



UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvenoot • your knowledge partner

Master's Applied Time-Series Econometrics

Session I: Introduction to Autoregressive Time Series

Lecturer:

Nico Katzke

nicokatzke@sun.ac.za



Department of Economics
(copyright)

DEPARTMENT OF
ECONOMICS



Info on the course...



- This course is intended to offer a highly practical take on time-series econometrics.
- We will evaluate macroeconomic data for the purposes of building sound forecasting measures.
- This course builds into the next (Financial Econometrics) in which we will be looking deeper into volatility modelling, portfolio optimization and risk assessment in R.



Idea



- The idea of this part of the course is to ensure you have a **firm intuitive grounding** of what we are trying to achieve by modelling time-series, particularly with a view to building forecasting models.
- After this part of the course you should be able to fit standard time-series models, understand conceptually the processes underlying more complex models and understand the processes underlying dynamic multivariate forecasting models.



Info on the course...



- In this part of the course we will be covering the following topics:
 - Univariate Autoregressive time-series
 - Box-Jenkins ARIMA modelling
 - Multivariate time-series analysis
 - Short run dynamics: Vars, autoregressive processes
 - Long run dynamics: VECMs, co-integration testing
 - System of equations modelling (conducting forecasts similar to that being done by reserve banks).



Info on the course



- We will then also be using the tutorial sessions to fit these models and see how they are practically applied.
- Very Important: You will be evaluated on both your ability to grasp the theory and your ability to fit the models and interpret their results
- You will write a **practical** and a **theoretic** part in the exam.



Info on the course



- All your notes and other information will be posted on my website (curiousquant.com) under courses.
- **Statistical packages used:**
- **Eviews** (as it is widely used in practice by time series econometricians and is a good interface to learn advanced techniques on).
- **Textbooks used are:**
- **Prescribed:**
 - Brooks, *Introductory Econometrics for Finance (2nd Ed)*. (chapters 5, 6, 7)
 - Ruppert, D. (2011). *Statistics and Data Analysis for Financial Engineering (Chapter 9)*
- **Recommended reading:**
 - Lutkepohl & Kratzig, *Applied Time Series Econometrics*
 - Enders, *Applied Econometric Time Series, 3rd ed.*



What we will be covering today



- Introduction to Time-Series
- Definition and notations
- Stationarity
- White Noise
- Difference and trend stationarity
- Testing for Unit Roots
- AR process
- MA process



Recap section: Types of Data



- In econometric analyses, we have three types of data:
- **Cross-sectional:** One / more variables collected at a **single point in time.**
- **Time series:** Data **collected over a period of time.** The frequency indicates at which time interval data is collected (monthly, daily, etc) – and variables should be of same frequency
- **Panel data: Combination of both TS & CS.** E.g. we might study the annual GDP (cross-sectional component) of SA over the last 20 years (TS component)



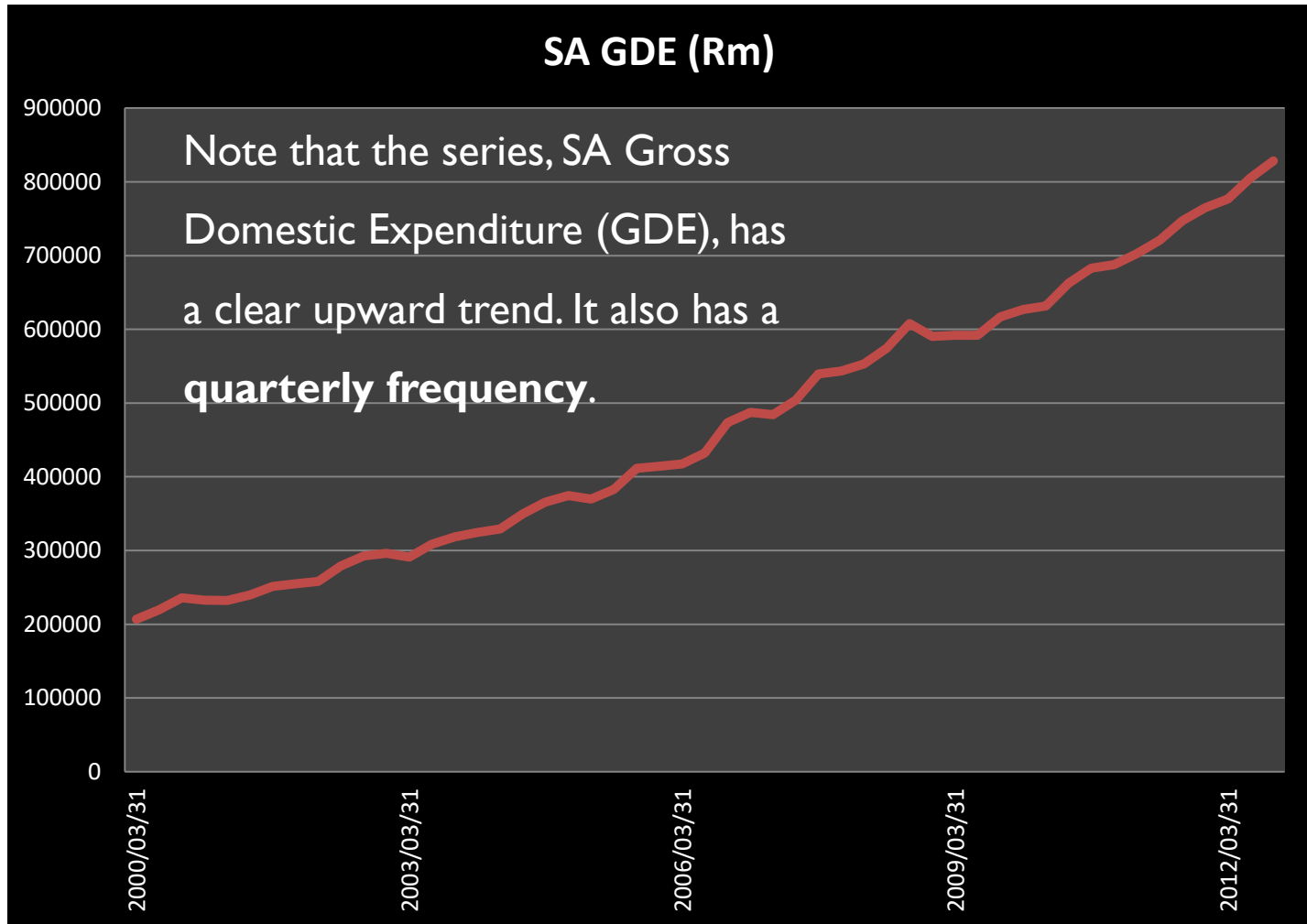
Types of Data



- **Continuous data:** Variable can take any value (e.g. 6.553%).
- **Discrete data:** Can only take on limited values (e.g. integers
– there are 23 people on the bus. There cannot ever be 23.446 people – thus the data is discrete (i.e. limited to whole numbers or integers)).
- For Time-Series, our data will mostly be continuous



Example of a Time Series





What is Time Series all about?



- As seen on the previous slide, a time series is a **chronological sequence** of observations taken of a particular variable – with adjacent observations often highly dependent.
 - Discovering and unpacking the nature of this dependence and the evolution of the series over time is of major importance.
- The task facing time-series econometricians is to develop TS models that can:
 - 1) **interpret the series at hand (understand the past).**
 - 2) **Test hypotheses of relationships** between it's own past and other variables' impact on it.
 - 3) **Use the model to forecast** values into the future



Basic Methodolgy



- An observed time series may not make much sense to the untrained (and even trained) eye at first.
- We are primarily interested in unpacking observed series into different parts that we *can* describe and accounted for.
- As such, a methodology has developed that aims to decompose a series of interest into its main parts:
 - **Trend-, Seasonal-, Cyclical-, and Irregular** components (where applicable.).
 - Thus we break the series down into parts we hope to say more about as the analysis deepens.



Components



- **Trend component** : Long term behaviour of the series.
- **Seasonal variations**: are periodic patterns in a time series that usually complete within a calendar year (or two) and are repeated yearly (or two yearly : the length now is arbitrary)
- **Cyclical components**: imply regular periods of up and down momentum around a longer term trend.
- The **irregular** (or **stochastic**) **component** represents the unexplained movement of the series over time, after we have accounted for the abovementioned components.
- This implies, in order to forecast a series, we split the explainable part of the series (T, S & C components) from the stochastic part, so that we are **only left with** unexplainable future shocks (ε) with expected outcomes of zero...



Stripped to the basics



- Suppose a time series $[X_t]$ is observed for period $t \in [0, \dots, 100]$. Suppose then the *equation of motion* broadly describing the series can be written as:

Trend : $T_t = \alpha + \beta_T(t)$

Seasonal : $S_t = \beta_S(\sin(t \cdot \pi/8))$

Irregular: $I_t = 0.8 \cdot I_{t-1} + \varepsilon_t$

- With : T_t value of the trend (and includes time as a regressor with a unit change at each period), S_t value of seasonal factor (having a sine function in this case) and I_t the value of the irregular component at time (t).
- Note that in this example, 80% of the irregular movement is described by the previous period's irregular disturbance + some error: $(\varepsilon_t) \rightarrow$ with (ε_t) the pure random disturbance in t .



Illustrating this in Eviews



- We can create a DGP in Eviews to show how the different components can make up a time-series.
- Setting the work file to monthly data with the range:

1995m01 – 2012m01

We can program the components as follows:

$$T_t = 0.05(t) + 0.007t^2 + u_t$$

$$S_t = 1.2 * \sin(1.4(t)) + u_t$$

$$I_t = 0.87 * I_{t-1} + u_t$$

$$C_t = \sin(0.15(t)) + 0.2 * u_t$$

With the time series then created as:

$$y_t = 100 + 5(S_t) + 0.22(T_t) + 3(I_t) + 5(C_t) + u_t$$



Illustrating this in Eviews

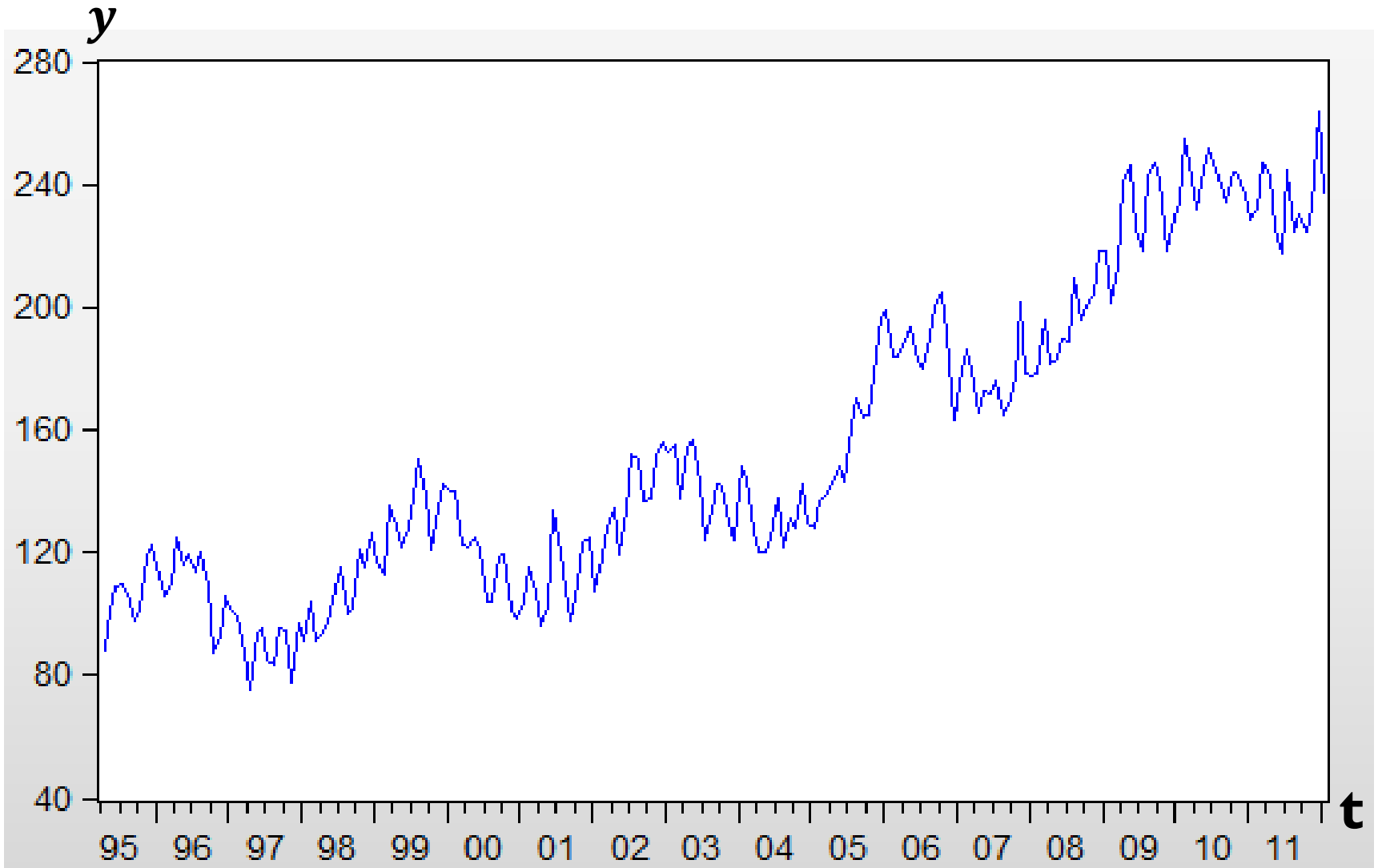


- I programmed the following DGP in Eviews to give you an idea of how the components in a time-series can be split:
- Open a new workfile: Unstructured / undated, with 100 observations
- Select **File – New – Program**
- Then enter the follow and select **run**:

```
series u=0.5*nrnd
smpl @first @first+2
series y1=0
series t=0
series s=0
series i=0
series c1=0
smpl @first +3 @last
t = 0.005*@trend + 0.007*@trend^2 + u
s= sin(1.4*@trend)+u
i= 0.97*i(-1) +u
c1= sin(0.15*@trend)+0.2*u
y1 = 0.2 + 7*s + 0.52*t + 3*i + 15*c1 + u
```

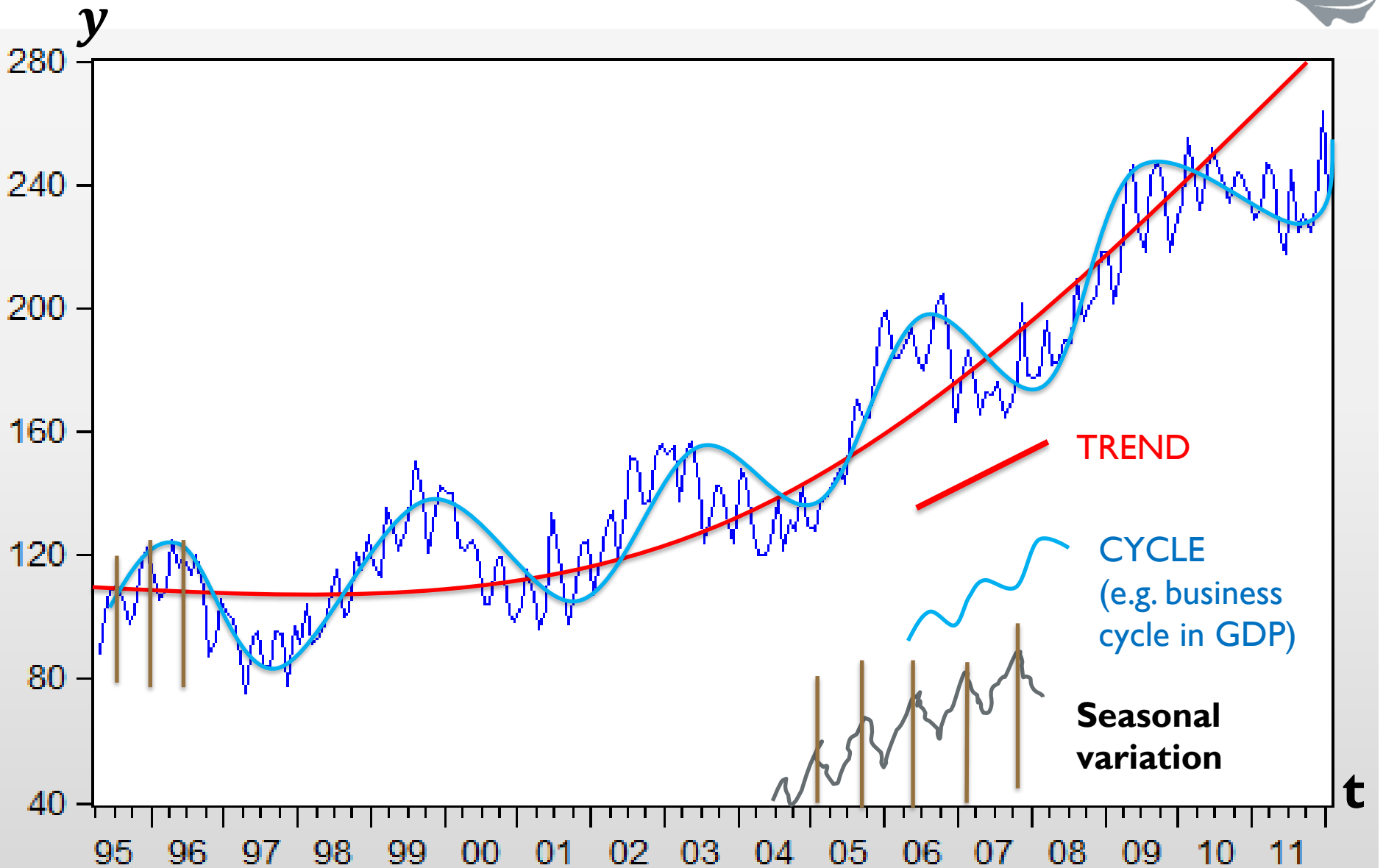


Illustrating this in Eviews



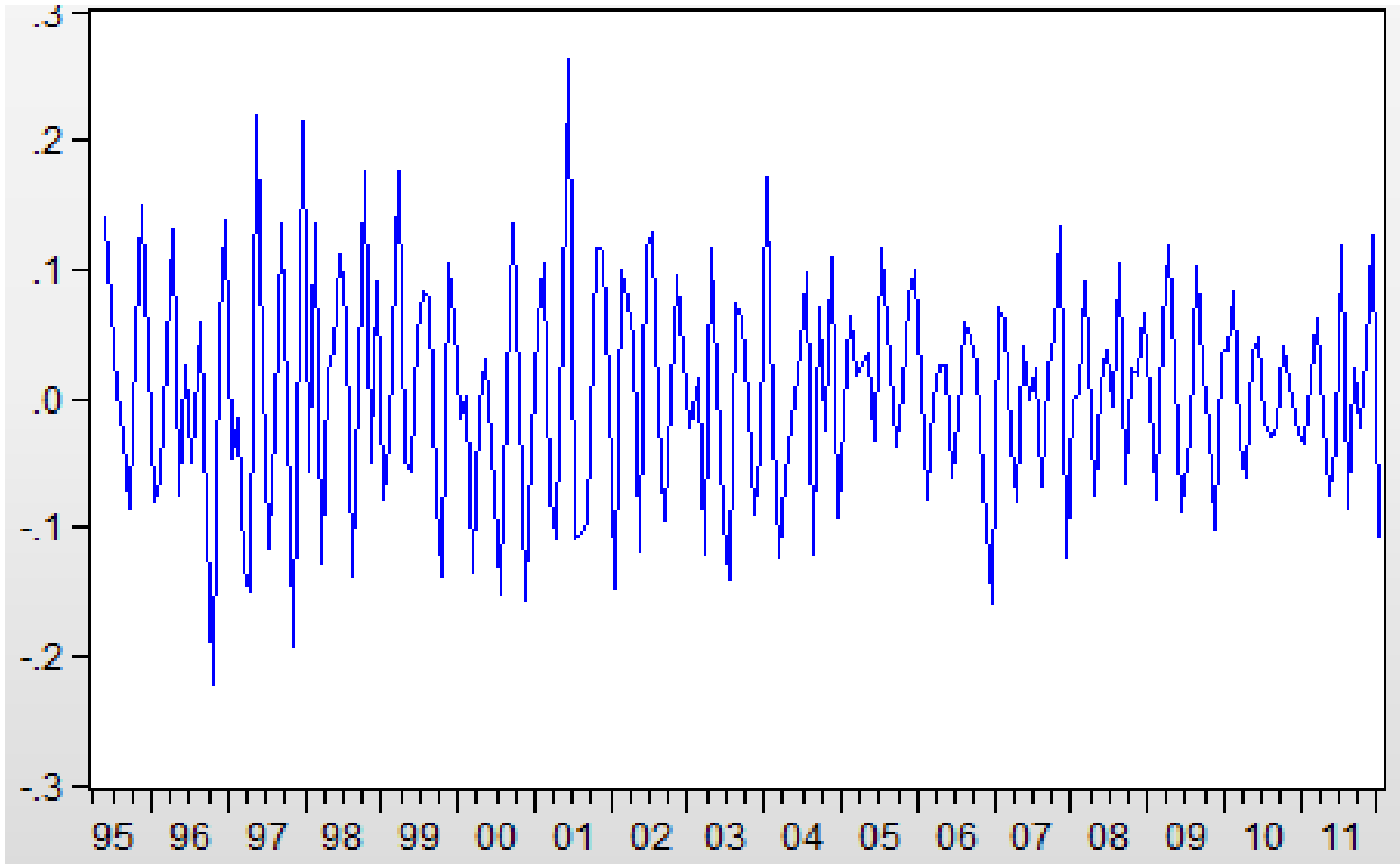


Example...





Residuals remaining...





UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

How do we explain the time-series behaviour?





Department of Economics

DEPARTMENT OF
ECONOMICS



Serial Autocorrelation



Consider the fact that normally for macroeconomic time-series data - the adjacent observations are **highly dependent** (or **autocorrelated**)... i.e. $\rightarrow [y_t \approx y_{t+1}]$

- To understand why this is the case— consider e.g. GDP. It is roughly equal to the previous period, plus some change in the last period.
 - This follows intuitively as any country's GDP does not vary wildly and exhibit completely independent adjacent observations:
 - GDP numbers will not be \$30'000 one period, and then \$700bn and then be – \$5trillion and so on. It will exhibit some trend along which it varies only slightly.



What we want to do then...



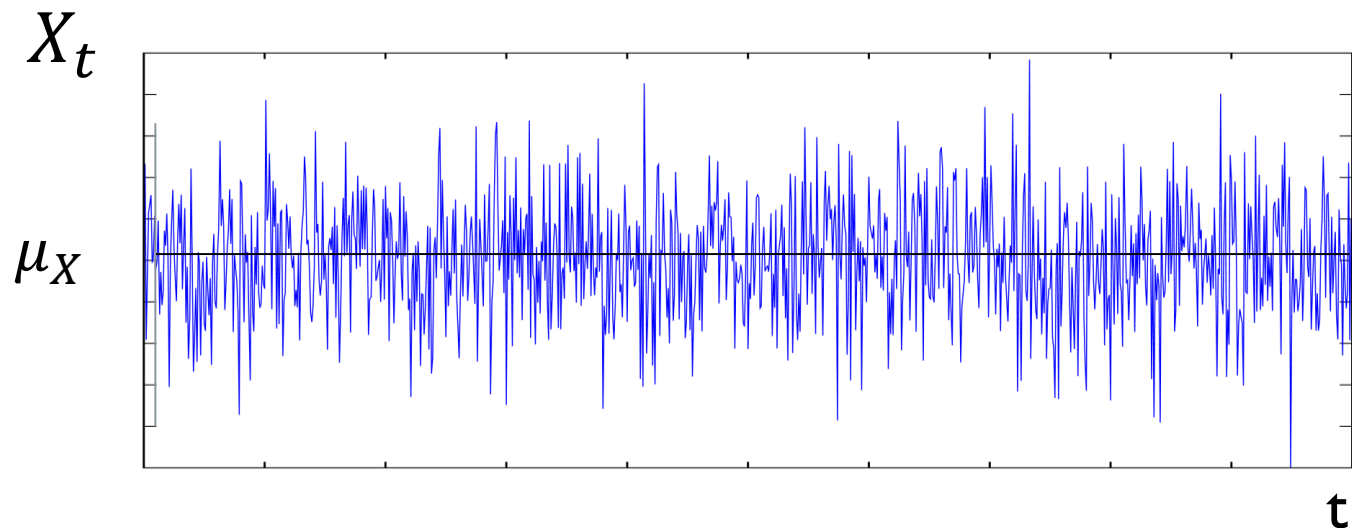
- Thus with time-series often being highly persistent, our aim is to cut through the persistence and explain what causes the data to change.
- In particular, we aim to understand both short run and long run dynamics of the system.
- This would therefore enable us to conduct forecasts and explain what the future holds in store...
- Let's consider some of the most important concepts relating to TS analysis.



Stationarity



- A stochastic time series process is regarded as **stationary** when it has time invariant **1st & 2nd order unconditional moments**. This implies that the series fluctuates roughly with a **constant variance** (2nd order moment) around a **constant mean** (1st order moment), and therefore does not exhibit a time varying **trend** over time.





Stationarity



- So what does stationary first and second moments imply?

Formally, stationarity implies:

- $E(y_t) = \mu \quad \forall t \in T$
- $E[(y_t - \mu_Y)(y_{t-h} - \mu_Y)] = \gamma_h, \quad \forall t \in T$

Thus for:

- $h = 0 \rightarrow E[(y_t - \mu)]^2 = \sigma_Y^2 \rightarrow$ *Constant **variance** over time*
- $h \neq 0 \rightarrow E[(y_t - \mu)(y_{t-h} - \mu)] = \gamma_h \rightarrow$ *constant **autocovariance** over time*

→ Thus both variance and covariances = independent of time



White-Noise



- White Noise can be defined formally as:

$$E(y_t) = 0 \quad (\text{zero mean over time})$$

$$\text{Var}(y_t) = \sigma^2 \quad (\text{constant variance over time})$$

$$E(y_t - \mu)(y_{t-k} - \mu) = \gamma_k = 0, \quad \forall k \neq t$$

Note from the last equation above that a White Noise series implies that the time series is not **serially correlated** (there is no correlation with the past), and it fluctuates **randomly around the zero mean**.

Notice that this is very much like the $i. i. d. N(0, \sigma^2)$ assumption of cross sectional work, accept now the iid part is for adjacent observations in the time series!



Striving for perfection...



- Although it is a rare sight to obtain a data set which even approximately resembles a stationary process, we can **make certain transformations** to our **series** to obtain an **approximately stationary series**.
- Remember that we are trying to **find a model** that can explain the evolution of y_t over time – using its own lags and & contemporary and lagged values of other variables...
- Having fitted a model → We then need, however, to ensure the residuals of the transformed series are **White Noise...**
- Let's consider this important feature of time series...



White Noise residuals... Why?



- Suppose we fit an explanatory model to describe the stationary process of y_t :

$$y_t = F(x, y, z) + \varepsilon_t$$

In order for our model's inference to be efficient and unbiased, **we need the residuals to be White-Noise residuals.**

$$\varepsilon_t \sim N(0, \sigma^2) \ \& \ \gamma_{t-k} = 0, \quad \forall k \ \& \ \forall t$$

This follows intuitively and similarly to the cross-sectional understanding of the error term:

- The **zero mean** is required, otherwise the parameters would be **biased** (notice that the β estimates are calculated as constant values assumed to hold over time)
- The **constant variance** over time is required, otherwise our parameters would not be **efficient** (much like the heteroskedasticity problem in cross-section, however here we have **conditional** heteroskedasticity – i.e. **heteroskedasticity conditional on the past**)
- The autocorrelation needs to be zero to ensure the trends over time has been controlled for – else our estimates would again be biased (downward bias if an upward trend is present in the residuals).



Why is stationarity important?



- The assumption of and testing for **Stationarity** is very important before we can fit models to explain the behaviour of certain time-series data.
 - Stationarity assumes a **constant time-invariant** mean, variance and autocovariance.
- But why so much focus on it?



Why is stationarity important?



- **Stationarity** influences the behaviour and properties of a series.
- Could lead to Non-sensical results when we model non-stationary data:
 - This is because the influence of a shock (or unanticipated occurrence) to nearly all macroeconomic time-series data **should** die away gradually over time.
 - However, non-stationary time-series can have shocks that do not die away (this is akin to a poor quarter in that 1963 impacts textile production even more intensely today than in 1964).
 - This can thus produce nonsensical model forecasts if errors have perpetual influences on future values.



Why is stationarity important?

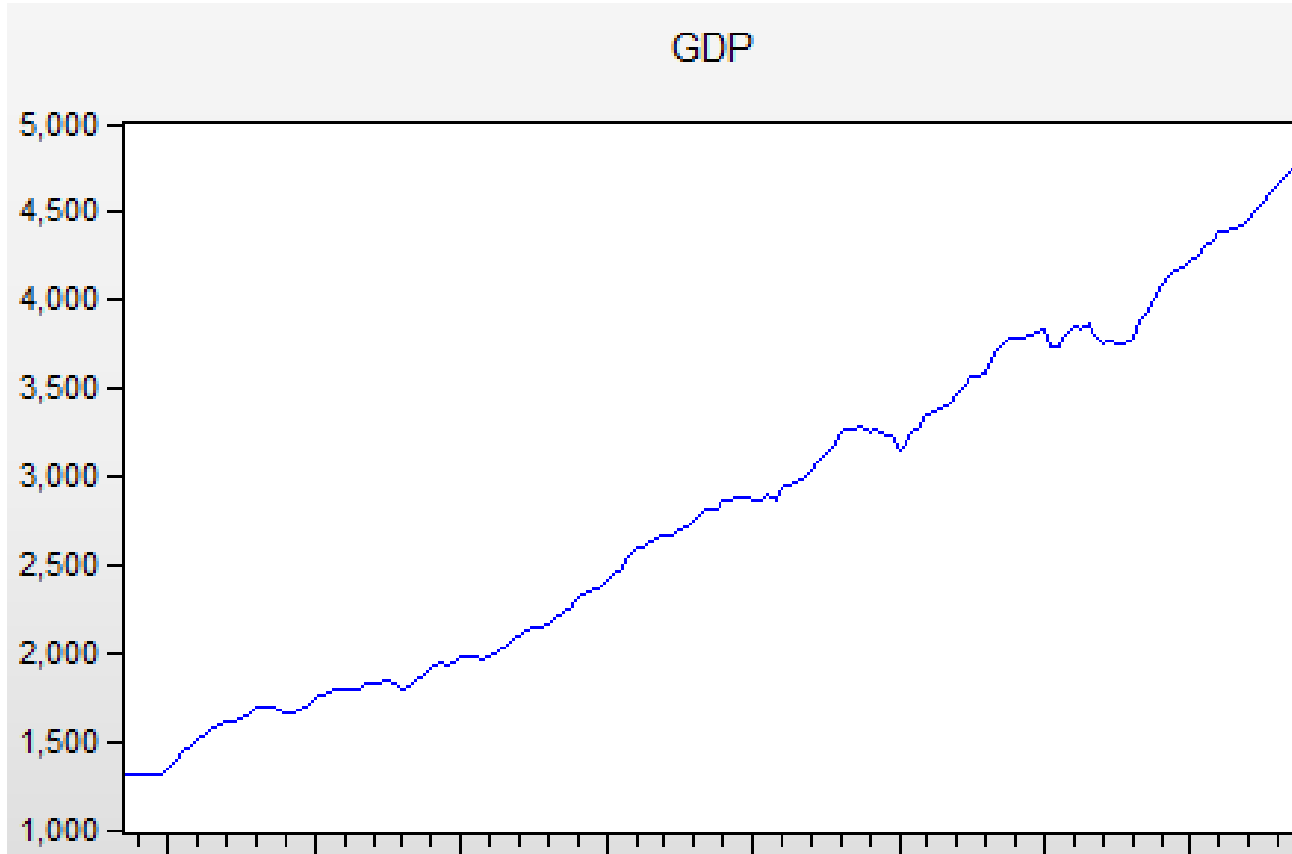


- **Spurious** regression results:
 - When standard regression techniques (such as OLS) is applied to non-stationary series – the result may show **falsely significant** findings.
 - Thus multivariate series / an autoregressive series may seem statistically significantly correlated with the past when in actual fact they are unrelated.
- The standard asymptotic assumptions do not hold
 - If non-stationary series are regressed – the *t stats*, R^2 , *F – stats*, etc. will not be accurate.



SO:

Suppose we have the following series:



Clearly this series is **not stationary**:

Its mean is **time dependent**.

(Stationarity requires a time-invariant mean...)



Types of non-stationarity



- It is not immediately clear which type of trend the above series exhibits.
- What we do know is that it needs to be controlled for before any further analysis on it can be done.

Such a trend can either be:

Stochastic trend (Difference stationary model):

$$y_t = \alpha + y_{t-1} + \varepsilon_t \quad (1)$$

Deterministic trend (Trend Stationary model):

$$y_t = \alpha + \beta(t) + \varepsilon_t \quad (2)$$

Suppose that in both cases the shocks (ε_t) are **WN**



Random Walk



- Note that any autoregressive series $[y_t = \alpha + \beta \cdot y_{t-1} + \varepsilon_t]$ where $\beta \geq 1 \rightarrow$ is a non-stationary series
 - Having $\beta > 1 \rightarrow$ implies an **explosive series** – meaning that a shock in t leads to an increasing influence of the shock over time.
 - $\beta = 1$ or *close to one* (we call it: “Beta being close to unity) is more likely and very common in time series.
 - This needs to be controlled for, otherwise the strong persistence of the series will make inferences nonsensical.
- If: $\beta \approx 1$, we say the series has a **Unit Root**. This needs to be controlled for by taking the **first difference**.



How do we know?

Testing for the Unit Roots...



- We should, however, be careful of not taking a first difference **if** the series is in fact a stationary process with the parameters lying within the unit circle ($\beta's < 1$) - then we run into problems of **overdifferencing**... This follows as suppose the true $\beta = 0.7$ and we take a first difference, it implies we forced the value of β to be = 1.
- The problem now is that running an OLS and fitting the parameters and checking for unit roots (i.e. to see if the parameters are very close to 1) – **will not yield satisfactory results** as the OLS techniques will not be appropriate in the presence of a unit root (spurious results) and may pick up only **strong persistence** (i.e. a high β) and not indicate a unit root ($\beta = 1$)
- One way testing for the presence of a unit root is to conduct a **Dickey Fuller test**.



Dickey Fuller Test



- The standard DF test basically tests (for the simplest AR(1) process):

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

$$H_0: \beta = 1 \quad vs \quad H_1: \beta < 1$$

If y_t is **not-stationary** (i.e. H_0 cannot be rejected), then the t – **stats** are not reliable and therefore standard OLS cannot be used.

D&F created asymptotic critical values for the DF test based on CPU simulations made easy for us in Eviews to check.

We can also test the following in eviews:

Intercept

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

Trend and intercept

$$y_t = \alpha + \beta y_{t-1} + \rho \cdot (t) + \varepsilon_t$$

None

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

With ε_t assumed WN



Augmented DF test



- Consider the series: GDP_t
- The Augmented version of the DF again tests the Hypothesis that: $\beta = 1$, but now *augments* the test to allow for some autoregressive behaviour of the residuals. This is a more robust test and the one which we will use most often.
- The *p* – *value* suggests the probability that the $H_0: \beta = 1$ cannot be rejected. **From the Eviews output:**

Null Hypothesis: GDP has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	1.444792	0.9991
Test critical values:		
1% level	-3.470934	
5% level	-2.879267	
10% level	-2.576301	

*MacKinnon (1996) one-sided p-values.

$p > 0.05$
thus the H_0
cannot be rejected,
thus there is
a unit root
present...



Testing for the presence of a Unit Root



- After taking the FD, we test for the presence of a unit root on the process: ΔGDP by again using the ADF:

Null Hypothesis: $D(GDP)$ has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-8.948867	0.0000
Test critical values:		
1% level	-3.470934	
5% level	-2.879267	
10% level	-2.576301	

- Thus as $p < 0.05$, we can reject the H_0 : $D(GDP)$ has a unit root.
- Thus the process: ΔGDP now does not have a unit root any more.
- We can now test for any further autoregressive dependence...



Other techniques of establishing URs



- EViews allows you to compute the **GLS-detrended** Dickey-Fuller (Elliot, Rothenberg, and Stock, 1996), Kwiatkowski, Phillips, Schmidt, and Shin (**KPSS**) Test (1992), Elliott, Rothenberg, and Stock Point Optimal (**ERS**, 1996), and Ng and Perron (**NP**, 2001) unit root tests.
- All of these tests are available as a view of a series.
- We will not discuss these measures in any further detail.



UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvenoot • your knowledge partner

DLOG Transformations

Let's briefly look at what this transformation – $D\log(\text{GDP})$ – implies.





Log transformations



- Taking the log of a time series will change the interpretation from **level changes** to **relative changes** in the series. **This is important because:**
- 1) For most of the macro and financial series we will be dealing with, **relative changes** are of more importance than level changes (e.g. rate of GDP change in % is easier to interpret than how it changed in millions of rand).
 - This is often interpreted as the logarithmic **return of a series / index / stock** in financial economics
- 2) In most macroeconomic and financial time series sets it will also be the case that changes in the series are dependent on the previous level...
 - Therefore, the series might display **larger fluctuations** as the **level** of the series **increases**.
 - Taking a Log of such an **increasingly fluctuating series** could **dampen** these increased **fluctuations** and move the series closer to stationarity.
- **NOTE : taking the $\log(Y_t)$ will not remove a trend nor remove a unit root if there is one of these present! – we still need to test Log(GDP) for both...**



Poor man's deflator



- Logging a series often has an effect very similar to **deflating**: it dampens exponential growth patterns and reduces perceived heteroskedasticity (i.e., stabilizes variance).
 - Logging is therefore considered the "poor man's deflator"
- Note that logging is not *exactly* the same as deflating--it does not *eliminate* an upward trend in the data--but it can straighten the trend out so that it can be better fitted by a linear model.
- NOTE: throughout I assume we use the natural logarithm, or $\ln(x) = \log_e(x)$



Size of the universe.... On log scale



- Log scales place a series / values into a different perspective. Consider e.g. the incomprehensible scale of the universe... in log base scale.
- The estimated mass of ordinary matter is then simply $\log(10^{53}) = 122.037 \dots$ although this is probably not a usefully interpretable value...





Log transformations



- 3) Earlier for illustration we implicitly assumed an **additive** model, where:

$$Y_t = trend_t + Seasonal_t + Cyclical_t + Irregular_t$$

- Specifying these components and / or taking d^{th} -order differences might remove such components in an attempt to leave only the *Irregular* component (which we hope is stationary, otherwise autoregressive techniques may be used).
- In particular, if we have a drift in the random walk process, taking a first-order difference might remove it.

- However, what if we have **multiplicative** models, with:
- $Y_t = trend_t * Seasonal_t * Cyclical_t * Irregular_t$
- ***Differencing this series would not remove the constant drift!!***



Log transformations



- Notice that by definition taking the log of the previously defined Y_t series implies that the model is made **additive** :

$$\text{Log}(Y_t) = \text{Log}(\text{trend}_t) + \text{Log}(\text{Seasonal}_t) + \text{Log}(\text{Cyclical}_t) + \text{Log}(\text{Irregular}_t)$$

- It is therefore a very useful transformation for a relatively complicated multiplicative time series... Taking the Difference now removes the time-constant factors...
 - *Just remember that logs can only be taken on positive values, hence a variable containing many negative values might lead to high levels of missing data generated and lost information!*



Log transformations - dlog



$$d\{Log(Y_t)\} = \cancel{Log(trend_t)} + \cancel{Log(Seasonal_t)} + \cancel{Log(Cyclical_t)} + Log(Irregular_t)$$

Because (simplifying for the sake of clarity):

$$dLog(trend_t) = \log(Trend_t) - \log(trend_{t-1}) \approx 0,$$

if we have a constant trend over time (same for seasonal and cyclical components)

Thus taking the Dlog(Y) implies removing a large part of the time dependent components (which typically have strong **persistence**) – leaving us only with (we hope) the irregular component



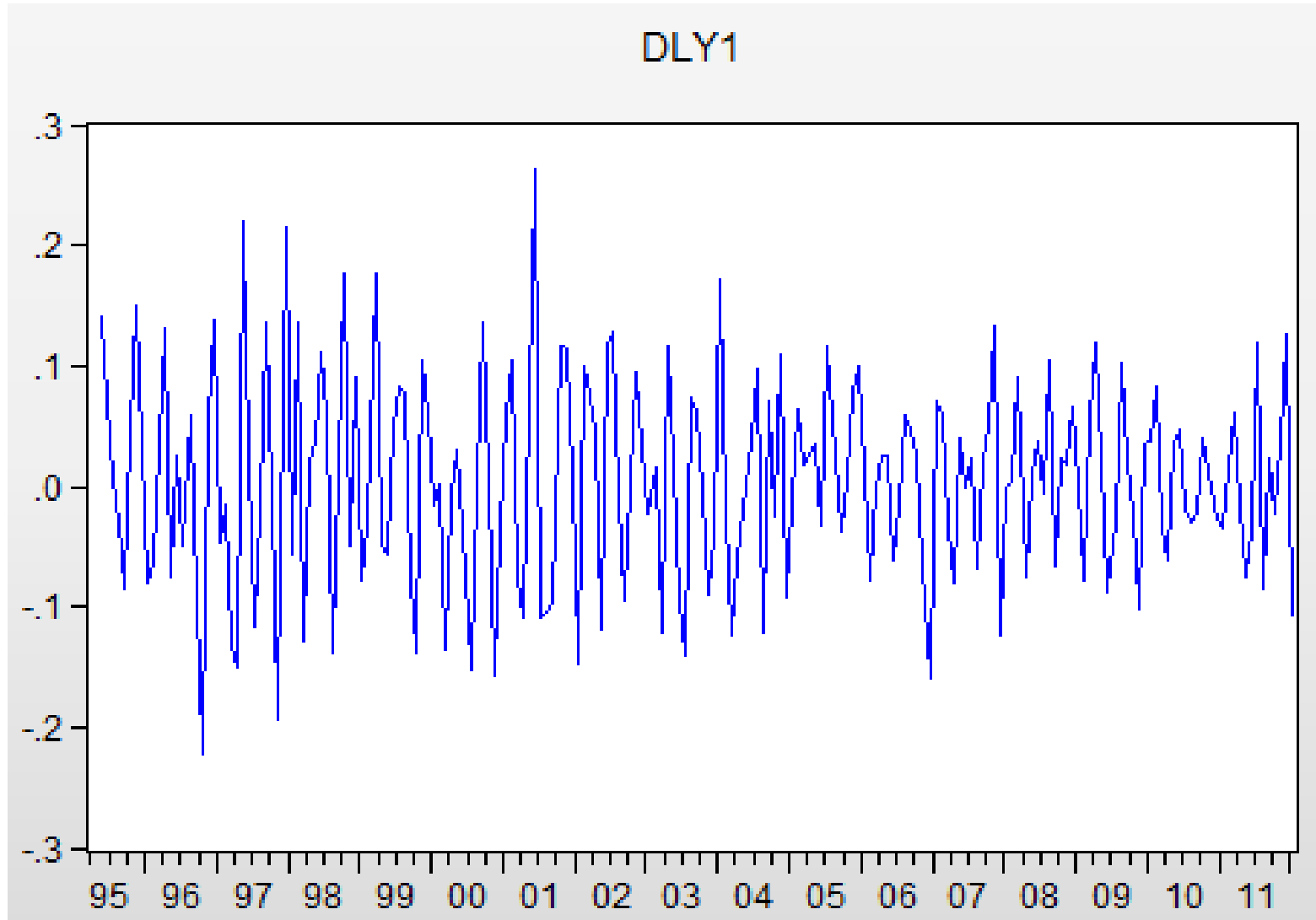
Dlog(Y_t)



- From the previous slide, if we took the $Log(Y_t)$, a potentially complicated multiplicative model now becomes an additive model in levels, and taking the Difference now might yield the series stationary.
 - Although it is a widely used measure, many regard it as a shortcut that may cause the modeller to lose potentially valuable information, by “buying” approximate stationarity in taking a Dlog.
 - We could also run into the problem of over-differencing if the series was indeed trend stationary or not strongly persistent. Nevertheless take note → it is widely used in practice!
- Note, however, that it might also be a useful transformation in terms of its interpretive value – as a measure of periodic changes (if that is what we are interested in, as opposed to understanding the levels).



Taking the $D\log(y_t)$ of our simulated data earlier...





Interpretation value of Dlog



- **First difference of LOG = percentage change:** When used in conjunction with differencing, logging converts absolute differences into relative (i.e., percentage) differences.
- Thus $D\log(Y)$ represents (approximately) the **percentage change** in Y from period to period.
 - Strictly speaking, the percentage change in Y at period t is defined as $\frac{Y_t}{Y_{t-1}} - 1$, which is only **approximately** equal to $D\log(Y(t))$...but the approximation is almost exact if the change is small, so for practical purposes it is often interpreted as being the same.
 - Test this by graphing: $D(Y_t/Y_{t-1})$ and $D\log(Y_t)$
- **In finance, we refer to $D\log(X)$ as the continuously compounded returns series...**



Fractal integrations



- There is also a somewhat different approach whereby differences are allowed to be in fractal forms.
- Thus the differencing factor might be $0 < d < 1$.
- This is often used in *long memory* models – and seeks to avoid the problem of over-differencing, whereby information is removed from the series.
- We will return to this later when we fit the **ARFIMA** model...



UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvenoot • your knowledge partner

Trend Stationarity

And smoothing techniques to forecast a series



Department of Economics

DEPARTMENT OF
ECONOMICS



Trend Stationarity



- For a **trend stationary process** (where we should actually have fitted a time trend) - taking a FD on the series introduces a **non-invertible MA** structure on the errors.
- For the above series, e.g., the following would happen if we **difference the series**:

$$y_t = \alpha + \beta(t) + \varepsilon_t$$

$$\Delta y_t = \alpha - \alpha + \beta(t) - \beta(t - 1) + \varepsilon_t - \varepsilon_{t-1}$$

$$\Delta y_t = \beta + \varepsilon_t - \varepsilon_{t-1}$$

- With: $[\varepsilon_t - \varepsilon_{t-1}] \rightarrow$ representing effectively a **Unit Root** autoregressive process created on the residuals.
- Thus we made the problem even worse!



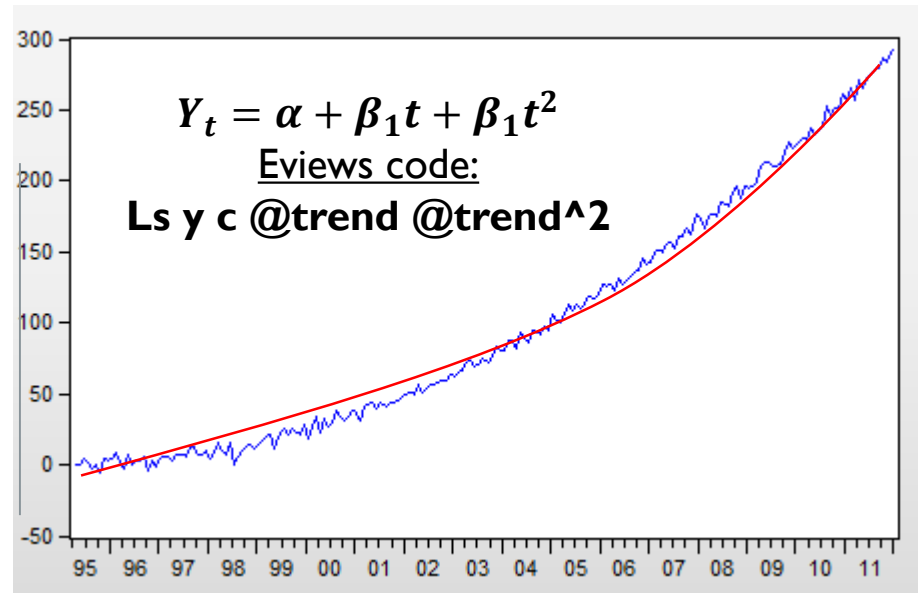
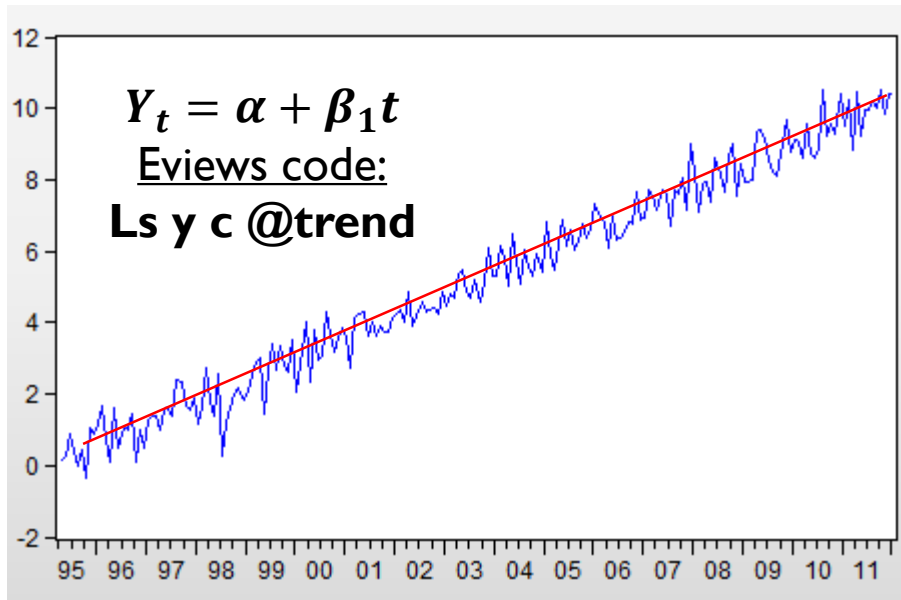
Trend Stationary process



- We therefore need to control for the deterministic trend of a series if it is trend stationary and not just blindly take the FD!!
- There are several techniques that can be used to de-trend a series when applicable.
- A linear trend over time may be fitted in eviews using:
@trend, and a non-linear trend as *@trend* ^2.



Fitting time-trends



- The trends above does not fit the data too well as its specification is too rigid. In a Bayesian context – the “model” does not learn from its past.
- Another method of de-trending the series above may be to let the estimates adjust as time changes (implying a non-constant time trend).
- This is known as smoothing techniques...



How do we know which it is...?



- How do we know whether a series with a non-constant mean has a **stochastic- (unit root)** or a **deterministic (time varying) trend?**
- This is a tough question and the answer is not as straight forward as using a test to tell us which to fit.
- We can include a trend component into the ADF test to see whether there is perhaps both a trend and difference stationary component – we could also fit both on the series separately and test which fits best (compare log-likelihoods)
- In practice, however, economists often assume (for simplicity) that macroeconomic series have **stochastic trends** and as such simply fit **Dlog transformations** on their data series...



UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvenoot • your knowledge partner

Smoothing Techniques



Department of Economics

DEPARTMENT OF
ECONOMICS



Smoothing techniques



- Sometimes a time series may not display a clear trend or seasonal component, but the **mean** trend of the series **changes slowly** over time in an unclear fashion → Allowing the mean (or level) of the series to change over time requires then a simple smoothing technique.
- This procedure allows the forecaster to update the estimate of the level of the time series so that it can incorporate the changing mean.
- It normally weights more recent observations with greater weight and older observations with successively smaller weights.



Simple Moving-Average



- The simplest form of smoothing is the simple Moving Average (equally weighted) model

$$Y_t = \frac{\{Y_{t-1} + Y_{t-2} + \dots + Y_{t-k}\}}{k}$$

This then follows that the forecast for the next period is a simple, equally weighted function of the past k observations.

Note: if $k = 1 \rightarrow$ this model is equivalent to a random walk model.

If $k = N \rightarrow$ this model is equivalent to the mean model.

The level k is chosen by minimizing forecasting errors.

- Moving Averages, although not accurate for longer term forecasts, are often used as a means of indicating the relative level of a series (Think, e.g., 200 day MA of a series)



Simple Exponential Smoothing (SES) techniques



- The main drawback of the Simple MA model above is that it weights past observations equally.
- Brown's SES model, in contrast, weighs past observations gradually less .
- Suppose we use μ_t to model changes in the stochastic series Y_t (i.e. μ_t is the estimate of the mean of Y at time t - we can refer to it as the **smoothed series**) a simplified exponential smoothing series may look as follows:

$$\mu_t = \alpha Y_t + (1 - \alpha) \cdot \mu_{t-1}$$

- α is called the **smoothing constant** and lies between $[0, 1]$.
- Implied the smoothed value of Y at time t is updated from its level at $t - 1$ (which is μ_{t-1}) **after observing Y_t** .
- From it, if the **mean is constant** over the entire series, $\alpha = 0$ should hold, so that the mean at t is equal to the mean at $t - 1$: i. e.: $\mu_t = \mu_{t-1}$
- If a series displays a **rapidly changing mean** over time, α would become larger.
- We use historical data to calculate α , hoping this smoothing technique would take accurately account for the changing stationary mean of the time series in the future.



Simple Exponential Smoothing (SES) techniques



- Using the SES model to forecast requires iterating forward the estimates:

$$\widehat{y}_{t+1} \approx f(y_t)$$



$$\widehat{y}_{t+2} \approx f(\widehat{y}_{t+1})$$



$$\widehat{y}_{t+3} \approx f(\widehat{y}_{t+2})$$

- Also, as will be seen later, the SES model is equivalent to an ARIMA(0,1,1) model without a constant term.
- Also, the SES model can be rewritten as (hence the name “exponential”):

$$y_t = \alpha \sum_{s=0}^{t-1} (1 - \alpha)^s \cdot Y_{t-s}$$



Double (linear) Smoothing SES



- If both the trend and the mean is varying slowly over time, it might be more accurate to model Brown's linear SES, or LES.
- It can be expressed as:

$$\mu_t = \alpha \cdot Y_t + (1 - \alpha) \cdot \mu_{t-1}$$

$$D_t = \alpha \cdot \mu_t + (1 - \alpha) \cdot D_{t-1}$$

- With μ_t the single smoothed series, D_t the double smoothed.
- Note this method still requires only one parameter to be estimated.
- Forecasts then lie on a linear trend with
- **Intercept:** $(2 \cdot \mu_t - D_t)$ & slope: $\frac{\alpha \cdot (\mu_t - D_t)}{1 - \alpha}$



Holt-Winters Multiplicative smoothing



- HW smoothing techniques use three parameters, estimating a linear time-trend and multiplicative seasonal component.

The smoothed series is then given by:

$$\hat{Y}_t = (i + b \cdot k) \cdot c_{t-k}$$

With: $i \rightarrow$ intercept (permanent)

$b \rightarrow$ trend

$a \rightarrow$ multiplicative seasonal component

$$i_t = \frac{\alpha(y_t)}{c_t(t-s)} + (1 - \alpha)(i(t-1) + b(t-1))$$

$$b_t = \beta\{i(t) - i(t-1)\} + (1 - \beta)b(t-1) \quad 0 < \alpha, \beta, \gamma < 1$$

$$c_t = \gamma \cdot \frac{y_t}{i(t-1)} + (1 - \gamma)c_t(t-s)$$

$S =$ Seasonal freq.



Holt-Winters Additive smoothing



- HW additive smoothing techniques are conceptually similar to the additive technique, but assume an additive seasonal variation.

The smoothed series is then given by:

$$\hat{Y}_t = (i + b.k) + c_{t-k}$$

$$i_t = i_t = \alpha y_t - c_t(t - s) + (1 - \alpha)(i(t - 1) + b(t - 1))$$

$$b_t = \beta\{i(t) - i(t - 1)\} + (1 - \beta)b(t - 1)$$

$$c_t = \gamma \cdot \{y_t - i(t - 1)\} - \gamma \cdot c_t(t - s) \quad 0 < \alpha, \beta, \gamma < 1$$

$S = \text{Seasonal freq.}$



Holt-Winters No seasonal



- For financial time-series analyses and higher frequency series, seasonal components might be largely redundant. The we can rather use Non-seasonal HW techniques:

$$\widehat{Y}_t = i + b.k$$

$$i_t = \alpha y_t + (1 - \alpha)(i(t - 1) + b(t - 1))$$

$$b_t = \beta \cdot \{i(t) - i(t - 1)\} + (1 - \beta)b(t - 1)$$

Forecasts are then computed as:

$$\widehat{Y}_{t+k} = i(T) + b(T)k$$



Holt-Winters Smoothing (summary)



- Holt-Winters Seasonal Smoothing is an extension of exponential smoothing that simultaneously estimates time-varying level, trend, and seasonal factors using recursive equations. (Thus, if you use this model, you would not first seasonally adjust the data.).
- The Winters seasonal factors can be either multiplicative or additive: normally you should choose the multiplicative option **unless you have logged the data.**
- Although the Winters model is clever and reasonably intuitive, it can be tricky to apply in practice: it has **three** smoothing parameters--alpha, beta, and gamma--for separately smoothing the level, trend, and seasonal factors: which must be estimated simultaneously.
- Determination of starting values for the seasonal indices can be done by applying the ratio-to-moving average method of seasonal adjustment to part or all of the series and/or by back-forecasting.



In Eviews



- Select a series and then choose Proc / Exponential smoothing / exponential smoothing. The following box appears:

Exponential Smoothing

Smoothing method # of params

- Single 1
- Double 1
- Holt-Winters - No seasonal 2
- Holt-Winters - Additive 3
- Holt-Winters - Multiplicative 3

Smoothed series

Series name for smoothed and forecasted values.

Estimation sample

Forecasts begin in period following estimation endpoint.

Cycle for seasonal

Alpha: (mean) E Enter number between 0 and 1, or E to estimate.

Beta: (trend) E

Gamma: (seasonal) E

OK Cancel

All the choices discussed are given. If the estimated smoothing parameters are close to unity – it implies the possible presence of a unit root (most recent

Observation = best estimate of future values...)



HP-Filter



- We can also apply Hodrick & Prescott's (1997) HP-filter.
- This is a more advanced technique used to obtain a **smoothed estimate** of the long-term trend component of a series.
- We will fit this in Eviews and interpret its results later – the technicalities of this technique is beyond the scope of this course, but it basically is a hybrid of the elementary smoothing technique discussed above.
- The next slide shows the math behind it.



HP-Filter



- For a time-series y_t and a non-negative smoothing parameter λ :

$$\sum_{t=1} (y_t - \tau_t)^2 + \lambda \sum_{t=2} (\tau_{t+1} - 2\tau_t + \tau_{t-1})^2$$

- Where τ is the smoothed series which results from minimizing the above equation;
- $\widehat{c}_T = y_t - \widehat{\tau}_t$ is the cyclical component.
- The above equation can then be rewritten into its FOC, and then minimized (most standard texts derive this – which falls outside the scope of this course for brevity).
- For this course, just know that the parameter λ controls the smoothness of the series, with $\lambda \rightarrow \infty$ **implying a straight line** and $\lambda \rightarrow 0$ **implying a trend that lies on the series itself**.

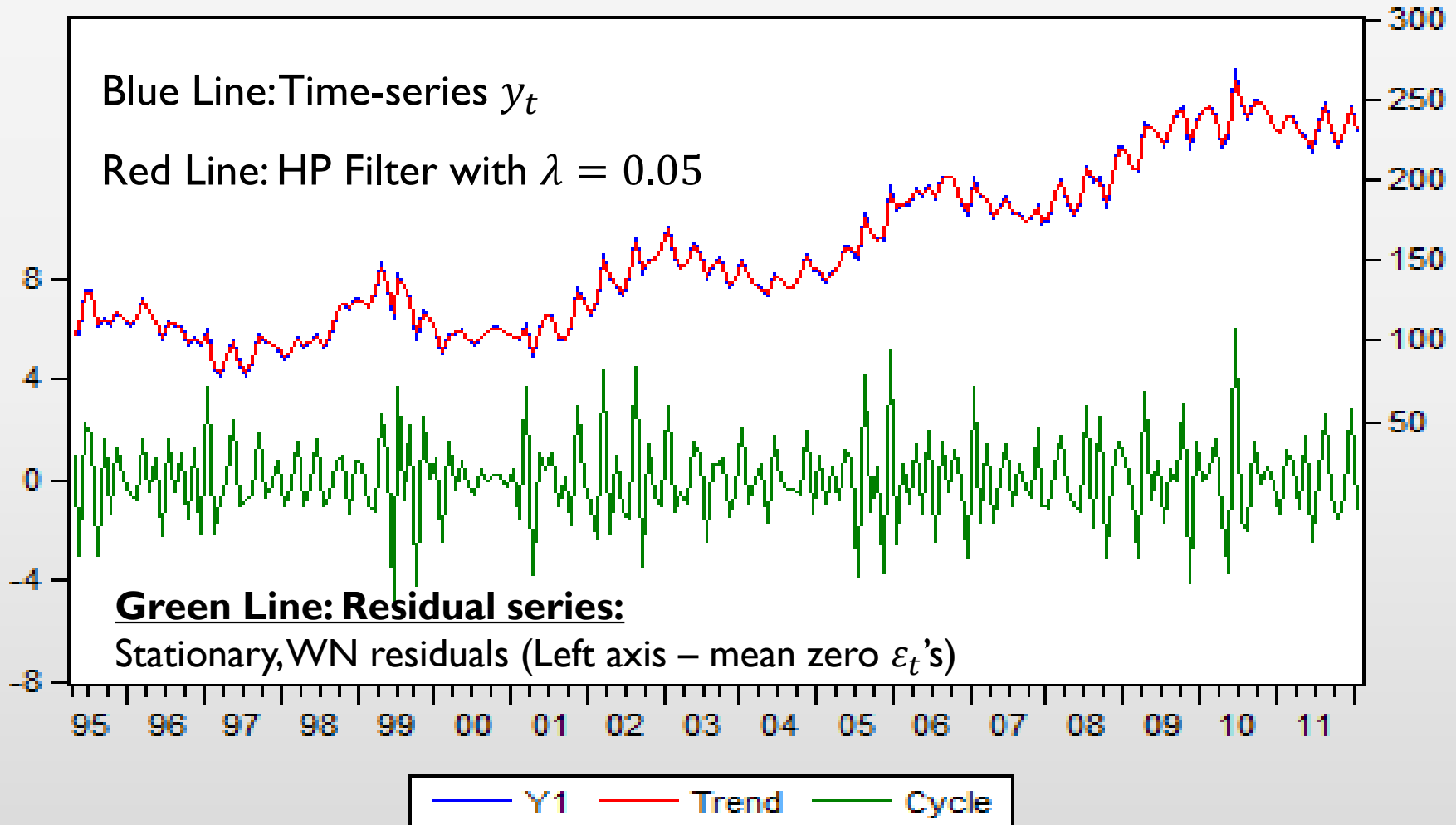
Let's look at 3 HP filters fitted on the DGP created data from earlier



HP Smoothing filter (de-trending GDP)



Hodrick-Prescott Filter (lambda=0.05)

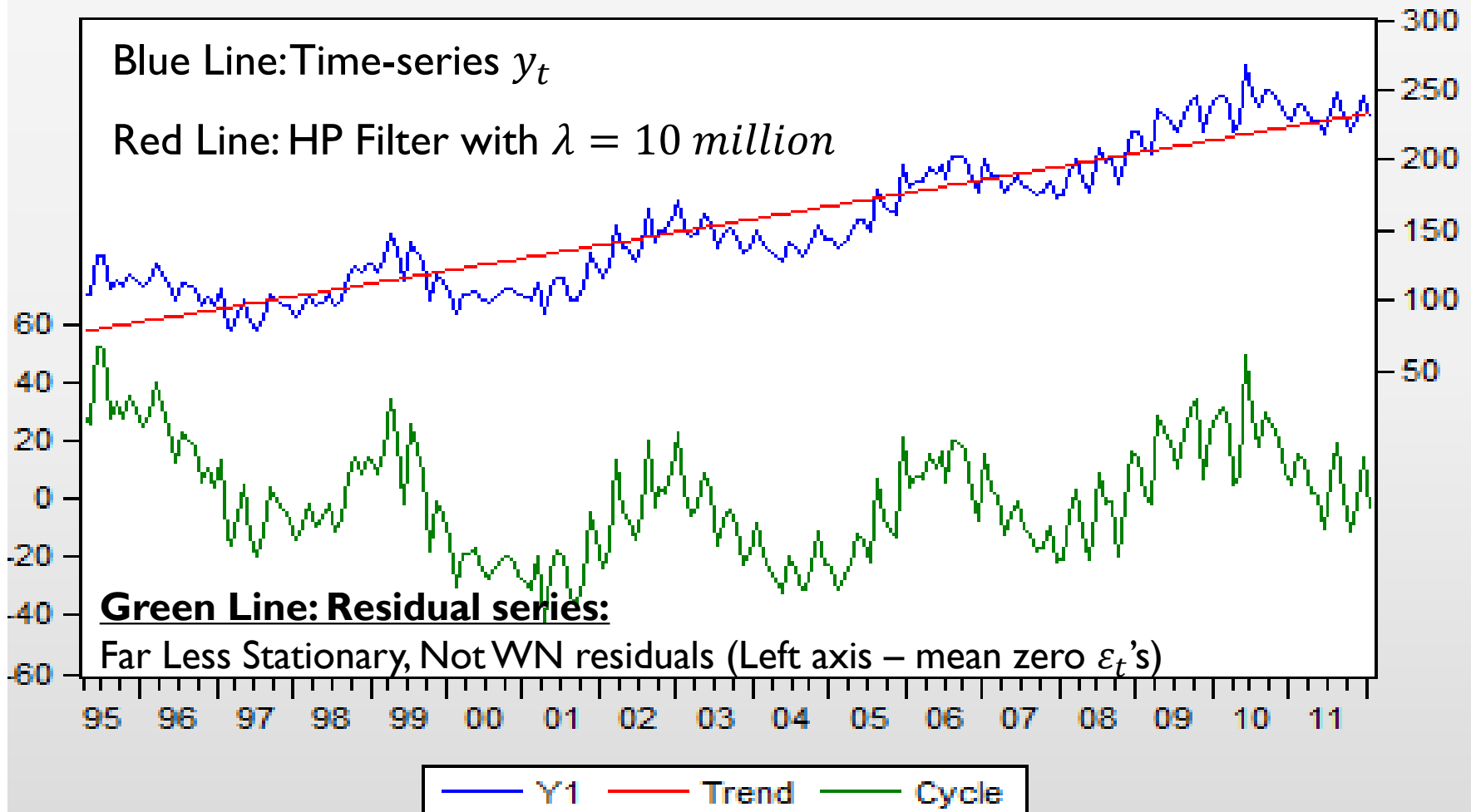




HP Smoothing filter (de-trending GDP)



Hodrick-Prescott Filter (lambda=100000000)

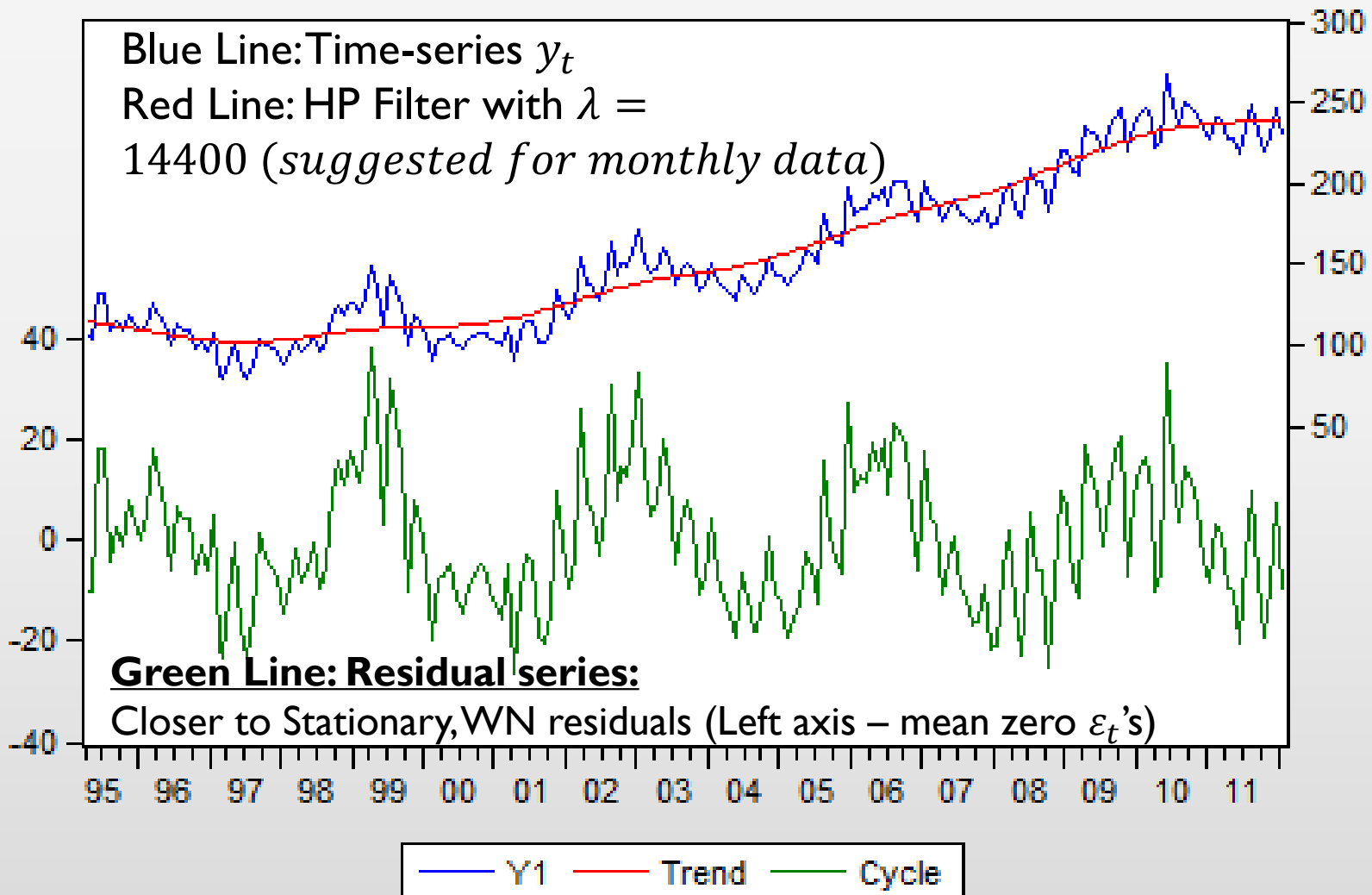




HP Smoothing filter (de-trending GDP)



Hodrick-Prescott Filter ($\lambda=14400$)





HP Filter with different Lambdas



- Notice that if Lambda is too small – it fits the data almost precisely and the errors are therefore at a minimum.
- This might seem like the best practice, but remember that we use time-series models to **forecast**... And fitting an HP with Lambda of 2 in this case – implies a perfect fit of the past, but is not at all useful in forecasting!!
- Convention: Daily = 100; Quarterly = 1600;
 Monthly = 14 400.



The end...



- Next session we will be looking deeper at fitting the Box-Jenkins ARIMA models and conducting simple forecasts.