\Big)^{-1}$.

t value is given by:

$$\frac{\hat{\phi}-1}{s_\phi} = \frac{T(\hat{\phi}-1)}{T s_\phi} \longrightarrow \frac{\frac{1}{2}\Big(W(1)^2-1\Big)/\int_0^1 W(r)^2 \mathrm{d}r}{\Big(\int_0^1 W(r)^2 \mathrm{d}r\Big)^{-1/2}} = \frac{\frac{1}{2}\Big(W(1)^2-1\Big)}{\Big(\int_0^1 W(r)^2 \mathrm{d}r\Big)^{1/2}},$$

which is not a normal distribution.

- **True Model $\boldsymbol{y_t = y_{t-1} + \epsilon_t}$ vs Estimated Model $\boldsymbol{y_t = \alpha + \phi y_{t-1} + \epsilon_t}$:** Under $\alpha = 0$ and $\phi = 1$, we estimate α and ϕ in the regression model:

$$y_t = \alpha + \phi y_{t-1} + \epsilon_t$$

OLSEs of α and ϕ are:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_t \\ \sum y_{t-1} y_t \end{pmatrix} = \begin{pmatrix} \alpha \\ \phi \end{pmatrix} + \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum \epsilon_t \\ \sum y_{t-1} \epsilon_t \end{pmatrix}$$

$$= \begin{pmatrix} \alpha \\ \phi \end{pmatrix} + \frac{1}{T \sum y_{t-1}^2 - (\sum y_{t-1})^2} \begin{pmatrix} \sum y_{t-1}^2 & -\sum y_{t-1} \\ -\sum y_{t-1} & T \end{pmatrix} \begin{pmatrix} \sum \epsilon_t \\ \sum y_{t-1} \epsilon_t \end{pmatrix}$$

$$= \begin{pmatrix} \alpha \\ \phi \end{pmatrix} + \frac{1}{T \sum y_{t-1}^2 - (\sum y_{t-1})^2} \begin{pmatrix} (\sum y_{t-1}^2)(\sum \epsilon_t) - (\sum y_{t-1})(\sum y_{t-1} \epsilon_t) \\ -(\sum y_{t-1})(\sum \epsilon_t) + T(\sum y_{t-1} \epsilon_t) \end{pmatrix}$$

$$
= \begin{pmatrix} \alpha \\ \phi \end{pmatrix} + \begin{pmatrix} \dfrac{(\sum y_t^2)(\frac{1}{T}\sum \epsilon_t) - \overline{y}\sum y_{t-1}\epsilon_t}{\sum y_t^2 - T\overline{y}^2} \\ \dfrac{-\overline{y}\sum \epsilon_t + \sum y_{t-1}\epsilon_t}{\sum y_t^2 - T\overline{y}^2} \end{pmatrix}
$$

Note that $\frac{1}{T}\sum y_{t-1} \approx \frac{1}{T}\sum y_t = \overline{y}$ and $\sum y_{t-1}^2 \approx \sum y_t^2$ for large T.

In the true model, $\alpha = 0$ and $\phi = 1$.

$$
\begin{pmatrix} \hat{\alpha} \\ \hat{\phi} - 1 \end{pmatrix} = \begin{pmatrix} \dfrac{(\sum y_t^2)(\frac{1}{T}\sum \epsilon_t) - \overline{y}\sum y_{t-1}\epsilon_t}{\sum y_t^2 - T\overline{y}^2} \\ \dfrac{-\overline{y}\sum \epsilon_t + \sum y_{t-1}\epsilon_t}{\sum y_t^2 - T\overline{y}^2} \end{pmatrix}
$$

For each element of the vector, we consider each term in the numerator and denominator.

- $\sum_{t=1}^{T} y_{t-1}\epsilon_t$:

Taking the square of $y_t = y_{t-1} + \epsilon_t$ on both sides, we obtain: $y_t^2 = y_{t-1}^2 + y_{t-1}\epsilon_t + \epsilon_t^2$.

Then, we can rewrite as: $y_{t-1}\epsilon_t = \frac{1}{2}(y_t^2 - y_{t-1}^2 - \epsilon_t^2)$ for $y_0 = 0$.

Taking a sum from $t = 1$ to T, we have:

$$\sum_{t=1}^{T} y_{t-1}\epsilon_t = \frac{1}{2}\sum_{t=1}^{T}(y_t^2 - y_{t-1}^2 - \epsilon_t^2) = \frac{1}{2}y_T^2 - \frac{1}{2}\sum_{t=1}^{T}\epsilon_t^2,$$

which is divided by $T\sigma^2$ on both sides, then we obtain:

$$\frac{1}{T\sigma^2}\sum_{t=1}^{T} y_{t-1}\epsilon_t = \frac{1}{2}\Big(\frac{y_T}{\sqrt{T}\sigma}\Big)^2 - \frac{1}{2\sigma^2}\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 \longrightarrow \frac{1}{2}W(1)^2 - \frac{1}{2}.$$

Note that $\frac{y_T}{\sqrt{T}\sigma} = W(1)$ and $\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 \longrightarrow \mathrm{E}(\epsilon_t^2) = \sigma^2$.

- $\overline{y}$:

Note that $\dfrac{1}{T}\sum_{t=1}^{T} y_t = \overline{y}$ and $\dfrac{1}{T}\sum_{t=1}^{T} y_{t-1} = \overline{y}$. We can rewrite $\overline{y}$ as follows:

$$\overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t = \sqrt{T}\sigma\frac{1}{T}\sum_{t=1}^{T}\frac{y_t}{\sqrt{T}\sigma}$$

which is rewritten as:

$$\frac{\overline{y}}{\sqrt{T}\sigma} = \frac{1}{T}\sum_{t=1}^{T}\frac{y_t}{\sqrt{T}\sigma} \longrightarrow \int_0^1 W(r)\mathrm{d}r$$

Note that $\dfrac{y_t}{\sqrt{T}\sigma} \longrightarrow W(r)$ as $\dfrac{t}{T} \longrightarrow r$.

- $\sum_{t=1}^{T} \epsilon_t$:

From $y_T = \sum_{t=1}^{T} \epsilon_t$, we have: $\dfrac{1}{\sqrt{T}\sigma} \sum_{t=1}^{T} \epsilon_t = \dfrac{y_T}{\sqrt{T}\sigma} = W(1)$.

- $\sum_{t=1}^{T} y_{t-1}^2$:

From $\sum_{t=1}^{T} y_{t-1}^2 \approx \sum_{t=1}^{T} y_t^2$, we obtain: $\dfrac{1}{T} \sum_{t=1}^{T} \Big(\dfrac{y_t}{\sqrt{T}\sigma}\Big)^2 \longrightarrow \int_0^1 W(r)^2 \mathrm{d}r$.

Note that $\dfrac{y_t}{\sqrt{T}\sigma} \longrightarrow W(r)$ as $\dfrac{t}{T} \longrightarrow r$.

Thus, $\hat{\phi} - 1 = \dfrac{\sum y_{t-1}\epsilon_t - \bar{y}\sum \epsilon_t}{\sum y_t^2 - T\bar{y}^2}$ is rerwitten as:

$$T(\hat{\phi} - 1) = \frac{\frac{1}{T\sigma^2}(\sum y_{t-1}\epsilon_t - \bar{y}\sum \epsilon_t)}{\frac{1}{T^2\sigma^2}(\sum y_t^2 - T\bar{y}^2)} = \frac{\frac{1}{T\sigma^2}\sum y_{t-1}\epsilon_t - \big(\frac{\bar{y}}{\sqrt{T}\sigma}\big)\big(\frac{1}{\sqrt{T}\sigma}\sum \epsilon_t\big)}{\frac{1}{T}\sum\big(\frac{y_t}{\sqrt{T}\sigma}\big)^2 - \big(\frac{\bar{y}}{\sqrt{T}\sigma}\big)^2}$$

$$\longrightarrow \frac{\frac{1}{2}\Big(W(1)^2 - 1\Big) - W(1)\int_0^1 W(r)\mathrm{d}r}{\int_0^1 W(r)^2\mathrm{d}r - \Big(\int_0^1 W(r)\mathrm{d}r\Big)^2}$$

Remember that OLSE of α is given by:

$$\hat{\alpha} = \alpha + \frac{(\sum y_t^2)(\frac{1}{T}\sum \epsilon_t) - \bar{y}\sum y_{t-1}\epsilon_t}{\sum y_t^2 - T\bar{y}^2}$$

Under $\alpha = 0$, $\hat{\alpha}$ is rewritten as follows:

$$\sqrt{T}\hat{\alpha} = \frac{\sigma\Big(\frac{1}{T}\sum\big(\frac{y_t}{\sqrt{T}\sigma}\big)^2\Big)\Big(\frac{1}{\sqrt{T}\sigma}\sum\epsilon_t\Big) - \sigma\Big(\frac{\bar{y}}{\sqrt{T}\sigma}\Big)\Big(\frac{1}{T\sigma^2}\sum y_{t-1}\epsilon_t\Big)}{\frac{1}{T}\sum\big(\frac{y_t}{\sqrt{T}\sigma}\big)^2 - \big(\frac{\bar{y}}{\sqrt{T}\sigma}\big)^2}$$

$$\longrightarrow \frac{\sigma W(1)\int_0^1 W(r)^2 \mathrm{d}r - \sigma\frac{1}{2}(W(1)^2 - 1)\int_0^1 W(r)\mathrm{d}r}{\int_0^1 W(r)^2 \mathrm{d}r - \Big(\int_0^1 W(r)\mathrm{d}r\Big)^2}$$

Thus, convergence speed of $\hat{\phi}$ is different from that of $\hat{\alpha}$.

Neither $\sqrt{T}\hat{\alpha}$ nor $T(\hat{\phi} - 1)$ are normal.

8.3 Serially Correlated Errors

Consider the case where the error term is serially correlated.

8.3.1 Augmented Dickey-Fuller (ADF) Test

Consider the following AR(p) model:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t, \qquad \epsilon_t \sim \text{iid}(0, \sigma^2),$$

which is rewritten as: $\qquad \phi(L)y_t = \epsilon_t$.

When the above model has a unit root, we have $\phi(1) = 0$, i.e., $\phi_1 + \phi_2 + \cdots + \phi_p = 1$.

The above AR(p) model is written as:

$$y_t = \rho y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots + +\delta_{p-1} \Delta y_{t-p+1} + \epsilon_t,$$

where $\rho = \phi_1 + \phi_2 + \cdots + \phi_p$ and $\delta_j = -(\phi_{j+1} + \phi_{j+2} + \cdots + \phi_p)$.

The null and alternative hypotheses are:

$$H_0: \ \rho = 1 \text{ (Unit root)},$$

$$H_1: \ \rho < 1 \text{ (Stationary)}.$$

Use the t test, where we have the same asymptotic distributions.

We can utilize the same tables as before.

Choose p by AIC or SBIC.

Use $N(0,1)$ to test $H_0: \ \delta_j = 0$ against $H_1: \ \delta_j \neq 0$ for $j = 1, 2, \cdots, p-1$.

Reference

Kurozumi (2008) “Economic Time Series Analysis and Unit Root Tests: Development and Perspective,” *Japan Statistical Society*, Vol.38, Series J, No.1, pp.39 – 57.

Download the above paper from:

http://ci.nii.ac.jp/vol_issue/nels/AA11989749/ISS0000426576_ja.html

Example of ADF Test

```
. gen time=_n

. tsset time
        time variable:  time, 1 to 516
                delta:  1 unit

. gen sexpend=expend-l12.expend
(12 missing values generated)

. corrgram sexpend

                                                  -1       0       1 -1       0       1
 LAG       AC       PAC      Q      Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.7177   0.7184   261.14  0.0000          |-----             |-----
2        0.7036   0.3895    512.6  0.0000          |-----             |---
3        0.7031   0.2817   764.23  0.0000          |-----             |--
4        0.6366   0.0456   970.94  0.0000          |-----             |
5        0.6413   0.1116   1181.1  0.0000          |-----             |
6        0.6267   0.0815   1382.2  0.0000          |-----             |
7        0.6208   0.0972     1580  0.0000          |----              |
8        0.6384   0.1286   1789.5  0.0000          |-----             |-
9        0.5926  -0.0205   1970.5  0.0000          |----              |
10       0.5847  -0.0014   2146.9  0.0000          |----              |
11       0.5658  -0.0185   2312.6  0.0000          |----              |
12       0.4529  -0.2570   2418.9  0.0000          |---             --|
13       0.5601   0.2318   2581.8  0.0000          |----              |-
14       0.5393   0.1095   2733.2  0.0000          |----              |
15       0.5277   0.0850   2878.4  0.0000          |----              |
```

```
. varsoc d.sexpend, exo(l.sexpend) maxlag(25)

   Selection-order criteria
   Sample:  39 - 516                                    Number of obs      =       478
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-4917.7				5.1e+07	20.5845	20.5914	20.6019
1	-4878.69	78.013	1	0.000	4.3e+07	20.4255	20.4358	20.4516
2	-4858.95	39.481	1	0.000	4.0e+07	20.3471	20.3608	20.382
3	-4858.46	.97673	1	0.323	4.0e+07	20.3492	20.3664	20.3928
4	-4855.44	6.0461	1	0.014	4.0e+07	20.3407	20.3613	20.3931
5	-4853.84	3.1904	1	0.074	4.0e+07	20.3383	20.3623	20.3993
6	-4851.58	4.5304	1	0.033	4.0e+07	20.333	20.3604	20.4027
7	-4847.61	7.942	1	0.005	3.9e+07	20.3205	20.3514	20.399
8	-4847.51	.20154	1	0.653	3.9e+07	20.3243	20.3586	20.4115
9	-4847.51	.00096	1	0.975	3.9e+07	20.3285	20.3662	20.4244
10	-4847.43	.16024	1	0.689	4.0e+07	20.3323	20.3735	20.437
11	-4831.38	32.094	1	0.000	3.7e+07	20.2694	20.3139	20.3828
12	-4818.46	25.834	1	0.000	3.5e+07	20.2195	20.2675	20.3416*
13	-4815.64	5.6341	1	0.018	3.5e+07	20.2119	20.2633	20.3427
14	-4813.98	3.321	1	0.068	3.5e+07	20.2091	20.264	20.3487
15	-4813.38	1.2007	1	0.273	3.5e+07	20.2108	20.2691	20.3591
16	-4810.57	5.6184	1	0.018	3.5e+07	20.2032	20.265	20.3603
17	-4808.7	3.7539	1	0.053	3.5e+07	20.1996	20.2647	20.3653
18	-4806.12	5.1557	1	0.023	3.4e+07	20.193	20.2616	20.3674
19	-4804.6	3.0319	1	0.082	3.4e+07	20.1908	20.2628	20.374
20	-4804.6	2.7e-05	1	0.996	3.5e+07	20.195	20.2704	20.3869
21	-4797.33	14.542	1	0.000	3.4e+07	20.1688	20.2476	20.3694
22	-4794.2	6.2571*	1	0.012	3.3e+07*	20.1598*	20.2422*	20.3692
23	-4793.42	1.5626	1	0.211	3.3e+07	20.1608	20.2465	20.3788
24	-4792.85	1.1533	1	0.283	3.3e+07	20.1625	20.2517	20.3893

```
  | 25 | -4792.78  .13518    1  0.713  3.4e+07   20.1664    20.259    20.402  |
  +-----------------------------------------------------------------------------+
   Endogenous:  D.sexpend
    Exogenous:  L.sexpend  _cons

. dfuller sexpend, lags(22)

Augmented Dickey-Fuller test for unit root         Number of obs   =       481

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.627            -3.442            -2.871            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.4689

. dfuller sexpend, lags(12)

Augmented Dickey-Fuller test for unit root         Number of obs   =       491

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.399            -3.441            -2.870            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.1420
```

$\Longrightarrow$ Unit root is detected.

8.4 Cointegration (共和分)

1. For a scalar y_t, when $\Delta y_t = y_t - y_{t-1}$ is a white noise (i.e., iid), we write $\Delta y_t \sim I(1)$.

2. **Definition of Cointegration:**

   Suppose that each series in a $g \times 1$ vector y_t is $I(1)$, i.e., each series has unit root, and that a linear combination of each series (i.e, $a'y_t$ for a nonzero vector a) is $I(0)$, i.e., stationary.

   Then, we say that y_t has a cointegration.

   a is called the cointegrating vector.

3. **Example:**

   Suppose that $y_t = (y_{1,t},\ y_{2,t})'$ is the following vector autoregressive process:

$$y_{1,t} = \phi_1 y_{2,t} + \epsilon_{1,t},$$

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

   Then,

$$\Delta y_{1,t} = \phi_1 \epsilon_{2,t} + \epsilon_{1,t} - \epsilon_{1,t-1}, \quad \text{(MA(1) process)},$$

$$\Delta y_{2,t} = \epsilon_{2,t},$$

where both $y_{1,t}$ and $y_{2,t}$ are $I(1)$ processes.

The linear combination $y_{1,t} - \phi_1 y_{2,t}$ is $I(0)$.

In this case, we say that $y_t = (y_{1,t},\ y_{2,t})'$ is cointegrated with $a = (1,\ -\phi_1)$.

$a = (1,\ -\phi_1)$ is called the cointegrating vector, which is not unique.

Therefore, the first element of a is set to be one.

8.5 Spurious Regression (見せかけ回帰)

1. Suppose that $y_t \sim I(1)$ and $x_t \sim I(1)$.

   For the regression model $y_t = x_t\beta + u_t$, OLS does not work well if we do not have the β which satisfies $u_t \sim I(0)$.

   $\Longrightarrow$ **Spurious regression (見せかけ回帰)**

2. Suppose that $y_t \sim I(1)$, y_t is a $g \times 1$ vector and $y_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix}$.

   $y_{2,t}$ is a $k \times 1$ vector, where $k = g - 1$.

Consider the following regression model:

$$y_{1,t} = \alpha + \gamma' y_{2,t} + u_t, \qquad t = 1, 2, \cdots, T.$$

OLSE is given by:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} T & \sum y'_{2,t} \\ \sum y_{2,t} & \sum y_{2,t} y'_{2,t} \end{pmatrix}^{-1} \begin{pmatrix} \sum y_{1,t} \\ \sum y_{1,t} y_{2,t} \end{pmatrix}.$$

Next, consider testing the null hypothesis $H_0: \ R\gamma = r$, where R is a $m \times k$ matrix ($m \le k$) and r is a $m \times 1$ vector.

The F statistic, denoted by F_T, is given by:

$$F_T = \frac{1}{m}(R\hat{\gamma} - r)' \left(s_T^2 \begin{pmatrix} 0 & R \end{pmatrix} \begin{pmatrix} T & \sum y'_{2,t} \\ \sum y_{2,t} & \sum y_{2,t} y'_{2,t} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ R' \end{pmatrix} \right)^{-1} (R\hat{\gamma} - r),$$

where

$$s_T^2 = \frac{1}{T-g} \sum_{t=1}^{T} (y_{1,t} - \hat{\alpha} - \hat{\gamma}' y_{2,t})^2.$$

When we have the γ such that $y_{1,t} - \gamma y_{2,t}$ is stationary, OLSE of γ, i.e., $\hat{\gamma}$, is not statistically equal to zero.

When the sample size T is large enough, H_0 is rejected by the F test.

3. Phillips, P.C.B. (1986) "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, Vol.33, pp.95 – 131.

Consider a $g \times 1$ vector y_t whose first difference is described by:

$$\Delta y_t = \Psi(L)\epsilon_t = \sum_{s=0}^{\infty} \Psi_s \epsilon_{t-s},$$

for ϵ_t an i.i.d. $g \times 1$ vector with mean zero , variance $\mathrm{E}(\epsilon_t \epsilon_t') = PP'$, and finite fourth moments and where $\{s\Psi_s\}_{s=0}^{\infty}$ is absolutely summable.

Let $k = g - 1$ and $\Lambda = \Psi(1)P$.

Partition y_t as $y_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix}$ and $\Lambda\Lambda'$ as $\Lambda\Lambda' = \begin{pmatrix} \Sigma_{11} & \Sigma_{21}' \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, where $y_{1,t}$ and Σ_{11} are scalars, $y_{2,t}$ and Σ_{21} are $k \times 1$ vectors, and Σ_{22} is a $k \times k$ matrix.

Suppose that $\Lambda\Lambda'$ is nonsingular,and define $\sigma_1^{*2} = \Sigma_{11} - \Sigma_{21}'\Sigma_{22}^{-1}\Sigma_{21}$.

Let L_{22} denote the Cholesky factor of Σ_{22}^{-1}, i.e., L_{22} is the lower triangular matrix satisfying $\Sigma_{22}^{-1} = L_{22}L_{22}'$.

Then, (a) – (c) hold.

(a) OLSEs of α and γ in the regression model $y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$, denoted by $\hat{\alpha}_T$ and $\hat{\gamma}_T$, are characterized by:

$$\begin{pmatrix} T^{-1/2}\hat{\alpha}_T \\ \hat{\gamma}_T - \Sigma_{22}^{-1}\Sigma_{21} \end{pmatrix} \longrightarrow \begin{pmatrix} \sigma_1^* h_1 \\ \sigma_1^* L_{22} h_2 \end{pmatrix},$$

where $\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} 1 & \int_0^1 W_2^*(r)'\mathrm{d}r \\ \int_0^1 W_2^*(r)\mathrm{d}r & \int_0^1 W_2^*(r)W_2^*(r)'\mathrm{d}r \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 W_1^*(r)\mathrm{d}r \\ \int_0^1 W_2^*(r)W_1^*(r)\mathrm{d}r \end{pmatrix}.$

$W_1^*(r)$ and $W_2^*(r)$ denote scalar and g-dimensional standard Brownian motions, and $W_1^*(r)$ is independent of $W_2^*(r)$.

(b) The sum of squared residuals, denoted by $\mathrm{RSS}_T = \sum_{t=1}^T \hat{u}_t^2$, satisfies

$$T^{-2}\mathrm{RSS}_T \longrightarrow \sigma_1^{*2} H,$$

where $H = \int_0^1 (W_1^*(r))^2 \mathrm{d}r - \left(\begin{pmatrix} \int_0^1 W_1^*(r)\mathrm{d}r \\ \int_0^1 W_2^*(r)W_1^*(r)\mathrm{d}r \end{pmatrix}' \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}\right)^{-1}.$

(c) The F_T test satisfies:

$$T^{-1}F_T \longrightarrow \frac{1}{m}(\sigma_1^* R^* h_2 - r^*)' \times \left(\sigma_1^{*2} H \begin{pmatrix} 0 & R^* \end{pmatrix} \begin{pmatrix} 1 & \int_0^1 W_2^*(r)'\mathrm{d}r \\ \int_0^1 W_2^*(r)\mathrm{d}r & \int_0^1 W_2^*(r)W_2^*(r)'\mathrm{d}r \end{pmatrix}^{-1} \begin{pmatrix} 0 & R^* \end{pmatrix}'\right)^{-1}$$

$$\times(\sigma_1^* R^* h_2 - r^*),$$

where $R^* = RL_{22}$ and $r^* = r - R\Sigma_{22}^{-1}\Sigma_{21}$.

Summary: <u>Spurious regression (見せかけの回帰)</u>

Consider the regression model: $y_{1.t} = \alpha + y_{2,t}\gamma + u_t$ for $t = 1, 2, \cdots, T$

and $y_t \sim I(1)$ for $y_t = (y_{1,t}, y_{2,t})'$.

(a) indicates that OLSE $\hat{\gamma}_T$ is not consistent.

(b) indicates that $s_T^2 = \dfrac{1}{T-g}\displaystyle\sum_{t=1}^{T}\hat{u}_t^2$ diverges.

(c) indicates that F_T diverges.

$\Longrightarrow$ It seems that the coefficients are statistically significant, based on the conventional t statistics.

4. **Resolution for Spurious Regression:**

Suppose that $y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$ is a spurious regression.

(1) Estimate $y_{1,t} = \alpha + \gamma' y_{2,t} + \phi y_{1,t-1} + \delta y_{2,t-1} + u_t$.

Then, $\hat{\gamma}_T$ is $\sqrt{T}$-consistent, and the t test statistic goes to the standard normal distribution under $H_0: \ \gamma = 0$.

(2) Estimate $\Delta y_{1,t} = \alpha + \gamma' \Delta y_{2,t} + u_t$. Then, $\hat{\alpha}_T$ and $\hat{\beta}_T$ are $\sqrt{T}$-consistent, and the t test and F test make sense.

(3) Estimate $y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$ by the Cochrane-Orcutt method, assuming that u_t is the first-order serially correlated error.

Usually, choose (2).

However, there are two exceptions.

(i) The true value of ϕ is not one, i.e., less than one.

(ii) $y_{1,t}$ and $y_{2,t}$ are the cointegrated processes.

In these two cases, taking the first difference leads to the misspecified regression.

5. **Cointegrating Vector:**

Suppose that each element of y_t is $I(1)$ and that $a'y_t$ is $I(0)$.

a is called a **cointegrating vector (共和分ベクトル)**, which is not unique.

Set $z_t = a'y_t$, where z_t is scalar, and a and y_t are $g \times 1$ vectors.

For $z_t \sim I(0)$ (i.e., stationary),

$$T^{-1}\sum_{t=1}^{T} z_t^2 = T^{-1}\sum_{t=1}^{T}(a'y_t)^2 \longrightarrow \mathrm{E}(z_t^2).$$

For $z_t \sim I(1)$ (i.e., nonstationary, i.e., a is not a cointegrating vector),

$$T^{-2}\sum_{t=1}^{T}(a'y_t)^2 \longrightarrow \lambda^2 \int_0^1 (W(r))^2\,\mathrm{d}r,$$

where $W(r)$ denotes a standard Brownian motion and λ^2 indicates variance of $(1-L)z_t$.

If a is not a cointegrating vector, $T^{-1}\sum_{t=1}^{T} z_t^2$ diverges.

$\Longrightarrow$ We can obtain a consistent estimate of a cointegrating vector by minimizing $\sum_{t=1}^{T} z_t^2$ with respect to a, where a normalization condition on a has to be imposed.

The estimator of the a including the normalization condition is super-consistent (T-consistent).

● Stock, J.H. (1987) "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica*, Vol.55, pp.1035 – 1056.

Proposition:

Let $y_{1,t}$ be a scalar, $y_{2,t}$ be a $k \times 1$ vector, and $(y_{1,t}, y'_{2,t})'$ be a $g \times 1$ vector, where $g = k + 1$.

Consider the following model:

$$y_{1,t} = \alpha + \gamma' y_{2,t} + z_t^*, \qquad \begin{pmatrix} z_t^* \\ u_{2,t} \end{pmatrix} = \Psi^*(L)\epsilon_t,$$
$$\Delta y_{2,t} = u_{2,t},$$

ϵ_t is a $g \times 1$ i.i.d. vector with $\mathrm{E}(\epsilon_t) = 0$ and $\mathrm{E}(\epsilon_t \epsilon_t') = PP'$.

OLSE is given by: $\begin{pmatrix} \hat{\alpha} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} T & \sum y_{2,t}' \\ \sum y_{2,t} & \sum y_{2,t} y_{2,t}' \end{pmatrix}^{-1} \begin{pmatrix} \sum y_{1,t} \\ \sum y_{1,t} y_{2,t} \end{pmatrix}$.

Define λ_1^*, which is a $g \times 1$ vector, and Λ_2^*, which is a $k \times g$ matrix, as follows:

$$\Psi^*(1)\, P = \begin{pmatrix} \lambda_1^{*\prime} \\ \Lambda_2^* \end{pmatrix}.$$

Then, we have the following results:

$$\begin{pmatrix} T^{1/2}(\hat{\alpha} - \alpha) \\ T(\hat{\gamma} - \gamma) \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & \left(\Lambda_2^* \int W(r) \mathrm{d}r\right)' \\ \Lambda_2^* \int W(r) \mathrm{d}r & \Lambda_2^* \left(\int (W(r))\,(W(r))'\, \mathrm{d}r \right) \Lambda_2^{*\prime} \end{pmatrix}^{-1} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix},$$

where $\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} \lambda_1^{*\prime} W(1) \\ \Lambda_2^* \left(\int W(r)\,(\mathrm{d}W(r))' \right) \lambda_1^* + \sum_{\tau=0}^{\infty} \mathrm{E}(u_{2,t} z_{t+\tau}^*) \end{pmatrix}$.

$W(r)$ denotes a g-dimensional standard Brownian motion.

1) OLSE of the cointegrating vector is consistent even though u_t is serially correlated.

2) The consistency of OLSE implies that $T^{-1} \sum \hat{u}_t^2 \longrightarrow \sigma^2$.

3) Because $T^{-1} \sum (y_{1,t} - \overline{y}_1)^2$ goes to infinity, a coefficient of determination, R^2, goes to one.

8.6 Testing Cointegration

8.6.1 Engle-Granger Test

$y_t \sim I(1)$

$y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$

- $u_t \sim I(0) \implies$ Cointegration
- $u_t \sim I(1) \implies$ Spurious Regression

Estimate $y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$ by OLS, and obtain $\hat{u}_t$.

Estimate $\hat{u}_t = \rho \hat{u}_{t-1} + \delta_1 \Delta \hat{u}_{t-1} + \delta_2 \Delta \hat{u}_{t-2} + \cdots + \delta_{p-1} \Delta \hat{u}_{t-p+1} + e_t$ by OLS.

ADF Test:

- $H_0: \ \rho = 1$ (Sprious Regression)
- $H_1: \ \rho < 1$ (Cointegration)

$\implies$ **Engle-Granger Test**

For example, see Engle and Granger (1987), Phillips and Ouliaris (1990) and Hansen (1992).

Asymmptotic Distribution of Residual-Based ADF Test for Cointegration

# of Refressors,	(a) Regressors have no drift				(b) Some regressors have drift			
excluding constant	1%	2.5%	5%	10%	1%	2.5%	5%	10%
1	−3.96	−3.64	−3.37	−3.07	−3.96	−3.67	−3.41	−3.13
2	−4.31	−4.02	−3.77	−3.45	−4.36	−4.07	−3.80	−3.52
3	−4.73	−4.37	−4.11	−3.83	−4.65	−4.39	−4.16	−3.84
4	−5.07	−4.71	−4.45	−4.16	−5.04	−4.77	−4.49	−4.20
5	−5.28	−4.98	−4.71	−4.43	−5.36	−5.02	−4.74	−4.46

J.D. Hamilton (1994), *Time Series Analysis*, p.766.