Model Regresi Untuk Data Deret Waktu (1)

Pika Silvianti, M.Si

pikasilvianti@apps.ipb.ac.id

Sumber: Hyndman, R.J and Athanasopoulos, G. 2013. Forecasting: principles and practice

https://www.otexts.org/fpp/4

Review model regresi

- **forecast variable y** is sometimes also called the regressand, dependent or explained variable.
- **predictor variable x** is sometimes also called the regressor, independent or explanatory variable.

Review model regresi

- Konsep dasarnya adalah bahwa kita meramalkan variabel y dengan asumsi y memiliki hubungan linear dengan variabel x.
- Model ini disebut regresi “sederhana" karena hanya digunakan satu variabel prediktor x
- Model regresi sederhana: $y = \beta_0 + \beta_1 x + \varepsilon$

Ringkasan

- x_i is observation i on variable x.
- $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ is the **simple linear model** with intercept β_0 and slope β_1. The error is denoted by ε_i.
- $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$ is the **estimated regression model** with intercept $\hat{\beta}_0$ and slope $\hat{\beta}_1$. The estimated error or **residual** is denoted by e_i.
- $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ is the fitted or estimated **regression line**; $\hat{y}_i$ is the **fitted value** corresponding to observation y_i.

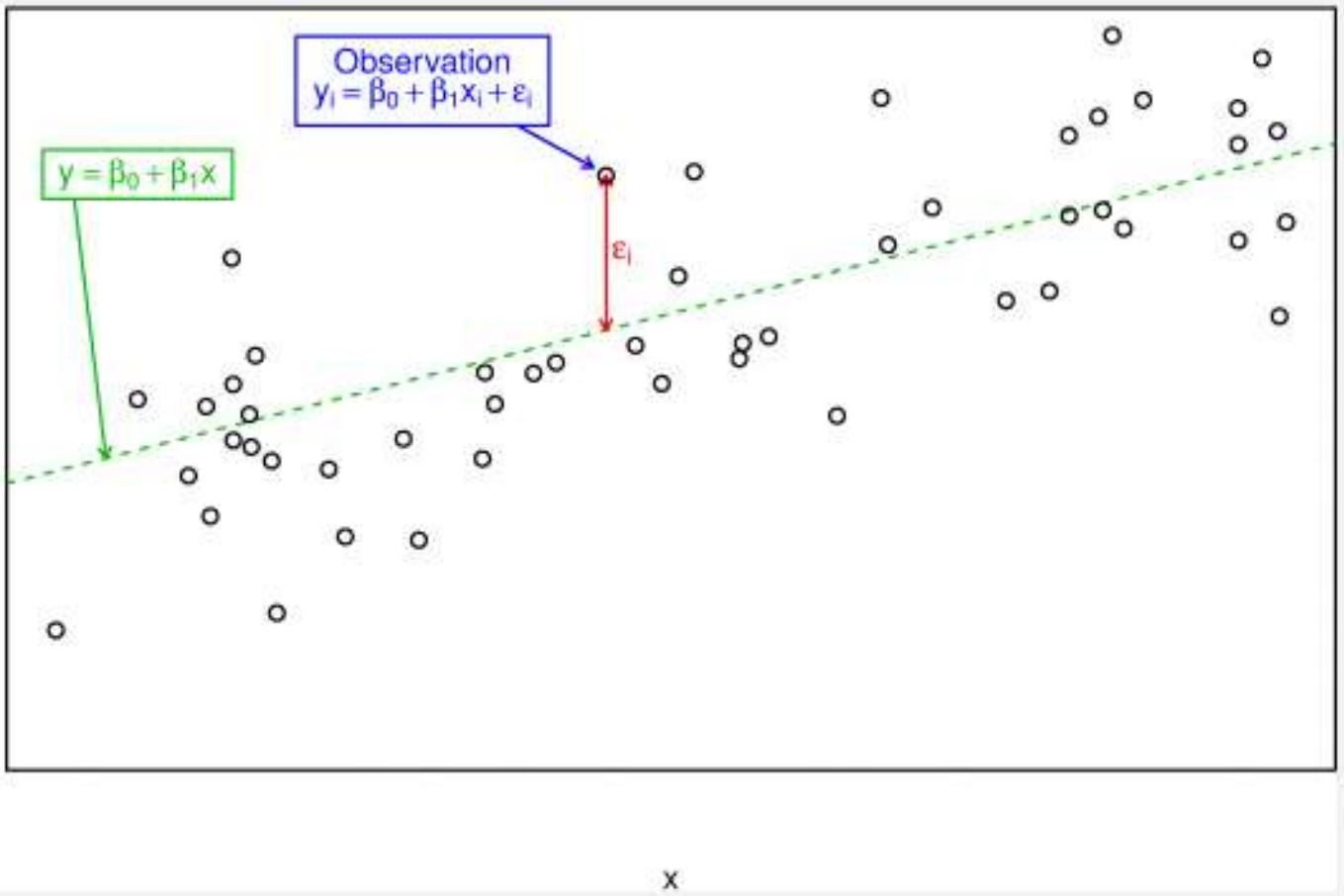


Figure 4.1: An example of data from a linear regression model.

Asumsi untuk error

- Nilai tengahnya 0; jika tidak, hasil ramalannya akan mengalami bias sistematik
- Tidak berkorelasi diri (autocorrelated); jika berkorelasi diri, maka ramalannya tidak efisien karena ada informasi lebih lanjut untuk dieksploitasi dalam data.
- tidak terkait dengan variabel prediktor; kalau tidak ,akan ada informasi lebih lanjut yang harus dimasukkan di bagian sistematis dari model.

Regresi dengan Data Deret Waktu

- Bila menggunakan regresi untuk prediksi(prediction), kita sering mempertimbangkan data time series dan ingin meramalkan(forecast) masa depan.
- Ada beberapa masalah yang timbul dengan data time series tapi tidak dengan data cross-sectional

Example 4.3 US consumption expenditure

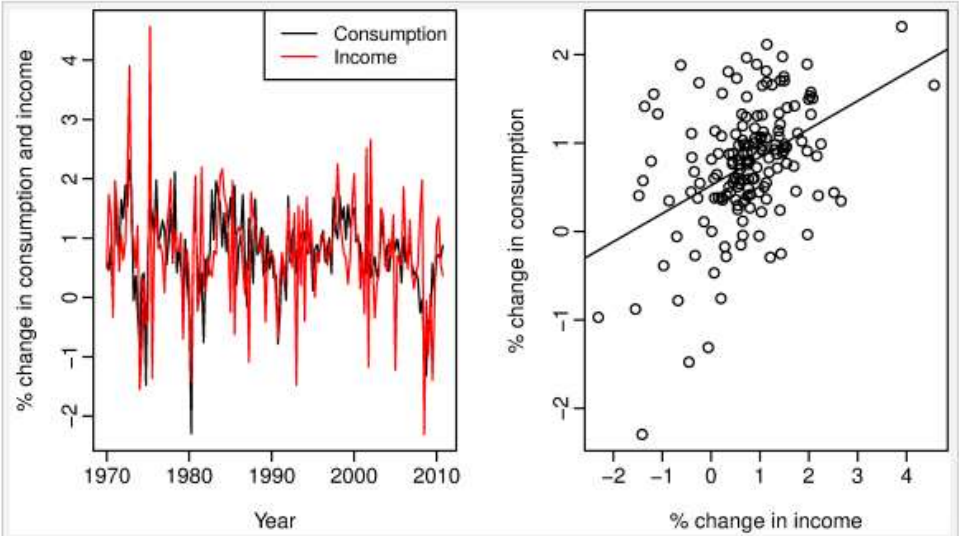


Figure 4.10: Percentage changes in personal consumption expenditure for the US.

- Figure 4.10 shows time series plots of quarterly percentage changes (growth rates) of real personal consumption expenditure (CC) and real personal disposable income (II) for the US for the period March 1970 to Dec 2010.
- $\hat{C} = 0{,}51 + 0{,}32I$
- these show that a 1% increase in personal disposable income will result to an average increase of 0.84% in personal consumption expenditure.
- **We are interested in forecasting consumption for the four quarters of 2011.**

- variabel prediktor (income dalam hal ini) diperlukan untuk dimasukkan ke dalam model regresi , tetapi nilai ini tidak diketahui sebelumnya.
- Salah satu solusi untuk masalah ini adalah dengan menggunakan " scenario based forecasting ".

Scenario based forecasting

- peramal berasumsi skenario yang mungkin untuk variabel predictor. Misalnya pembuat kebijakan AS mungkin ingin meramalkan konsumsi jika ada pertumbuhan 1%income untuk setiap kuartal pada tahun 2011. Atau penurunan 1% income untuk masing-masing kuartal.
- Peramalan yang dihasilkan dihitung dan ditunjukkan pada Gambar 4.11.

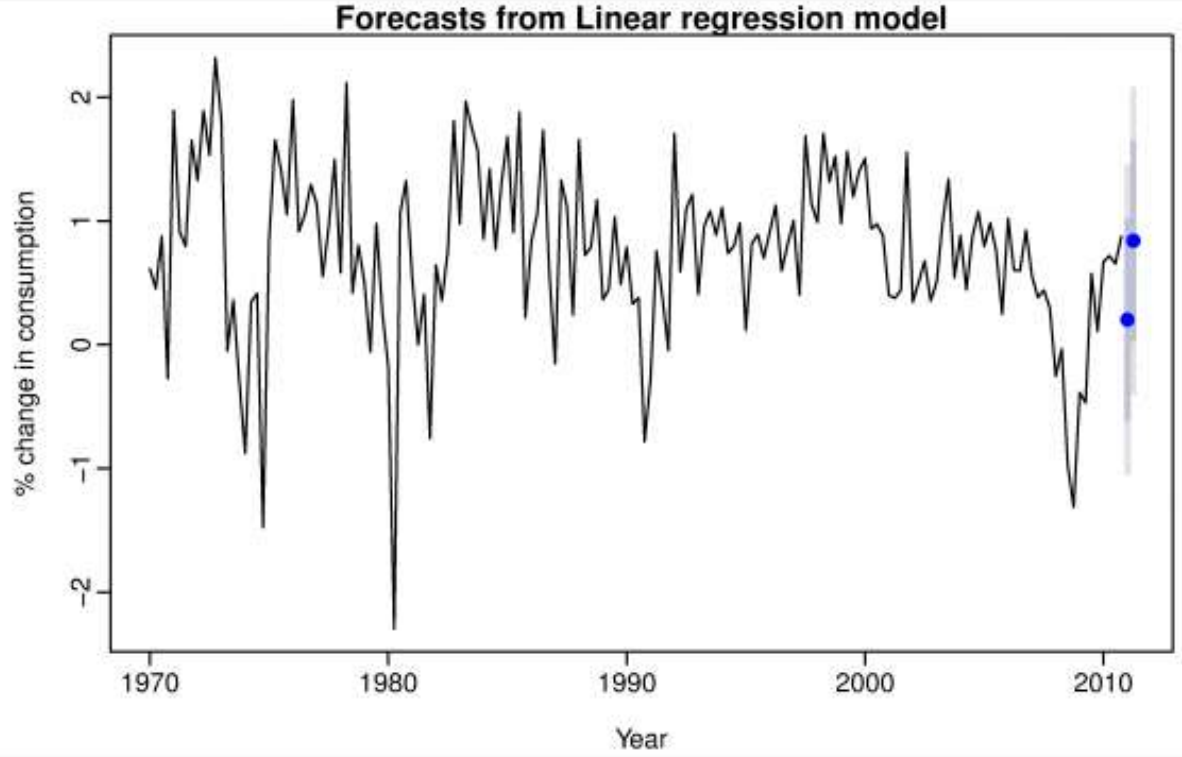


Figure 4.11: Forecasting percentage changes in personal consumption expenditure for the US.

R code

```
fcast <- forecast(fit.ex3, newdata=data.frame(income=c(-1,1)))
plot(fcast, ylab="% change in consumption", xlab="% change in income")
```

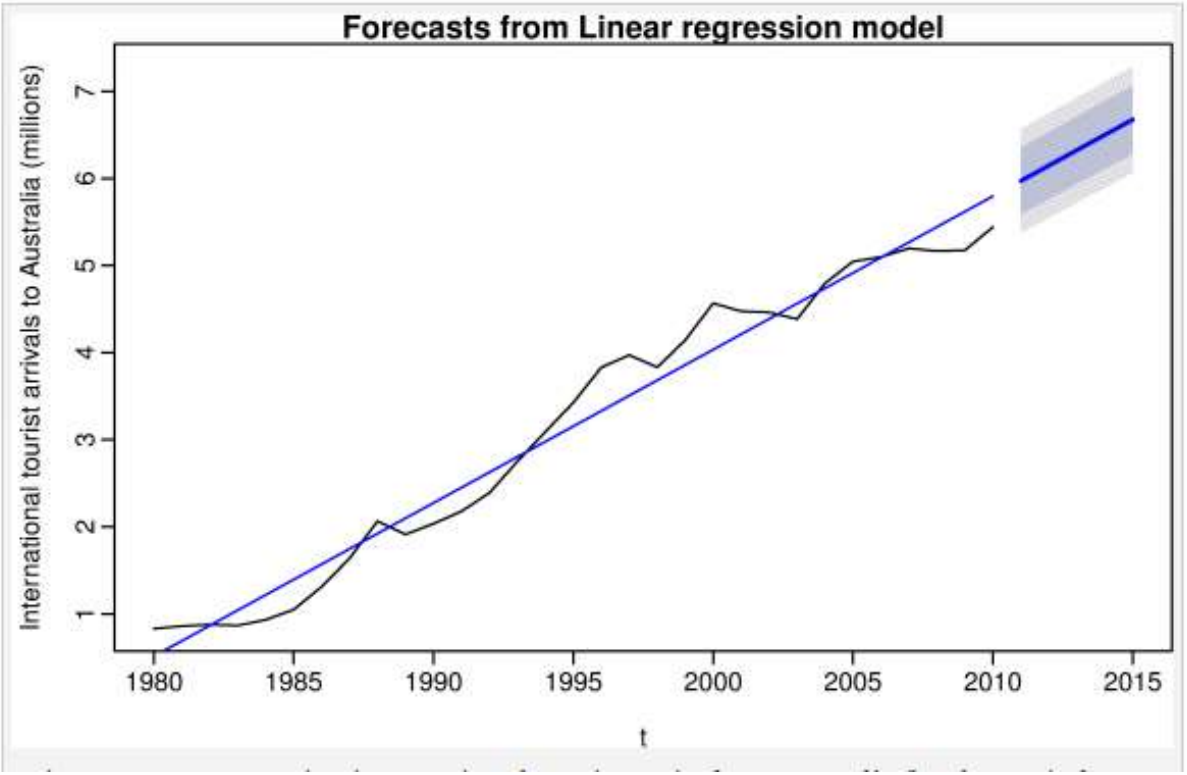
- Selang peramalan untuk for scenario based forecasts **tidak termasuk ketidakpasti**an terkait dengan nilai-nilai masa depan variabel prediktor. Mereka menganggap nilai prediktor yang diketahui sebelumnya.
- Pendekatan alternatif adalah dengan menggunakan **peramalan asli** untuk variabel prediktor. Misalnya, pendekatan time series murni dapat digunakan untuk menghasilkan peramalan untuk variabel prediktor atau **perkiraan yang diterbitkan oleh beberapa sumber lain** seperti lembaga pemerintah dapat digunakan.

Ex-ante versus ex-post forecasts

- *Ex ante forecasts* are those that are made **using only the information that is available in advance**. For example, ex ante forecasts of consumption for the four quarters in 2011 should only use information that was available *before* 2011. These are the only genuine forecasts, made in advance using whatever information is available at the time.
- *Ex post forecasts* are those that are made **using later information on the predictors**. For example, ex post forecasts of consumption for each of the 2011 quarters may use the actual observations of income for each of these quarters, once these have been observed. These are not genuine forecasts, but are useful for studying the behaviour of forecasting models.

Example 4.4 Linear trend

Figure 4.12 shows a time series plot of aggregate tourist arrivals to Australia over the period 1980 to 2010 with the fitted linear trend line $\hat{y}_t = 0.3375 + 0.1761t$. Also plotted are the point and forecast intervals for the years 2011 to 2015.



Residual autocorrelation

- Regresi data derte waktu → residual pasti memiliki korelasi diri
- Dugaan model melanggar asumsi tidak ada autokorelasi dalam residual,
- peramalan mungkin tidak efisien - ada beberapa informasi yang tersisa yang harus dimanfaatkan untuk mendapatkan peramalan yang lebih baik.
- Peramalan dari model dengan residual yg berkorelasi diri masih tidak bias, dan juga tidak "salah", tetapi biasanya akan memiliki selang prediksi yang lebih besar.

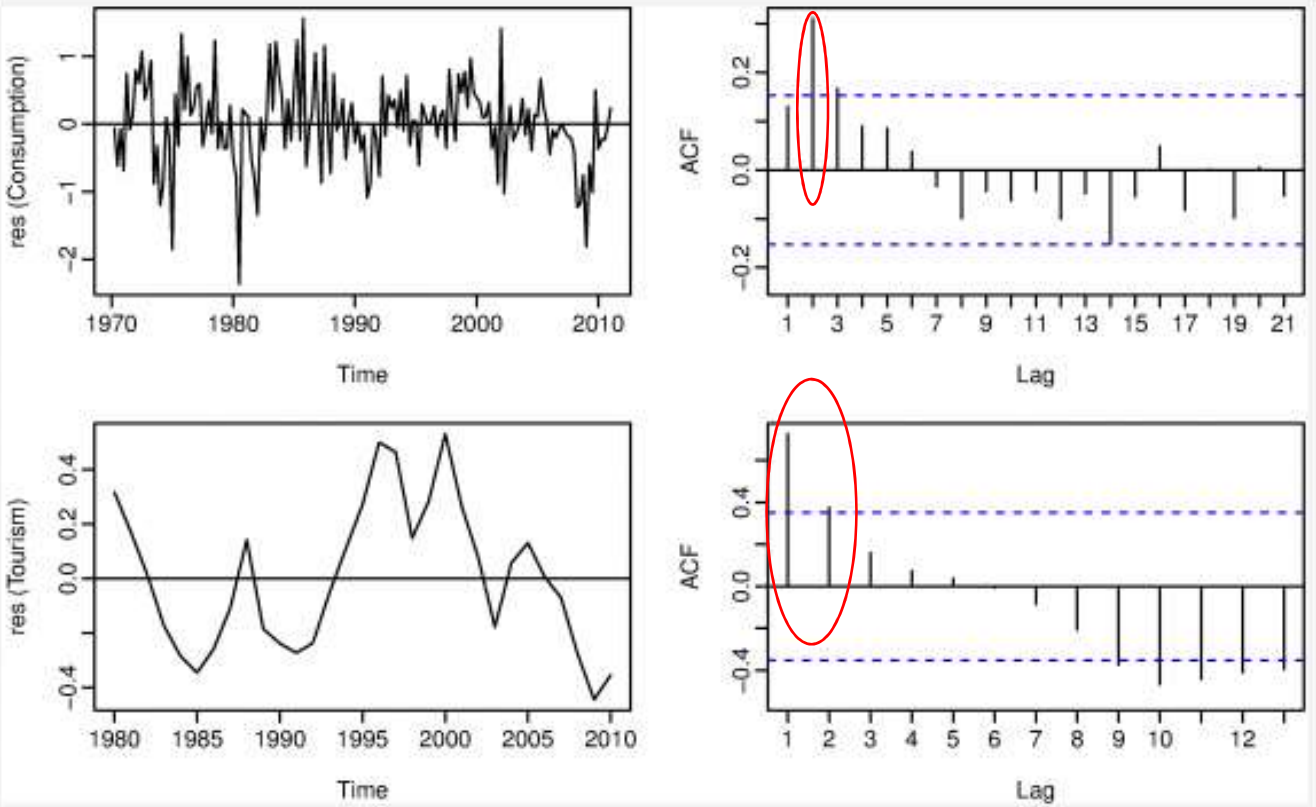


Figure 4.13: Residuals from the regression models for Consumption and Tourism. Because these involved time series data, it is important to look at the ACF of the residuals to see if there is any remaining information not accounted for by the model. In both these examples, there is some remaining autocorrelation in the residuals.

Exercise

Day	1	2	3	4	5	6	7	8	9	10	11	12
Mwh	16.3	16.8	15.5	18.2	15.2	17.5	19.8	19.0	17.5	16.0	19.6	18.0
temp	29.3	21.7	23.7	10.4	29.7	11.9	9.0	23.4	17.8	30.0	8.6	11.8

Electricity consumption was recorded for a small town on 12 randomly chosen days. The following maximum temperatures (degrees Celsius) and consumption (megawatt-hours) were recorded for each day.

a) Plot the data and find the regression model for Mwh with temperature as an explanatory variable. Why is there a negative relationship?
b) Produce a residual plot. Is the model adequate? Are there any outliers or influential observations?
c) Use the model to predict the electricity consumption that you would expect for a day with maximum temperature 10∘10∘ and a day with maximum temperature 35∘35∘. Do you believe these predictions?
d) Give prediction intervals for your forecasts. The following R code will get you started:

R COde

R code

```
plot(Mwh ~ temp, data=econsumption)
fit  <- lm(Mwh ~ temp, data=econsumption)
plot(residuals(fit) ~ temp, data=econsumption)
forecast(fit, newdata=data.frame(temp=c(10,35)))
```