Applied Time Series Econometrics

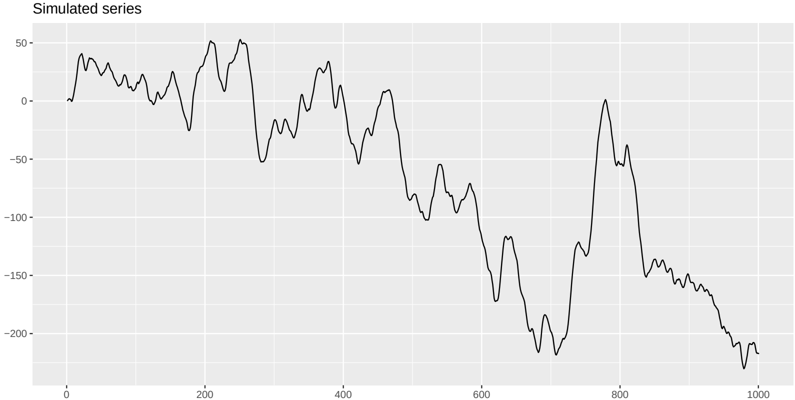
Oxana Malakhovskaya, NRU HSE

October 15, 2019

Simulated series

Simulate $y3$:

$$y_t = 2.3y_{t-1} - 1.7y_{t-2} + 0.4y_{t-3} + \nu_t$$
$$\Delta y_t = 1.3\Delta y_{t-1} - 0.4\Delta y_{t-2} + \nu_t$$



tseries::adf.test function: output

```
library(tseries)
adf.test(y3, k = 10)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  y3
## Dickey-Fuller = -3.3992, Lag order = 10, p-value = 0.053
## alternative hypothesis: stationary
```

- The adf.test function modifies the data before performing the ADF test. First it automatically detrends the data; then it recenters the data, giving it a mean of zero.
- If either detrending or recentering is undesireable for the application in hand, we use `ur.df()` instead.

urca::ur.df function: output

```r
library(urca)
urtest1 <- ur.df(y3, lags = 10, type = 'trend')
summary(urtest1)
```

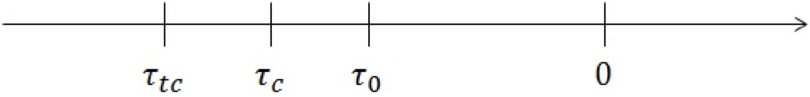
Value of test-statistic is: -2.3431 2.1577 2.7919

Critical values for test statistics:

	1pct	5pct	10pct
tau3	-3.96	-3.41	-3.12
phi2	6.09	4.68	4.03
phi3	8.27	6.25	5.34

Deterministic terms in ADF-test: why do we care?

- The limiting distribution depends on the included deterministic terms.
- Different critical values are used when a constant and/or linear trend are included.
- Seasonal dummies do not modify the limiting distribution.



τ_{tc} - critical value for a model when both a trend and a constant is included

τ_c - critical value for a model when a constant is included

τ_0 - critical value for a model when neither trend nor constant are included

Deterministic terms in ADF-test: how to choose?

How do we determine which type of the model to test?

1. Using economic reasoning

If a linear trend is assumed in the DGP for y_t, only a constant should be added as a regressor in a model for Δy_t because:

$$\text{if } y_t = \mu_0 + \mu_1 t + x_t \text{ then } \Delta y_t = \mu_1 + \Delta x_t$$

Similarly, if just a constant is assumed in the DGP for y_t, then the model for Δy_t is tested without any deterministic terms

$$\text{if } y_t = \mu_0 + x_t \text{ then } \Delta y_t = \Delta x_t$$

2. According to an iterative procedure

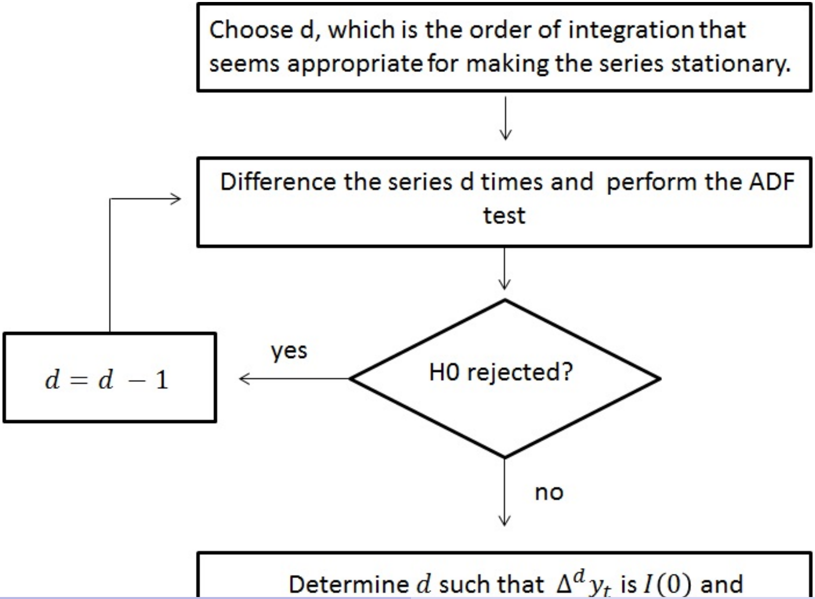
There is a number of different iterative procedures (e.g. Dolado, Jenkinson and Sosvilla-Rivero(1990)) to determine which deterministic terms to include. These procedures are not popular now.

AR order choice

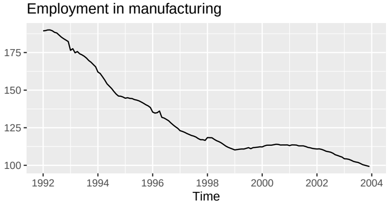
- The choice of the number of lags may be based on information criteria or a sequental testing procedure may be used that eliminates insignificant coefficients sequentially starting from some high-order model.
- `adf.test()` does not have an option of selecting a lag according to the criteria. If we do not define the number of lags (k) explicitely, it is computed automatically as $k = trunc(T-1)^{1/3}$.
- `ur.df()` has an option to take AIC or BIC into account. In this case the option `lags` determines the maximum number of lags considered.

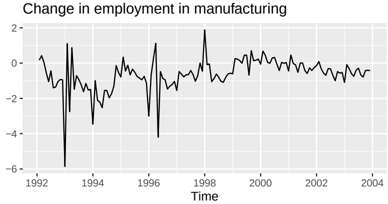
```
ur.df(y3, type = 'trend', lags = 12, selectlags = "BIC")
```

Order of integration: Pantula(1989) principle



Unit root testing:example





- Assume we are not sure that $\Delta empl$ is stationary and we take $d = 2$.
- If $empl_t$ has a linear trend and a constant, then $\Delta empl_t$ has only a constant and $\Delta^2 empl_t$ has neither trend nor constant.

Unit root testing:example(2)

```r
model1 <- ur.df(diff(empl,differences = 2), type = "none",
                lags = 6, selectlags = "BIC")
summary(model1)
```

Value of test-statistic is: -10.67

Critical values for test statistics:

	1pct	5pct	10pct
tau1	-2.58	-1.95	-1.62

We reject H_0 even at 1% significance level.

Unit root testing:example(3)

We do the ADF test for $\Delta empl_t$ in a model with drift.

```
model2 <- ur.df(diff(empl), type = "drift",
                 lags = 6, selectlags = "BIC")
summary(model2)
```

Value of test-statistic is: -2.7405 3.7641

Critical values for test statistics:

	1pct	5pct	10pct
tau2	-3.46	-2.88	-2.57
phi1	6.52	4.63	3.81

We do not reject H_0 at 5% significance level.

KPSS test: the idea of the test

- Another possibility for investigating the integration properties of a series y_t is to test the H_0 that DGP is stationary.
- Assume that there is no linear trend term, the point of departure is a process as follows:

$$y_t = x_t + z_t,$$

where x_t is a random walk, $x_t = x_{t-1} + \upsilon_t, \upsilon \sim iid(0, \sigma_\upsilon^2)$ and z_t is a stationary process.

$$H_0 : \sigma_\upsilon^2 = 0$$
$$H_1 : \sigma_\upsilon^2 > 0$$

- If H_0 holds,y_t is composed of a constant and the stationary process z_t; hence, y_t is stationary.

KPSS test: statistics

Kwiatkowski et al.(1992) proposed the following test statistics:

$$KPSS = \frac{1}{T^2} \sum_{t=1}^{T} \frac{S_t^2}{\hat{\sigma}_\infty^2},$$

where $S_t = \sum_{j=1}^{t} \hat{w}_j$ with $\hat{w}_j = y_j - \bar{y}$ and $\hat{\sigma}_\infty^2$ is an estimator of

$$\sigma_\infty^2 = lim_{T \to \infty} T^{-1} Var \left(\sum_{t=1}^{T} z_t \right)$$

If y_t is I(1), the numerator will grow without bounds, causing the statistic to become large for large sample sizes.

KPSS: test statistics computation

- Kwiatkowski et al.(1992) proposed a non-parametric estimator of σ^2_∞ as follows:

$$\hat{\sigma}^2_\infty = \frac{1}{T}\sum_{t=1}^{T}\hat{w}_t^2 + 2\sum_{j=1}^{l_q}\nu_j\left(\frac{1}{T}\sum_{t=j+1}^{T}\hat{w}_t\hat{w}_{t-j}\right),$$

where l_q is a lag truncation parameter and it is equal to $(qT/100)^{1/4}$ and $\nu_j = 1 - \frac{j}{l_q+1}$. In applications q is usually equal 4 (for quarterly series) or 12 (for monthly series).

KPSS: estimation and possible trend in DGP

- If a deterministic trend is suspected in the data-generating process (DGP), the point of depature is a DGP, which includes such a term:

$$y_t = \mu_1 t + x_t + z_t$$

and the $\hat{w}_t$ are residuals from a regression

$$y_t = \mu_0 + \mu_1 t + w_t$$

- The limiting distribution of the test statistics under H_0 in this case is different from the case without a trend term.
- Anyway, the distribution is not standard, and tabulated critical values are used.

tseries:: kpss.test

To implement the KPSS test, we use the `kpss.test()` function from `tseries` package or the `ur.kpss()` from urca.

```
kpss.test(y3, null = "Trend", lshort =TRUE)
# null = "Trend" means we test a model with a liner
# deterministic trend, lshort = TRUE  means
#lower value of lag truncation parameter.
```

KPSS Test for Level Stationarity

data: y3

KPSS Level = 10.6, Truncation lag parameter = 7, p-value = 0.01

urca:: ur.kpss

`ur.kpss()` tests the model with the same options as follows:

```
kpss_y3 <- ur.kpss(y3, type = "tau", lags = "short")
summary(kpss_y3)
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: tau with 7 lags.
##
## Value of test-statistic is: 0.4226
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
```

KPSS test for a real series (1)

```r
kpss_empl0 <- ur.kpss(empl, type = "tau", lags = "long")
summary(kpss_empl0)
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: tau with 13 lags.
##
## Value of test-statistic is: 0.2753
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
```

KPSS test for a real series (2)

```r
kpss_empl1 <- ur.kpss(diff(empl), type = "mu",
                      lags = "long")
summary(kpss_empl1)
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 13 lags.
##
## Value of test-statistic is: 0.5882
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

KPSS test for a real series (3)

```r
kpss_empl2 <- ur.kpss(diff(diff(empl)), type = "mu",
                      lags = "long")
summary(kpss_empl2)
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 13 lags.
##
## Value of test-statistic is: 0.0776
##
## Critical value for a significance level of:
##                  10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

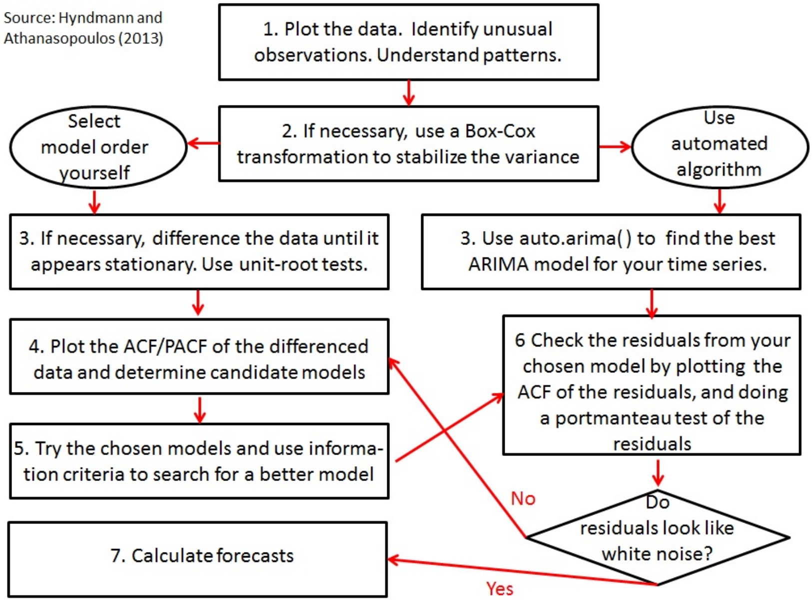
KPSS and ADF tests

- Idealy, if a series y_t is I(0), the ADF test should reject the nonstationarity null hypothesis, whereas the KPSS test should not reject the stationarity null hypothesis.
- Such an ideal result is not always obtained for various reasons.

Results of H_0 testing for the `Employment in Manufacturing` series.

Series	ADF	KPSS
empl	-	Reject
$\Delta empl$	Not reject	Reject
$\Delta^2 empl$	Reject	Not reject

Algorithm of working with TS (see also PS7)



Seasonal ARIMA models: definition

- In some cases we have to combine both seasonal and non-seasonal autoregressive terms or/and moving average terms or/an differences.

$$ARIMA(p, d, q)(P, D, Q)_m \quad (1)$$

(p, d, q)- non-seasonal part of the model $(P, D, Q)_m$ - seasonal part of the model

Seasonal ARIMA models: example

For example, $ARIMA(1,1,1)(1,1,1)_4$ means that the DGP is as follows:

$$\underbrace{(1-\phi_1 L)}_{A}\underbrace{(1-\Phi_1 L^4)}_{B}\underbrace{(1-L)}_{C}\underbrace{(1-L^4)}_{D}y_t = \underbrace{(1+\psi_1 L)}_{E}\underbrace{(1+\Psi_1 L^4)}_{F}\nu_t$$

A: non-seasonal AR(1)

B: seasonal AR(1)

C: non-seasonal difference

D: seasonal difference

E: non-seasonal MA(1)

F: seasonal MA(1)

Seasonal ARIMA models: example(2)

If we multiply all the factors out:

$$(1-\phi_1 L-\Phi_1 L^4+\phi_1\Phi_1 L^5)(1-L-L^4+L^5)y_t=$$
$$=(1+\psi_1 L+\Psi_1 L^4+\psi_1\Psi_1 L^5)\nu_t$$
$$(1-(\phi_1+1)L+\phi_1 L^2-(1+\Phi_1 L^4)-(\phi_1\Phi_1-\Phi_1-\phi_1-1)L^5-$$
$$-(\phi_1\Phi_1+\phi_1)L^6+\Phi_1 L^8-(\phi_1\Phi_1+\Phi_1)L^9+\phi_1\Phi_1 L^{10})y_t=$$
$$=(1+\psi_1 L+\Psi_1 L^4+\psi_1\Psi_1 L^5)\nu_t$$
$$y_t=(\phi_1+1)y_{t-1}-\phi_1 y_{t-2}+(1+\Phi_1)y_{t-4}-(\phi_1\Phi_1+\Phi_1+\phi_1+$$
$$+1)y_{t-5}+(\phi_1\Phi_1+\phi_1)y_{t-6}-\Phi_1 y_{t-8}+(\phi_1\Phi_1+\Phi_1)y_{t-9}-$$
$$-\phi_1\Phi_1 y_{t-10}=\nu_t+\psi_1\nu_{t-1}+\Psi_1\nu_{t-4}+\psi_1\Psi_1\nu_{t-5}$$

ACF and PACF of seasonal ARIMA

The seasonal part of an AR or MA model will be seen in the seasonal lags of the ACF and PACF.

$ARIMA(0,0,0)(0,0,1)_{12}$ will show:

- a spike at lag 12 in the ACF but no other significant spikes.
- exponential decay at seasonal lags (that is, at lags 12, 24...) in the PACF.

$ARIMA(0,0,0)(1,0,0)_{12}$ will show:

- exponential decay in the seasonal lags of the ACF.
- a single significant spike at lag 12 in the PACF.

Forecasting: an algorithm

Point forecast (see Pratctice session 7)

1. Rearrange ARIMA so y_t is on LHS and all other terms on RHS.
2. Rewrite the equation by replacing t by $T + h$
3. On the RHS, replace future observarions by their forecasts, future errors by zero, and past errors by corresponding residuals.

Q: What changes in case of nonstationary series forecast?

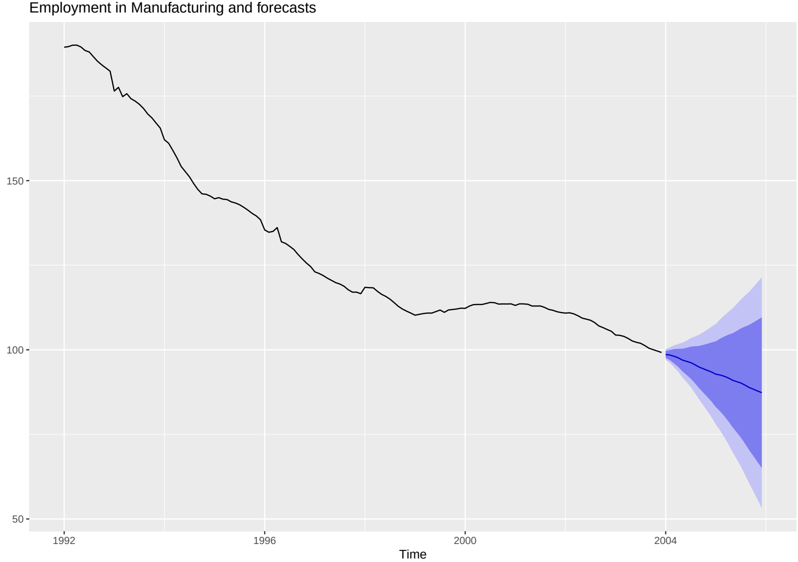
A: Nothing but the representation!

Example: ARIMA(p,1,q)

$$\begin{aligned}\Delta y_t &= \hat{\mu} + \hat{\phi}_1 \Delta y_{t-1} + \hat{\phi}_2 \Delta y_{t-2} + \cdots + \hat{\phi}_p \Delta y_{t-p} + \nu_t + \\ &\quad + \hat{\psi}_1 \nu_{t-1} + \cdots + \hat{\psi}_q \nu_{t-q} \\ y_t &= \hat{\mu} + (\hat{\phi}_1 + 1) y_{t-1} - (\hat{\phi}_1 - \hat{\phi}_2) y_{t-2} - \cdots - (\hat{\phi}_{p-1} - \hat{\phi}_p) y_{t-p} - \\ &\quad - \hat{\phi}_p y_{t-p-1} + \nu_t + \hat{\psi}_1 \nu_{t-1} + \cdots + \hat{\psi}_q \nu_{t-q}\end{aligned}$$

Example with Seasonal ARIMA

See code.



Literature

1. Hyndman, R. J. and Athanasopoulos, G. (2013) Forecasting: principles and practice, Otexts
2. Lütkepohl, H., and Krätzig, M. (2004) Applied Time Series Econometrics, New York, Cambridge University Press
3. Martin, V., Hurn, S., and HarriS,D. (2013) Econometric Modelling with Time Series: Specification, Estimating and Testing, New York, Cambridge University Press