Time Series: Week 5 Lecture Slides

Inference with units roots

- So far we have discussed stochastic trends, unit roots and how to transform your data.
- But not how to test for unit roots and how to do inference in their presence.
- Goal:
  - For you to know that it is possible to conduct inference in the presence of unit roots and have some basic building blocks to do so.
  - More important, how to **test for unit roots** so you know whether your data is stationary – much more important in practise.
- In the end this is not a statistics course, you do not need to know the details of the FCLT and Brownian motions.

Inference in models with trends

- Recall: detrend or keep the trend but change the inference
  - Deterministic trends $\Rightarrow$ detrend

$$y_t = \text{Time Trend} + \underbrace{\text{Stationary Component}}_{\text{ARMA}}$$

$$\text{Time Trend} \begin{cases} \text{Nonlinear trend (HP Filter)} \\ \text{Polynomial trend} \end{cases}$$

  - Stochastic trends $\Rightarrow$ detrend by taking the difference (=growth rates) $\Rightarrow$ model Δy_t which is stationary $\Rightarrow$ standard inference (ARIMA model).
  - If you want to keep trend inference non standard.

Inference in models with unit roots I

- AR(1):

$$y_t = \phi y_{t-1} + \varepsilon_t$$
$$y_0 = 0$$

  where $\varepsilon_t \sim i.i.d.\mathcal{N}(0, \sigma^2)$.
- Estimator of ϕ is

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

- "Discontinuity" in asymptotic distribution:
  - if $|\phi| < 1 \Rightarrow \sqrt{T}(\hat{\phi} - \phi) \to \mathcal{N}(0, 1 - \phi^2)$.
  - If $\phi = 1 \Rightarrow T(\hat{\phi} - 1) \to \frac{\frac{1}{2}[W(1)^2 - 1]}{\int_0^1 W(r)^2 dr}$ where $W(t)$ is a Brownian motion.

Inference in models with trends II

- Again:
  - if $|\phi| < 1 \Rightarrow \sqrt{T}(\hat{\phi} - \phi) \rightarrow \mathcal{N}(0, 1 - \phi^2)$.
  - If $\phi = 1 \Rightarrow T(\hat{\phi} - 1) \rightarrow \frac{\frac{1}{2}[W(1)^2 - 1]}{\int_0^1 W(r)^2 dr}$ where $W(t)$ is a Brownian motion.
- Note that OLS estimator in both cases is consistent
- In fact, in the unit root case it is **super-consistent =** converges to the true value at rate T instead of the standard $\sqrt{T}$ rate that applies in the stationary case

Brownian motions

- **Definition**: A Brownian motion $W(r)$, $r \in [0, 1]$ is a **continuous** stochastic process such that:
  - W(0)=0.
  - For any $0 \leq r_1 < r_2 < ... < r_k \leq 1$ the **increments** $W(r_j) - W(r_{j-1})$ are **independent** and normally distributed, that is:

$$W(r_j) - W(r_{j-1}) \sim \mathcal{N}(0, r_j - r_{j-1})$$

FCLT instead of CLT

- LLN and CLT no longer valid. To derive asymptotic distributions in the presence of unit roots use instead the functional central limit theorem **(FCLT)**
- FCLT: Let $\varepsilon_t \sim iid(0, \sigma^2)$, then

$$\frac{\sqrt{T}\frac{1}{T}\sum_{t=1}^{[Tr]}\varepsilon_t}{\sigma} \to^d W(r)$$

for all $r \in [0, 1]$

$$\begin{aligned} [Tr] &= \text{integer part} \\ r &= \text{fraction of the total sample} \end{aligned}$$

- **Intuition**: comes from the fact that the increments of the partial sums $\sqrt{T}\frac{1}{T}\sum_{t=1}^{[Tr]}\varepsilon_t$ are sample means of ε's which by the CLT are normal $\Rightarrow \sqrt{T}\frac{1}{T}\sum_{t=1}^{[Tr]}\varepsilon_t$ behaves like a process with normal increments $\to$ Brownian motion

Unit root tests

- In practice, a unit root test is one of the first thing you do with each time series (always plot them first!)
- Simplest, most common test is the Augmented Dickey Fuller (ADF) test
- Simplest case

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

- $H_0 : \phi = 1$ against $H_1 : |\phi| < 1$

$$t_{\text{ADF}} = T(\hat{\phi} - 1) \rightarrow \frac{\frac{1}{2}[W(1)^2 - 1]}{\int_0^1 W(r)^2 dr}$$

under $\mathrm{H}_0 \rightarrow$ the critical value is the quantile (e.g. 95%) of the function of Brownian motion, which is obtained by simulation

Unit root test for AR(p) model

- More generally, for an AR(p)

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t$$

- Rewrite as

$$y_t = \rho y_{t-1} + \alpha_1 \Delta y_{t-1} + ... + \alpha_p \Delta y_{t-p+1} + \varepsilon_t$$

$$\rho = \phi_1 + ... + \phi_p$$

- E.g., AR(2)

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \\ &= (\phi_1 + \phi_2) y_{t-1} - \phi_2 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \\ &= (\phi_1 + \phi_2) y_{t-1} - \phi_2 \Delta y_{t-1} + \varepsilon_t \end{aligned}$$

Augmented Dickey Fuller test

- General AR(p):

$$y_t = \rho y_{t-1} + \alpha_1 \Delta y_{t-1} + ... + \alpha_p \Delta y_{t-p+1} + \varepsilon_t$$

where $\rho = \phi_1 + ... + \phi_p$

- If there is a unit root, $\phi_1 + ... + \phi_p = 1$
- ADF test is a test of $\text{H}_0 : \rho = 1$ against H_1: $\rho < 1$
- Test statistic based on the OLS estimator of ρ
- Critical values come from a non-standard distributions (function of Brownian motion different from the one we have seen for AR(1)).

Where we are in the course outline

1. Introduction: time series data and examples of empirical questions
2. Regression with time series data
3. Univariate models
   - 3.1 Models of conditional mean
   - 3.2 Relaxing the assumption of stationarity
     - 3.3.1 Dealing with trends
     - 3.3.2 **Tests for structural breaks in parameters**
     - 3.3.3 **Modelling time variation in parameters**
   - 3.3 Model selection

Structural breaks

- Violation of stationarity due to changes in model's parameters
- E.g.,

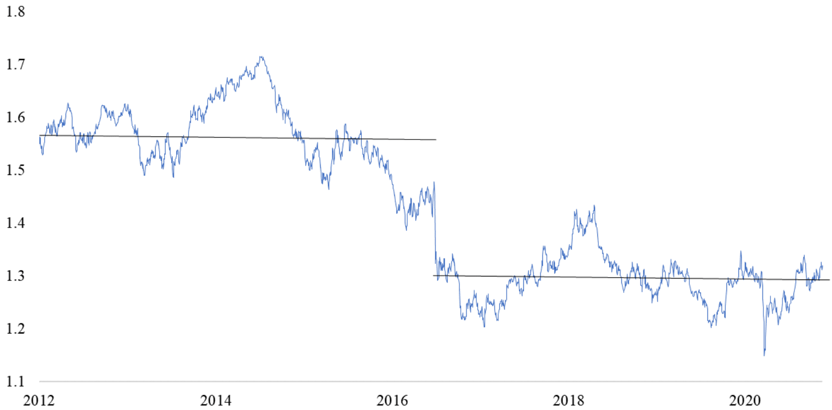
$$
\begin{aligned}
y_t &= c + \phi y_{t-1} + \varepsilon_t, \varepsilon_t \sim iidN(0, \sigma^2), \\
\theta &= (c, \phi, \sigma^2)
\end{aligned}
$$

- Stationarity $\rightarrow \theta$ constant
- Structural break: at least one component of θ changes at some date (break date)

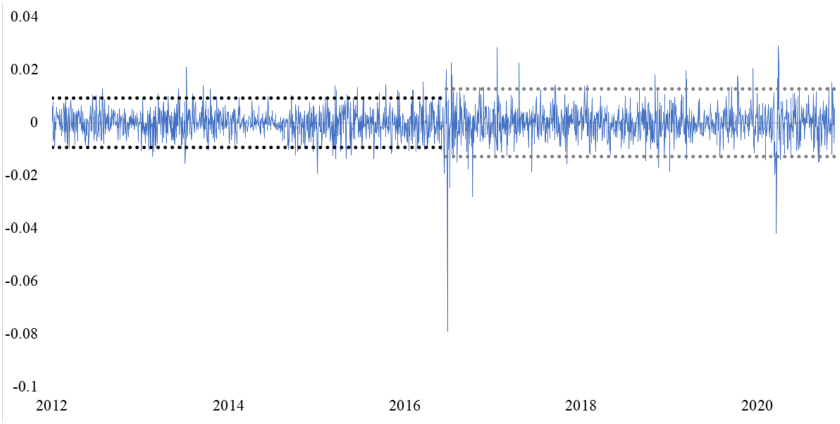
Structural breaks

- Possibilities:
  - Break in $c \rightarrow$ change in unconditional mean $E(y_t) = \frac{c}{1-\phi}$
  - Break in $\phi \rightarrow$ change in persistence and unconditional mean $E(y_t) = \frac{c}{1-\phi}$
  - Break in $\sigma^2 \rightarrow$ change in unconditional variance $Var(y_t) = \frac{\sigma^2}{1-\phi^2}$ ("great moderation")

Structural Breaks in Level: GBPUSD



Structural Breaks in Volatility: GBPUSD



A longer view

Questions

1. Test for a break at known date T_0.
2. Test for break at unknown date T_0.
3. Estimate time of break T_0.
4. Multiple breaks.

Chow Test: Test with known break date (I)

- Chow's (1960) test.

$$\text{Model } y_t = x_t\beta + \varepsilon_t \tag{1}$$

  where x_t is a $(1 \times k)$ row vector.
  - **Example** AR(1): $x_t = (1 \ y_{t-1})$
- Estimate (1) before and after T_0 and test if coefficients change
- Regress

$$\begin{pmatrix} y_B \\ y_A \end{pmatrix} = \begin{pmatrix} x_B & 0 \\ 0 & x_A \end{pmatrix} \begin{pmatrix} \beta_B \\ \beta_A \end{pmatrix} + \begin{pmatrix} \varepsilon_B \\ \varepsilon_A \end{pmatrix}$$

- This allows $\beta_B \neq \beta_A$

Chow Test: Test with known break date (II)

- Test the hypothesis H_0 of **no break**

$$H_0 : \beta_B = \beta_A \Rightarrow \mathbb{I}_k \beta_B - \mathbb{I}_k \beta_A = \mathbf{0}$$

  where $\mathbb{I}_k$ is the $(k \times k)$ identity matrix

- Simple test of linear hypotheses. Let $\delta = \begin{pmatrix} \beta_B \\ \beta_A \end{pmatrix}$ be a $(2k \times 1)$ column vector $\rightarrow$ Wald test of the hypothesis H$_0$: $R\delta = \mathbf{0}$ (in our case $R = [\mathbb{I}_k - \mathbb{I}_k]$, so k restrictions) is

$$F = (R\widehat{\delta})'(R\widehat{Var}(\widehat{\delta})R')^{-1}(R\widehat{\delta}) \sim \chi^2_k$$

- If reject $\Rightarrow$ there is a structural break at T_0

Test for unknown break date

- Andrews (1993), Andrews and Ploberger (1994)

- **Intuition**: Consider Chow's test over all possible break dates in the interior 15% to 85% of the sample. Then consider the **largest** (worst case) Chow's test statistic and find the distribution under H_0: no break
- Note that χ^2 critical value is not valid when break date unknown
- Andrews finds the critical value of the max$F$ under the null hypothesis (nonstandard distribution but critical values can be tabulated – software will do it for you.)
- If $\max F >$ critical value there is **at least** one break. Where?

Estimation of Break Date

- Bai (1994). Idea: estimate break date T_0 by OLS
- For each t in the interior of the sample, say $t = T_{15\%}, ..., T_{85\%}$, estimate

$$\begin{pmatrix} y_B \\ y_A \end{pmatrix} = \begin{pmatrix} x_B & 0 \\ 0 & x_A \end{pmatrix} \begin{pmatrix} \beta_B \\ \beta_A \end{pmatrix} + \underbrace{\begin{pmatrix} \varepsilon_B \\ \varepsilon_A \end{pmatrix}}_{\varepsilon}$$

- Estimate T_0 as the value that minimises the sum of squared residuals.

$$\hat{T}_0 = \underset{t}{\text{argmin}}\{\varepsilon'\varepsilon\}$$

- Bai shows that $\hat{T}_0$ corresponds to the max F date only when regression is linear + homoskedastic.

Test for multiple breaks

- Bai and Perron (1998) sequential procedure

1. Test for one break by Andrews $\max F$ test H_0 : no break
2. If reject, estimate $\hat{T}_0$ by Bai's method.
3. Split sample into before and after $\hat{T}_0$ and repeat steps $1-2$. in each subsample.
4. Continue until fail to reject H_0 : no break.

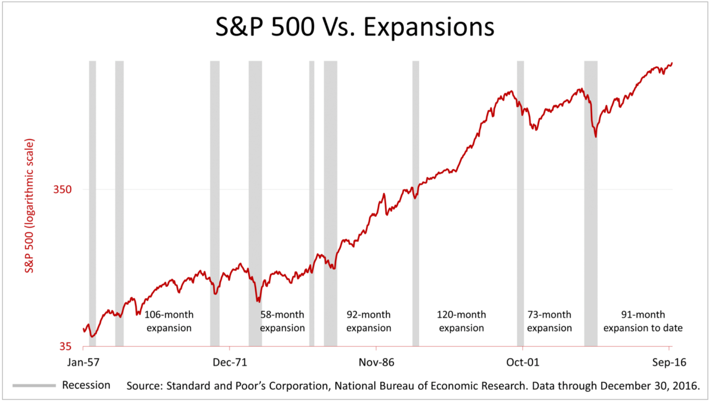
- **Justification**: The SSR has **local minima** around break dates
- **Possible issues**: cannot detect breaks in the beginning and end of the sample. In particular recent breaks would be the most useful to detect $\Rightarrow$ end-of-sample break testing considered by Andrews (2005) but difficult to implement.

Time variation

- We discussed how to test for structural breaks.
- But how should you model time variation when it comes analysing time series?
- Goal: How to model variation in parameters.
- Big topic, particular focus on threshold autoregressions.

Remember this example?

- Are stock price dynamics different during booms and recessions?



Modelling time variation

- Model parameter change as varying smoothly over time or depending on regimes/states of the economy
- Three main examples:
  1. Threshold autoregression (TAR) (parameters depend on **observable regimes**).
  2. Markov-Switching model (parameters depend on **latent regimes**)
  3. Time-varying parameters models (**parameters are latent** and vary smoothly over time)
- We won't cover the tools to model latent variables (state space models).
- In this course we only focus on TAR.

Threshold autoregressions (TAR)

- Parameters modelled as depending on observable regimes. E.g. TAR(1)

$$
\begin{aligned}
y_t &= \alpha_1 + \phi_1 y_{t-1} + \varepsilon_t \quad \text{if} \quad W_t > c \\
y_t &= \alpha_2 + \phi_2 y_{t-1} + \varepsilon_t \quad \text{if} \quad W_t \leq c
\end{aligned}
$$

- c is the **threshold**, W_t is the **observable** forcing variable
  - bull or bear market
  - business cycle indicator
  - changes in policy framework
- Parameters $\theta = (\alpha_1, \alpha_2, \phi_1, \phi_2, \sigma^2, c)$:
- Can be generalised to multiple thresholds.

Estimating threshold autoregressions

- If c known, estimate by OLS using dummies, e.g. TAR(1)

$$y_t = 1\left(W_t > c\right)\left(\alpha_1 + \phi_1 y_{t-1}\right) + 1\left(W_t \leq c\right)\left(\alpha_2 + \phi_2 y_{t-1}\right) + u_t$$

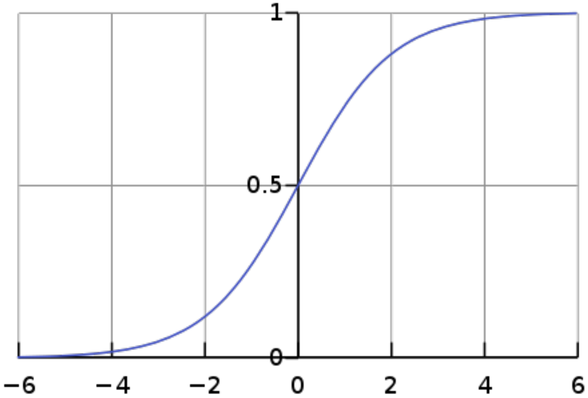
- If c is unknown, the model is nonlinear in the parameters $\Rightarrow$ cannot use OLS but must use MLE. Observation likelihood:

$$\begin{aligned} f_{y_T|y_{T-1},\ldots,y_1} &= \left(\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp(-\frac{\left(y_t - \alpha_1 - \phi_1 y_{t-1}\right)^2}{2\sigma_1^2})\right)^{1(W_t>c)} \times \\ &\quad \left(\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp(-\frac{\left(y_t - \alpha_2 - \phi_2 y_{t-1}\right)^2}{2\sigma_2^2})\right)^{1(W_t\leq c)} \end{aligned}$$

Smooth transistion threshold autoregressions

- Issue with TARs is that we make a hard choice regarding which regime we are in: 1,0 indicator.
- Smooth transition replace the indicator with a function:
  - If $W_t \in \mathbb{R}^h$, then use $F : \mathbb{R}^h \to [0, 1]$
  - Typical choice of F is a logistic function: $F(W_t) = (1 + \exp(\kappa_0(W_t - \kappa_1)))^{-1}$
  - Intuition: F is the probability of being in a regime.
- Estimation the same as a TAR. Just replace the indicator functions with $F(W_t)$ and $1 - F(W_t)$.
  - If F is known (or parameters calibrated) use OLS.
  - If F has parameters to be estimated use MLE.
- Possible to include multiple variables in W_t

Logistic fuction



Where we are in the course outline

Introduction

- We have discussed a variety of different models for your data.
- How to choose between them?
- Goal: Simple model selection techniques
- Then sum up the course so far: a cheat sheet for time series analysis.

Model selection in time series

1. Hypothesis testing
2. Information criteria

- Which one to use depends on whether models are **nested** (= one can be obtained from the other by imposing parameter restrictions) or **non-nested**
  - **Example:** nested

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \text{ vs}$$
$$y_t = \phi_1 y_{t-1} + \varepsilon_t$$

$\Rightarrow$ choice of models is equivalent to testing

$$H_0 : \phi_2 = 0 \Rightarrow \text{t-test}$$

or non-nested ($y_t = \phi y_{t-1} + \varepsilon_t$ vs $y_t = \varepsilon_t + \theta \varepsilon_{t-1}$) $\Rightarrow$ use information criteria

Information criteria

- General expression for a model with k parameters and sample size T.
$$IC = -\frac{\mathcal{L}(k)}{T} + \frac{k}{T} f(T)$$
where $\mathcal{L}(k)$ is the likelihood and $f(T)$ a penalty function.
- you want to find the minimum of IC. **Intuition**: as k grows, $\mathcal{L}(k) \uparrow$ but it must go up enough to compensate for penalty term $\frac{k}{T} f(T)$.
- Tradeoff goodness of fit (likelihood) againt overfitting (penalty function)
  - **Most common ICs**:
$$\begin{array}{ll} \text{AIC (Akaike IC)} & f(T) = 2 \\ \text{BIC (Bayesian IC)} & f(T) = \log(T) \end{array}$$
  - BIC penalizes more than AIC $\Rightarrow$selects smaller models (useful for principle of **parsimony**)

Other tests still needed

- Just because you have selected an AR(1) over an MA(1) doesn't mean you have selected the right model.
- The post-estimation tests such as stable parameters and uncorrelated residuals are still needed.
- And you should ensure your data is stationary in the first place and that the ACF/PCF justify a first order model.
- Is an ARMA model even right? How about other covariates that could help predict the data?

So you want to do some time series analysis?

1. First pick your question and that should guide you in picking time series of interest y_t and potential covariates x_t.
2. Ex-ante inspection of the data:
   - Plot the data, look for outliers, trends. Summary statistics.
   - Formal unit root tests.
3. Deal with trends (if any): model deterministic trends or difference the data.
4. Choose a model
   - what models fit your question? Think about candidates
   - Before estimation – ACF/PCF for lag orders
   - Post estimation – hypothesis testing/model selection
5. Post estimation diagnostics/testing
   - What are the right standard errors?
   - Properties of the residuals?
   - Are the parameters stable?

- These are all steps you know how to take now.
- With Raffaella you will see more complicated models but the same principles apply.