Forecasting Economic and Financial Time Series
Week 3: Trends, deterministic and stochastic, cycles and seasonals

Ron Smith
MSc/PGCE Option, EMS Birkbeck

Spring 2019

Components of a time series 1

- Common to split the movements of a variable to be forecast into unobserved: trend; cycle, seasonal components and any other irregular movements.
- **Trend:** slow, long-run evolution, linked to factors such as technology, institutions, preferences and demographics, τ_t
- **Seasonal** deterministic fixed frequency cycles, usually time of year but may be time of day in high frequency financial time series. s_t (Problems with Easter and Ramadan, not fixed frequency).
- **Cycle:** Irregular medium term (2-10 year) movements in the economy, c_t. Longer Kondratieff cycles of 50 years?
- **Irregular** or random component, ε_t.

Components of a time series 2

- Components may be additive or multiplicative (additive in logs).

$$
\begin{aligned}
y_t &= \tau_t + s_t + c_t + \varepsilon_t \\
y_t &= \tau_t \times s_t \times c_t \times \varepsilon_t \\
\ln y_t &= \ln \tau_t + \ln s_t + \ln c_t + \ln \varepsilon_t
\end{aligned}
$$

- Infinite number of observationally equivalent ways to decompose an observed series into unobserved components, e.g. trend and remainder: $y_t = \tau_t + \varepsilon_t$. How do you choose?

Types of Trend

- We will begin with **deterministic trends** which trends evolve in a perfectly predictable way, which can often be very useful in practice

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

estimated on data $t = 1, 2, ..., T$ and used to forecast $T + 1, T + 2, ..., T + h$

- **Stochastic trends** arise from difference stationary processes

$$\begin{aligned} y_t - y_{t-1} &= \alpha + \varepsilon_t \\ y_t &= \alpha t + \sum_{j=0}^{t-1} \varepsilon_{t-i} + y_0 \end{aligned}$$

- **Filters** like the Hodrick-Prescott filter fit smoothly varying trends to data
- There is also the **Kalman filter** time-varying parameter model which can be used to extract a trend

Structural Time-Series Models

- These use the Kalman Filter to achieve this decomposition: Andrew Harvey *Forecasting, structural time series models and the Kalman filter*, CUP 1989 and program STAMP part of Oxmetrics suite.
- Trend, τ_t, cycle c_t

$$\begin{aligned} y_t &= \tau_t + c_t + \omega_t \\ \tau_t &= \tau_{t-1} + \beta_t + \eta_t \\ \beta_t &= \beta_{t-1} + v_t \\ \phi(L) c_t &= \varepsilon_t \end{aligned}$$

- Shocks: $(\omega_t, \eta_t, v_t, \varepsilon_t)$ mean zero normally distributed, independent of each other and over time. Only small changes not big jumps or outliers.
- Deterministic trend: $\beta_t = \beta$, $E(\eta_t^2) = 0$. $\phi(L)$ polynomial in lag operator $\left(1 - \phi_1 L - \phi_2 L^2 ... - \phi_p L^p\right)$

Deterministic Trends

- Assuming a linear trend model, at time $T+h$

$$y_{T+h} = \beta_0 + \beta_1(T+h) + \varepsilon_{T+h}$$

  a **point forecast** of y_{T+h} made at time T $h-$steps-ahead is the conditional expectation:

$$E(y_{T+h} \mid \Im_T) = \widehat{y}_{T+h,T} = \widehat{\beta}_0^T + \widehat{\beta}_1^T(T+h)$$

  with coefficients $\widehat{\beta}_0^T$ estimated using a sample $t = 1, 2, \ldots T$; from $y_t = \widehat{\beta}_0^T + \widehat{\beta}_1^T t + \widehat{\varepsilon}_t^T$, superscript to indicate sample used.
- Note we need two subscripts for $y^f_{T+h,T} = E(y_{T+h} \mid \Im_T)$, when the forecast is for $(T+h)$ and when it is made (T). No standard notation, e.g. may be ${}_T y_{T+h}$.

Point Forecast

In constructing the **point** forecast:

- We use estimates of the parameters from the available sample
- $(T + h)$ is known as it is a artificially constructed time variable, but other exogenous variables need to be forecast
- ε_{T+h} is not known, therefore we replace it with its expectation using information available up to time T. Under the assumption that ε is an independent; zero mean; random error, the optimal forecast of ε_{T+h} for any future period is 0.
- Often judgemental adjustment involves a forecast of ε_{T+h}, if the forecasters information set is larger than that of the model, e.g. includes qualitative information.

Interval and density forecasts

- For an **interval forecast**, we might also assume that $\varepsilon_t \tilde{} N(0, \sigma^2)$; and the sample is large enough that the normal approximation to the t is OK.
- Then a 95% interval forecast, **ignoring parameter uncertainty**, is $\widehat{y}_{T+h,T} \pm 1.96s$, where the estimated standard error of regression is $s = \sqrt{\sum_t \left(\widehat{\varepsilon}_t^T\right)^2 / T - 2}$.
- To make a **density forecast,** we might again use the normality assumption, so that the density forecast is $N(\widehat{y}_{T+h,T}, s^2)$.
- Alternatively esimate using some other distribution, e.g. $t(dof)$ or simulate drawing from the empirical frequency distribution of residuals or past forecast errors.

Variance of forecast allowing for parameter uncertainty

- Estimates using data up to T are $\widehat{\alpha}^T$ and $\widehat{\beta}^T$, the forecast of Y_{T+1}, made at time T using forecast of X_{T+1} made at T is

$$\widehat{Y}_{T+1,T} = \widehat{\alpha}^T + \widehat{\beta}^T X_{T+1,T}$$

- The estimated variance of the **forecast** is

$$\begin{aligned} V(\widehat{Y}_{T+1,T}) &= V(\widehat{\beta}_0^T) + X^2_{T+1,T} V(\widehat{\beta}_1^T) + 2X_{T+1,T} Cov(\widehat{\beta}_0^T, \widehat{\beta}_1^T) \\ &= s^2(\frac{1}{T} + (X_{T+1,T} - \overline{X})^2 / \sum_{t=1}^{T} (xX - \overline{X})^2). \end{aligned}$$

Variance increases the further $X_{T+1,T}$ is from $\overline{X}$ with confidence interval is is $\widehat{Y}_{T+h,T} \pm 1.96\sqrt{V(\widehat{Y}_{T+1,T})}$.

Variance of forecast error allowing for parameter uncertainty

- The forecast error is

$$\widehat{u}_{T+1,T} = Y_{T+1} - \widehat{Y}_{T+1} = u_{T+1} + (\alpha - \widehat{\alpha}^{T}) + (\beta - \widehat{\beta}^{T})X_{T+1,T}$$

- The estimated variance of the **forecast errors** is

$$V(\widehat{\varepsilon}^{f}_{T+1}, T) = s^2(1 + \frac{1}{T} + (X_{T+1} - \overline{X})^2 / \sum_{t=1}^{T}(x_t - \overline{x})^2).$$

Allowing for parameter uncertainty in forecast variances

To derive these results simplify the notation and consider the bivariate linear regression model with the usual assumptions

$$\begin{aligned} Y_t &= \alpha + \beta X_t + u_t, \;\; t = 1, 2, ..., T, \\ E(u_t) &= 0, \;\; E(u_t^2) = \sigma^2; \;\; E(u_t u_{t-i}) = 0. \end{aligned}$$

$$\begin{aligned} V(\widehat{\alpha}) &= \sigma^2 \sum X_t^2 / \left(T \sum (X_t - \overline{X})^2\right) \\ V(\widehat{\beta}) &= \sigma^2 / \left(\sum (X_t - \overline{X})^2\right) \\ Cov(\widehat{\alpha}, \widehat{\beta}) &= -\sigma^2 \overline{X} / \sum (X_t - \overline{X})^2 \end{aligned}$$

Define the forecast as

$$\widehat{Y}_{T+1} = \widehat{\alpha} + \widehat{\beta} X_{T+1}$$

with variance of forecast

$$V(\widehat{Y}_{T+1}) = E(\widehat{Y}_{T+1} - Y_{t+1})^2 = E\left((\widehat{\alpha} - \alpha) + (\widehat{\beta} - \beta) X_{T+1}\right)^2$$

Allowing for parameter uncertainty in forecast variances

So the variance of the forecast $V(\widehat{Y}_{T+1})$ is given by

$$
\begin{aligned}
& V(\widehat{\alpha}) + X_{T+1}^2 V(\widehat{\beta}) + 2X_{T+1} Cov(\widehat{\alpha}, \widehat{\beta}) \\
= \; & \frac{\sigma^2 \sum X_t^2}{T \sum (X_t - \overline{X})^2} + X_{T+1}^2 \frac{\sigma^2}{\sum (X_t - \overline{X})^2} - 2X_{T+1} \frac{\sigma^2 \overline{X}}{\sum (X_t - \overline{X})^2} \\
= \; & \frac{\sigma^2}{T \sum (X_t - \overline{X})^2} \left(\sum X_t^2 + T X_{T+1}^2 - 2X_{T+1} \sum X_t \right) \\
= \; & \frac{\sigma^2}{T \sum (X_t - \overline{X})^2} \left(\sum (X_t - \overline{X})^2 + T (X_{T+1} - \overline{X})^2 \right) \\
= \; & \sigma^2 \left(\frac{1}{T} + \frac{(X_{T+1} - \overline{X})^2}{\sum (X_t - \overline{X})^2} \right)
\end{aligned}
$$

To get from line 3 to 4 add and subtract $\overline{X}^2$.

Variance of a forecast in multiple regression

- Consider model $y = X\beta + u$, where y is $T \times 1$, X $T \times k$ β $k \times 1$ with estimates $\widehat{\beta} = (X'X)^{-1}X'y$ and $V(\widehat{\beta}) = E(\widehat{\beta} - \beta)(\widehat{\beta} - \beta)' = \sigma^2(X'X)^{-1}$.
- The scalar forecast for $T+1$ is $y^f_{T+1,T} = x'_{T+1}\widehat{\beta}$ with variance conditional on x'_{T+1} of

$$
\begin{aligned}
V(y^f_{T+1,T}) &= E(x'_{T+1}\widehat{\beta} - x'_{T+1}\beta)(x'_{T+1}\widehat{\beta} - x'_{T+1}\beta)' \\
&= x'_{T+1}E(\widehat{\beta} - \beta)(\widehat{\beta} - x'_{T+1}\beta)x_{T+1} \\
&= \sigma^2 x'_{T+1}(X'X)^{-1}x_{T+1}
\end{aligned}
$$

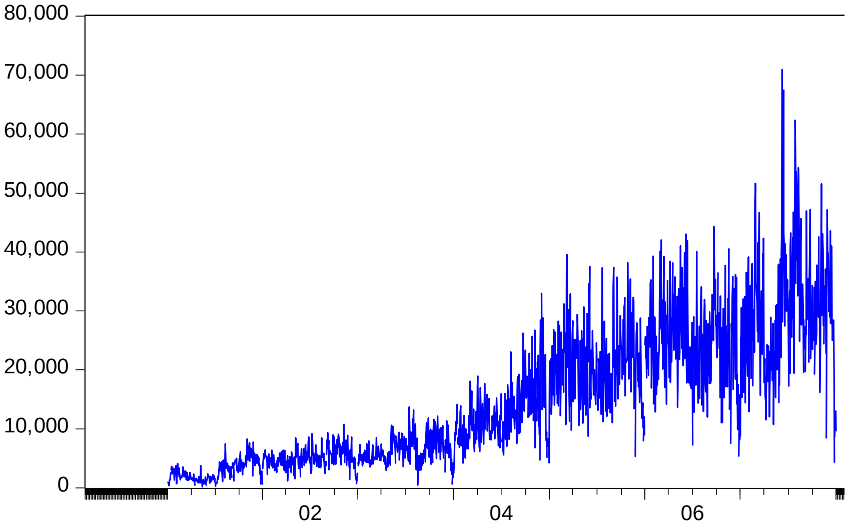
- The variance of the forecast errors is

$$V(y^f_{T+1,T} - y_{T+1}) = \sigma^2\left(1 + x'_{T+1}(X'X)^{-1}x_{T+1}\right)$$

Examples

- Inspecting graphs is often a good way to determine a model.
- We will consider a number of different series fitting trends and seasonals.
- Beginning with trading volume of US Treasuries.

Trading vol 10Y treasuries, 02/01/2001-31/12/2007



Modeling and Forecasting Trend

- Does not look linear, try a quadratic

$$v_t = \beta_0 + \beta_1 t + \beta_2 t^2 + u_t$$

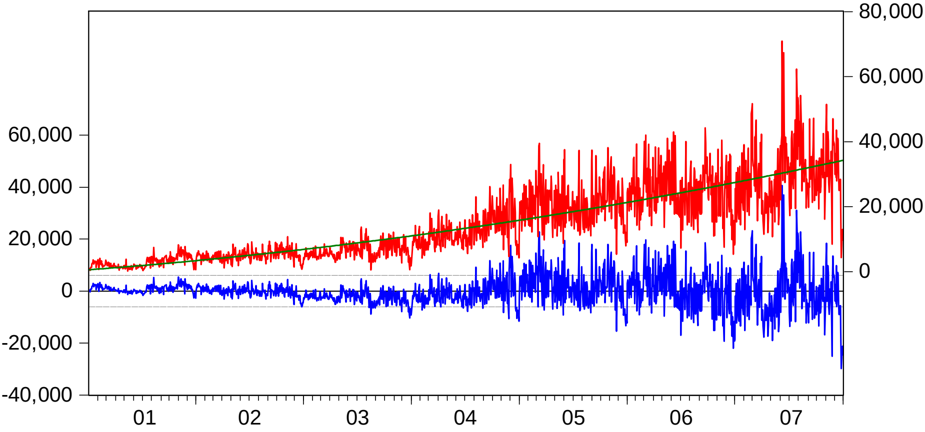
  where the variable t is a time trend or time dummy, which equals 1 in the first period of the sample, 2 in the second period, and so on.
- A variety of shapes are possible depending on the signs of β_1 and β_2. Both positive gives a slope with increasing slope of the trend.

Dependent Variable: V
Method: Least Squares
Date: 10/12/15 Time: 14:30
Sample (adjusted): 1/02/2001 12/31/2007
Included observations: 1747 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1303.208	688.5355	-1.892724	0.0586
@TREND	7.678429	1.421369	5.402138	0.0000
@TREND^2	0.005327	0.000639	8.335562	0.0000

R-squared	0.722611	Mean dependent var	14730.51
Adjusted R-squared	0.722293	S.D. dependent var	11530.05
S.E. of regression	6076.099	Akaike info criterion	20.26383
Sum squared resid	6.44E+10	Schwarz criterion	20.27321
Log likelihood	-17697.45	Hannan-Quinn criter.	20.26730
F-statistic	2271.596	Durbin-Watson stat	1.080570
Prob(F-statistic)	0.000000		

Volume, quadratic trend and residual



Modeling and Forecasting Trend

- Both t and t^2 are significant, but there is a lot of serial correlation left. DW=1.08. quadratic trend is still more U-shaped than the volume data - which we can see in the residuals.
- Try a nonlinear trend, take the logarithm of NYSE volume data , the trend of which appears to be approximately linear - so non-linear in levels but linear in logarithms - referred to as **exponential trend** or **log-linear trend** - very common in finance/economics.
- If the trend is characterised by constant growth rate β_1, then we can write:

$$v_t = \beta_0 e^{\beta_1 t + u_t}$$

so the trend is nonlinear (exponential) function of time in levels, but in logarithms we have:

$$\ln(v_t) = \ln(\beta_0) + \beta_1 t + u_t$$

therefore $\ln(v_t)$ is a linear function of time.

Logarithms of economic variables are often used since

1. prices and quantities are non-negative so the logs are defined and predicted values are always positive. What to do with zero observations?
2. the coefficients in log-log regressions can be interpreted as elasticities, % change in the dependent variable in response to a 1% change in the independent variable, so the units of measurement of the variables do not matter
3. in many cases errors are proportional to the variable, so the variance is more likely to be constant in logs,
4. the logarithms of economic variables are often closer to being normally distributed
5. the change in the logarithm is approximately equal to the growth rate and exponential growth $Y_t \approx Y_0 e^{gt} \approx Y_0(1+g)^t$ very common.
6. lots of interesting hypotheses can be tested in logarithmic models as linear restrictions.

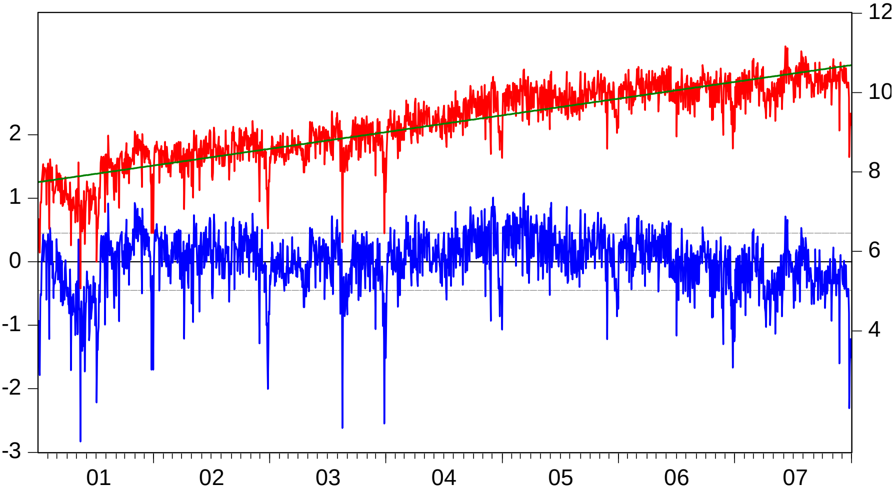
Log Volume on linear trend

Dependent Variable: LV
Method: Least Squares
Date: 10/12/15 Time: 14:32
Sample (adjusted): 1/02/2001 12/31/2007
Included observations: 1747 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.382758	0.025611	288.2607	0.0000
@TREND	0.001686	2.13E-05	79.02084	0.0000

R-squared	0.781583	Mean dependent var	9.219225
Adjusted R-squared	0.781458	S.D. dependent var	0.962264
S.E. of regression	0.449844	Akaike info criterion	1.241313
Sum squared resid	353.1179	Schwarz criterion	1.247571
Log likelihood	-1082.287	Hannan-Quinn criter.	1.243627
F-statistic	6244.294	Durbin-Watson stat	0.810867
Prob(F-statistic)	0.000000		

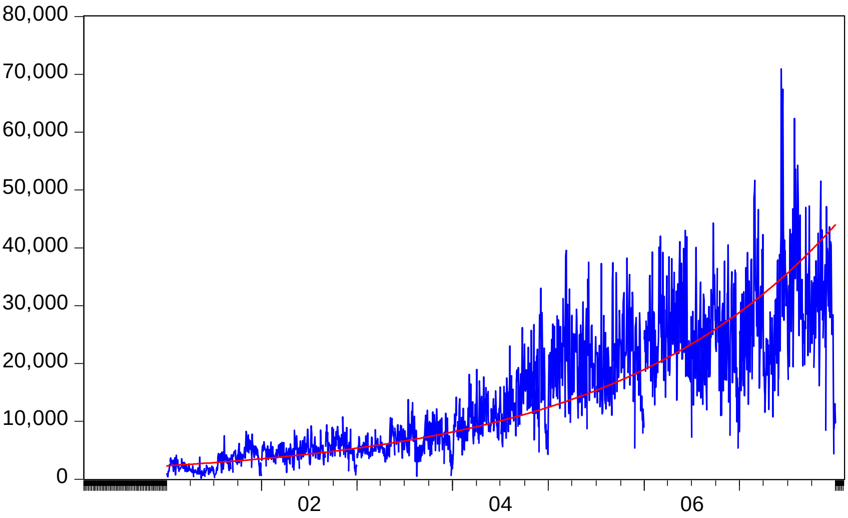
Log volume, linear trend and residual



Modeling and Forecasting Trend

- If you use the log volume data with a linear trend, the fit looks quite good. Maybe a break around 2005.
- Note R squared not comparable between linear and logarithmic dependent variables.
- Still a lot of serial correlation DW=0.8 If you use HAC standard errors t statistic on trend is 9.8 not 79.
- Equivalently you could look at the levels of the data with an exponential trend superimposed:
- Genr: ulv=resid, flv=lv-ulv, fv=exp(flv)

Volume and trend from log model



Transformed fitted values not unbiased

Dependent Variable: V
Method: Least Squares
Date: 10/12/15 Time: 15:55
Sample (adjusted): 1/02/2001 12/31/2007
Included observations: 1747 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2903.262	254.1577	11.42308	0.0000
FV	0.835557	0.014036	59.52921	0.0000

R-squared	0.670053	Mean dependent var	14730.51
Adjusted R-squared	0.669864	S.D. dependent var	11530.05
S.E. of regression	6624.874	Akaike info criterion	20.43619
Sum squared resid	7.66E+10	Schwarz criterion	20.44245
Log likelihood	-17849.02	Hannan-Quinn criter.	20.43851
F-statistic	3543.727	Durbin-Watson stat	0.908469
Prob(F-statistic)	0.000000		

Log to linear assuming log-normal

Consider model for $t = 1, 2, ..., T$

$$\ln y_t = \beta' x_t + \varepsilon_t$$

and forecast for $T + h$ conditional on x_{T+h}

$$\widehat{y}^a_{T+h,T} = \exp(\widehat{\beta}' x_{T+h})$$

the problem is that $E(y_{T+h} \mid x_{T+h}) \neq \exp E(\ln(y_t) \mid x_{T+h})$ assuming ε_t normal, correction gives

$$\widehat{y}^b_{T+h,T} = \exp(\widehat{\beta}' x_{T+h} + s^2/2)$$

which is a predictor for the conditional mean, but $\widehat{y}^a_{T+h,T}$ as a predictor for the conditional median may be more useful.

Log to linear not assuming log-normal

Consider forecast conditional on x^* for known β

$$\ln y^* = \beta' x^* + \varepsilon^*$$

$$\begin{aligned} E(y \quad | \quad x^*) &= E\left[\exp\left(\beta' x^* + \varepsilon^* \mid x^*\right)\right] \\ &= \exp(\beta' x^*) E\left[\exp\left(\varepsilon^* \mid x^*\right)\right] \end{aligned}$$

if you want to avoid normality assumption estimate

$$\widehat{E}\left[\exp\left(\varepsilon^* \mid x^*\right)\right] = h^* = T^{-1} \sum_{t=1}^{T} \exp(\widehat{\varepsilon}_t)$$

then the **smearing estimator** for prediction is

$$\widehat{y}^c_{T+h,T} = h^* \exp(\widehat{\beta}' x_{T+h}).$$

Can be used for other transformations.

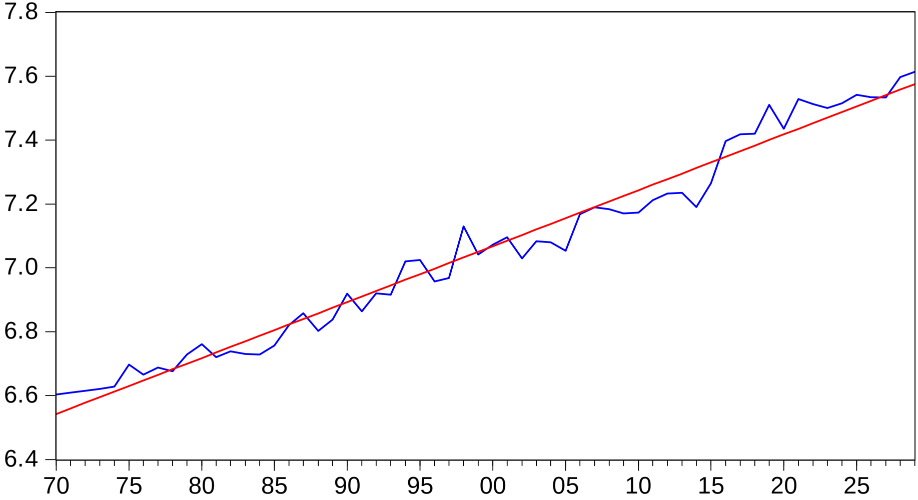
Sir Alec Cairncross

- - A trend is a trend is a trend.
  - But the question is will it bend?
  - Will it alter its course
  - through some unforeseen force
  - and come to a premature end?

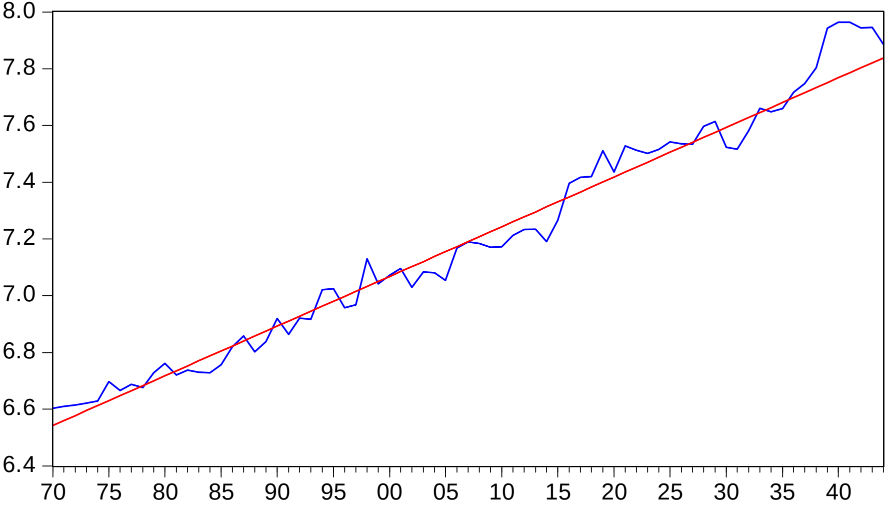
A Variant

- - A trend is a trend is a trend
  - So why should it alter or bend?
  - But the wise planner knows that the further it goes the nearer it is to the end.
- Below we estimate a linear trend on data for observations 70-129
  Then see how it does as we extend the period

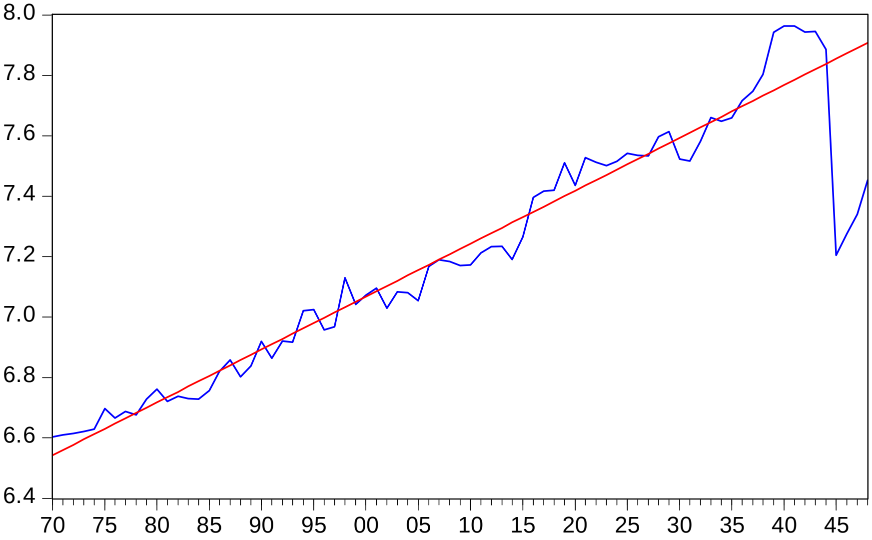
Observations 70-129 and fitted trend



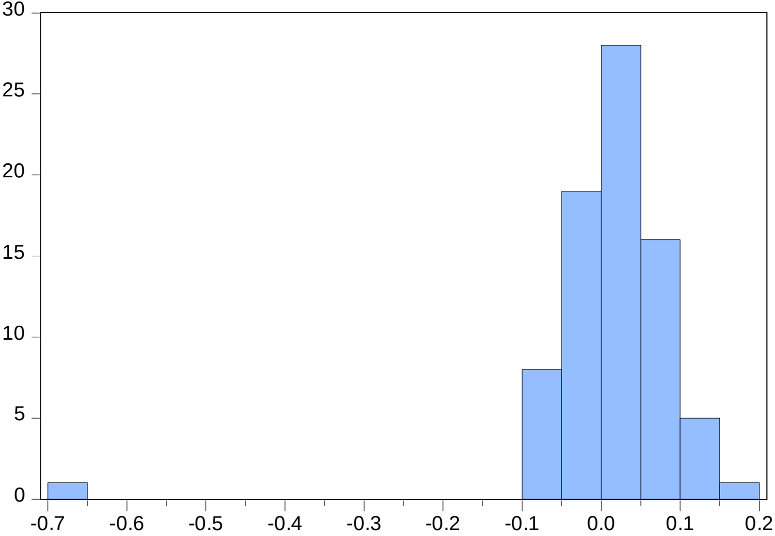
Observations 70-144 and 70-129 trend



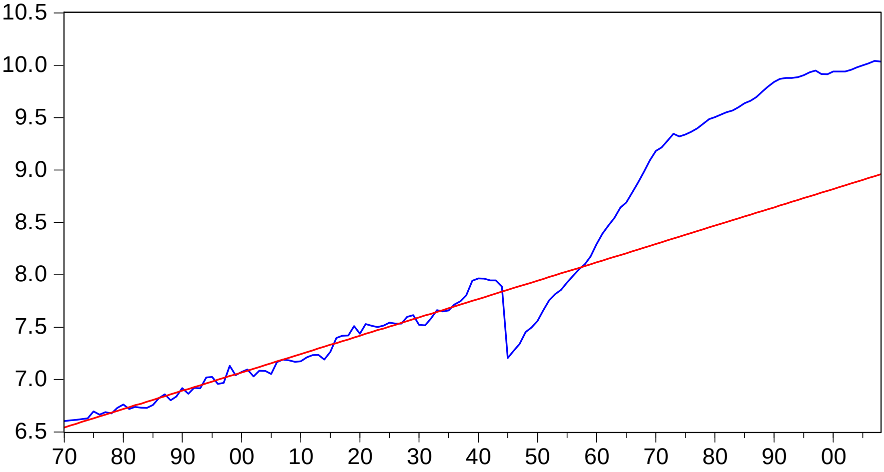
Observations 70-148



Tail risk: growth rate JB stat 4246



Japanese PC GDP 1870 2008 and 1870-1929 trend



Hodrick-Prescott filter 1

- Two sided filter that chooses a smoothed series to minimise a weighted average of squared deviation from trend and squared second difference of trend, $\Delta^2 s_t$

$$\sum_{t=1}^{T}(y_t - s_t)^2 + \lambda \sum_{t=1}^{T}\left[(s_{t+1} - s_t) - (s_t - s_{t-1})\right]^2$$

- $\lambda = \infty$ corresponds to a linear trend, $\lambda = 0$ to the original series.
- This uses $\Delta^2 s_t$ general case uses $\Delta^m s_t$. EViews lets you choose either λ or the power.
- HP suggested power $=2$: $\lambda = 1600$ for quarterly data, $\lambda = 100$ for annual, Ravn and Uhlig (REStat 2002, 371-376) suggest power equal 4 (general case $2m$) giving a much lower value, 6.25, for annual data. Comparison on Japanese data 1920-50 below.

Hodrick-Prescott filter 2

- Very popular despite the fact that it may not remove stochastic trends and this can induce spurious cycles.
- Being two sided, except at the end, it is no use for forecasting. One sided versions can be got by recursively changing end points or by using forecasts.
- Originally due to Whittaker (1923) Phillips and Jin (2015) Business Cycles, trend elimination and the HP filter, Cowles Foundation Discussion Paper 2005 discuss its properties and looks at the effect of the 2007-8 crisis on the US using different trends.

James D Hamilton (2017) Why you should never use the HP filter

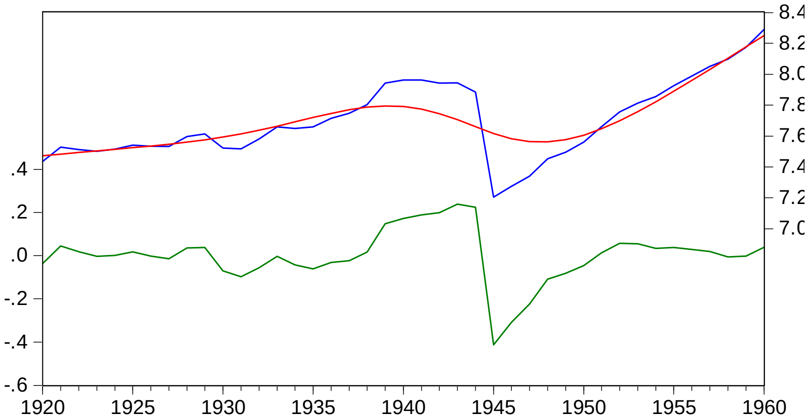
.

- Produces series with spurious dynamics that have no basis in the underlying DGP
- A one sided version does not solve these problems
- Formalising the problem produces values for λ far below 1600 for quarterly data
- There's a better alternative: a regression of y_{t+h} on the four (for quarterly) most recent values as of date t provides a robust alternative to detrending that achieves the objectives sourght by HP users with none of its drawbacks. Corresponds to a Beveridge Nelson, long horizon, trend in some cases. Linear projection.

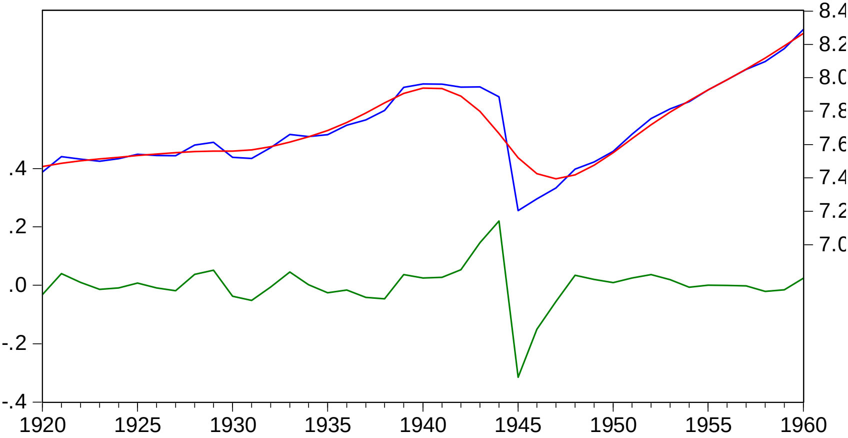
Estimating HP filter

- In Eviews
    - Click on the variable, LJ
    - Click proc on the menu of the box that comes up
    - Choose Hodrick Prescott filter
    - Name smoothed series and set lambda or power

HP Filter lambda 100



HP Filter lambda 6.25



Modeling and Forecasting Seasonality 1

- A seasonal pattern is one that regularly repeats, e.g. every year - so not present in annual data or lower frequency.
- There are within day seasonal patterns in high frequency financial data.
- Let s be the number of observations in a year. So $s = 4$ for quarterly data, $s = 12$ for monthly data (though months not the same length), $s \approx 52$ if we have weekly data.
- Lot of non-systematic seasonality: Easter and Ramadan do not always fall in the same quarter, number of working days in the month varies etc.
- Can use seasonally adjusted series for some purposes, but may need to forecast the unadjusted series and so need to model seasonality

- Seasonal dummy variables, where we assume four seasons are:

$$
\begin{aligned}
D_1 &= (1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, \ldots.) \\
D_2 &= (0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, \ldots.) \\
D_3 &= (0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, \ldots.) \\
D_4 &= (0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, \ldots.)
\end{aligned}
$$

Modeling and Forecasting Seasonality 2

- The pure seasonal dummy model is:

$$y_t = \sum_{i=1}^{4} \gamma_i D_{it} + \varepsilon_t$$

- We have a different intercept for each season, each γ_i are seasonal factors. Can't include an intercept, causes exact multicollinearity.
- If we include $s - 1$ seasonal dummies and an intercept, the intercept coefficient is that for the omitted season, e.g. quarter 1, the coefficients of the dummies give the difference from the omitted season.

$$y_t = \gamma_1 + \sum_{i=2}^{4} (\gamma_i - \gamma_1) D_{it} + \varepsilon_t$$

- This parameterisation facilitates testing differences between seasons.

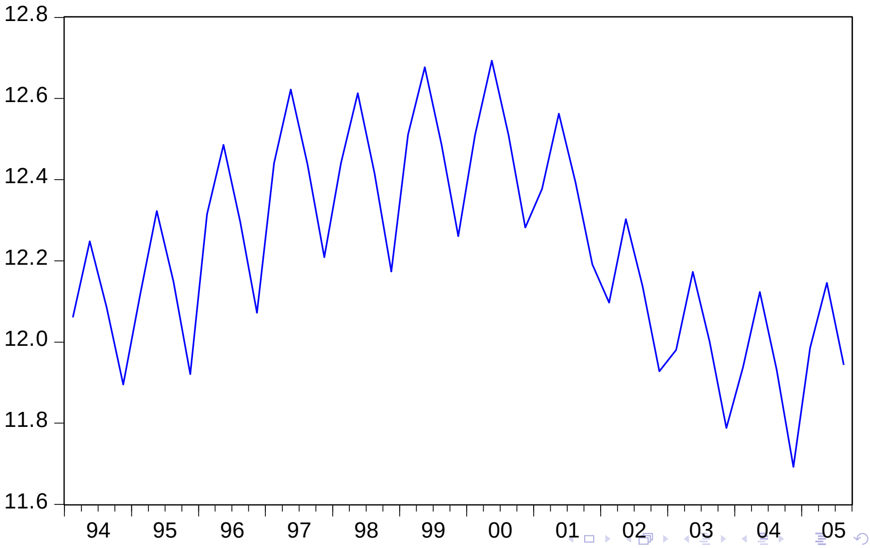
Modeling and Forecasting Seasonality 3

- You might also include the trend in the regression - seasonal pattern as well as an upward trend, also you might include calendar effects e.g. holiday variation, HDV, trading day variation, TDV (Stock market data for example, Easter dummy)
- So the full forecasting model, which gives the **point forecast** would be:

$$\begin{aligned}
\widehat{y}_{T+h,T} &= \widehat{\beta}_1(T+h) + \sum_{i=1}^{s} \widehat{\gamma}_i D_{i,T+h} + \sum_{i=1}^{v_1} \widehat{\delta}_i^{HD} HDV_{i,T+h} \\
&\quad + \sum_{i=1}^{v_2} \widehat{\delta}_i^{TD} TDV_{i,T+h}
\end{aligned}$$

- **Interval** and **density** forecast are computed in exactly the same manner as when dealing with the trends only case

Example Log Coca production in Columbia



4 Seasonals and a Quadratic Trend

Dependent Variable: LD
Method: Least Squares
Date: 10/17/15 Time: 17:46
Sample: 1994Q1 2005Q3
Included observations: 47

Variable	Coefficient	Std. Error	t-Statistic	Prob.
@SEAS(1)	12.00795	0.045011	266.7776	0.0000
@SEAS(2)	12.19557	0.045572	267.6135	0.0000
@SEAS(3)	12.02064	0.046045	261.0624	0.0000
@SEAS(4)	11.78750	0.047908	246.0426	0.0000
@TREND	0.041758	0.003976	10.50193	0.0000
@TREND^2	-0.001031	8.36E-05	-12.33296	0.0000

R-squared	0.870517	Mean dependent var	12.23285
Adjusted R-squared	0.854726	S.D. dependent var	0.246480
S.E. of regression	0.093945	Akaike info criterion	-1.773462
Sum squared resid	0.361855	Schwarz criterion	-1.537273
Log likelihood	47.67637	Hannan-Quinn criter.	-1.684583
Durbin-Watson stat	0.601107		

Constant, 3 seasonals (Q4 base) and a Quadratic trend

Dependent Variable: LD
Method: Least Squares
Date: 10/17/15 Time: 17:40
Sample: 1994Q1 2005Q3
Included observations: 47

Variable	Coefficient	Std. Error	t-Statistic	Prob.
@SEAS(1)	0.220444	0.039317	5.606785	0.0000
@SEAS(2)	0.408070	0.039299	10.38379	0.0000
@SEAS(3)	0.233133	0.039317	5.929537	0.0000
C	11.78750	0.047908	246.0426	0.0000
@TREND	0.041758	0.003976	10.50193	0.0000
@TREND^2	-0.001031	8.36E-05	-12.33296	0.0000

R-squared	0.870517	Mean dependent var	12.23285
Adjusted R-squared	0.854726	S.D. dependent var	0.246480
S.E. of regression	0.093945	Akaike info criterion	-1.773462
Sum squared resid	0.361855	Schwarz criterion	-1.537273
Log likelihood	47.67637	Hannan-Quinn criter.	-1.684583
F-statistic	55.12878	Durbin-Watson stat	0.601107
Prob(F-statistic)	0.000000		

Constant, 3 seasonals (Q1 base) and a Quadratic trend, can drop SEAS(3)

Dependent Variable: LD
Method: Least Squares
Date: 10/17/15 Time: 17:54
Sample: 1994Q1 2005Q3
Included observations: 47

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12.00795	0.045011	266.7776	0.0000
@SEAS(2)	0.187626	0.038366	4.890370	0.0000
@SEAS(3)	0.012690	0.038406	0.330407	0.7428
@SEAS(4)	-0.220444	0.039317	-5.606785	0.0000
@TREND	0.041758	0.003976	10.50193	0.0000
@TREND^2	-0.001031	8.36E-05	-12.33296	0.0000

R-squared	0.870517	Mean dependent var	12.23285
Adjusted R-squared	0.854726	S.D. dependent var	0.246480
S.E. of regression	0.093945	Akaike info criterion	-1.773462
Sum squared resid	0.361855	Schwarz criterion	-1.537273
Log likelihood	47.67637	Hannan-Quinn criter.	-1.684583
F-statistic	55.12878	Durbin-Watson stat	0.601107
Prob(F-statistic)	0.000000		

Seasonal alternatives

- Seasonally adjust data with e.g. X-13 in EViews.
- Dummy variables
- Seasonal lags, e.g. lag 4 in quarterly data
- Trigonometric functions, e.g. Harvey, *Forecasting Structural Time Series Models and the Kalman Filter*.
- Seasonal exponential smoothing, see Diebold Holt-Winters smoothing to allow for seasonality

Holt Winters (Exponential) Smoothing, estimates a local mean

- Initialise at time $t = 1 : \bar{y}_1 = y_1$
- Update: $\bar{y}_t = \alpha y_t + (1 - \alpha)\bar{y}_{t-1}$
- Forecast: $\bar{y}_{T+h,T} = \bar{y}_T$
- Note
$$\bar{y}_t = \sum_{j=0}^{t-1} \alpha(1-\alpha)^j y_{t-j}$$
- Choose $0 < \alpha < 1$ to reflect the relative importance of signal in y_t (large α) relative to noise (small α).

Holt Winters smoothing allowing for trend

- Initialise at time $t = 2 : \bar{y}_2 = y_2; \;\; F_2 = y_2 - y_1$
- Update

$$\begin{aligned} \bar{y}_t &= \alpha y_t + (1-\alpha)(\bar{y}_{t-1} + F_{t-1}) \\ F_t &= \beta(\bar{y}_t - \bar{y}_{t-1}) + (1-\beta)F_{t-1}, \\ t &= 3, 4, ..., T \end{aligned}$$

- Forecast: $y_{T+h,T} = \bar{y}_T + hF_t$

Holt Winters smoothing allowing for seasonal period s

- Initialise at $t = s : \bar{y}_s = \sum_{t=1}^{s} y_t / s$; $F_s = 0$; $G_j = y_j - \bar{y}_s$, $j = 1, 2, .., s$
- Update for $s + 1, s + 2, ..., T$

$$
\begin{aligned}
\bar{y}_t &= \alpha(y_t - G_{t-s}) + (1-\alpha)(\bar{y}_{t-1} + F_{t-1}) \\
F_t &= \beta(\bar{y}_t - \bar{y}_{t-1}) + (1-\beta)F_{t-1}, \\
G_t &= \gamma(y_t - \bar{y}_t) + (1-\gamma)G_{t-s}
\end{aligned}
$$

- Forecast

$$
\begin{aligned}
y_{T+h,T} &= \bar{y}_T + hF_T + G_{T+h-s};\ h = 1, 2, ..., s \\
y_{T+h,T} &= \bar{y}_T + hF_T + G_{T+h-2s};\ h = s+1, s+2, ..., 2s
\end{aligned}
$$

Smoothing

- Optimal forecasts for some processes
- Only produce point forecasts not confidence intervals
- Basis of the Riskmetrics approach
- Need to determine parameters: trial and error
- Easily automated and often works very well

Trigonometric Functions (Fourier Series)

- Monthly seasonal component for $m_t = 1, 2, .., 12$

$$s_t = \sum_{j=1}^{6} \left[\xi_j \cos\left(\frac{2\pi j}{12} m_t\right) + \xi_j^* \sin\left(\frac{2\pi j}{12} m_t\right) \right]$$

- $\frac{2\pi j}{12}$ are seasonal frequencies and ξ_j and ξ_j^* estimated parameters. Equivalent to 12 dummy variables, but brings out the cyclical nature.
- Frequency $2\pi/12$, fundamental frequency, period of 12 months, others harmonics, waves with periods less than a year.
- More parsimonious represenation often works well, e.g. just fundamental frequencies

$$s_t = \xi \cos\left(\frac{2\pi j}{12} m_t\right) + \xi^* \sin\left(\frac{2\pi j}{12} m_t\right)$$

- Can make ξ_j and ξ_j^* time varying with Kalman Filter.

Deterministic and stochastic trends and cycles

- Trend stationary

$$\begin{aligned} y_t &= \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \gamma t + \varepsilon_t \\ -1 &< \rho_1 + \rho_2 < 1 \end{aligned}$$

- Difference stationary

$$\begin{aligned} y_t &= \alpha + y_{t-1} + \varepsilon_t \\ &= \alpha t + \sum_{j=0}^{t-1} \varepsilon_{t-j} + y_0 \end{aligned}$$

deterministic trend, stochastic trend and initial condition.

Exercises week 3: Direct and indirect forecasts

Assume that the data are measured as deviations from the mean to remove intercepts, and over a sample $t = 1, 2, ..., T$ you estimate by least squares an AR1 model

$$y_t = \rho_a y_{t-1} + \varepsilon_t^a \tag{1}$$

and a 3 period ahead linear projection

$$y_t = \rho_b y_{t-3} + \varepsilon_t^b. \tag{2}$$

(1) What are your estimates for ρ_a and ρ_b?
(2) What are your forecasts for three periods ahead by each method as a function on y_T.
(3) Using (1) compare the forecasts. Under what circumstances would they be the same.

Forecasting with EViews 1.

Use EViews file Electricity. When you forecast make sure there is space in the program for the forecasts. The data ends in 2017, but in the spreadsheet has years up to 2024. Set the sample as beginning in 2000 and ending in 2024.

- We will forecast the logs. Plot INSTALLED ELECTRICTY CAPACITY (MEGAWATT) and its logarithm LC. Notice LC is much more linear.
- Try to give your variabls simple names that you can remember. You cannot use C as a variable in EViews since that is reserved for the constant.
- Quick estimate and equation put in LC C @TREND LC(-1). Notice the trend is not signifcant, but LC(-1) is very significant. In the file are two other models using just LC(-1) or just @TREND with a constant, together with their forecasts. Although the models are quite different the forecasts are very similar lcf and lcft.

Forecasting with EViews 2.

- Click Forecast. Change the forecast name from LCF to LCFA. In SE(optional) type LCSA. Note we choose a dynamic forecast and allow for coefficient uncertainty in the standard error calculation.
- You could adjust the sample for forecasts and SEs over just the forecast period rather than the whole period.
- OK will get you a picture of the forecast and its confidence interval over the whole sample, together with some forecast evaluation stats: RMSE etc.
- Name the equation and close it.

Forecasting with EViews 3.

- Create confidence intervals for the high estimate use: quick, generate, HLCFA=LCFA+2*LCSA . OK. for the low estimate use: quick, generate, LLCFA=LCFA-2*LCSA . OK.
- Plot LLCFA LCFA HLCFA, will give you the forecast and confidence intervals
- Look at the other forecasts and standard errors. Although the forecasts from a linear trend, LCFT and random walk with drift, LCF are almost identifcal the confidence interval for the random walk is much larger.

Holt Winters smoothing allowing for trend 1

Initialise at time $t = 2 : \bar{y}_2 = y_2; \;\; F_2 = y_2 - y_1$
Update

$$\begin{aligned} \bar{y}_t &= \alpha y_t + (1-\alpha)(\bar{y}_{t-1} + F_{t-1}) \\ F_t &= \beta(\bar{y}_t - \bar{y}_{t-1}) + (1-\beta)F_{t-1}, \\ t &= 3, 4, ..., T \end{aligned}$$

Forecast: $y_{T+h,T} = \bar{y}_T + hF_t$
Year is in column A, 1870-2008.
Japanese per-capita GDP, JYPC is in column B.
We are going to forecast Y the log of JYPC in column C.

Holt Winters smoothing allowing for trend 2

Y_BAR in column D is the local mean,
F is the trend in column E. We use 1870 and 1871 to initialise these. So Y_BAR in 1871 = Y in 1871.
Y_BAR and F use the parameters. alpha and beta, which are set to 0.3 to start and put in cells C1 and E1. They can be changed.
Y(t+3,t) is the forecast for t+3 made in t. This starts in 1874 using information from 1871.
The last column just repeats Y to make it easier to graph against the forecast.
1. Plot Y(t+3,t) against Y. Comment on the pattern. To what extent does it capture structural breaks and changes in trend.
2. Calculate the Mean Square Error of Forecast over the period 1874-2008.
3. Experiment with changing alpha and beta. What do you think are optimum values?
4. Calculate Y(t+2,t).