PO12Q: Introduction to Quantitative Political Analysis II

Week 10 - Worksheet Solutions

Dr Florian Linke

F.Linke@warwick.ac.uk

1 | Exercises with the app

1. **Base fit**
   a) The fitted line is the OLS regression $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.
   b) The estimated slope $\hat{\beta}_1$ will be close to the true β_1 on average, but not exactly equal due to sampling variability (random noise ε).

2. **Heteroscedasticity on/off**
   a) With the slider at 0, the p-value is large, showing no evidence of heteroscedasticity.
   b) As heteroscedasticity increases, `SE (OLS)` underestimates variability while `SE (HC3)` corrects for it. The HC3 p-value is more conservative, sometimes turning insignificant when OLS still shows significance.

3. **Reading the bowtie guide**
   a) The dashed curve shows how the conditional variance of ε changes with x.
   b) As the curve steepens (variance grows with x), the Breusch–Pagan test statistic increases, and its p-value drops, indicating heteroscedasticity.

4. **Functional form: keeping the same sample**
   a) With $\beta_2 = 0$, the data match the base sample; the linear fit coincides with the true model.
   b) For $\beta_2 > 0$, the scatter curves upward (U-shape), and the F-test p-value becomes small, indicating nonlinearity.
   c) For $\beta_2 < 0$, the curve bends downward (inverted U), and the RESET test also flags misspecification.

5. **Comparing linear vs quadratic coefficients**
   a) The coefficient on x^2 corresponds to curvature.
   b) It becomes significant when the true curvature (β_2) is large enough. This aligns with the F-test and the dashed quadratic overlay, which fits the data better.

6. **Multicollinearity: creating instability**
   a) VIF values rise sharply toward infinity as $\rho_x \to 1$, showing inflated standard errors.
   b) The coefficient estimate for β_1 changes drastically depending on whether x_2 is included. This instability illustrates how collinearity makes coefficients unreliable.
   c) Collinearity does not bias OLS coefficients; instead, it increases their variance, making estimates unstable and inference weak.

7. **Equation comparison panel**
   a) For nonlinearity, the card shows only the linear fit because the app is demonstrating how misspecification occurs when a quadratic term is omitted.
   b) For collinearity, the model is written as $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$, highlighting that interpretation of each coefficient depends heavily on the other due to overlap between x_1 and x_2.

2 | Testing the Classical Linear Assumptions

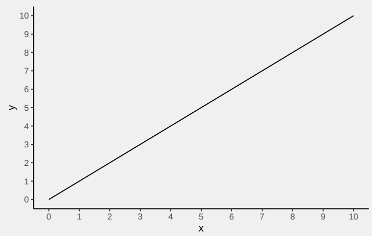
see RScript.

3 | The Gauss-Markov Theorem

2. According to Malinvaud, the assumption that $E(u_i|X_i) = 0$ is quite important. To see this, consider the PRF: $Y = \beta_1 + \beta_2 X_i + u_i$. Now consider two situations: (1) $\beta_1 = 0$, $\beta_2 = 1$ and $E(u_i) = 0$; and (2) $\beta_1 = 1$, $\beta_2 = 0$ and $E(u_i) = X_i - 1$. Now take the expectation of the PRF conditional upon X in those two cases, and see if you agree with Malinvaud about the significance of the assumption $E(u_i|X_i) = 0$.

Situation 1

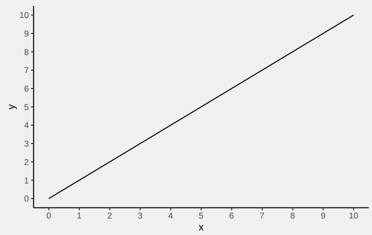
$$
\begin{aligned}
Y &= \beta_0 + \beta_1 X_i + \epsilon_i \\
&= 0 + X_i + 0 \\
&= X_i
\end{aligned} \tag{1}
$$



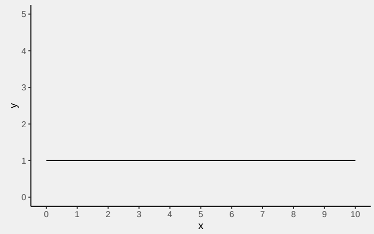
Situation 2

$$
\begin{aligned}
Y &= \beta_0 + \beta_1 X_i + \epsilon_i \\
&= 1 + 0 + X_i - 1 \\
&= X_i
\end{aligned} \tag{2}
$$

So what we get is:



What we **should** get is:



Conclusion

- The second PRF is the same as the first, even though – according to the betas – it should be a flat line
- This is only possible, because the error term is not zero and has an interdependence with X_i.