

ECON 4230 Intermediate Econometric Theory
Solutions to Problem Set #4

#1. Gujarati and Porter, 5th Edition, Exercise 10.27. In part (b), make sure to use the simple correlation coefficient(s) and VIF(s) to defend your answer.

SOLUTIONS:

Table 10.13 gives data on imports, GDP, and the Consumer Price Index (CPI) for the United States over the period 1975-2005. You are asked to consider the following model:

$$\ln \text{Imports}_t = \beta_1 + \beta_2 \ln \text{GDP}_t + \beta_3 \ln \text{CPI}_t + u_t$$

(a) Estimate the parameters of this model using the data given in the table.

$$\hat{\ln \text{Imports}}_t = \underset{(0.27)}{1.409} + \underset{(0.183)}{1.85} \ln \text{GDP}_t - \underset{(0.285)}{0.873} \ln \text{CPI}_t$$

with $R^2 = 0.992$ and standard errors in parentheses.

(b) Do you suspect that there is multicollinearity in the data?

Table 1 displays the correlation table of the explanatory variables and the VIF is displayed in table 2.

Table 1. Pairwise Correlations

	lnGDP	lnCPI
lnGDP	1.000	0.993
lnCPI	0.993	1.000

Table 2. VIFs

VIFs	
lnGDP	69.313
lnCPI	69.313

The VIF is much higher than 5; there is strong multicollinearity in the data.

(c) Regress:

1. $\ln \text{Imports}_t = A_1 + A_2 \ln \text{GDP}_t$
2. $\ln \text{Imports}_t = B_1 + B_2 \ln \text{CPI}_t$
3. $\ln \text{GDP}_t = C_1 + C_2 \ln \text{CPI}_t$

On the basis of these regressions, what can you say about the nature of multicollinearity in the data?

Table 3. Regression 1 results

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.002	0.214	9.341	0
lnGDP	1.293	0.025	51.831	0

with $R^2 = 0.9893$.

Table 4. Regression 2 results

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.578	0.348	10.281	0
lnCPI	1.986	0.073	27.394	0

with $R^2 = 0.9628$.

Table 5. Regression 1 results

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.172	0.167	7.033	0
lnCPI	1.546	0.035	44.509	0

with $R^2 = 0.9856$.

The first two regressions do not have a direct relationship to the multicollinearity because multicollinearity refers to collinearity between the explanatory variables, not the dependent variable. Regression 3 shows a high R^2 , which indicates that CPI and GDP have a very high correlation. Note also that

$$\frac{1}{1 - R^2} = \frac{1}{1 - 0.9856} = 69.313$$

which gives the VIF from part (b).

- (d) Suppose there is multicollinearity in the data but $\hat{\beta}_2$ and $\hat{\beta}_3$ are individually significant at the 5 percent level and the overall F test is also significant. In this case should we worry about the collinearity problem?

The estimators are still unbiased under multicollinearity. It does, however, make the standard errors larger and could as a result make $\hat{\beta}_2$ and $\hat{\beta}_3$ insignificant. But if the results show that the individual estimates are significant, then multicollinearity is not usually considered a serious problem.

#2. Gujarati and Porter, 5th Edition, Exercise 11.15. Parts (a) and (b) only. Then test for heteroscedasticity and comment on the results. Also, provide GLS and White's adjusted OLS estimates. Again, comment on the results.

SOLUTIONS:

Table 11.7 gives data on 81 cars about MPG (averages miles per gallons), HP (engine horsepower), VOL (cubic feet on cab space), SP (top speed, miles per hour), and WT (vehicle weight in 100 lbs.).

(a) Consider the following model:

$$MPG_i = \beta_1 + \beta_2 SP_i + \beta_3 HP_i + \beta_4 WT_i + u_i$$

(b) Estimate the parameters of this model and interpret the results. Do they make economic sense?

The estimated model is

$$\hat{MPG}_i = 189.96 - 1.272 SP_i + 0.39 HP_i - 1.903 WT_i$$

(22.529) (0.233) (0.076) (0.186)

with $R^2 = 0.883$ and standard errors are in parentheses.

We would think that all variables have a negative impact on the MPG because a higher top speed, a more powerful engine, and a higher weight should hurt fuel efficiency. The estimated coefficients on both WT and SP match expectations, but HP has an unexpected positive influence on the fuel efficiency.

(b) Would you expect the error variance in the preceding model to be heteroskedastic? Why?

The first test for heteroscedasticity is to think through the problem. It is not clear exactly how the explanatory variables should be related to the error variance. Since there is not a clear intuitive explanation, we turn to the data.

In Figure 1, the residuals are plotted against the predicted values and each explanatory variable.

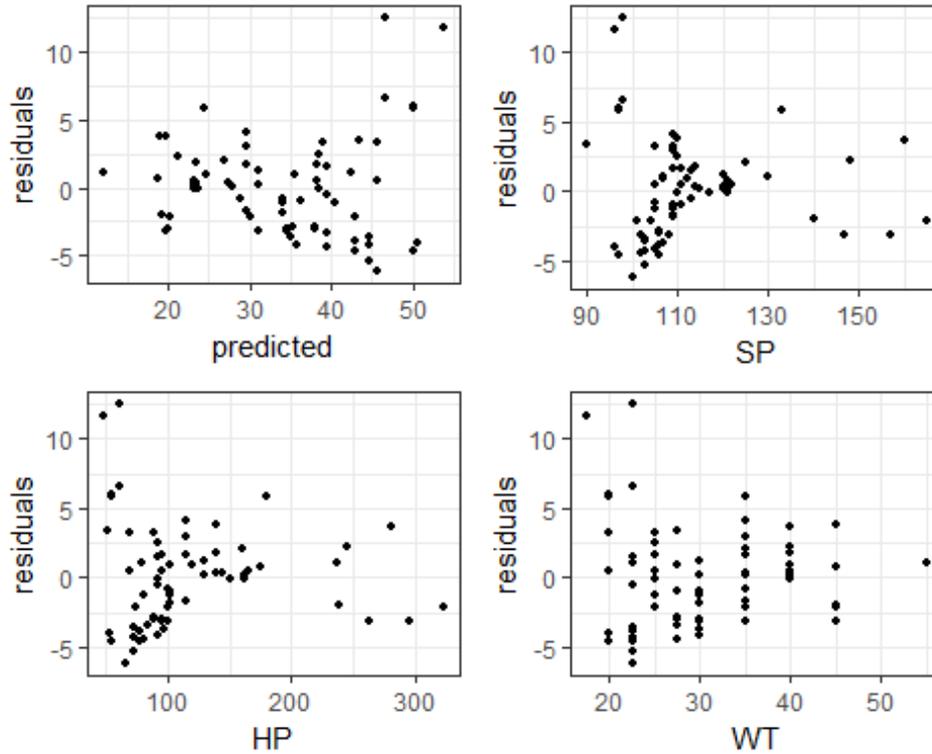


Figure 1. Residual Plots against the Explanatory Variables

The residuals clearly show higher variation at the lower end of each individual variable.

Next we will formally test if there is heteroskedasticity using the Breusch-Pagan test.

Step 1: Run the main OLS regression:

$$\widehat{MPG}_i = 189.96 - 1.272SP_i + 0.39HP_i - 1.903WT_i$$

(22.529)
(0.233)
(0.076)
(0.186)

Step 2: Find estimate of error variance:

$$\hat{\sigma}^2 = \sum \hat{u}_i^2 / (81 - 4) = 12.305$$

Step 3:

Divide the squared residuals \hat{u}_i obtained from the regression in Step 1 by $\hat{\sigma}^2$ to construct the variable p_i .

Step 4:

Assuming that p_i are linearly related to $X_i (= Z_i)$, we obtain the regression

$$\hat{p}_i = 53.565 - 0.52SP_i + 0.169HP_i - 0.45WP_i$$

with

$$R^2 = 0.315 \text{ and ESS} = 93.84.$$

Step 5: Form the test statistic:

$$\theta = \frac{1}{2}(ESS) = 46.92$$

Assuming the χ^2 distribution with 3 degrees of freedom, we find that for 3 degrees of freedom the 1 percent critical χ^2 value is 11.345. Thus we reject the null hypothesis of homoskedasticity.

What follows is the comparison of OLS, OLS with White's adjusted standard errors, and GLS estimation.

Table 6. OLS with Regular Standard Errors

variable	estimate	std.error	statistic	p.value
(Intercept)	189.960	22.529	8.432	0
SP	-1.272	0.233	-5.455	0
HP	0.390	0.076	5.121	0
WT	-1.903	0.186	-10.259	0

Table 7. OLS with White's Adjusted Standard Errors

variable	estimate	std.error	statistic	p.value
(Intercept)	189.960	33.906	5.603	0.000
SP	-1.272	0.336	-3.784	0.000
HP	0.390	0.109	3.589	0.001
WT	-1.903	0.285	-6.676	0.000

Feasible GLS Estimation:

Start by assuming that $\sigma_i^2 = \sigma^2 \times WT_i^2$. Then transform the model and estimate:

$$\frac{MPG_i}{\sigma_i} = \hat{\beta}_1^* \left(\frac{1}{\sigma_i}\right) + \hat{\beta}_2^* \left(\frac{SP_i}{\sigma_i}\right) + \hat{\beta}_3^* \left(\frac{HP_i}{\sigma_i}\right) + \hat{\beta}_4^* \left(\frac{WT_i}{\sigma_i}\right) + \frac{\hat{u}_i}{\sigma_i}$$

where the transformed errors are now homoscedastic.

OLS regression on the transformed model gives the following results:

Table 8. Feasible GLS Results

variable	estimate	std.error	statistic	p.value
(Intercept)	210.213	23.754	8.849	0
SP	-1.469	0.247	-5.939	0
HP	0.461	0.083	5.565	0
WT	-2.106	0.203	-10.369	0

Note that the feasible GLS estimates, as well as the standard errors, have changed.

#3. Build and estimate a model to predict Wyoming's unemployment rate for 2019.

SOLUTIONS:

Many different models and autocorrelation results are possible.