Modeling and Forecasting Trends

Ka-fu WONG
University of Hong Kong

February 1, 2019

Contents

1. Background

- The unobserved components approach to modeling and forecasting economic time series assumes that the typical economic time series, y_t, is made up of the sum of three independent components
  - a time trend component
  - a seasonal component
  - an irregular or cyclical component.

$$y_t = \text{time trend} + \text{seasonal} + \text{cyclical} = T_t + S_t + C_t$$

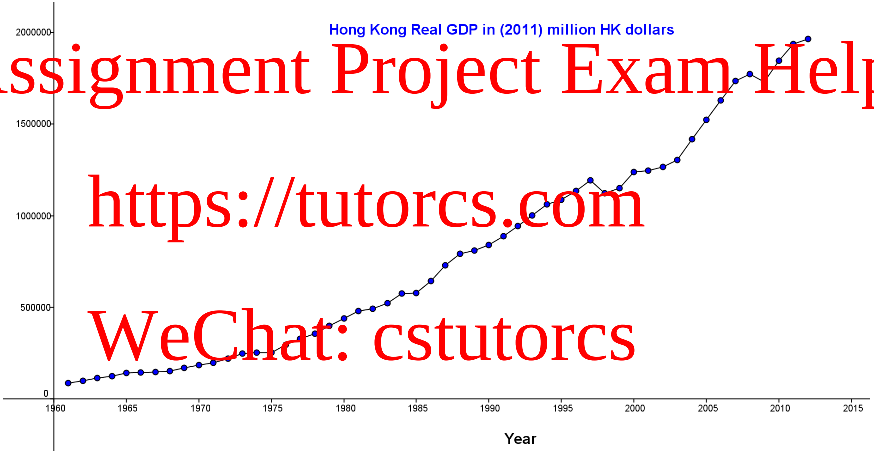
$$y_t = \text{time trend} + \text{seasonal} + \text{cyclical} = T_t + S_t + C_t$$

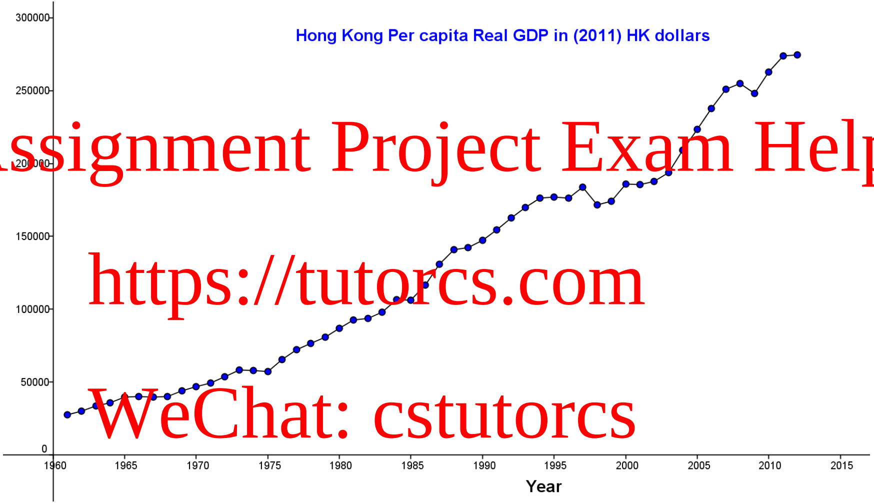
- The time trend refers to
  - *the "long-run" average behavior of the series*.
- The seasonal refers to
  - *the annual predictable cyclical behavior of the series* associated with weather patterns, holiday patterns, etc.
- The cyclical component refers to
  - *the remainder of the series after the trend and seasonal have been accounted for*.

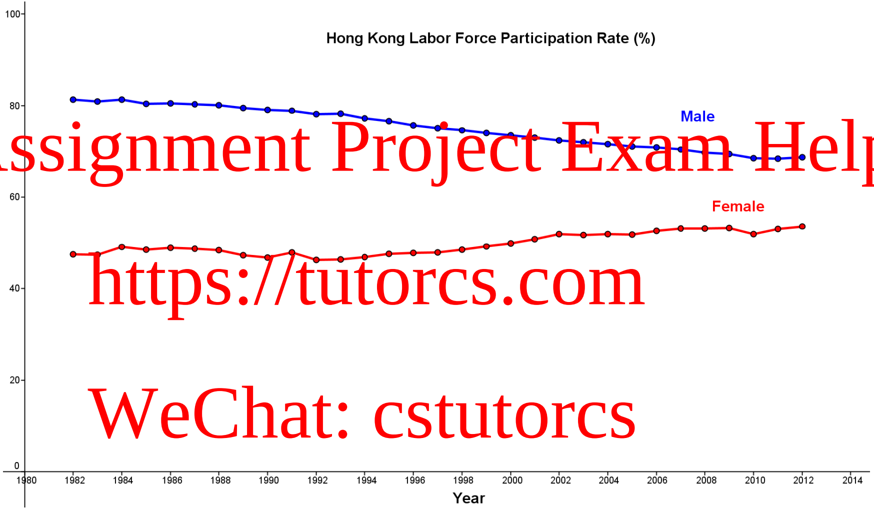
- The assumption that these components are determined independently means that each component is determined and influenced by its own set of forces and, consequently, each component can be studied separately.
- The approach is called an "unobserved components" approach because
  - we do *not directly* observe each of the three components; we only get to observe their sum.
- Our job will be to model and estimate the various components and use these estimates as the basis for forecasting the components and their sum.

- Whether the assumption underlying the unobserved components approach, that the trend, seasonal, and cyclical components are determined independently, is plausible or not is *debatable* and is, in fact, an issue of some controversy among economists.
- For example, many macroeconomists argue that economic growth (trend) and the business cycle (cyclical) are determined by a common set of forces.

2. Example of Data with Trend




Hong Kong Per capita Real GDP in (2011) HK dollars
300000
250000
200000
150000
100000
50000
0
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
Year


Hong Kong Labor Force Participation Rate (%)
Male
Female
Year
0
20
40
60
80
100
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014

3. Modeling the Trend

- If we look at Hong Kong's per capita real GDP time series or any one of your time series at *annual frequency*, the first thing that stands out is
  - the obvious tendency of the series to grow (or, in some cases, to fall) over time.
  - That is, it is immediately apparent from the time series plot that the average change in the GDP series is positive (or, in some cases such as men's labor participation rate, negative).
- This tendency is the series's *trend*.

3.1. **Linear trend model.**

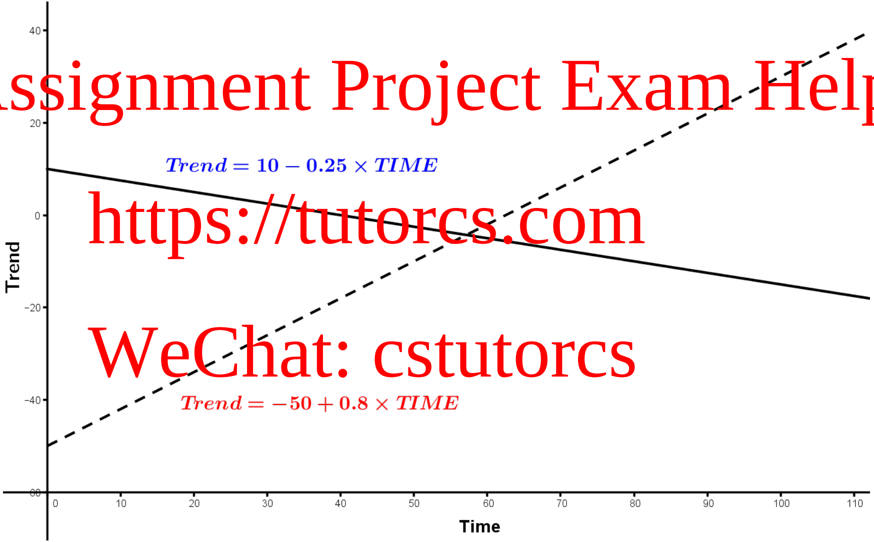
- The simplest model of the time trend is the linear trend model –

$$T_t = \beta_0 + \beta_1 t, \qquad t = 1, 2, ..., T$$

- The trend component is a straight line with intercept β_0 and slope β_1. And, $T_1 = \beta_0 + \beta_1$, $T_2 = \beta_0 + 2\beta_1$, ..., $T_T = \beta_0 + T\beta_1$.

- Note that $\beta_1 = dT_t/dt$ and $\beta_1 = T_t - T_{t-1}$. So,
  - $\beta_1 > 0$ if y has a positive trend and
  - $\beta_1 < 0$ if y has a negative trend.
- The intercept, as is often the case in econometric models, *does not have a meaningful interpretation* and its sign can be positive or negative, regardless of the trend's sign.

Examples of upward and downward trends



3.2. Polynomial trend model.

- In some cases, a linear trend is inadequate to capture the trend of a time series. A natural generalization of the linear trend model is the polynomial trend model –

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + ... + \beta_p t^p, \qquad t = 1, ..., T$$

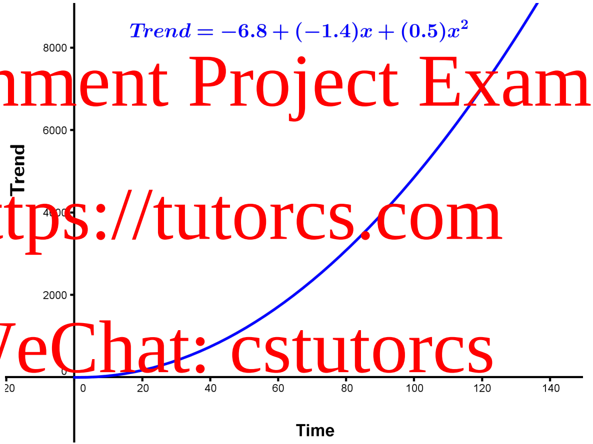
  where p is a positive integer.

- Note that the linear trend model is a special case of the polynomial trend model ($p = 1$).
- For economic time series we almost never require $p > 2$. That is, if the linear trend model is not adequate, the quadratic trend model will usually work:

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2$$

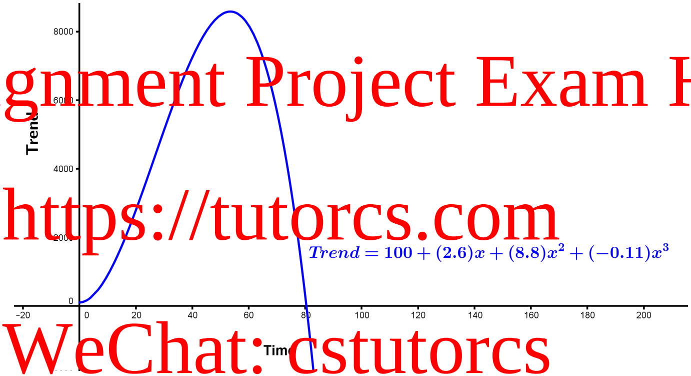
- In the quadratic model, $dT_t/dt = \beta_1 + 2t\beta_2$

Examples of Quadratic Trends



QuadraticTrend00.ggb

Examples of Cubic Trends



CubicTrend00.ggb

4. The Log Linear Trend Model

Another alternative to the linear trend model is the log linear trend model, which is also called the exponential trend model:

$$T_t = \beta_0 \exp(\beta_1 t)$$

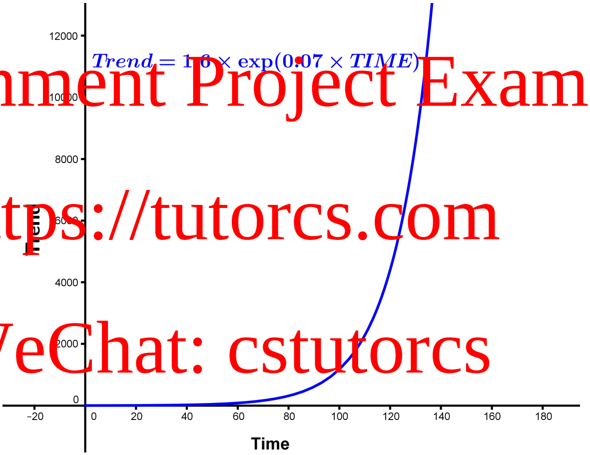
or, taking natural logs on both sides,

$$\log(T_t) = \log(\beta_0) + \beta_1 t$$

so that the log of the trend component is linear.
Note that for the log linear trend model

$$\begin{aligned}
\beta_1 &= \log(T_t) - \log(T_{t-1}) \\
&= \log(T_t/T_{t-1}) \\
&= \log\{[T_{t-1} + (T_t - T_{t-1})]/T_{t-1}\} \\
&= \log[1 + (T_t - T_{t-1})/T_{t-1}] \\
&\approx (T_t - T_{t-1})/T_{t-1} = \% \text{ change in } T_t
\end{aligned}$$

Examples of Exponential Trends

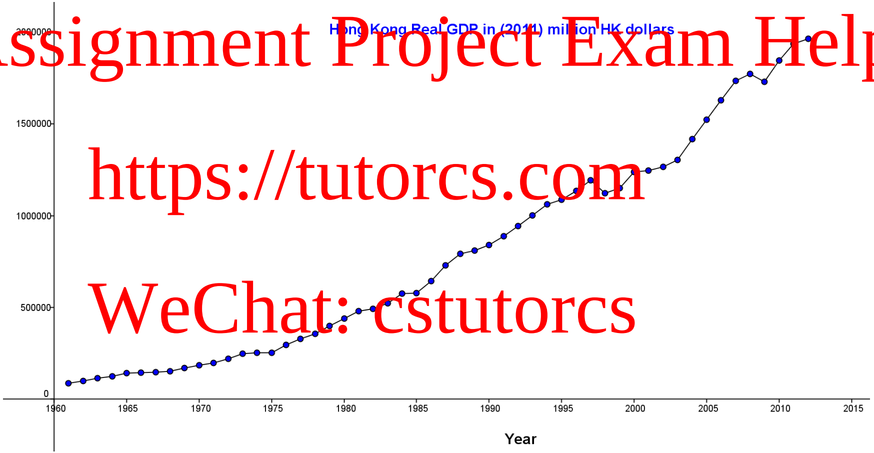


ExponentialTrend00.ggb

5. Which trend model to use?

- Knowing the differences among these models can help us decide whether the linear, quadratic or log linear trend model is more appropriate for our data.
  - In the linear trend model the change in T is constant over time.
  - In the quadratic trend model the change in T has a linear trend.
  - In the log linear trend model the growth rate that is constant over time.
- However, in practice, it is not always obvious by simply looking at the time series plot which form the trend model should take – linear, log linear, quadratic? Other?
- *Practice and experience are the most helpful*

What kind of trend model would best fit the data?



Which trend model to use?

6. All Deterministic Trend Models

- Note that in all of these models, the trend is deterministic, i.e., perfectly forecastable. For instance, in the linear trend model, the time trend at time t is modeled as

$$T_t = \beta_0 + \beta_1 t$$

  Note that the subscript "t" in T_t is the time at which the trend is evaluated. The capital T denotes the Trend.

- Given a trend model and the parameters, it is straightforward to produce the forecast of the time trend at time $T + h$, while standing at time $t = T$.

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

  $T + h$ is the time at which the trend is evaluated. The capital T in T_{T+h} denotes the Trend.

- (Later in the course we will talk about *stochastic trend* models, in which the trend of the series is not perfectly forecastable.)

- The parameters of the trend model are almost always unknown — even if we correctly specify the shape *of* the trend (linear, quadratic, exponential, ...). So, in practice, we will have to estimate these parameters.
- Estimation based on a sample of data will introduce errors (called sampling or estimation error) into our trend forecasts.

Beware of sampling errors!

7. Estimating the Trend Model

- Our assumption at this point is that our time series, y_t, can be modeled as

$$y_t = T_t(\theta) + \epsilon_t$$

where

  - T_t is one of the trend models we discussed earlier,
  - θ is the set of parameters; $\theta = (\beta_0, \beta_1)$ in a linear trend model.
  - ϵ_t denotes the other factors (i.e. the seasonal and cyclical components) that determine y_t.

- Since θ is unknown, it is natural to estimate the trend model via the least squares approach (based on quadratic loss)

$$\hat{\theta} = \arg\min_{\theta} \sum_{t=1}^{T} (y_t - T_t(\theta))^2$$

7.1. Models that can be estimated by OLS.

- Linear Trend model:

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0,\beta_1} \sum_{t=1}^{T} [y_t - (\beta_0 + \beta_1 t)]^2$$

- Quadratic Trend model:

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \arg\min_{\beta_0,\beta_1,\beta_2} \sum_{t=1}^{T} \left[y_t - (\beta_0 + \beta_1 t + \beta_2 t^2)\right]^2$$

- Log-Linear Trend model:

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0,\beta_1} \sum_{t=1}^{T} [\ln(y_t) - (\ln(\beta_0) + \beta_1 t)]^2$$

7.2. **Models that have to be estimated numerically by Nonlinear LS.**

- Exponential Trend Models

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0,\beta_1} \sum_{t=1}^{T} [y_t - \beta_0 \exp(\beta_1 t)]^2$$

Question: What is the difference between linear trend models and exponential trend models?

8. Property of the Ordinary Least Squares Estimators

- Under the assumptions of the unobserved components model, the OLS estimator of the linear and quadratic trend models is
  - unbiased,

$$E(\hat{\theta}_T) = \theta$$

  - consistent,

$$\text{plim} \quad \hat{\theta}_T = \theta$$

  and
  - asymptotically efficient

$$\lim_{T \to \infty} Var(\hat{\theta}_T) \leq \lim_{T \to \infty} Var(\hat{\theta}'_T)$$

- Standard regression procedures can be applied to test hypotheses about the β's and construct interval estimates.
  - This is true even though the regression errors will generally be *serially correlated and heteroskedastic*. (Why?)

9. Forecasting the Trend

- Once we have specified a trend model, standing at time T,
  - When θ is known, our forecast of the h-step ahead trend component of y will simply be

$$T_{T+h}(\theta)$$

  - When θ is unknown, we can estimate it. Our forecast of the h-step ahead trend component of y will simply be

$$T_{T+h}(\hat{\theta})$$

We would like to forecast y_{T+h} based on all information available at time T.

- Assume that the trend is linear.

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

- If we know the true parameters, the part $\beta_0 + \beta_1(T+h)$ can be forecasted perfectly.
- Can we forecast ϵ_{T+h}? Sometimes YES. Sometimes NO.
  - NO when ϵ_t is known to be an independent zero-mean random noise.

If ϵ_t is an i.i.d. sequence with zero mean, then

$$\begin{aligned} E(\epsilon_{T+h} \mid \Omega_T) &= E(\epsilon_{T+h}) \quad \because \text{ Independence} \\ &= E(\epsilon_t) \quad \because \text{ Identically Distributed} \\ &= 0 \quad \because \text{ zero mean} \end{aligned}$$

where Ω_T is used to denote "information available at time T".

For the time being, assume ϵ_t to be an independent zero-mean random noise.

9.1. **When parameters are known.**

- Forecast :

$$E(y_{T+h} \mid \Omega_T) = y_{T+h,T} = \beta_0 + \beta_1(T+h)$$

- Forecast error:

$$\begin{aligned} e_{T+h,T} &= y_{T+h} - y_{T+h,T} \\ &= \beta_0 + \beta_1(T+h) + \epsilon_{T+h} - [\beta_0 + \beta_1(T+h)] \\ &= \epsilon_{T+h} \end{aligned}$$

This forecast error ϵ_{T+h} cannot be avoided even if we know the model parameters. Thus, ϵ_{T+h} is often called fundamental uncertainty!

9.2. When parameters are unknown:

- Forecast

$$E(y_{T+h} \mid \Omega_T) = \hat{y}_{T+h,T} = \hat{\beta}_0 + \hat{\beta}_1(T+h)$$

That is, the unknown parameters are substituted by the estimate from the OLS regression.

- Forecast error:

$$\begin{aligned} e_{T+h,T} &= y_{T+h} - \hat{y}_{T+h,T} \\ &= \beta_0 + \beta_1(T+h) + \epsilon_{T+h} - \left[\hat{\beta}_0 + \hat{\beta}_1(T+h)\right] \\ &= \epsilon_{T+h} + (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)(T+h) \end{aligned}$$

That is, the forecast error consists of the part of fundamental uncertainty and a part that is due to estimation uncertainty.

forecast uncertainty = fundamental uncertainty + estimation uncertainty

10. Density forecast

10.1. **When parameters are known.** The distribution of the forecast error will simply be the distribution of ϵ_{T+h}. That is, for any real number c,

$$Prob(e_{T+h,T} < c) = Prob(y_{T+h} - y_{T+h,T} < c) = Prob(\epsilon_{T+h} < c)$$

Further assume the ϵ's are i.i.d. $N(0, \sigma^2)$, while continuing to ignore parameter uncertainty. Then the density forecast will be that

$$e_{T+h,T} = y_{T+h} - y_{T+h,T} = \epsilon_{T+h} \sim N(0, \sigma^2)$$

and

$$\begin{aligned} y_{T+h} &= y_{T+h,T} + \epsilon_{T+h} \\ &\sim N(y_{T+h,T}, \sigma^2) \end{aligned}$$

Note that this density forecast depends on the unknown parameter σ^2. To make the density forecast operational, we can replace σ^2 with an unbiased and consistent estimator,

$$y_{T+h} \sim N(y_{T+h,T}, \hat{\sigma}^2)$$

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} \epsilon_t^2, \qquad \epsilon_t = y_t - \beta_0 - \beta_1 t$$

Note: Divided by T, not $T-2$ because there is no loss of degree of freedom.

10.2. **When parameters are unknown.** The distribution of the forecast error will be more spread out than the distribution of ϵ_{T+h}.

$$e_{T+h,T} = y_{T+h} - \hat{y}_{T+h,T} = \beta_0 + \beta_1(T+h) + \epsilon_{T+h} - \left[\hat{\beta}_0 + \hat{\beta}_1(T+h)\right]$$

$$= \epsilon_{T+h} + (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)(T+h)$$

Under usual assumptions, the forecast error due to parameter uncertainty is asymptotically normal.

$$(\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)(T+h) \overset{A}{\sim} N(0, .)$$

Thus, $e_{T+h,T}$ will be asymptotically normal.

$$e_{T+h,T} \sim N(0, \sigma_e^2)$$

$$y_{T+h} \sim N(\hat{y}_{T+h,T}, \sigma_e^2)$$

The unknown variance can be estimated by

$$\hat{\sigma}_e^2 = \frac{1}{T-2}\sum_{t=1}^{T} e_t^2, \qquad e_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 t$$

Note: Divided by $T-2$, not T because there is a loss of two degrees of freedom due to the estimation of β_0 and β_1.

Then we act as though y_{T+h} is distributed as $N(\hat{y}_{T+h,T}, \hat{\sigma}_e^2)$. Or equivalently

$$Z \equiv \frac{y_{T+h} - \hat{y}_{T+h,T}}{\hat{\sigma}_e} \sim N(0,1)$$

$$Prob(y_{T+h} - \hat{y}_{T+h,T} < c) = Prob\left(Z < \frac{c}{\hat{\sigma}_e}\right)$$

- Further, we can construct (symmetric) interval forecasts of y_{T+h} according to:

$$\hat{y}_{T+h,T} \pm \hat{\sigma}_e Z_{1-\alpha/2}$$

is a $(1-\alpha) \times 100\%$ forecast interval for y_{T+h}, where $Z_{1-(\alpha/2)}$ is the $(1-(\alpha/2)) \times 100$ percentile of the $N(0,1)$ distribution.

- For example, if $\alpha = .05$ then we obtain a 95-percent forecast interval for y_{T+h},

$$\hat{y}_{T+h,T} \pm 1.96\hat{\sigma}_e$$

since 1.96 is the 97.5 percentile of the $N(0,1)$.

- Often, we write the 95-percent forecast interval for y_{T+h},

$$\hat{y}_{T+h,T} \pm 2\hat{\sigma}_e$$

as an approximation, because ....

because in small samples

$$Z \equiv \frac{y_{T+h} - \hat{y}_{T+h,T}}{\hat{\sigma}_e}$$

is unlikely normal (partly due to the estimated $\hat{\sigma}_e$) but some other distribution like "Student-t".

10.3. **Interpretation of 95% forecast interval:**

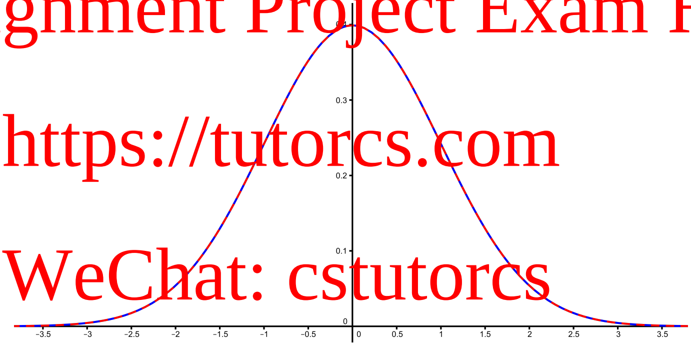
- Imagine that we have the chance to see 1000 samples generated using a similar DGP (data generating process).

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

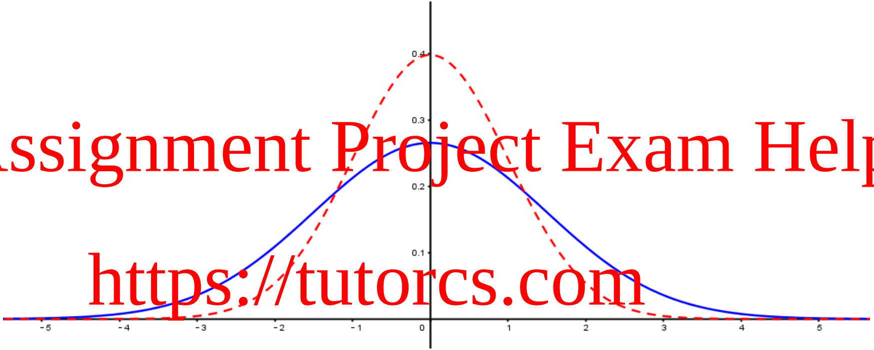
- If, for each of the example, we estimate the parameter, produce the 95% forecast interval as described, we would expect 950 of the 1000 intervals (i.e., 95% of them) will turn out to include the actual value of y_{T+h}.

R001_ForecastInterval01.R

Observe how the additional parameter uncertainty will change the density forecast.



ParameterUncetainty01.ggb



ParameterUncetainty01.ggb

11. Selecting Forecasting Models

11.1. R-square as a criteria.

- Consider the mean squared error (MSE)

$$MSE \equiv \frac{\sum_{t=1}^{T} e_t^2}{T}$$

where T is the sample size, $e_t = y_t - \hat{y}$ and $\hat{y}$ is a predicted value of regression model, say $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 t$ based on a linear trend model.

- Note that models with smallest MSE is also the model with smallest sum of squared residuals, because scaling the sum of squared residuals with a constant $(1/T)$ will not change the ranking.

- Consider the R-square, a typical measure of goodness of fit of a regression model

$$R^2 = 1 - \frac{\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2} = 1 - \frac{MSE}{\sum_{t=1}^{T}(y_t - \bar{y})^2/T}$$

- Because $\sum_{t=1}^{T}(y_t - \bar{y})^2/T$ depends only on data but not on model, models with the largest R-square is also the model with the smallest MSE.

R^2(Model A) $>$ R^2(Model B) $\Longleftrightarrow$MSE(Model A) $<$ MSE(Model B)

- Perfect fit happens when $MSE = 0$. $MSE = 0$ if and only if $R^2 = 1$.

- Although the R-square (R^2) may be a good measure of in-sample fit, it is not a useful measure for out-of-sample fit.
- The reason is that adding an additional regressor in the model always results in an R-square no less than the one with less regressors.

$$R^2(\text{Model A + variable}) \;>\; R^2(\text{Model A})$$

$$\Downarrow$$

$$R^2(\text{Trend model of order } p+1) > R^2(\text{Trend model of order } p)$$

- This effect of obtaining an arbitrary R^2 by adding more explanatory is often known as *in-sample over-fitting or data mining*.

MSE is a biased estimator of out-of-sample h-step-ahead prediction error variance. Indeed, as discussed earlier, forecast error consists of two parts

(1) Fundamental uncertainty (unavoidable even if we know the parameters)
(2) Parameter uncertainty (increases with the number of parameters in the model)

The simplistic use of MSE as a criteria of model fitness ignores the impact of parameter uncertainty.

Beware of the added forecast uncertainty due to added parameters!

11.2. **Adjusted R-square.**

- Adjusted R-square accounts for the parameter uncertainty by penalizing the addition of parameters via the adjustment of degree of freedom.

$$\begin{aligned}\bar{R}^2 &= 1 - \frac{\sum_{t=1}^{T} e_t^2/(T-k)}{\sum_{t=1}^{T}(y_t-\bar{y})^2/(T-1)} \\ &= 1 - \frac{(\frac{T}{T-k})MSE}{\sum_{t=1}^{T}(y_t-\bar{y})^2/(T-1)} \\ &= 1 - \frac{s^2}{\sum_{t=1}^{T}(y_t-\bar{y})^2/(T-1)}\end{aligned}$$

- Again, $\sum_{t=1}^{T}(y_t-\bar{y})^2/(T-1)$ depends only on data but not on model, models with the largest R-square is also the model with the smallest s^2.

$\bar{R}^2$(Model A) $>$ $\bar{R}^2$(Model B) $\Longleftrightarrow s^2$(Model A) $<$ s^2(Model B)

- Because $T/(T-k)$ increases with the number of parameters (k) in the model, adding an additional regressor needs not raise the Adjusted R^2 or lower the s^2. Adding an additional regressor raises the Adjusted R^2(or lower s^2) only if the additional regressor reduces MSE more than its contribution to the penalty factor $T/(T-k)$.
- In short, if we were to use this criterion, we would be equivalently choosing the model to minimize s^2.

$$M^* = \arg\min_{M} \quad s^2(M)$$

11.3. **AIC and SIC.** AIC and SIC are two alternatives that will penalize the number of parameters included in a model.

(1) Akaike information criterion (AIC):

$$AIC = e^{\left(\frac{2k}{T}\right)} \frac{\sum_{t=1}^{T} e_t^2}{T}$$

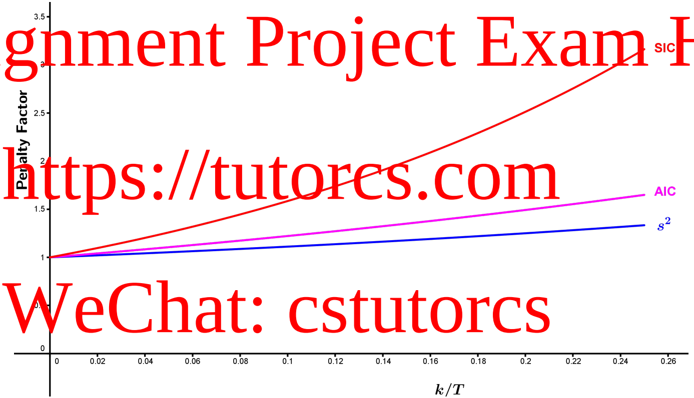
$$M^* = \arg\min_{M} \quad AIC(M)$$

(2) Schwarz information criterion (SIC):

$$SIC = T^{\left(\frac{2k}{T}\right)} \frac{\sum_{t=1}^{T} e_t^2}{T}$$

$$M^* = \arg\min_{M} \quad SIC(M)$$

The variation of criteria with k/T



PenaltyFactor01.ggb

11.4. **Desirable properties of model selection criteria.**

Consistency:

A model selection criterion is consistent if the following conditions are met:

- When the true model – i.e. the data-generating process (DGP) – is among the models considered, the probability of selecting the true DGP approaches 1 as the sample size gets large.
- When the true model is not among those considered, so that it is impossible to select the true DGP, the probability of selecting the best approximation to the true DGP approaches 1 as the sample size gets large.

Asymptotically efficiency:

An asymptotically efficient model selection criterion chooses a sequence of models, as the sample size gets large, whose 1-step-ahead forecast error variances approach the one that would be obtained using the true model with known parameters at a rate at least as fast as that of any other model selection criteria.

	Consistent?	Asymptotically efficient?
Adjusted R^2	No	No
AIC	No	Yes
SIC	Yes	No

- Usually AIC and SIC suggest the same model.
- When AIC and SIC suggest different models, we usually choose the model selected by SIC because the SIC often suggests a more parsimonious model (i.e., smaller number of parameters).

Keep it small and simple!

Note on statistical packages.

Some software packages may report variants of AIC and SIC, which are ranking preserved transformation of AIC and SIC. The common ones are

- $\ln(AIC) = \ln(MSE) + 2k/T$
- $\ln(SIC) = \ln(MSE) + (2k/T) \times \ln(T)$

$$x_1 > x_2 \iff \ln(x_1) > \ln(x_2)$$

That is, if

$$M_1^* = \arg\min_M \; AIC(M) \quad \text{and} \quad M_2^* = \arg\min_M \; \ln\left[AIC(M)\right]$$

we must have

$$M_1^* = M_2^*$$

Caution!

Some students mistakenly compare absolute values of AIC to choose the best model, i.e.,

$$M_3^* = \arg\min_M |AIC(M)| \quad \text{and} \quad M_4^* = \arg\min_M |\ln[AIC(M)]|$$

Yes, we must have

$$M_3^* = M_1^*$$

because $AIC(M)$ is positive and thus taking absolute value on has no impact.

But, we NEED NOT have

$$M_4^* = M_1^*$$

because $\ln(AIC(M))$ can be negative and taking absolute value may change the ranking of the criterion and hence the model selection.

In fact, if $\ln(AIC(M))$ are negative for all model M considered, the model selection critieria

$$M_4^* = \arg\min_M |\ln[AIC(M)]|$$

will give us the worst model!


y
x
y = ln(x)

12. Out-of-sample fitting

Suppose we have a data sample $y_1, \ldots, y_T$ and we are about to produce one-step-ahead forecast.

- Break it up into two parts (where $n \ll T$):
  - $y_1, \ldots y_{T-n}$ (first $T-n$ observations)
  - $y_{T-n+1}, \ldots, y_T$ (last n observations)

OutofSamplefit01.ggb

- Fit the shortened sample, $y_1, \ldots, y_{T-n}$ to various trend models :
  - linear, quadratic, log linear, exponential.

For each estimated trend model, forecast $y_{T-n+1}, \ldots, y_T$ and compute the forecast errors: $e_1, \ldots, e_n$.

Fixed Scheme.

$$e_1 = y_{T-n+1} - \hat{y}_{T-n+1,T-n}$$

$$e_2 = y_{T-n+2} - \hat{y}_{T-n+2,T-n}$$

$$\ldots$$

$$e_n = y_{T-n+n} - \hat{y}_{T-n+n,T-n}$$

Recursive Scheme.

$$
\begin{aligned}
e_1 &= y_{T-n+1} - \hat{y}_{T-n+1,T-n} \\
e_2 &= y_{T-n+2} - \hat{y}_{T-n+2,T-n+1} \\
&\vdots \\
e_n &= y_{T-n+n} - \hat{y}_{T-n+n,T-n+n-1}
\end{aligned}
$$

- Compare the errors across the various models
  - time series plots (of the forecasts and actual values of $y_{T-n+1}, \ldots, y_T$; of the forecast errors)
  - tables of the forecasts, actuals, and
  - mean squared prediction errors (MSPE)

$$MSPE = \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

- Choose the trend model with the smallest MSPE.

AIC/SIC are in-sample criteria. MSPE is an out-of-sample criterion!

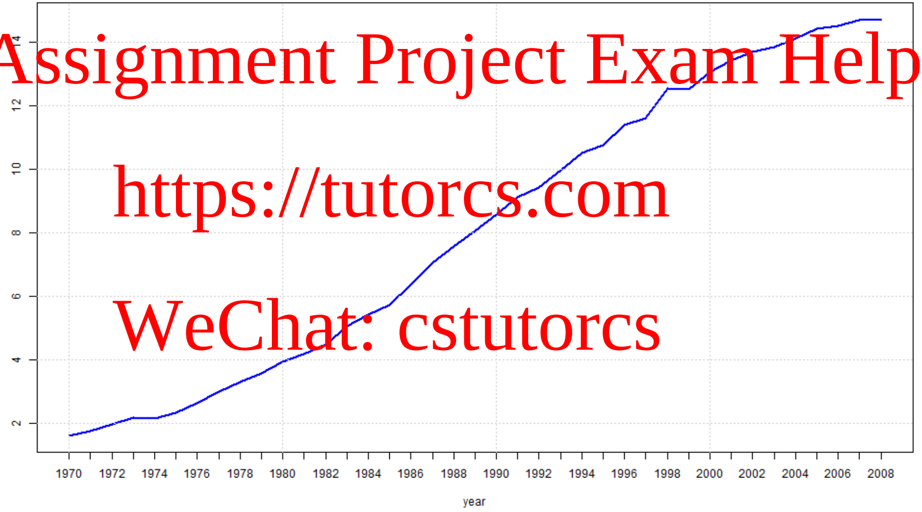
- The advantage of this approach is that we are actually comparing the trend models in terms of their *out-of-sample* forecasting performance.
- A disadvantage is that the comparison is based on models fit over *$T-n$ observations* rather than the T observations we have available.
- Note that if you do use this approach and, for example, settle on the quadratic model, then when you proceed to construct your forecasts for $T+1$, ... you should use the quadratic model fit to the full T observations in your sample.

13. Application – Hong Kong Electricity Consumption

- Annual domestic electricity Consumption data (in Terajoule) from 1970 to 2012, a total of 43 observations, were obtained from the Hong Kong Census and Statistics Department website.[1]
  - Domestic = Total - Exports
- We save four observations (2009 to 2012) for checking the accuracy of our model out of sample.
  - i.e., estimation and model selection uses only 39 observations (1970-2008)

[1] http://www.censtatd.gov.hk/hkstat/sub/sp90.jsp?tableID=127&ID=89&productType=9

Electricity Consumption (10,000 Terajoule)

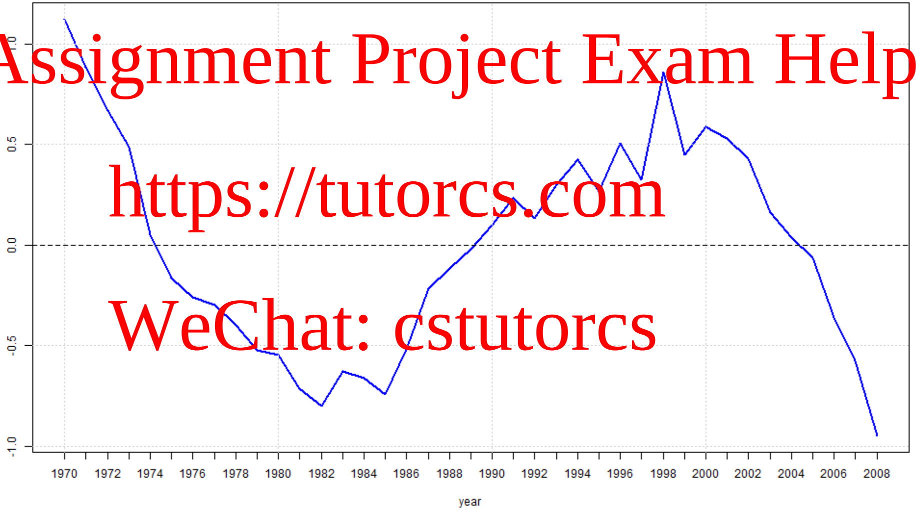


13.1. **Linear Trend Model.**

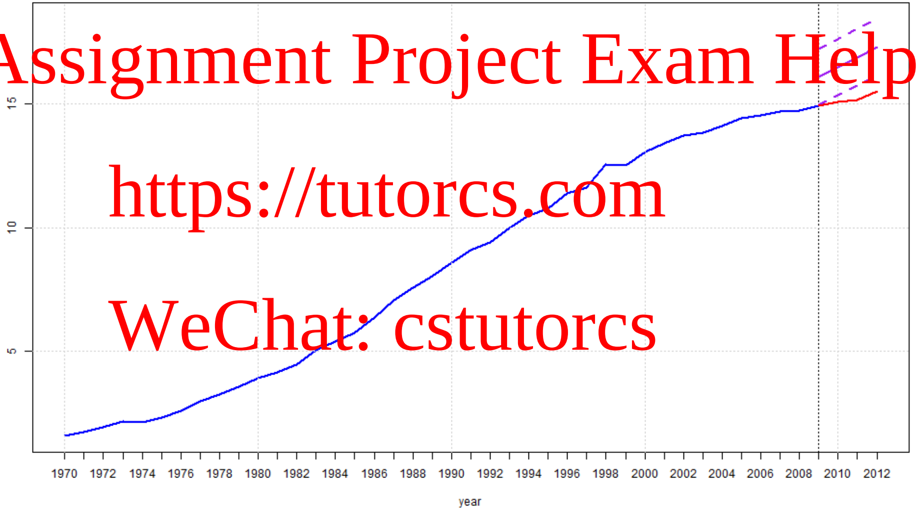
	coefficient	Std. Error	t-Stat	p-value
(Intercept)	0.07683	0.173638	0.443	0.661
TIME	0.400151	0.007566	52.887	<2e-16

Residual standard error: 0.5318 on 37 degrees of freedom
Multiple R-squared: 0.9869; Adjusted R-squared: 0.9866
F-statistic: 2797 on 1 and 37 DF, p-value: $< 2.2e\text{-}16$
AIC=65.367; SIC=70.357

Residuals (Linear Trend Model)



Forecast (Linear Trend Model)

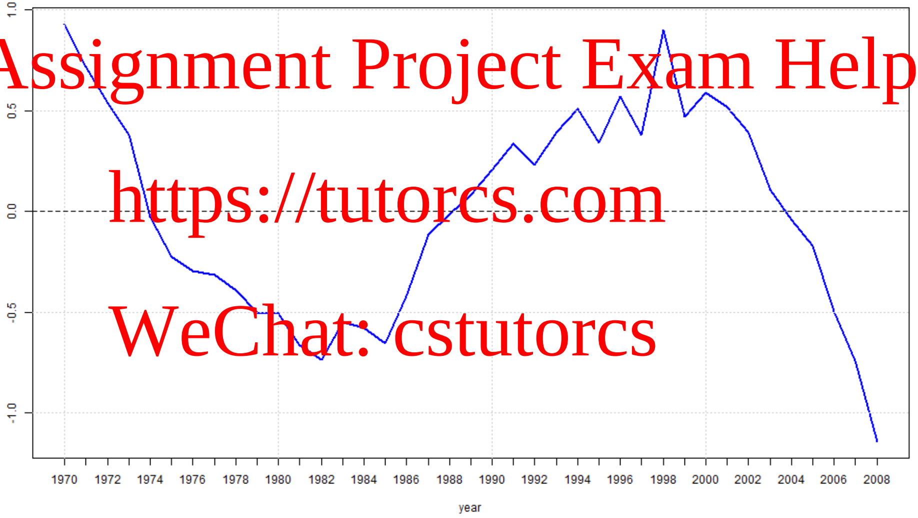


13.2. **Quadratic Trend Model.**

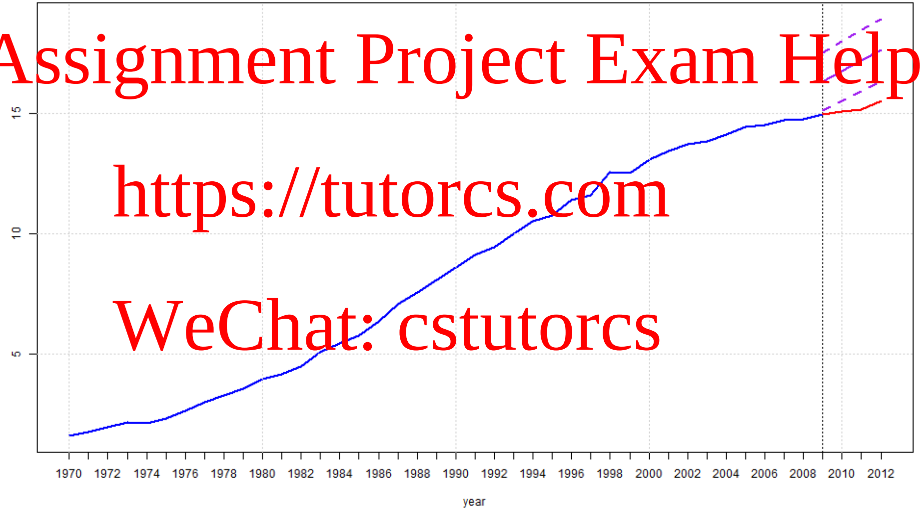
	coefficient	Std. Error	t-Stat	p-value
(Intercept)	0.303786	0.2683343	1.132	0.265
TIME	0.3669454	0.0309374	11.861	5.39E-14
TIME2	0.0008301	0.0007501	1.107	0.276

Residual standard error: 0.5302 on 36 degrees of freedom
Multiple R-squared: 0.9874, Adjusted R-squared: 0.9867
F-statistic: 1408 on 2 and 36 DF, p-value: < 2.2e-16
AIC=66.062; SIC=72.716

Residuals (Quadratic Trend Model)



Forecast (Quadratic Trend Model)

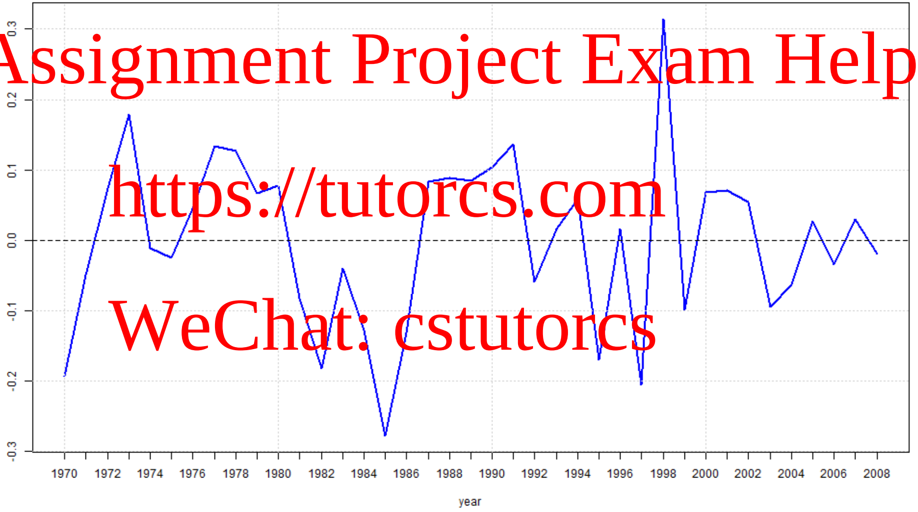


13.3. **Cubic Trend Model.**

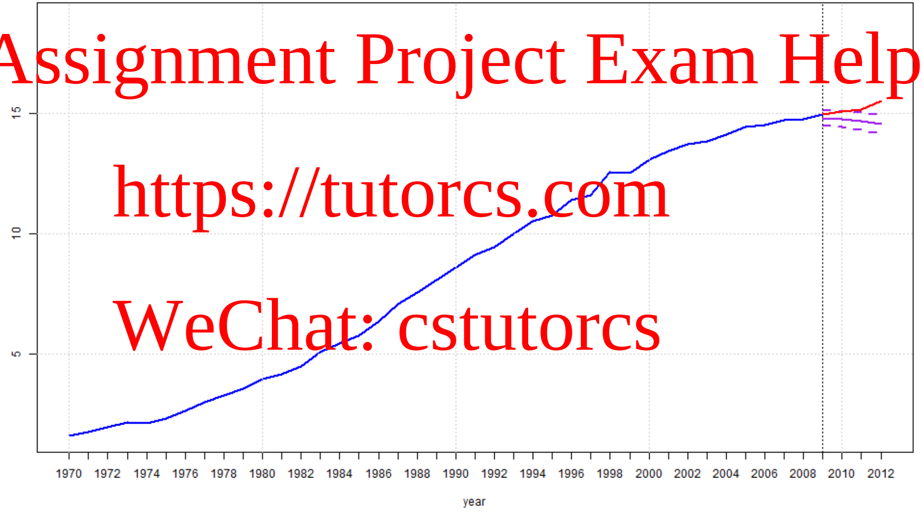
	coefficient	Std. Error	t-Stat	p-value
(Intercept)	1.83E+00	8.81E-02	20.81	$< 2e-16$
TIME	-6.48E-02	1.88E-02	-3.44	0.00152
TIME2	2.75E-02	1.09E-03	25.28	$< 2e-16$
TIME3	-4.44E-04	1.79E-05	-24.85	$< 2e-16$

Residual standard error: 0.1246 on 35 degrees of freedom
Multiple R-squared: 0.9993; Adjusted R-squared: 0.9993
F-statistic: 1.721e+04 on 3 and 35 DF, p-value: $< 2.2e-16$
AIC=-46.02; SIC=-37.703

Residuals (Cubic Trend Model)



Forecast (Cubic Trend Model)



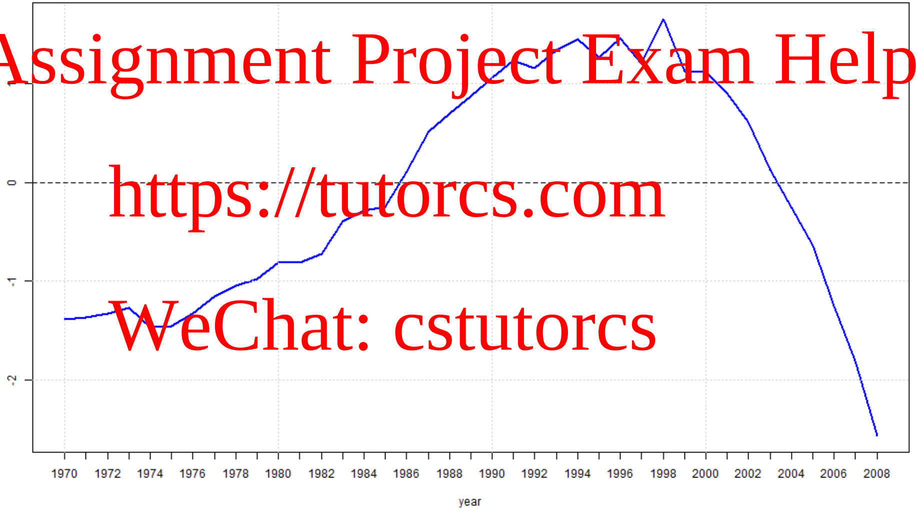
13.4. **Exponential Trend Model.**

$$y_t = \exp(A + B \times TIME_t) + \epsilon_t$$

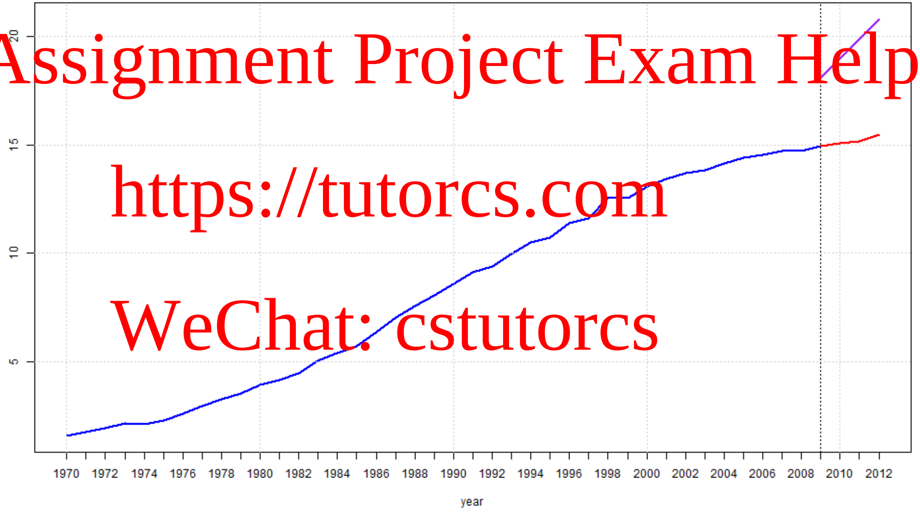
	Coefficient	Std. Error	t-Stat	p-value
A	1.048769	0.07427	14.12	<2e-16
B	0.046192	0.002398	19.26	<2e-16

Residual standard error: 1.179 on 37 degrees of freedom
Number of iterations to convergence: 7
Achieved convergence tolerance: 1.535e-06
AIC=127.453; SIC=132.448

Residuals (Exponential Trend Model)



Forecast (Exponential Trend Model)



13.5. **Comparison.**

	In-sample criteria		Out-of-sample
	AIC	SIC	MSPE
Linear Trend	65.367	70.357	2.387
Quadratic Trend	66.062	72.716	3.339
Cubic Trend	-46.02	-37.703	0.311
Exponential Trend	127.458	132.448	18.926

- The best model chosen based on in-sample criteria (AIC and SIC) also yield the best out-of-sample performance (measured in terms of MSPE).
- *No forecast interval will be produced for Exponential Trend models.*