04 Time series with unit root
Part II

Andrius Buteikis, andrius.buteikis@mif.vu.lt
http://web.vu.lt/mif/a.buteikis/

Recap: Time series Types

Time series can be divided into two categories

- ▶ **T**rend **S**tationary (**TS**) series:

  $\phi(L)Y_t = \mu_t + w_t$, where $\mu_t = T_t + S_t$, w_t - stationary, e.g. $ARMA(p, q)$

  - ▶ The mean component is **deterministic**;
  - ▶ The original series, Y_t, is non-stationary (if $\mu_t \neq const$). Once we estimate and remove the deterministic component, the residuals $\widehat{e}_t = Y_t - \widehat{\mu}_t$ are **stationary**.
  - ▶ **TS** series revert to the trend in the long run (the effects of shocks are eventually eliminated). Forecast intervals have constant width.

- ▶ **D**ifference **S**tationary (**DS**) series:

  $$\phi(L)(1 - L)^d Y_t = \alpha + \theta(L)\epsilon_t$$

  - ▶ The mean component is **stochastic**;
  - ▶ The original series, Y_t, is non-stationary. Furthermore, estimation and removal of a deterministic component does not produce a stationary series. On the other hand, **Differencing** produces a stationary series.
  - ▶ **DS** series never recover from shocks to the system (the effects of shocks are permanent). Forecast intervals grow over time.

Recap: The ARIMA(p, d, q) Process

If the series of dth differences is a stationary $\mathrm{ARMA(p,q)}$ process, then the series Y_t is the dth order integrated series and denoted by $I(d)$.

Then, we say that Y_t is described by an autoregressive *integrated* moving average model of order *p, d, q*, i.e. an **ARIMA(p,d,q)** model:

$$\Phi(L)(1-L)^d Y_t = \alpha + \Theta(L)\epsilon_t$$

The following are variations, which we can use to describe the process[1]:

- $\Delta^d Y_t$ is described by a **stationary** $\mathrm{ARMA(p,q)}$ process;
- Y_t is described by a (*stationary*) $\mathrm{ARIMA(p,d,q)}$ process;
- The $\mathrm{ARMA(p+d,q)}$ process Y_t has a **unit root** of multiplicity d.

The symbol *I(0)* is used to denote a *stationary* series.

---

[1][Click here for a quick summary of ARIMA models]

Recap: Rewriting an AR(p) Model for Differences

If we add and subtract Y_{t-1} and various combinations of ϕ_j and Y_{t-j} from a process $Y_t \sim \mathrm{AR(p)}$:

$$Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \epsilon_t, \quad \epsilon_t \sim WN(0, \sigma^2)$$

we get an equivalent *AR(p)* model, written in differences:

$$\Delta Y_t = \alpha + \rho Y_{t-1} + \gamma_1 \Delta Y_{t-1} + ... + \gamma_{p-1} \Delta Y_{t-p+1} + \epsilon_t$$

where:

$$\rho = \phi_1 + ... + \phi_p - 1$$
$$\gamma_j = -[\phi_{j+1} + ... + \phi_p], \quad j = 1, ..., p-1$$

Note that both equations are AR models. Then:

▶ $\rho = 0$ implies that the $\mathrm{AR(p)}$ time series **contains** a unit root.
▶ $-2 < \rho < 0$ implies that the $\mathrm{AR(p)}$ series is stationary.

If $\rho = 0$, then our equation involves only differences of Y_{t-j} - i.e. an $\Delta Y_t \sim \mathrm{AR(p-1)}$ process:

$$\Delta Y_t = \alpha + \gamma_1 \Delta Y_{t-1} + ... + \gamma_{p-1} \Delta Y_{t-p+1} + \epsilon_t$$

Remarks

- We say that the $\mathrm{ARMA}(p+1, q)$ process Y_t has a **unit root** if **at least one** of the $p+1$ autoregressive roots $\phi(x) = 0$ equals 1.

- A drifting unit root series exhibits trend behavior.

- Unfortunately, for any finite amount of data there is a deterministic and stochastic trend that fits the data equally well (Hamilton, 1994).[2]

- In order to asses the presence of a stochastic trend in an observed series one should carry out unit root tests.

- Remember that a stationary $\mathrm{ARMA}(p, q)$, or a stationary $ARIMA(p, d, q)$ can be written as an infinite order polynomial, $\mathrm{AR}(\infty)$, which can be approximated by a finite order $\mathrm{AR}(k)$ process.

- The popular `Dickey-Fuller` unit root test always interprets the process under consideration as an AR process, which can be generalized to

$$\Delta Y_t = \alpha + \delta t + \rho Y_{t-1} + \gamma_1 \Delta Y_{t-1} + ... + \gamma_{p-1} \Delta Y_{t-p+1} + \epsilon_t$$

which we can use to test the hypothesis for a unit root, $H_0 : \rho = 0$.

---

[2]Hamilton, J. D. *Time Series Analysis*. Princeton, NJ: Princeton University Press, 1994.

AR(p) order selection for ΔY_t

Sequential procedure (based on coefficient significance):

1. Estimate an $AR(p_{\max})$ (begin with a fairly high order, say $p_{\max} \approx 5$):

$$\Delta Y_t = \alpha + \delta t + \rho Y_{t-1} + \gamma_1 \Delta Y_{t-1} + ... + \gamma_{p_{\max}} \Delta Y_{t-p_{\max}} + \epsilon_t$$

2. If the $p_{\max}$-th lag coefficient, $\phi_{p_{\max}}$, is insignificant, reduce the model order and estimate $AR(p_{\max} - 1)$;
3. If the $p_{\max} - 1$-th lag coefficient, $\phi_{p_{\max}-1}$, is insignificant, further reduce the model lag order to $p_{\max} - 2$ etc. until the largest lag is significant.

A useful rule of thumb for determining $p_{\max}$, suggested by Schwert $(1989)^3$, is

$$p_{\max} = \left\lfloor 12 \cdot \left(\frac{T}{100} \right)^{1/4} \right\rfloor$$

where $\lfloor \cdot \rfloor$ denotes the integer part (i.e. the *floor* value).
Alternatively, instead of significance, we could use the information criteria, like AIC, or BIC.

---

[3]Schwert, W. (1989). *Test for Unit Roots: A Monte Carlo Investigation*, Journal of Business and Economic Statistics, 7, 147-159.

For example, examine the connection between Y_t and ΔY_t process lag order - it is easy to verify that the process

$$Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \delta t + w_t$$

can be expressed as

$$\Delta Y_t = \alpha + (\phi_1 + \phi_2 - 1)Y_{t-1} - \phi_2 \Delta Y_{t-1} + \delta t + \epsilon_t$$

where $\phi_1 + \phi_2 - 1 =: \rho$.

Testing for the significance of the trend coefficient, δ

4. Once the right lag order is determined, look at the trend coefficient $\widehat{\delta}$ - if it is insignificant - drop it from the model.

Unit Root Hypothesis

Next, the null hypothesis:

$$H_0 : \text{ the process has a unit root}$$

can be restated as:

$$H_0 : \rho = 0$$

Augmented Dickey-Fuller (ADF) test

1. Determine the AR lag order for ΔY_t;
2. Determine whether the trend, δ, is significant;
3. Test the null hypothesis $H_0 : \rho = 0$ by checking whether $\widehat{\rho}$ is significant. The significance of $\widehat{\rho}$ is tested in a different ways:

   3.1 If the final model contains the term $\delta \cdot t$ - the 5% Dickey-Fuller critical value is ≈ -3.45. If the t-ratio of the term Y_{t-1}, (**NOT** ΔY_{t-1}) is **less** than -3.45, then the null hypothesis H_0 : the process has a unit root is **rejected** and we conclude that the AR process is **stationary**.

   3.2 If the final model does **not** contain a trend term, the Dickey-Fuller critical value ≈ -2.89. If the t-statistic of $\widehat{\rho}$ is **more negative** than -2.89, we **reject the null hypothesis of a unit root** and conclude that the series is stationary.

Note: the ADF test removes all the structural effects (autocorrelation) in the time series and then tests using the same procedure as the original Dickey-Fuller (DF) test, which only considered cases when Y_t was an $\mathrm{AR}(1)$ process.

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests

An alternative test for unit root testing. The main differences from the ADF are:

- ▶ The null hypothesis is for the *absence* of a unit root -
  $H_0 : H_0 :$ the process does not have a unit root;
- ▶ The absence of a unit root is not a proof of stationarity, but of **trend-stationarity**.

Dickey-Fuller test is based on a null hypothesis that a unit root *exists*, which biases results towards more differencing of the series.

Philips-Perron (PP) test

It builds on the Dickey-Fuller test:

- ▶ Like the ADF test, the PP test addresses the issue that the Y_t process might have a higher order autocorrelation;
- ▶ While the ADF introduces additional lags of ΔY_{t-j}, the PP test instead makes a non-parametric correction to the $t$test statistic.
- ▶ The test is robust with respect to unspecified autocorrelation and heteroscedasticity in the disturbance process of the test equation.

- **KPSS**-type tests are intended to *complement* unit root tests, such as the ADF tests. By testing **both** the unit root hypothesis and the stationarity hypothesis, one can distinguish series that appear to be stationary, series that appear to have a unit root, and series for which the data are not sufficient to be sure whether they are stationary or integrated.
- According to Davidson and MacKinnon (2004)[4], the **PP** test performs worse in finite samples than the ADF test (we examine whether this always holds true in the next slide).

There are a number of ways to carry out the tests in R:

- ADF: using `dynlm::dynlm(...)` to estimate the models manually and carry out an ADF unit root test;
- ADF: using `urca::ur.df(...)` or `tseries::adf.test(...)` to carry out the ADF test of unit root;
- KPSS: using `urca::ur.kpss(...)` (tau - trend, mu - drift) or `tseries::kpss.test(...)` to carry out the KPSS test of trend stationarity;
- PP: using `stats::PP.test(...)`, `urca::ur.pp()` or `tseries::pp.test(...)` to carry out the PP test of unit root;

---

[4]Davidson, Russell, MacKinnon, James G. (2004). *Econometric Theory and Methods*. New York: Oxford University Press. p. 623.

Test Comparison for simulated data

Proportion of unit root tests, which indicated unit root, based on AR(1) coefficient values for sample size N = 100 with MC = 1000 samples generated
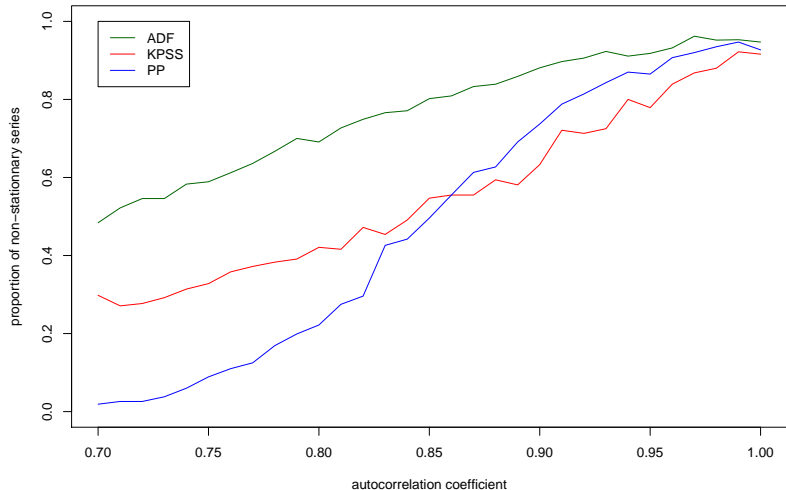
The ADF test behaves quite poorly for smaller samples (from `tseries` package) - it concluded that roughly 50% of samples exhibit a unit root, when the $\mathrm{AR}(1)$ coefficient $\phi_1 = 0.7$.

Recap: Seasonal ARIMA models

A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models we have seen so far: $SARIMA(p, d, q)(P, D, Q)_S$:

$$\Phi(L^S)\phi(L)(1 - L)^d(1 - L^S)^D(Y_t - \mu) = \Theta(L^S)\theta(L)\epsilon_t$$

where d is the integration order of the non-seasonal part of the model, and D is the integration order of the seasonal part of the model and:

▶ The non-seasonal components are:

$$AR: \phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$$
$$MA: \theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$$

▶ The seasonal components are:

$$\text{Seasonal AR: } \Phi(L^S) = 1 - \Phi_1 L^S - ... - \Phi_p L^{S \cdot P}$$
$$\text{Seasonal MA: } \Theta(L^S) = 1 + \Theta_1 L^S + ... + \Theta_q L^{S \cdot Q}$$

See the previous lecture slides - they contain an example of how either only differencing, or only seasonally differencing the series still results in a presence of a unit root.

Part II: Box-Jenkins method

Box-Jenkins modelling involves identifying an appropriate ARIMA process, fitting it to the data, and then using the fitted model for forecasting.

Since ARIMA model processes are a very rich class of models, which usually make it is possible to find an adequate model to fit the data, the Box-Jenkins modelling approach is quite attractive.

It is recommended to modify the original Box-Jenkins iterative three-step procedure by adding a preliminary data preparation stage and a final model forecasting stage.

Step (1): Data Preparation

This step involves variable transformations and differencing.
In order to determine, whether transformations are necessary,
visualizations may be employed:

- ▶ Time series plots, i.e. the run-sequence plots, which can help identify
  trend, seasonality and heteroskedasticity;
- ▶ ACF plots should be examined - non-stationarity is often indicated
  by an autocorrelation plot with very slow decay.

From the plots, we may determine which transformations and differencing
are needed:

- ▶ Variable transformations (logarithms, square roots, etc.) can be used
  to stabilize the variance (e.g. if a series exhibits exponential growth);
- ▶ The data are (non-seasonally and/or seasonally) differenced until
  there are no obvious trend or seasonality patterns left.

After we have verified that the transformed data is stationary, we more
on to the second step.

Step (2): Model Selection

Once stationarity and seasonality have been addressed, the next step is to identify the order (i.e. the lag orders p and q) of the autoregressive and moving average terms which might provide a good fit to the data. To select the order, we do the following:

- ▶ Examine the ACF and PACF plots of the **transformed** series to determine a possible lag order;
- ▶ Try various combinations of p and q and select the combination, which produces the smallest AIC, or BIC criterion value.

Step (3): Parameter Estimation

The main approaches to fitting Box-Jenkins models are nonlinear least squares or maximum likelihood estimation (which is usually preferred). The parameter estimation procedures are usually implemented in statistical software and are readily available.

Step (4): Model Diagnostics

This step involves testing the model assumptions to identify whether the model in question is adequate. Usually this step involves the error term - the residuals should be a *white noise* process - constant mean, constant variance and no autocorrelation. this is usually done via time series plots, ACF and PACF plots as well as autocorrelation significance tests, like the $\mathrm{Jlung - Box}$ test.

If these assumptions are not satisfied, one needs to go back to Step (2) and select a different model (this is assuming that the transformations in Step (1) were adequate).

Step (5): Forecasting

This is usually not only just the last step, but also the goal of the model. Once the model is selected, estimated and verified to be adequate - one should compute the forecasts.

Cross-validation (out-of-sample testing)

Furthermore, in order to asses whether the forecasts are adequate for the data, we can carry out cross validation on the data. For example:
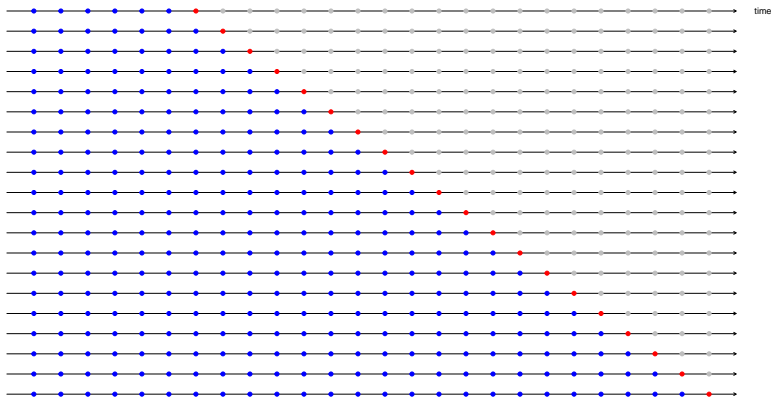

In this chart

▶ the blue observations form the training sets;

▶ the red observations form the test sets (one observation per set);

The forecast accuracy is computed by averaging over the test sets.

In short, we repeat Steps (3) and (5) on smaller subsets, by **re-estimating the model on each subset and calculating a one-step ahead forecast**. These tests not only verify the robustness of your model, but also allow examining whether the parameters are stable (i.e. whether their signs and magnitude remain unchanged throughout different samples).

- ▶ In this procedure, we use the initial dataset to create a series of testing sets. The corresponding training set consists only of observations that occurred prior to the observation that forms the test set. Thus, no future observations can be used in constructing the forecast.

- ▶ With time series forecasting, one-step forecasts may not be as relevant as multi-step forecasts. In this case, the cross-validation procedure can be modified so that the test sets include more than one observation.

Example: 4-step-ahead forecast cross validation.
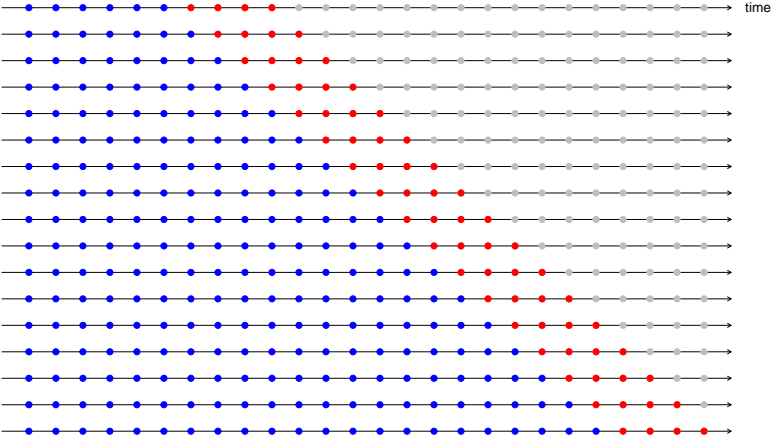
The underlying Box-Jenkins methodology provides a convenient framework, which is applicable to a wide variety of statistical modelling applications. It allows one to think about the data and find appropriate statistical model(-s) which can be used to help answer various relevant questions about the data.

k- fold Cross validation on cross-sectional data

The original sample is randomly partitioned into k equal sized subsamples, with one subsample acting as the *testing/validation* set and $k - 1$ subsamples acting as the *training set*. The process is then repeated k times, with each of the k subsamples used exactly *once* as the validation data.