Unit Roots, Stochastic Trends, ARIMA Forecasting Models

Ka-fu WONG

University of Hong Kong

January 17, 2020

Contents

1. A quick review of autoregressive models

Recall that the simple AR(1) model with zero mean:

$$
\begin{aligned}
(1.1) \qquad y_t &= \rho y_{t-1} + \epsilon_t \qquad \epsilon_t \sim WN(0, \sigma^2) \\
y_t - \rho y_{t-1} &= \epsilon_t \\
(1 - \rho L) y_t &= \epsilon_t
\end{aligned}
$$

Stationarity requires the root of $(1 - \rho x) = 0$ has absolute value larger than 1, i.e., $|\rho| < 1$. With this condition, the Wold representation of y_t exists.

$$
(1.2) \qquad y_t = \sum_{i=0}^{\infty} \rho^i \epsilon_{t-i} = \epsilon_t + \rho \epsilon_{t-1} + \rho^2 \epsilon_{t-2} + \ldots
$$

Stationarity, as illustrated in this AR(1) process, has several important implications.

(1) The impact of an innovation in time t, ϵ_t, on future y_t, i.e., y_{t+h}, diminishes *towards zero* as the horizon h increases. Or, if y_t has a non-zero mean, y_{t+h}, diminishes *towards the unconditional mean* of y_t as the horizon h increases.

(2) The unconditional variance of y_t is finite.

$$
\begin{aligned}
V(y_t) &= V\left(\sum_{i=0}^{\infty} \rho^i \epsilon_{t-i}\right) = \sum_{i=0}^{\infty} V(\rho^i \epsilon_{t-i}) \\
&= \sum_{i=0}^{\infty} \rho^{2i} V(\epsilon_{t-i}) = \sum_{i=0}^{\infty} \rho^{2i} \sigma^2 \\
&= \frac{\sigma^2}{1-\rho^2} < \infty
\end{aligned}
$$

A higher order AR process have similar properties. An AR(p) model with zero mean:

$$\begin{aligned} y_t &= \rho_1 y_{t-1} + \rho_2 y_{t-2} + \dots + \rho_p y_{t-p} + \epsilon_t \\ y_t - \rho_1 y_{t-1} - \rho_2 y_{t-2} - \dots - \rho_p y_{t-p} &= \epsilon_t \\ (1 - \rho_1 L - \rho_2 L^2 - \dots - \rho_p L^p) y_t &= \epsilon_t \end{aligned}$$

Stationarity requires the roots of

$$(1 - \rho_1 x - \rho_2 x^2 - \dots - \rho_p x^p) = 0$$

to have absolute value (or modulus for complex roots) larger than 1.

With this condition, the Wold representation of y_t exists.

(1.3) $$y_t = \sum_{i=0}^{\infty} b_i \epsilon_{t-i} = b_0 \epsilon_t + b_1 \epsilon_{t-1} + b_2 \epsilon_{t-2} + \dots$$

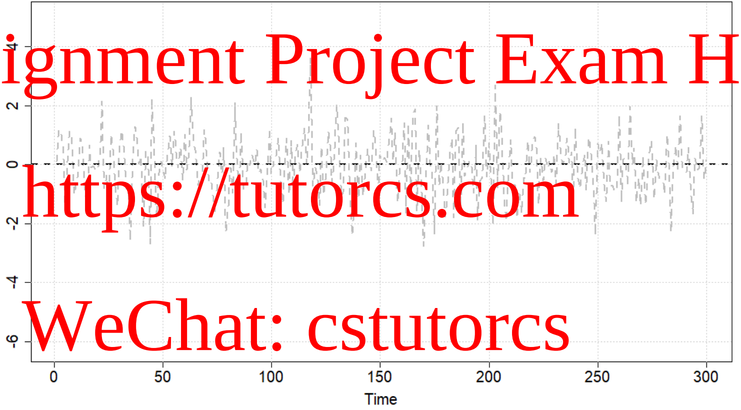
where b_0 equals to 1 and b_i will be a function of the ρ_j's.

Again, stationarity has several important implications.

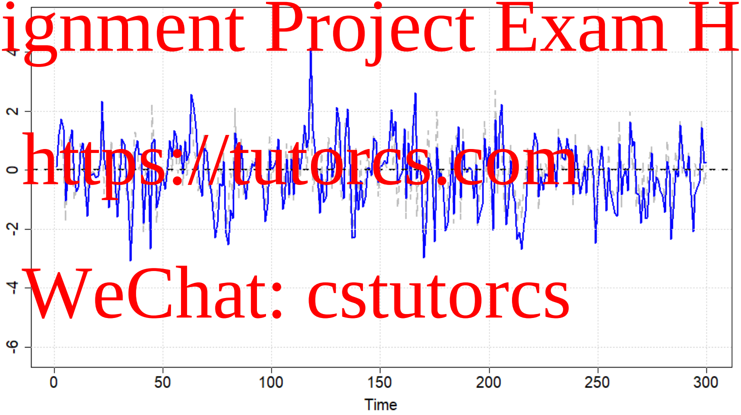
(1) The impact of an innovation in time t, ϵ_t, on future y_t, i.e., y_{t+h}, diminishes *towards zero* as the horizon h increases. Or, if y_t has a non-zero mean, y_{t+h}, diminishes *towards the unconditional mean* of y_t as the horizon h increases.

(2) The unconditional variance of y_t is finite.

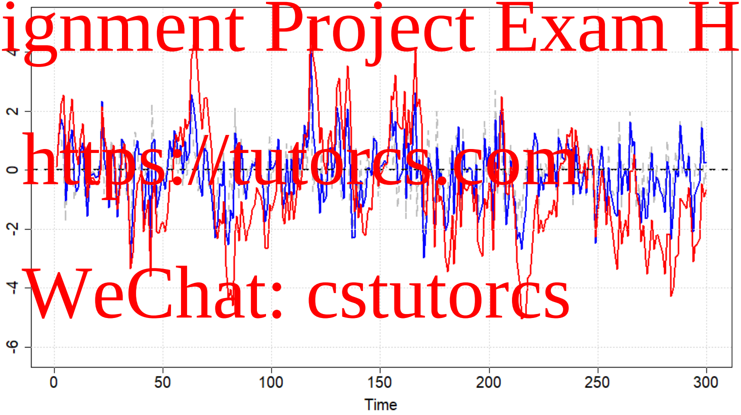
White Noise: $y_t = \epsilon_t$, $\epsilon_t \sim WN(0, 1)$



Grey dotted line: $y_t = \epsilon_t$, $\epsilon_t \sim WN(0,1)$
Blue solid line: $y_t = 0.5y_{t-1} + \epsilon_t$, $\epsilon_t \sim WN(0,1)$

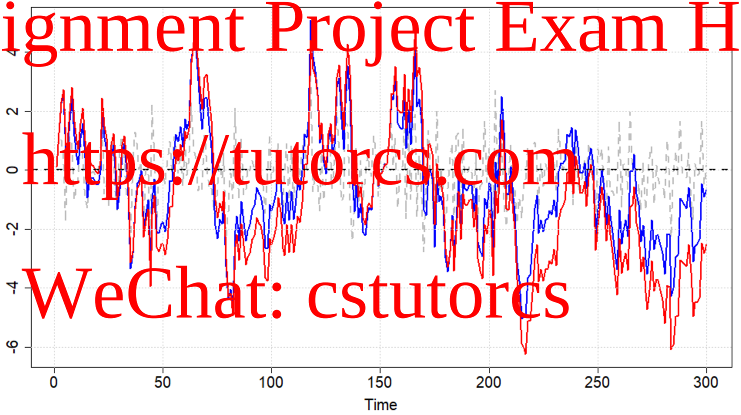


Blue solid line: $y_t = 0.5y_{t-1} + \epsilon_t$, $\epsilon_t \sim WN(0,1)$
Red solid line: $y_t = 0.9y_{t-1} + \epsilon_t$, $\epsilon_t \sim WN(0,1)$



Blue solid line: $y_t = 0.9y_{t-1} + \epsilon_t$, $\epsilon_t \sim WN(0,1)$

Red solid line: $y_t = 0.95y_{t-1} + \epsilon_t$, $\epsilon_t \sim WN(0,1)$



2. Random walk as a unit root

Random walk has the following form:

$$y_t = y_{t-1} + \epsilon_t \qquad (2.1)$$

The optimal forecast of y_{t+1} given time t information is

$$\begin{aligned} E\{y_{t+1} \mid y_t, y_{t-1}, \ldots\} &= E\{y_t + \epsilon_{t+1} \mid y_t, y_{t-1}, \ldots\} \\ &= E(y_t \mid y_t, y_{t-1}, \ldots) + E(\epsilon_{t+1} \mid y_t, y_{t-1}, \ldots) \\ &= y_t \end{aligned}$$

The optimal forecast of y_{t+2} given time t information is

$$\begin{aligned} E\{y_{t+2} \mid y_t, y_{t-1}, \ldots\} &= E\{y_{t+1} + \epsilon_{t+2} \mid y_t, y_{t-1}, \ldots\} \\ &= E(y_{t+1} \mid y_t, y_{t-1}, \ldots) + E(\epsilon_{t+2} \mid y_t, y_{t-1}, \ldots) \\ &= y_t \end{aligned}$$

Similarly, the optimal forecast of y_{t+h} given time t information is

$$E\{y_{t+h} \mid y_t, y_{t-1}, \ldots\} = y_t.$$

Whatever happen in time t (a shock that changes y_t) has *permanent effect* on future y_{t+h} and hence its forecast.

RW is considered as a naive and lazy forecasting model. Using RW as a forecasting model, there is no need to estimate any regression model before we make a forecast. As examples of short horizon forecast, we would use

(1) today's closing as a forecast of tomorrow's Hang Seng Index closing.
(2) today's yen-dollar rate as a forecast of tomorrow's yen-dollar exchange rate.
(3) today's oil price as a forecast of tomorrow's oil price.
(4) today's gold price as a forecast of tomorrow's gold price.
(5) the last target federal funds rate as a forecast of the target federal funds rate after the next FOMC meeting.
(6) past one year's GDP growth rate as a forecast of next year's GDP growth rate.
(7) past one year's inflation rate as a forecast of next year's inflation rate.

Even if we may not consider RW naive in the short horizon forecast, it would definitely sound *naive* when we use the model to perform *long-horizon forecast*, say 10-year ahead.

Comparing equations (1.1) and (2.1), we can easily see that RW is like $AR(1)$ with $\rho = 1$.

$$y_t = \rho y_{t-1} + \epsilon_t \quad \text{versus} \quad y_t = y_{t-1} + \epsilon_t$$

Interpreting RW in the AR(1) framework, for RW the root of $(1 - \rho x) = 0$ takes a value of 1, a *unit root*. That is why RW is also known as unit root.

The root (x) in $(1 - \rho x) = 0$ equals to '1' $\iff \rho = 1$
(The root (x) in $(1 - \rho x) = 0$ equals to "-1" $\iff \rho = -1$)

So, one may think that RW is a limiting case of AR(1) when ρ approaches 1 and thus they should have *similar properties*.

That turns out to be *WRONG*!

3. Drastic difference between stationary AR(1) and unit root

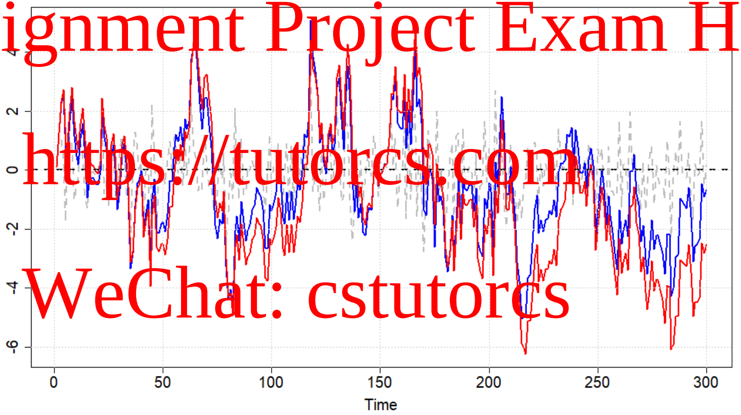
Table 3.1 shows that the dependence of the forecast $E(y_{t+h} \mid y_t, y_t - 1, \ldots)$ on y_t at different horizons for AR(1) and unit root processes.

Table 3.1. The forecast $E(y_{t+h} \mid y_t, y_t - 1, \ldots)$ at different horizons

	$\rho = 0.5$	$\rho = 0.9$	$\rho = 0.99$	$\rho = 0.999$	$\rho = 1$
$h = 1$	$0.5y_t$	$0.9y_t$	$0.99y_t$	$0.999y_t$	y_t
$h = 5$	$0.031y_t$	$0.59y_t$	$0.951y_t$	$0.995y_t$	y_t
$h = 10$	$0.001y_t$	$0.349y_t$	$0.904y_t$	$0.99y_t$	y_t
$h = 100$	0.000	0.000	$0.366y_t$	$0.905y_t$	y_t
$h = 1000$	0.000	0.000	0.000	$0.368y_t$	y_t
$h = 10000$	0.000	0.000	0.000	0.000	y_t

The difference becomes visible in the long horizons. As long as $|\rho| < 1$, y_t *would not help forecast* y_{t+h} eventually, at *long horizons*.

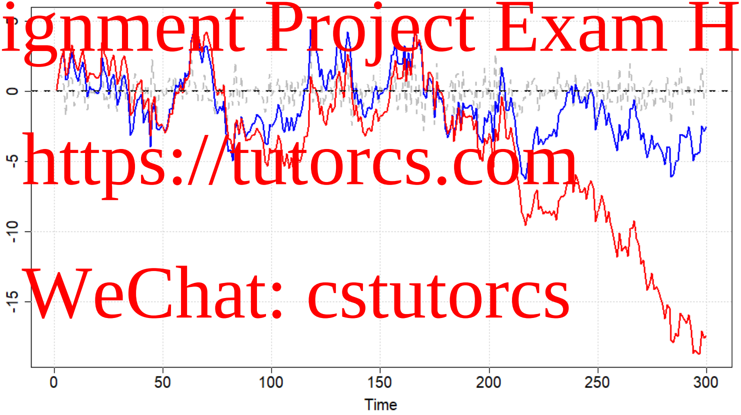
Blue solid line: $y_t = 0.5y_{t-1} + \epsilon_t$, $\epsilon_t \sim WN(0,1)$
Red solid line: $y_t = 0.9y_{t-1} + \epsilon_t$, $\epsilon_t \sim WN(0,1)$



Drastic difference between stationary AR(1) and unit root

Blue solid line: $y_t = 0.95y_{t-1} + \epsilon_t$, $\epsilon_t \sim WN(0,1)$

Red solid line: $y_t = y_{t-1} + \epsilon_t$, $\epsilon_t \sim WN(0,1)$



4. Non-stationary of unit root process

Given a random walk, we can use repeat substitutions to obtain

$$
\begin{aligned}
y_t &= y_{t-1} + \epsilon_t \\
&= y_{t-2} + \epsilon_{t-1} + \epsilon_t \\
&= y_{t-3} + \epsilon_{t-2} + \epsilon_{t-1} + \epsilon_t \\
&\cdots \quad \cdots \\
&= y_0 + \epsilon_1 + \epsilon_2 + \ldots + \epsilon_{t-1} + \epsilon_t \\
&= y_0 + \sum_{i=1}^{t} \epsilon_i
\end{aligned}
$$

where y_0 is the initial value of the process.

It is easy to see that

(1) $E(y_t) = y_0$, i.e., the unconditional mean depends on the initial value of the process – even for large t.

$$
\begin{aligned}
E(y_t) &= E\left(y_0 + \sum_{i=1}^{t} e_i\right) \\
&= E(y_0) + \sum_{i=1}^{t} E(e_i) \\
&= y_0
\end{aligned}
$$

(2) $Var(y_t) = t\sigma^2$, i.e., variance increases with time.

$$
\begin{aligned}
Var(y_t) &= Var\left(y_0 + \sum_{i=1}^{t} \epsilon_i\right) \\
&= Var\left(\sum_{i=1}^{t} \epsilon_i\right) \\
&= \sum_{i=1}^{t} Var(\epsilon_i) + \text{a bunch of covariances between } \epsilon_i \text{ and } \epsilon_j,\ i \neq j \\
&= t\sigma^2 + 0
\end{aligned}
$$

(3) $\lim_{t\to\infty} Var(y_t) = \infty$, i.e., the variance approaches infinity as t increases.

Clearly, a unit root process is *not covariance stationary.*

A unit root process is *not covariance stationary* *because*

- Unconditonal mean depends on the initial condition/value.
- Unconditional variance depends on time, and approaches infinity as t increases

5. Random walk with drift

By adding a constant term to the random walk process, we will have a random walk with drift.

$$y_t = \underset{\substack{\uparrow \\ \text{drift}}}{\delta} + y_{t-1} + \epsilon_t$$

$$\begin{aligned} y_t &= \delta + y_{t-1} + \epsilon_t \\ y_t - y_{t-1} &= \delta + \epsilon_t \\ (1 - L)y_t &= \delta + \epsilon_t \\ \Delta y_t &= \delta + \epsilon_t \end{aligned}$$

Again we can use repeat substitution to obtain

$$
\begin{aligned}
y_1 &= \delta + y_0 + \epsilon_1 \\
y_2 &= \delta + y_1 + \epsilon_2 = 2\delta + y_0 + \epsilon_2 + \epsilon_1 \\
y_3 &= \delta + y_2 + \epsilon_3 = 3\delta + y_0 + \epsilon_3 + \epsilon_2 + \epsilon_1 \\
y_t &= \delta + y_{t-1} + \epsilon_t = t\delta + y_0 + \epsilon_t + \epsilon_{t-1} + ... + \epsilon_1
\end{aligned}
$$

or

$$
y_t = t\delta + y_0 + \sum_{i=1}^{t} \epsilon_t
$$

Thus, the random walk with drift model implies a linear time trend, $t\delta$.

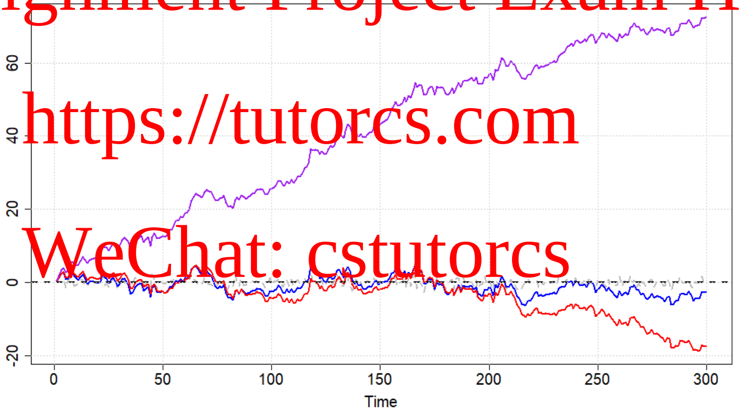
- This time trend *would not result if* we have $y_t = \delta + \rho y_{t-1} + \epsilon_t$ and $|\rho| < 1$.
- This time trend is different from the conventional deterministic trend model such as $y_t = \delta t + \epsilon_t$.

To distinguish it from the deterministic trend model, random walk with drift is often called a model of *stochastic trend*.

Blue solid line: $y_t = 0.95 y_{t-1} + \epsilon_t$, $\epsilon_t \sim WN(0,1)$

Red solid line: $y_t = y_{t-1} + \epsilon_t$, $\epsilon_t \sim WN(0,1)$

Purple solid line: $y_t = 0.3 + y_{t-1} + \epsilon_t$, $\epsilon_t \sim WN(0,1)$



It is easy to see that

(1) $E(y_t) = t\delta + y_0$, i.e., the unconditional mean depends on the initial value of the process.

(2) $Var(y_t) = t\sigma^2$, i.e., variance increases with time.

(3) $\lim_{t\to\infty} Var(y_t) = \infty$, i.e., the variance approaches infinity as t increases.

Thus, a *random walk with drift* process is *not covariance stationary*.

6. ARMA with Unit Root

While a unit root in y_t is not stationary, Δy_t, the difference of y_t is *stationary*.

$$\Delta y_t = y_t - y_{t-1} = \epsilon_t$$

Δy_t is in fact ARMA(0,0).
Sometimes, we want to talk about the process of y_t, and we say y_t is ARIMA(0,1,0).

AR	I	MA
0	1	0

The middle "integration" order "1" is used to tell readers that y_t has a unit root, and the "first" difference of y_t (Δy_t) will be stationary.

Consider an AR(2) process

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \epsilon_t$$

We can rewrite it as

$$(1 - \rho_1 L - \rho_2 L^2) y_t = \epsilon_t$$

or

$$(1 - c_1 L)(1 - c_2 L) y_t = \epsilon_t$$

Stationarity requires $| c_1 | < 1$ and $| c_2 | < 1$, or the roots of $(1 - c_1 x)(1 - c_2 x) = 0$ is larger than 1 in absolute values.

If one of the roots is equal to 1, we have one unit root, say, $c_1 = 1$, we can rewrite

$$
\begin{aligned}
(1 - c_2 L)(1 - L)y_t &= \epsilon_t \\
(1 - c_2 L)\Delta y_t &= \epsilon_t
\end{aligned}
$$

Thus, Δy_t, the difference of y_t is *stationary*. In fact, Δy_t is $ARMA(1, 0)$, or y_t is $ARIMA(1, 1, 0)$.

If y_t has two unit roots, i.e., $c_1 = 1$ and $c_2 = 1$, we have

$$
\begin{aligned}
(1 - L)(1 - L)y_t &= \epsilon_t \\
\Delta\Delta y_t &= \epsilon_t
\end{aligned}
$$

or

$$\Delta^2 y_t = \epsilon_t$$

Thus, $\Delta^2 y_t$, the "second" difference of y_t is *stationary*. In fact, $\Delta^2 y_t$ is $ARMA(0, 0)$, or y_t is $ARIMA(0, 2, 0)$. Again, the middle "integration" order "2" is used to tell readers that y_t has two unit roots.

Generally, consider an $AR(p)$ process

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + ... + \rho_p y_{t-p} + \epsilon_t$$

We can check stationarity by solving for the roots of

$$1 - \rho_1 x - \rho_2 x^2 - ... - \rho_p x^p = 0$$

If there is exactly one unit root, Δy_t, the first difference of y_t is *stationary*. And, Δy_t is $ARMA(p-1, 0)$, or y_t is $ARIMA(p-1, 1, 0)$.

If there is exactly two unit roots, $\Delta^2 y_t$, the second difference of y_t is *stationary*. And, $\Delta^2 y_t$ is $ARMA(p-2, 0)$, or y_t is $ARIMA(p-2, 2, 0)$.

If there is exactly d unit roots, $\Delta^d y_t$, the d-th difference of y_t is *stationary*. And, $\Delta^d y_t$ is $ARMA(p-d, 0)$, or y_t is $ARIMA(p-d, d, 0)$.

Remark:
Often, when the data is found non-stationary (say, by estimating a AR(p) and computing the roots), we will take the first difference of the data. If the first differenced data is found non-stationary, we will take an additional difference. If the second differenced data is found non-stationary, we will continue to take an additional difference.
That is, take n-th difference, until the n-th differenced data appears stationary.

It is easy to generalize the discussion to ARMA(p,q) models.

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + ... + \rho_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... \theta_q \epsilon_{t-q}$$

We can check stationarity by solving for the roots of

$$1 - \rho_1 x - \rho_2 x^2 - ... - \rho_p x^p = 0$$

If there is exactly one unit root, Δy_t, the first difference of y_t is *stationary*. And, Δy_t is $ARMA(p-1, q)$, or y_t is $ARIMA(p-1, 1, q)$.

If there is exactly two unit roots, $\Delta^2 y_t$, the second difference of y_t is *stationary*. And, $\Delta^2 y_t$ is $ARMA(p-2, q)$, or y_t is $ARIMA(p-2, 2, q)$.

If there is exactly d unit roots, $\Delta^d y_t$, the d-th difference of y_t is *stationary*. And, $\Delta^d y_t$ is $ARMA(p-d, q)$, or y_t is $ARIMA(p-d, d, q)$.

7. Similarity of $ARIMA(p, 1, q)$ to random walk

$ARIMA(p, 1, q)$ processes are appropriately made stationary by differencing.

Shocks (ϵ_t) to ARIMA(p,1,q) processes have *permanent* effects.

- Hence, shock persistence means that optimal forecasts even at very long horizons *do not completely revert to a mean or a trend*.
  (Shocks (ϵ_t) to ARIMA(p,0,q) processes have *transitory* effects.)

The variance of an ARIMA(p,1,q) process grows without bound as time progresses.

- Uncertainty associated with our forecasts grows with horizon of our forecast.

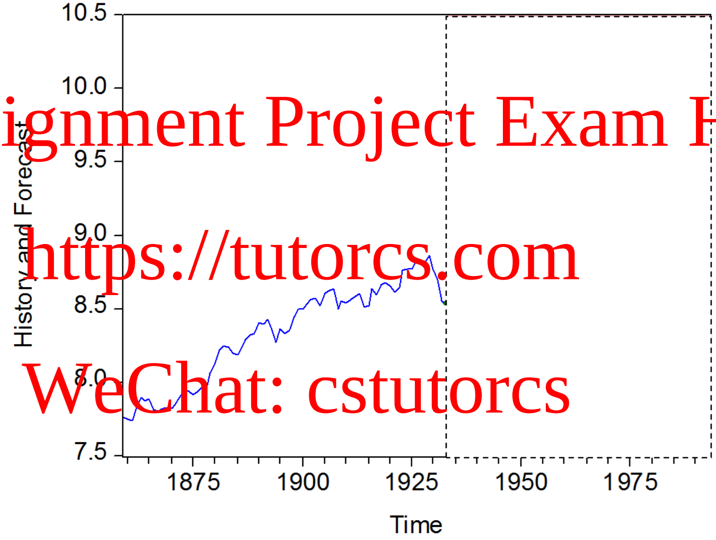
- Width of our interval forecast grows without bound with the horizon of our forecast.

8. Application: Forecast of US GDP per capita trend

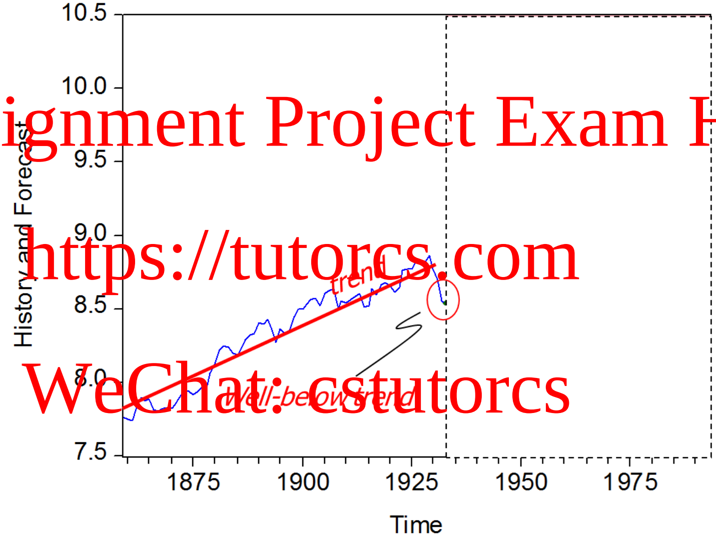
- Estimation: 1859-1933 US GDP per capita
  (data from http://www.measuringworth.com/usgdp/?q=hmit/gdp)
- Forecast: 1934-1993 US GDP per capita
- Two types of models
  - $ARMA(p, q)$: model selection criteria suggest $p = 2$, $q = 0$, i.e., $ARMA(2, 0)$
  - $ARIMA(p, d, q)$: model selection criteria suggest $AR(1)$ in difference (i.e., $y_t - y_{t-1}$) with drift, i.e., $ARIMA(1, 1, 0)$

Note: The example is drawn from Diebold and Senhadji (1996).[1]

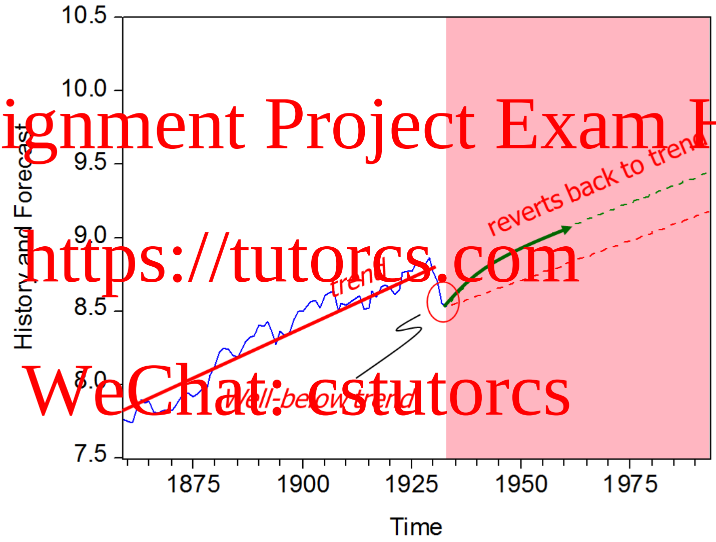
[1] Diebold, Francis X. and Abdelhak S. Senhadji (1996): "The Uncertain Unit Root in Real GDP: Comment," *American Economic Review*, 86: 1291-1298.



Application: Forecast of US GDP per capita trend



Application: Forecast of US GDP per capita trend



Application: Forecast of US GDP per capita trend

Application: Forecast of US GDP per capita trend

9. Detecting Unit Roots

Random walk may be viewed as a limiting case of AR(1).

The Correlogram (ACF and PACF) looks like that of AR(1), and thus cannot be used as a test of unit root.

Several statistical tests are available.

- Augmented Dickey–Fuller test
- Dickey–Fuller test
- Phillips–Perron test
- KPSS test
- Zivot–Andrews test

The most common ones are Dickey–Fuller (DF) test or augmented Dickey–Fuller (ADF) test.

9.1. Detecting unit root without drift.
A unit root process is a limiting case of $AR(1)$, which AR coefficient $\rho = 1$.

$$y_t = \rho y_{t-1} + \epsilon_t \qquad \text{versus} \qquad y_t = y_{t-1} + \epsilon_t$$

To test

$$H_0\text{: AR coefficient } \rho = 1 \qquad \text{versus} \qquad H_1\text{: AR coefficient } \rho < 1$$
(stationary alternative)

regress y_t on y_{t-1} and compute the test statistic in the usual way.

$$\hat{\tau} = \frac{\hat{\rho} - 1}{s\sqrt{\frac{1}{\sum_{t=2}^{T} y_{t-1}^2}}}$$

where

$$s = \sqrt{\frac{\sum_{t=2}^{T} \hat{\epsilon}_t}{T - 2}}$$

Sometimes, we would like to rewrite the model as

$$y_t - y_{t-1} = (\rho - 1)y_{t-1} + \epsilon_t$$
$$\Delta y_t = \beta y_{t-1} + \epsilon_t$$

To test

$$H_0\colon \text{ AR coefficient } \rho = 1 \qquad \text{versus} \qquad H_1\colon \text{ AR coefficient } \rho < 1$$

is almost equivalent to test

$$H_0\colon \beta = (\rho - 1) = 0 \qquad \text{versus} \qquad H_1\colon \beta = (\rho - 1) < 0$$

To conduct the test, regress Δy_t on y_{t-1} and compute the test statistic in the usual way.

Almost all statistical softwares will compute *automatically* the corresponding t-statistic for us.

While the test looks like the usual t-test, we *cannot use the usual t-test critical values* for our test, because it turns out that the t-statistic

$$\hat{\tau} = \frac{\hat{\rho} - 1}{s\sqrt{\frac{1}{\sum_{t=2}^{T} y_{t-1}^2}}}$$

does not have the usual Student-t distribution. That is, the p-value spitted out automatically by the OLS regression procedure of usual statistical software is not to be trusted.

The statistics has shown to have a special distribution, called the *Dickey-Fuller* distribution. That is, we have to use critical values based on Dickey-Fuller distribution, not Student-t distribution.

9.2. Detecting unit root *with drift*.

When the null is unit root with drift $y_t = \alpha + y_{t-1} + \epsilon_t$, it seems logical to consider the alternative of $y_t = \alpha + \rho y_{t-1} + \epsilon_t$. That is

$$y_t = \alpha + \rho y_{t-1} + \epsilon_t \quad \text{versus} \quad y_t = \alpha + y_{t-1} + \epsilon_t$$

We can rewrite the model as

$$\begin{aligned}
y_t &= \alpha + \rho y_{t-1} + \epsilon_t \\
y_t &= \mu(1-\rho) + \rho y_{t-1} + \epsilon_t \\
y_t - \mu &= \rho y_{t-1} - \rho\mu + \epsilon_t \\
(y_t - \mu) &= \rho(y_{t-1} - \mu) + \epsilon_t
\end{aligned}$$

Under null of unit root, $\rho = 1$, and *in addition*, $\rho = 1$ implies $\alpha = 0$.
(Note the inconsistency of unit root with drift?)

If μ is known, we can simply subtract μ from y_t and it reduces to the simple case without drift. We can then estimate
$(y_t - \mu) = \rho(y_{t-1} - \mu) + \epsilon_t$ and test $H_0 : \rho = 1$

Since μ is not known, we have to estimate $y_t = \alpha + \rho y_{t-1} + \epsilon_t$ and test $H_0 : (\alpha, \rho) = (0, 1)$.

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

If we rewrite the model as

$$\begin{aligned} y_t - y_{t-1} &= \alpha + (\rho - 1)y_{t-1} + \epsilon_t \\ &= \alpha + \beta y_{t-1} + \epsilon_t \end{aligned}$$

we would be testing $\beta = 0$ and $\alpha = 0$. Thus, some researchers will consider two possible null hypotheses in conducting unit root test.

- $H_0 : \beta = 0$
- $H_0 : \beta = 0, \alpha = 0$

Again, the test statistic will have a non-standard distribution. So, we have to refer to the Dickey-Fuller distribution.

9.3. Detecting unit root with drift (stochastic versus deterministic trend).

$$y_t = \alpha + \beta TIME_t + \rho y_{t-1} + \epsilon_t \quad \text{versus} \quad y_t = \alpha + y_{t-1} + \epsilon_t$$

We can rewrite the model as

$$y_t = \alpha + \beta TIME_t + \rho y_{t-1} + \epsilon_t$$

$$y_t = (1-\rho)(a + bTIME_t) + \rho y_{t-1} + \epsilon_t$$

$$y_t - (a + bTIME_t) = \rho\left[y_{t-1} - (a + bTIME_t)\right] + \epsilon_t$$

$$y_t - (a + bTIME_t) = \rho\left[y_{t-1} - (a + bTIME_{t-1})\right] - \rho b + \epsilon_t$$

Under null of unit root, $\rho = 1$.

If a and b are known, we can simply subtract $(a + bTIME_t)$ from y_t and it reduces to the simple case without drift and trend.

Since a and b are not known, we have to estimate $y_t = \alpha + \beta TIME_t + \rho y_{t-1} + \epsilon_t$ and test $\rho = 1$.

If we are rewrite the model as

$$\begin{aligned} y_t - y_{t-1} &= \alpha + \beta TIME_t + (\rho - 1) y_{t-1} + \epsilon_t \\ &= \alpha + \beta TIME_t + \gamma y_{t-1} + \epsilon_t \end{aligned}$$

we would be testing $\gamma = 0$.
Testing $\gamma = 0$ is equivalent to testing

Null	vs	Alternative
$y_t = \alpha + y_{t-1} + \epsilon_t$		$y_t = \alpha + \beta TIME_t + \rho y_{t-1} + \epsilon_t$, $\lvert\rho\rvert < 1$

Note that when the nonzero α (drift) in the null specification is consistent with the time trend in data (y_t).

Some researchers will consider the additional two hypotheses when a *deterministic trend* is included:

- $H_0 : \rho = 1, \alpha = 0$ (or $H_0 : \gamma = (\rho - 1) = 0, \alpha = 0$) ,

Null	vs	Alternative
$y_t = y_{t-1} + \epsilon_t$		$y_t = \alpha + \beta TIME_t + \rho y_{t-1} + \epsilon_t$, $\lvert\rho\rvert < 1$

- and sometimes $H_0 : \rho = 1, \alpha = 0, \beta = 0$ (or $H_0 : \gamma = (\rho - 1) = 0, \alpha = 0, \beta = 0$)

Null	vs	Alternative
$y_t = y_{t-1} + \epsilon_t$		$y_t = \alpha + \beta TIME_t + \rho y_{t-1} + \epsilon_t$, $\lvert\rho\rvert < 1$

(Note that if the null is $y_t = y_{t-1} + \epsilon_t$, we should not see a time trend in y_t.)

Again, the test statistic will have a non-standard distribution. So, we have to refer to the Dickey-Fuller distribution.

9.4. Detecting unit root with higher order autoregressive dynamics versus $AR(2)$.

We can rewrite the AR(2)

$$y_t + \rho_1 y_{t-1} + \rho_2 y_{t-2} = \epsilon_t$$

as

$$y_t = -\rho_1 y_{t-1} - \rho_2 y_{t-2} + \epsilon_t$$

$$y_t = -\rho_1 y_{t-1} - \rho_2 y_{t-1} + \rho_2 y_{t-1} - \rho_2 y_{t-2} + \epsilon_t$$

$$y_t = -(\rho_1 + \rho_2) y_{t-1} + \rho_2 (y_{t-1} - y_{t-2}) + \epsilon_t$$

$$y_t = \beta_1 y_{t-1} + \beta_2 (y_{t-1} - y_{t-2}) + \epsilon_t$$

where

$$\beta_1 = -(\rho_1 + \rho_2), \quad \beta_2 = \rho_2$$

To test for unit root, regress y_t on y_{t-1} and $(y_{t-1} - y_{t-2})$, and test the null of $\beta_1 = 1$, using the Dickey-Fuller distribution.

If we rewrite the model as

$$
\begin{aligned}
y_t - y_{t-1} &= (\beta_1 - 1) y_{t-1} + \beta_2 \left(y_{t-1} - y_{t-2} \right) + \epsilon_t \\
&= \gamma y_{t-1} + \beta_2 \left(y_{t-1} - y_{t-2} \right) + \epsilon_t \\
\Delta y_t &= \gamma y_{t-1} + \beta_2 \Delta y_{t-1} + \epsilon_t
\end{aligned}
$$

we would be testing $\gamma = 0$.

9.5. **Detecting unit root with higher order autoregressive dynamics versus $AR(3)$.**
We can rewrite the AR(3)

$$y_t + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \rho_3 y_{t-3} = \epsilon_t$$

as

$$y_t = \beta_1 y_{t-1} + \beta_2 \left(y_{t-1} - y_{t-2}\right) + \beta_3 \left(y_{t-2} - y_{t-3}\right) + \epsilon_t$$

To test for unit root, regress y_t on y_{t-1}, $(y_{t-1} - y_{t-2})$ and $(y_{t-2} - y_{t-3})$, and test the null of $\beta_1 = 1$, using the Dickey-Fuller distribution.
If we rewrite the model as

$$\begin{aligned} y_t - y_{t-1} &= (\beta_1 - 1) y_{t-1} + \beta_2 \left(y_{t-1} - y_{t-2}\right) + \beta_3 \left(y_{t-2} - y_{t-3}\right) + \epsilon_t \\ &= \gamma y_{t-1} + \beta_2 \left(y_{t-1} - y_{t-2}\right) + \beta_3 \left(y_{t-2} - y_{t-3}\right) + \epsilon_t \\ \Delta y_t &= \gamma y_{t-1} + \beta_2 \Delta y_{t-1} + \beta_3 \Delta y_{t-2} + \epsilon_t \end{aligned}$$

we would be testing $\gamma = 0$.

9.6. Detecting unit root with higher order autoregressive dynamics versus $AR(p)$.

We can rewrite the AR(p)

$$y_t + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \dots + \rho_p y_{t-p} = \epsilon_t$$

as

$$y_t = \beta_1 y_{t-1} + \beta_2 \left(y_{t-1} - y_{t-2}\right) + \dots + \beta_p \left(y_{t-p+1} - y_{t-p}\right) + \epsilon_t$$

To test for unit root, regress y_t on y_{t-1}, $(y_{t-1} - y_{t-2})$ to $(y_{t-p+1} - y_{t-p})$, and test the null of $\beta_1 = 1$, using the Dickey-Fuller distribution.
If we rewrite the model as

$$\begin{aligned} y_t - y_{t-1} &= (\beta_1 - 1) y_{t-1} + \beta_2 \left(y_{t-1} - y_{t-2}\right) + \dots + \beta_p \left(y_{t-p+1} - y_{t-p}\right) + \epsilon_t \\ &= \gamma y_{t-1} + \beta_2 \left(y_{t-1} - y_{t-2}\right) + \dots + \beta_p \left(y_{t-p+1} - y_{t-p}\right) + \epsilon_t \\ \Delta y_t &= \gamma y_{t-1} + \beta_2 \Delta y_{t-1} + \dots + \beta_p \Delta y_{t-p+1} + \epsilon_t \end{aligned}$$

we would be testing $\gamma = 0$.

9.7. **Detecting unit root with higher order autoregressive dynamics *with trend*.**

We can rewrite

$$(y_t - a - bTIME_t) - \sum_{j=1}^{p} \rho_j (y_{t-j} - a - bTIME_{t-j}) = \epsilon_t$$

as

$$y_t = k_1 + k_2 TIME_t + \beta_1 y_{t-1} + \sum_{j=2}^{p} \beta_p (y_{t-j+1} - y_{t-j}) + \epsilon_t$$

where

$$k_1 = a\left(1 + \sum_{i=1}^{p} \rho_i\right) - b \sum_{i=1}^{p} i\rho_i$$

and

$$k_2 = b\left(1 + \sum_{i=1}^{p} \rho_i\right)$$

Under the null of unit root with drift, we would test $\beta_1 = 1$.

Under the null of unit root without drift, we would test

- $H_0 : \beta_1 = 1, k_1 = 0$, and
- sometimes $H_0 : \beta_1 = 1, k_1 = 0, k_2 = 0$

Again, to test the null of $\beta_1 = 1$, and those k_1 and k_2, using the Dickey-Fuller distribution.

If we rewrite the model as

$$y_t - y_{t-1} = k_1 + k_2 TIME_t + (\beta_1 - 1) y_{t-1} + \sum_{j=2}^{p} \beta_p \left(y_{t-j+1} - y_{t-j} \right) + \epsilon_t$$

$$\Delta y_t = k_1 + k_2 TIME_t + \gamma y_{t-1} + \sum_{j=2}^{p} \beta_p \Delta y_{t-j+1} + \epsilon_t$$

we would be testing $\gamma = 0$, or $\gamma = 0, k_1 = 0, k_2 = 0$

9.8. **Detecting unit root with general ARMA models.**

- Complicated as any MA terms will map into infinite AR terms. But in practice we can include small number of AR terms as an approximation.
- As an *approximation*, *add many AR terms* and still use the standard Dickey-Fuller distribution.

10. Difference or not?

In certain respects, the most important part of unit root theory for forecasting concerns estimation, not testing. It's important for forecasters to understand the effects of unit roots on consistency and small-sample bias. Such understanding on the one hand leads to the insight that at least asymptotically *we're probably better estimating forecasting models in levels with trends included* because then we'll get an accurate approximation to the dynamics in the data regardless of the true state of the world, unit root or no unit root. If there's no unit root, then of course it's desirable to work in levels; if there is a unit root, then the estimated largest root will converge appropriately to unity, and at a fast rate. On the other hand, *differencing is appropriate only in the unit root case, and inappropriate differencing can be harmful, even asymptotically*.

(Francis Diebold, *Elements of Forecasting*)

11. Forecasting the Yen/Dollar Exchange Rate

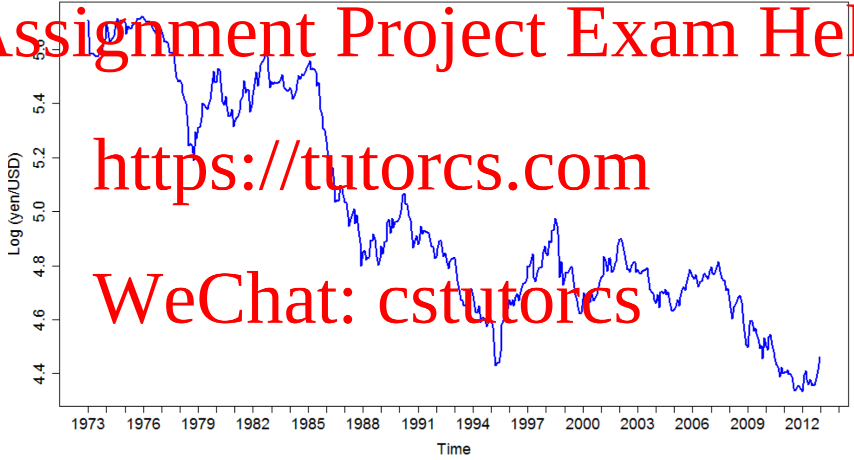
- y = yen/dollar series
- Work with Levels:
  - $\ln(y)$ or
  - Differences: $\Delta \ln(y)$
- Data: 1973.01 to 2014.12, *end-of-month* data (not monthly average), from http://research.stlouisfed.org/fred2/series/DEXJPUS#
- Estimation: 1973.01 – 2012.12
- Out-of-sample Forecast: 2013.01 – 2014.12

12. Level data

12.1. **Preliminary data analysis.**

- First, we will do some preliminary data analysis.
  - plot time series data
  - ACF and PACF

FIGURE 12.1. Log JPY/USD Rate (1971-2012)



Level data

Figure 12.2. ACF of Log JPY/USD Rate

FIGURE 12.3. PACF of Log JPY/USD Rate

12.2. **Box-Pierce Test of White Noise.** Obviously, if there series is white noise, we have nothing else to do but to report the non-forecastability.
Here, we adopt the simple Box-Pierce Test:

$$Q_{BP} = \sum_{k=1}^{m} \hat{\rho}_k^2 \sim \chi^2(m)$$

Setting $m = 21 \approx \sqrt{480}$, we obtain

- $\hat{Q}_{BP} = 3212.427$, df $= 21$, p-value $< 2.2\text{e-}16$

The white noise null is obviously rejected at all conventional level of significance.

12.3. **Model selection.** We consider ARMA model with linear trend, with maximum AR order (p) and MA order (q) of 5.

TABLE 12.1. p-value of Box Test

	q = 0	q = 1	q = 2	q = 3	q = 4	q = 5
p = 0	0	0	0	0	0	0
p = 1	0.0900	0.1608	0.2045	0.3023	0.3633	0.4252
p = 2	0.1707	0.2192	0.2449	0.3229	0.3732	0.5848
p = 3	0.2212	0.2393	0.6898	0.6879	0.3208	0.3601
p = 4	0.2665	0.2665	0.2539	0.4155	0.4102	0.7076
p = 5	0.2663	0.2685	0.7888	0.6905	0.2709	0.5736

TABLE 12.2. AIC and SIC of Log JPY/USD Rate

AIC	q = 0	q = 1	q = 2	q = 3	q = 4	q = 5
p = 0	-359.2032	-900.1638	-1229.3996	-1440.2734	-1548.9953	-1643.7729
p = 1	-1912.3681	-1911.4145	-1911.0683	-1910.9277	-1909.4812	-1908.0462
p = 2	-1911.5457	-1911.2376	-1910.1545	-1909.146	-1907.5734	-1907.5175
p = 3	-1911.3974	-1909.9885	-1914.6008	-1913.0301	-1905.2847	-1903.5307
p = 4	-1910.4738	-1908.4732	-1906.2927	-1915.9402	-1914.1297	-1912.9009
p = 5	-1908.4775	-1906.4963	-1911.3525	-1909.1295	-1903.9339	-1909.5526

SIC	q = 0	q = 1	q = 2	q = 3	q = 4	q = 5
p = 0	-346.6819	-883.4687	-1208.5306	-1415.2307	-1519.7788	-1610.3826
p = 1	-1895.6729	-1890.5455	-1886.0256	-1881.7112	-1876.0909	-1870.4821
p = 2	-1890.6767	-1886.1949	-1880.938	-1875.7557	-1870.0093	-1865.7796
p = 3	-1886.3547	-1880.772	-1881.2106	-1875.466	-1863.5469	-1857.6191
p = 4	-1881.2572	-1875.083	-1868.7286	-1874.2024	-1868.218	-1862.8155
p = 5	-1875.0872	-1868.9322	-1869.6147	-1863.2179	-1853.8485	-1855.2934

Box test suggests we cannot reject white noise residuals as long as the model include an AR term.

Conclusion from Box, AIC and SIC: AR(1) with trend.

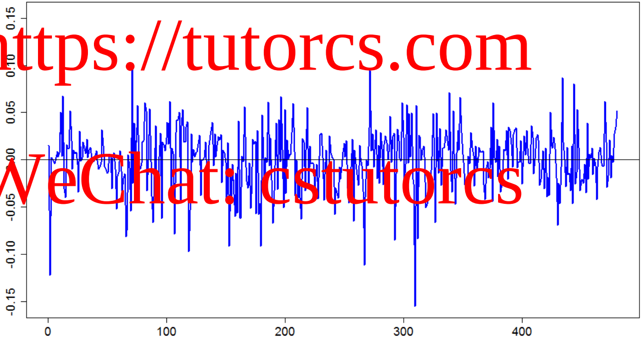
$$y_t = \beta_0 + \gamma_1 y_{t-1} + \beta_1 t + \epsilon_t$$

Table 12.3. Log JPY/USD Rate, best-fitting AR(1) with deterministic trend model

Variable	Coefficient	Standard Errors
Intercept	5.6357	0.1152
AR1	0.9791	0.0086
Trend	-0.0026	0.0004

12.4. **Residual check.** The residuals show no strong persistence.

FIGURE 12.4. Log JPY/USD Rate, best-fitting AR(1) with deterministic trend model, residual plot



12.5. **Forecast.** Note that the forecast exchange rate is decreasing with forecast horizon, essentially converging to the linear trend.

Figure 12.5. Log JPY/USD Rate, History and Forecast, AR(1) in levels with linear trend

FIGURE 12.6. Log JPY/USD Rate, History and Long-horizon Forecast, AR(1) in levels with linear trend

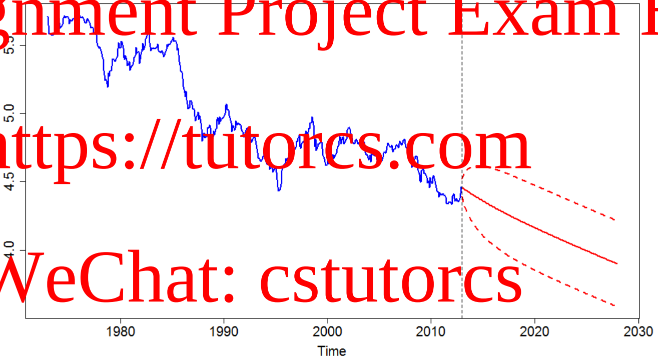
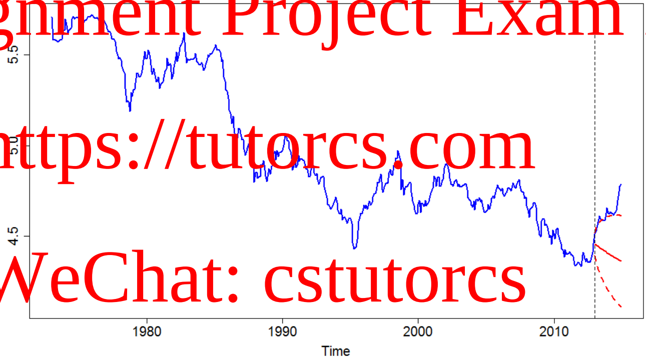


FIGURE 12.7. Log JPY/USD Rate, History, Forecast, and Realization, AR(1) in levels with linear trend



13. Augmented Dickey-Fuller Unit Root Test

$$\Delta y_t = \gamma y_{t-1} + \sum_{j=1}^{k} \Delta y_{t-j} + \epsilon_t$$

$$\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{j=1}^{k} \Delta y_{t-j} + \epsilon_t$$

$$\Delta y_t = \alpha + \gamma y_{t-1} + \beta t + \sum_{j=1}^{k} \Delta y_{t-j} + \epsilon_t$$

Table 13.1. Unit Root Test against the alternative of AR(1)

Variable	Coef.	Std.E.	Coef.	Std.E.	Coef.	Std.E.
Intercept			2.18E-02	1.87E-02	1.06E-01	5.09E-02
y_{t-1}	-5.49E-04	3.00E-04	-4.89E-03	3.74E-03	-1.95E-02	9.04E-03
Trend					-4.66E-05	2.62E-05
DF test						
t-stat	-1.8338		-1.3059		-2.159	
p-value	0.06761		0.5724		0.5108	

TABLE 13.2. Unit Root Test against alternative of AR(2)

Variable	Coef.	Std.E.	Coef.	Std.E.	Coef.	Std.E.
Intercept			1.70E-02	1.86E-02	1.08E-01	5.05E-02
y_{t-1}	-4.70E-04	2.97E-04	-3.85E-03	3.71E-03	-1.97E-02	8.96E-03
Δy_{t-1}	3.92E-02	4.52E-02	4.06E-02	4.52E-02	4.90E-02	4.53E-02
Trend					-5.03E-05	2.60E-05
DF test						
t-stat	-1.5836		-1.0377		-2.1949	
p-value	0.1116		0.6726		0.4956	

14. Differenced Data

14.1. Preliminary data analysis.

- First, we will do some preliminary data analysis.
  - plot time series data
  - ACF and PACF

FIGURE 14.1. Plot of Differenced Log JPY/USD Rate (1971-2012)

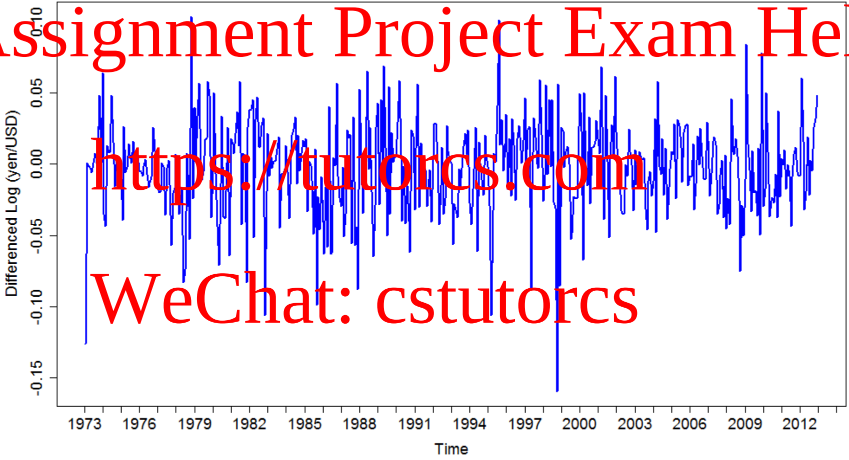
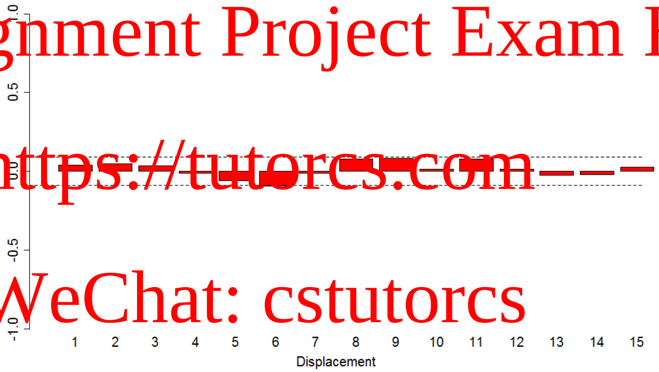


FIGURE 14.2. ACF of Differenced Log JPY/USD Rate

Figure 14.3. PACF of Differenced Log JPY/USD Rate



14.2. **Box-Pierce Test of White Noise.** Here, we adopt the simple Box-Pierce Test on the first difference of log exchange rate:

$$Q_{BP} = \sum_{k=1}^{m} \hat{\rho}_k^2 \sim \chi^2(m)$$

Setting $m = 21 \approx \sqrt{480}$, we obtain
$\hat{Q}_{BP} = 29.8071$, df = 21, p-value =0.09593

The white noise null is rejected at 10% level of significance but not at 5% and 1%. This result suggests that we may choose ARIMA(0,1,0) model of the level data, or the ARMA(0,0) of the differenced data.

14.3. **Model selection.**

TABLE 14.1. p-value of Box Test (Differenced Log JPY/USD Rate)

	q = 0	q = 1	q = 2	q = 3	q = 4	q = 5
p = 0	0.0954	0.1511	0.1795	0.2326	0.2441	0.3609
p = 1	0.1574	0.1842	0.1945	0.1772	0.2865	0.5793
p = 2	0.1873	0.1761	0.0924	0.1488	0.1998	0.8029
p = 3	0.2049	0.2221	0.1552	0.1351	0.2187	0.1932
p = 4	0.2087	0.2986	0.2429	0.1652	0.1773	0.2379
p = 5	0.3536	0.508	0.8363	0.2547	0.2428	0.0497

TABLE 14.2. AIC and SIC of Differenced Log JPY/USD Rate

AIC	q = 0	q = 1	q = 2	q = 3	q = 4	q = 5
p = 0	-1910.1332	-1908.8196	-1908.0154	-1907.205	-1905.3261	-1904.5182
p = 1	-1908.8905	-1907.7687	-1906.5527	-1904.0224	-1903.7878	-1904.6694
p = 2	-1908.1568	-1908.0108	-1902.1975	-1900.9416	-1900.5459	-1907.0197
p = 3	-1906.7613	-1907.76	-1901.052	-1900.4727	-1904.151	-1899.698
p = 4	-1904.8048	-1903.9363	-1904.3559	-1900.1902	-1903.762	-1899.0594
p = 5	-1904.3889	-1904.0262	-1907.5	-1904.3722	-1899.1125	-1892.3261

SIC	q = 0	q = 1	q = 2	q = 3	q = 4	q = 5
p = 0	-1901.7857	-1896.2983	-1891.3202	-1886.3361	-1880.2834	-1875.3017
p = 1	-1896.3692	-1891.0736	-1885.6837	-1878.9797	-1874.5713	-1871.2791
p = 2	-1891.4617	-1887.1419	-1877.1548	-1871.7251	-1867.1556	-1869.4556
p = 3	-1885.8924	-1882.7173	-1871.8355	-1867.0825	-1866.587	-1857.9601
p = 4	-1879.7621	-1874.7198	-1870.9656	-1862.6262	-1862.0241	-1853.1477
p = 5	-1875.1724	-1870.6359	-1869.9359	-1862.6343	-1853.2008	-1842.2407

Box test suggests we cannot reject white noise residuals for all ARMA combination, including ARMA(0,0) on the differenced data.

Conclusion from Box, AIC and SIC: ARIMA(0,1,0).

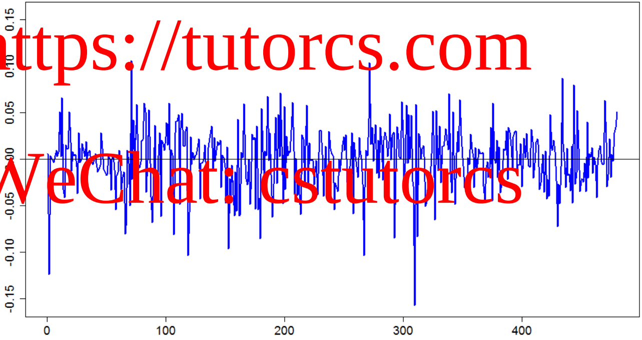
$$\Delta y_t = \alpha + \epsilon_t$$

TABLE 14.3. Regression results of Log JPY/USD Rate, ARIMA(0,1,0).

Variable	Coefficient	Standard Errors
Intercept	-0.0026	0.0015

14.4. **Residual check.** The residuals look like white noise and show no strong persistence.

Figure 14.4. Log JPY/USD Rate, ARIMA(0,1,0) residual plot



14.5. **Forecast.** Note the wider forcast interval when compared with the one based on ARMA model of log exchange rate.

FIGURE 14.5. Log JPY/USD Rate, History and Forecast, ARIMA(0,1,0)

FIGURE 14.6. Log JPY/USD Rate, History and Long-horizon Forecast, ARIMA(0,1,0)

FIGURE 14.7. Log JPY/USD Rate, History, Forecast, and Realization, ARIMA(0,1,0)

15. Concluding remarks:

The two sets of forecasts based on the two different assumption (without and with unit roots) do not differ much. This is not suprising because if there is indeed unit root, the AR(1) model should be able to approximate it well. And, indeed, the AR(1) coefficient is very close to one.

16. R packages used

16.1. Package "forecast".

```
Arima(x, order=c(0,0,0), seasonal=c(0,0,0), xreg=NULL,
include.mean=TRUE, include.drift=FALSE, include.constant,
lambda=model$lambda, transform.pars=TRUE, fixed=NULL,
init=NULL, method=c("CSS-ML","ML","CSS"), n.cond,
optim.control=list(), kappa=1e6, model=NULL)
```

```
forecast(object, h=ifelse(object$arma[5]>1,2*object$arma[5],10),
level=c(80,95), fan=FALSE, xreg=NULL, lambda=object$lambda,
bootstrap=FALSE, npaths=5000, ...)
```

16.2. Package "fUnitRoots".

```
adfTest(x, lags = 1, type = c("nc", "c", "ct"), title = NULL,
description = NULL)
```