

# Computer Assignment 2

*Uppsala University*

*Department of Statistics*

*Course: Econometrics, Fall 2019*

*Author: Claes Kock, Yuchong Wu, Mayara Latrech*

*Date: 29/10/2019*

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Variables</b>	<b>4</b>
<b>3</b>	<b>Tests for assumptions</b>	<b>5</b>
<b>4</b>	<b>Task 1a</b>	<b>6</b>
4.1	Model 1 . . . . .	6
4.1.1	Estimation of Model 1 . . . . .	6
4.1.2	Tests for model 1 . . . . .	7
4.2	Model 2 . . . . .	10
4.2.1	Estimation of Model 2 . . . . .	10
4.2.2	Tests for model 2 . . . . .	11
4.3	Model 3 . . . . .	14
4.3.1	Estimation of Model 3 . . . . .	14
4.3.2	Tests for model 3 . . . . .	15
4.4	Model 4 . . . . .	18
4.4.1	Estimation of Model 4 . . . . .	18
4.4.2	Tests for model 4 . . . . .	19
4.5	Model 5 . . . . .	22
4.5.1	Estimation of Model 5 . . . . .	22
4.5.2	Tests for model 5 . . . . .	23
<b>5</b>	<b>Task 1b</b>	<b>26</b>
5.1	Comparison of BIC and AIC for all five models. . . . .	27
5.2	Choosing the best model . . . . .	27
5.3	T-test . . . . .	29
5.4	F-test . . . . .	31
<b>6</b>	<b>Task 1c</b>	<b>33</b>
<b>7</b>	<b>Conclusion</b>	<b>36</b>

# 1 Introduction

Our task for this homework assignment is to study the demand for chicken in the United States between the years of 1960 and 1982. To do this we are given a dataset containing relevant data. This data can be found in the book “Basic Econometrics 5th Edition” (by Damodar N. Gujarati, and Dawn C. Porter). The data contains several variables which represents certain economic factors which are of interest to our study. Our task is to compare several models that try to explain the consumption of chicken, find how much they explain, and try to determine which model would be the most appropriate to use if we wanted to estimate the consumption of chicken. Some of the variables we use are per capita income and various real prices of food commodities.

## 2 Variables

Table 1: Description of our variables and the economic factors they represent.

Variable name	Represented Value
Y	Per capita consumption of chicken
X2	Real disposable income per capita
X3	Real detail price of chicken
X4	Real detail price of pork
X5	Real detail price of beef
X6	Composite real price of chicken substitutes

For ease of understanding, we need to change the names of the variables, in order for the names to better reflect what the variables represent:

Table 2: Renamed the previous variables in order to make variable names more accurate.

Variable name	Represented Value
CapChicken	Per capita consumption of chicken
RCapInc	Real disposable income per capita
RChicken	Real detail price of chicken
RPork	Real detail price of pork
RBeef	Real detail price of beef
RSub	Composite real price of chicken substitutes

From microeconomic theory we know that the demand for chicken is a function of:

1. the real income of consumers,
2. real price of the commodity,
3. real price of competing or complementary commodities.

The task is to estimate, evaluate and choose one of the models. This is in order to determine which model is the most appropriate model to use in order to estimate the demand for chicken in the US 1960-1982.

We have the following models:

$$\ln Y_t = \alpha_1 + \alpha_2 \ln X_t + \alpha_3 \ln X_{3t} + u_t \quad (1)$$

$$\ln Y_t = \gamma_1 + \gamma_2 \ln X_{2t} + \gamma_3 \ln X_{3t} + \gamma_4 \ln X_{4t} + u_t \quad (2)$$

$$\ln Y_t = \lambda_1 + \lambda_2 \ln X_{2t} + \lambda_3 \ln X_{3t} + \lambda_4 \ln X_{5t} + u_t \quad (3)$$

$$\ln Y_t = \theta_1 + \theta_2 \ln X_{2t} + \theta_3 \ln X_{3t} + \theta_4 \ln X_{4t} + \theta_5 \ln X_{5t} + u_t \quad (4)$$

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{6t} + u_t \quad (5)$$

### 3 Tests for assumptions

To determine that our assumptions are correct, we are going to test all of our models using several tests. To test for normality (is the data normally distributed?), we are looking at the graphs of the residuals, as well as using the Jarque-Bera test. To determine heteroscedasticity (if the variance of the error term is constant or not) we compare the standardised residuals to the fitted residuals, as well as use the Breusch-Pagan test. Finally we calculate the variance inflation factor (VIF) to check how high the risk for multicollinearity is. Finally we are using AIC and BIC to compare all the models in order to find the best model to use.

## 4 Task 1a

Our first task is to estimate our models, meaning that we test how well they fit the data that we are trying to analyse. In order to do this we are going to perform critical evaluations on all of the five models. We need our models to be justified from a theoretical perspective, based on economical theory - what do our models try to represent? We also make use of statistical tests in order to determine how well the models explain the data, and check if the assumptions we make regarding the variables of the models are confirmed or not.

### 4.1 Model 1

#### 4.1.1 Estimation of Model 1

Our first model is:

$$\ln CapChicken_t = \alpha_1 + \alpha_2 \ln RCapInc_t + \alpha_3 \ln RChicken_t + u_t$$

Mathematically, it, like all our models, is the logarithmic consumption of chicken for the year  $t$ . In this model the per capita consumption is a function of the per capita income and the price of chicken for the year  $t$ . We can assume that an increase in per capita income would lead to more consumption of chicken ( $\alpha_2 > 1$ ), whereas an increase in price of chicken would lead to less consumption, according to basic supply and demand ( $\alpha_3 < 1$ ), holding all other variables constant.

Table 3: Regression Results for Model 1

	<i>Dependent variable:</i>
	log(CapChicken)
log(RCapInc)	0.452*** (0.025)
log(RChicken)	-0.372*** (0.063)
Constant	2.033*** (0.116)
Observations	23
R <sup>2</sup>	0.980
Adjusted R <sup>2</sup>	0.978
Residual Std. Error	0.028 (df = 20)
F Statistic	491.868*** (df = 2; 20)
Note:	*p<0.1; **p<0.05; ***p<0.01

Estimate of Model 1:

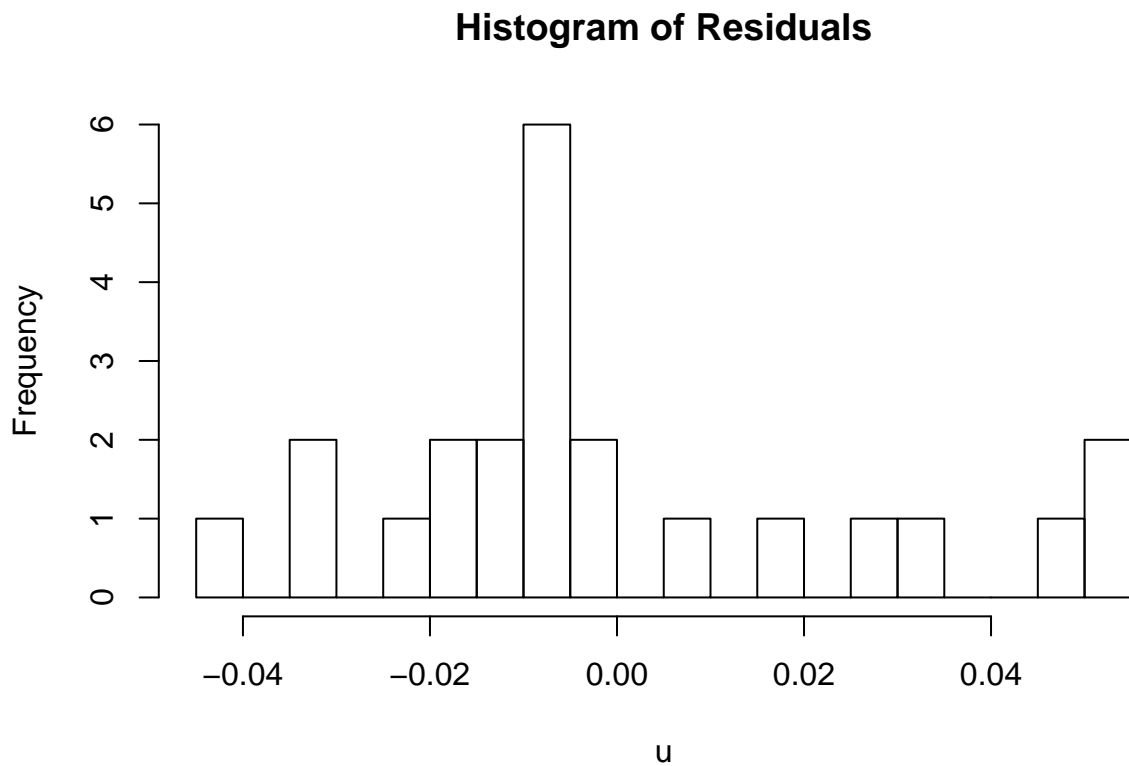
$$\ln CapChicken_t = 2.033 + 0.452 \ln RCapInc_t - 0.372 \ln RChicken_t$$

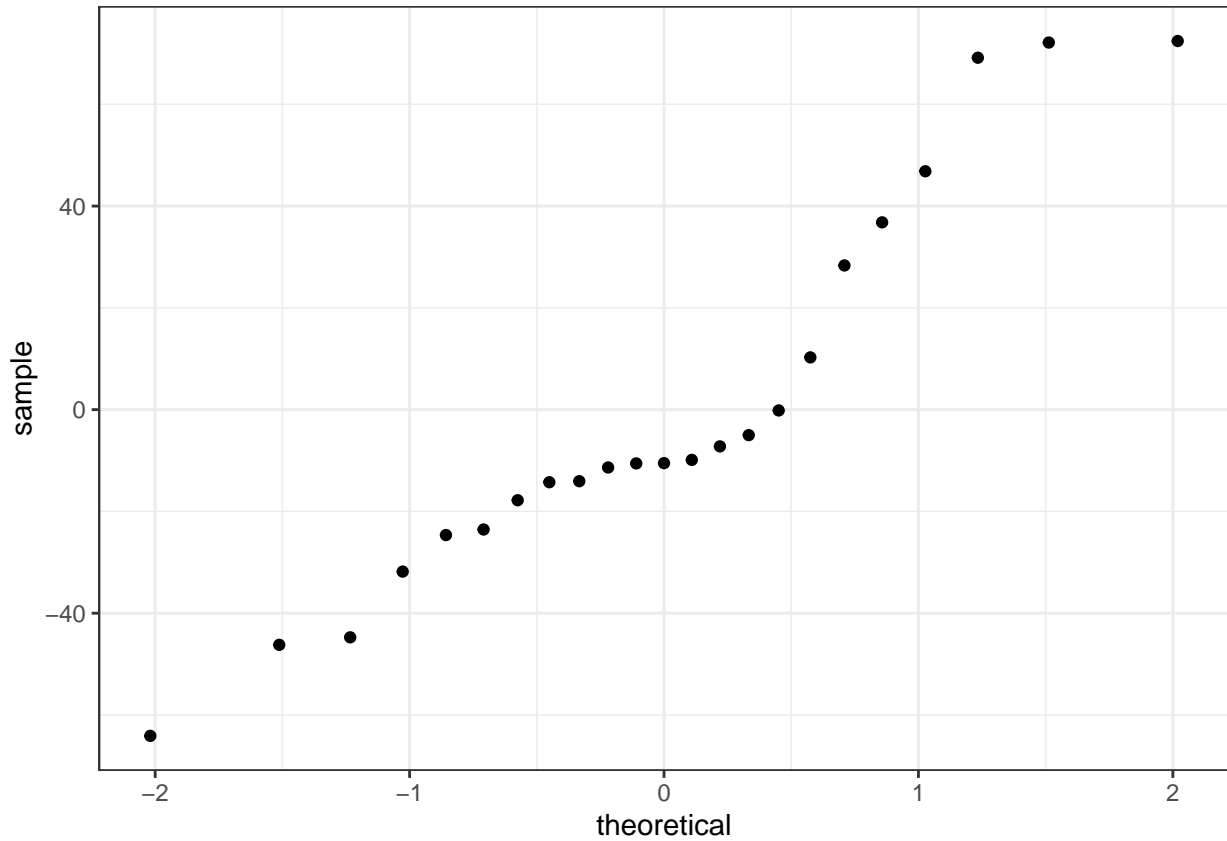
From the perspective of  $R^2 = 0.98$ , the model explains 98% of the variation of the dependent variable. Above, we can also see that per capita income and the price of chicken are significant variables for this model. (*significance level: 0.05*) The null hypothesis is  $\alpha_2 = \alpha_3 = \alpha_4 = 0$ . From the outcome of the F-test the p-value is much lower than  $\alpha = 0.05$ , and thus we reject the null hypothesis.

According to this model, the per capita consumption of chicken for any year  $t$  is determined by the per capita income plus the real price of chicken. This might not be representative of the actual consumption of chicken per capita, as the model ignores the price of other products that fill a similar household role, such as pork and beef. If the price of chicken is higher than one of these, the estimated demand for chicken could theoretically decrease, due to consumers instead buying other sorts of meat.

A relative increase of RCapInc, representing real capita income, would result in a relative increase to the per capita consumption of chicken. A decrease in the price of chicken would lead to an increase in the per capita consumption of chicken, holding all other variables constant. To summarize, we find that this model fulfills our assumptions that  $(\alpha_2 > 0)$  and  $(\alpha_3 < 0)$ . We are unsure, however, if the model gives a complete picture of the consumption of chicken, due to only containing two explanatory variables.

#### 4.1.2 Tests for model 1





According to the statistics theory, if the model were normal, we should have that the sample quantiles coincide with the theoretical quantiles of the normal distribution (dots should be along a diagonal line). However, this is **not** the case.

The model seems to fulfill the normality assumption according to the QQ-plot.

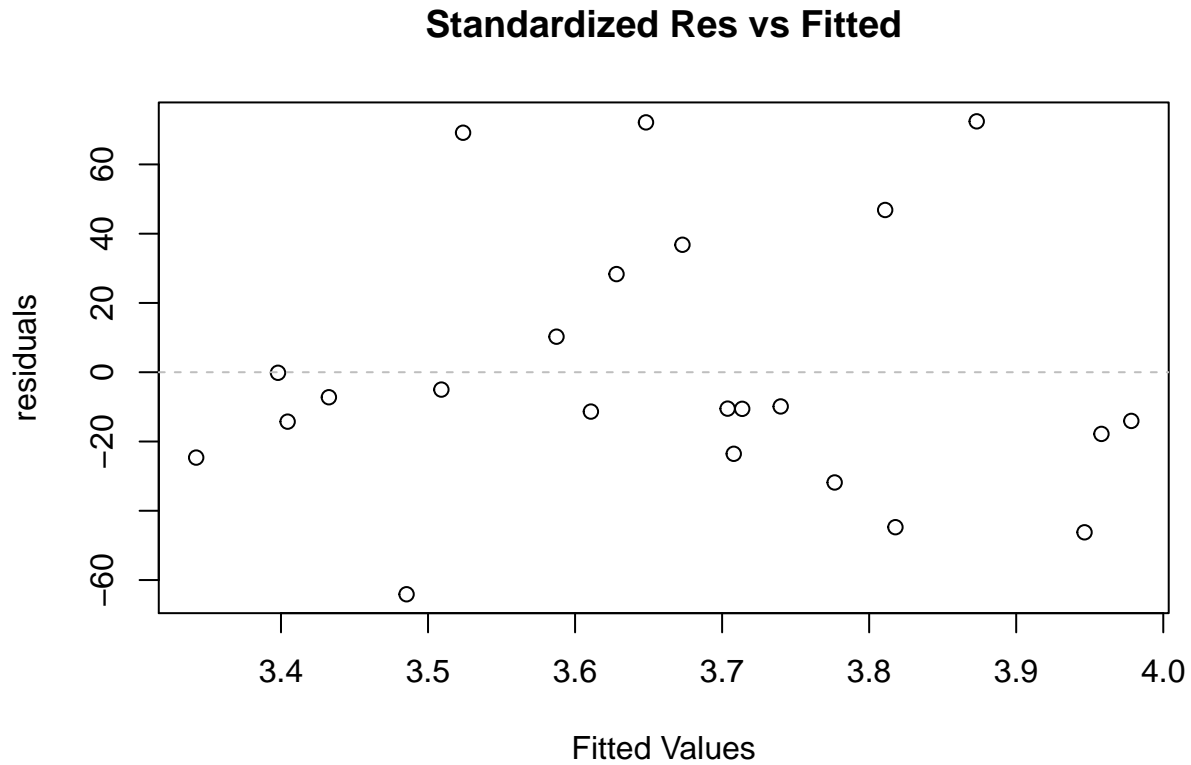
To make it more clear, we perform a JB-test.

Table 4: Jarque-Bera test for normality

Model	JB	p_value
1	1.676	0.176

Since the p-value is bigger than 0.05, we can not reject  $H_0$ .





We don't see any issue with the residuals for model 1 - the residuals are not following any kind of pattern which could indicate heteroscedasticity.

Table 5: studentized Breusch-Pagan test

	Model	BP	df	p_value
BP	1	1.005	2	0.605

The Breusch-Pagan test, or BP-test, tests for heteroscedasticity. If the p-value  $< 0.05$  then the error term of the model is heteroscedastic. Since the p-value is higher than 0.05, the error term of the model is not heteroscedastic.

Table 6: Variance Inflation Factors for the Model

	VIF
$\log(\text{RCapInc})$	5.65
$\log(\text{RChicken})$	5.65

Since  $\text{VIF} < 10$ , we do not observe any risk of multicollinearity.

## 4.2 Model 2

### 4.2.1 Estimation of Model 2

Our second model is:

$$\ln CapChicken_t = \gamma_1 + \gamma_2 \ln RCapInc_t + \gamma_3 \ln RChicken_t + \gamma_4 \ln RPork_t + u_t$$

In this model the per capita consumption of chicken is a function of the per capita income of the consumer, plus the the price of chicken, plus the price of pork. Here we can assume that an increase in the price of chicken and a decrease in the price of pork will lead to less consumption of chicken due to consumers buying more pork. We can also connect level of income to what kind of meat consumers might purchase.

If the base level of income is quite low then an increase of income will lead to an increase in the amount of chicken that consumers purchase. If the income decreases we assume that the consumer will buy more pork instead, since we assume it to be generally cheaper than chicken, holding all other variables constant. We assume income to have a positive effect on chicken consumption, so, ( $\gamma_2 > 0$ ). We also assume that the price of chicken should have a negative effect on consumption ( $\gamma_3 < 0$ ), and pork price should have a postive effect on chicken consumption ( $\gamma_4 > 0$ ), holding all other variables constant.

Table 7: Regression Results for Model 2

	<i>Dependent variable:</i>
	log(CapChicken)
log(RCapInc)	0.406*** (0.045)
log(RChicken)	-0.439*** (0.083)
log(RPork)	0.107 (0.088)
Constant	2.125*** (0.138)
Observations	23
R <sup>2</sup>	0.982
Adjusted R <sup>2</sup>	0.979
Residual Std. Error	0.027 (df = 19)
F Statistic	336.181*** (df = 3; 19)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Estimate of Model 2:

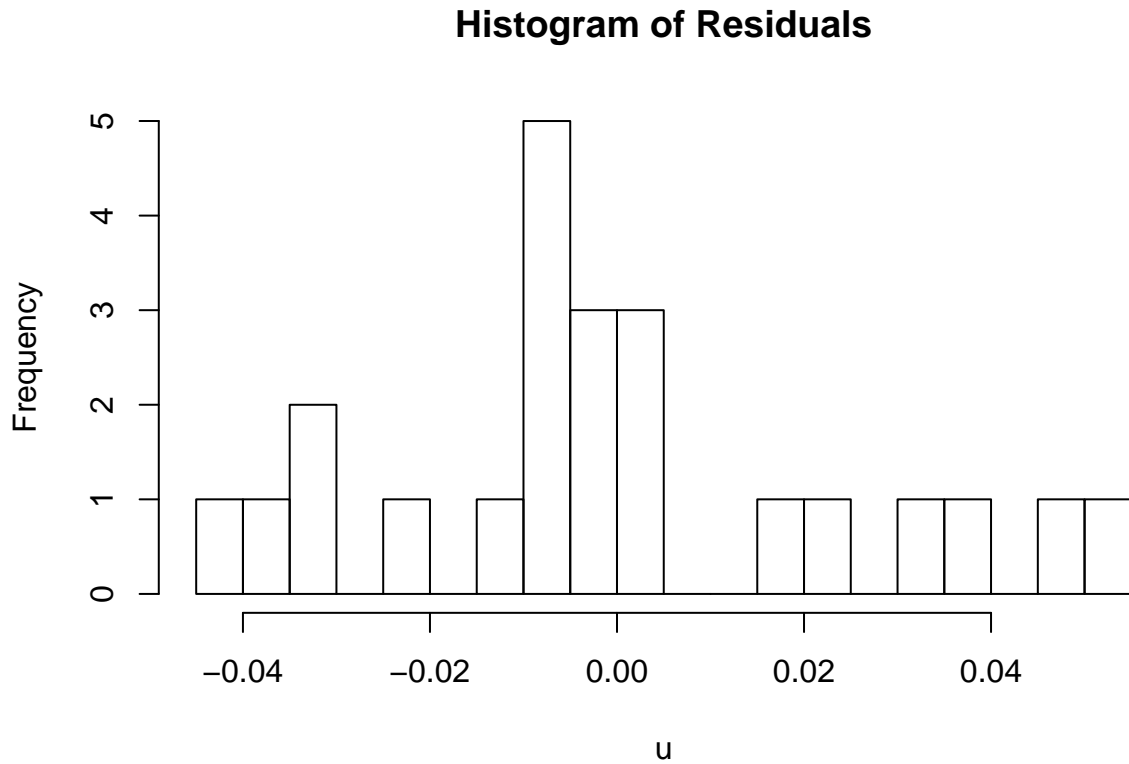
$$\ln CapChicken_t = 2.125 + 0.406 \ln RCapInc_t - 0.439 \ln RChicken_t - 0.107 \ln RPork_t + u_t$$

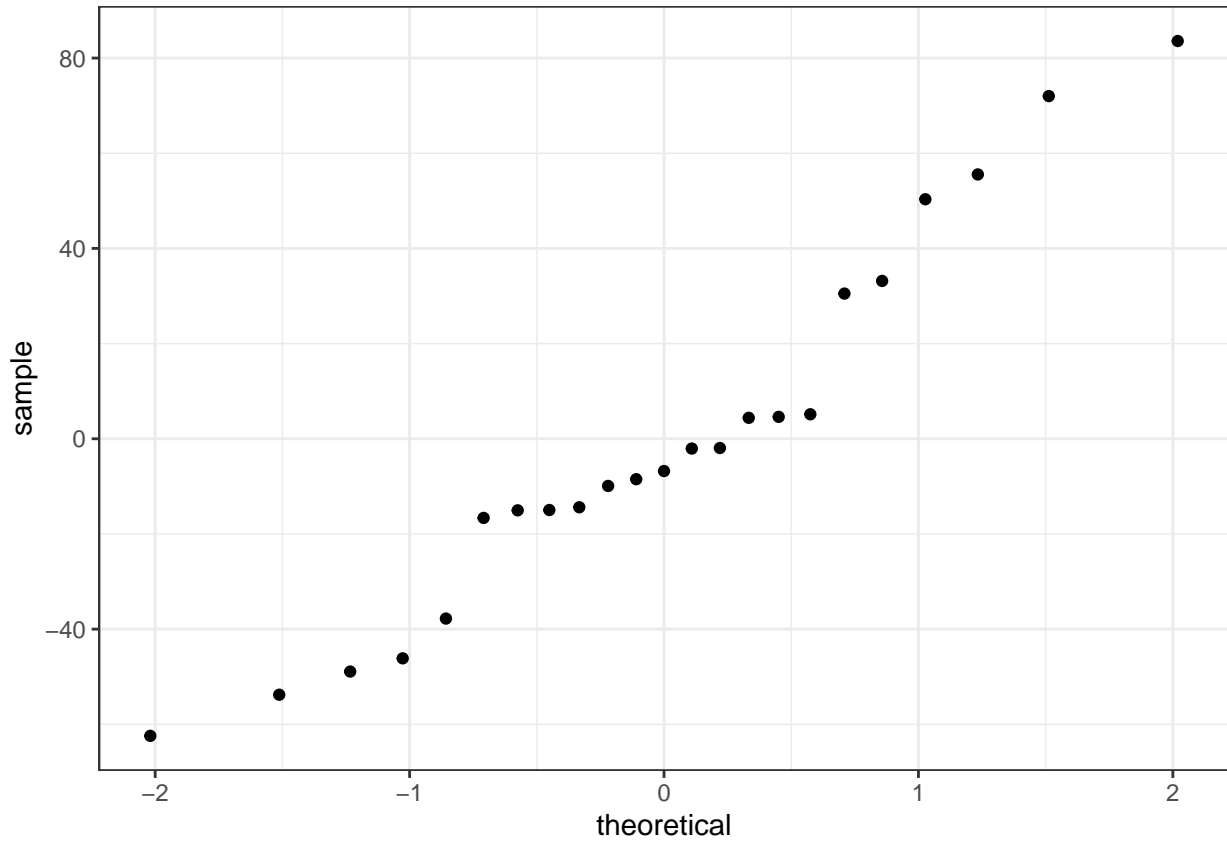
From the perspective of  $R^2 = 0.982$ , the model explains 98.2% of the variation of the dependent variable. In the above table, we can see that all variables except RPork are significant for this model. (*significance level: 0.05*) The null hypothesis is  $\gamma_2 = \gamma_3 = \gamma_4 = 0$ . From the outcome of the F-test the p-value is much lower than  $\alpha = 0.05$ , and thus we reject the null hypothesis.

A relative increase in per capita income will, according to this model, result in a relative increase in the consumption of chicken. An increase in the price of chicken will decrease the consumption of chicken. An increase in the price of pork will also decrease the consumption of chicken.

If we look at our model, the price of the chicken and the income per capita both have a bigger effect on the consumption of chicken compared to the price of pork. We can see that the real per capita income has a positive effect on the consumption of chicken and that the real price of chicken and pork has a negative effect.

#### 4.2.2 Tests for model 2





According to the statistics theory, if the model were normal, we should have that the sample quantiles coincide with the theoretical quantiles of the normal distribution (dots should be along a diagonal line). However, this is **not** the case.

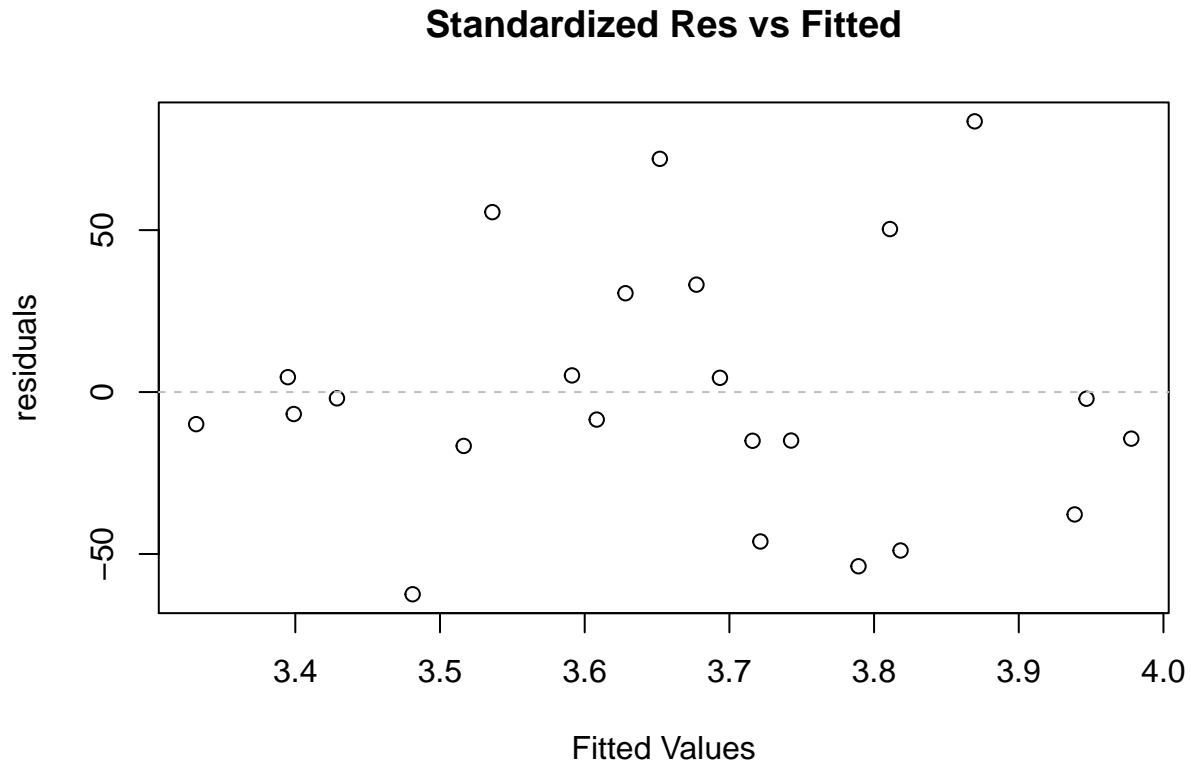
The model seems to fulfill the normality assumption according to the QQ-plot.

To make it more clear, we perform a JB-test.

Table 8: Jarque-Bera test for normality

Model	JB	p_value
1	1.676	0.176

Since the p-value is bigger than 0.05, we can not reject  $H_0$ .



We don't see any issue with the residuals for model 2 - the residuals are not following any kind of pattern which could indicate heteroscedasticity.

Table 9: studentized Breusch-Pagan test

	Model	BP	df	p_value
BP	2	4.565	3	0.207

The Breusch-Pagan test, or BP-test, tests for heteroscedasticity. If the p-value  $< 0.05$  then the error term of the model is heteroscedastic. Since the p-value is higher than 0.05, the error term of the model is not heteroscedastic.

Table 10: Variance Inflation Factors for the Model

	VIF
log(RCapInc)	19.02
log(RChicken)	9.97
log(RPork)	32.49

Since  $VIF > 10$ , we do observe a risk of multicollinearity for model 2.

### 4.3 Model 3

#### 4.3.1 Estimation of Model 3

Our third model is:

$$\ln CapChicken_t = \lambda_1 + \lambda_2 \ln RCapInc_t + \lambda_3 \ln RChicken_t + \lambda_4 \ln RBeef_t + u_t$$

In this model, the per capita consumption of chicken is a function of the per capita income of the consumer, plus the price of chicken, plus the price of beef. Here we can assume that an increase in the price of chicken and a decrease in the price of beef will lead to less consumption of chicken due to consumers buying more beef since beef is a substitute for chicken.

We can also connect the level of income to what kind of meat consumers might purchase. If the base level of income is high enough then an increase in income will lead to a decrease in the amount of chicken that consumers purchase. If the income increases we assume that the consumer will buy more beef instead, since we assume it to be generally more preferable than chicken. We assume income to have a negative effect on chicken consumption, so, ( $\gamma_2 < 0$ ), holding all other variables constant. We also assume that the price of chicken should have a negative effect on consumption ( $\gamma_3 < 0$ ), and beef price should have a positive effect on chicken consumption ( $\gamma_4 > 0$ ), holding all other variables constant.

Table 11: Regression Results for Model 3

	<i>Dependent variable:</i>
	log(CapChicken)
log(RCapInc)	0.441*** (0.052)
log(RChicken)	-0.381*** (0.076)
log(RBeef)	0.021 (0.092)
Constant	2.039*** (0.122)
Observations	23
R <sup>2</sup>	0.980
Adjusted R <sup>2</sup>	0.977
Residual Std. Error	0.028 (df = 19)
F Statistic	312.419*** (df = 3; 19)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Estimate of Model 3:

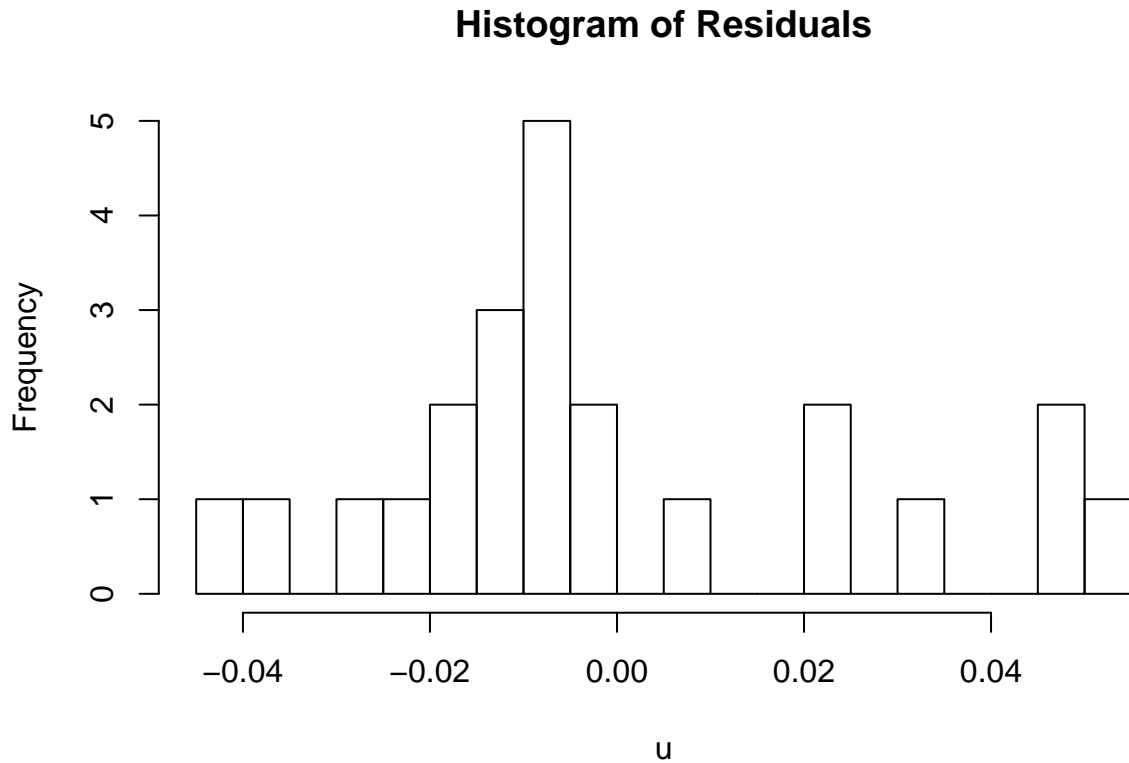
$$\ln CapChicken_t = 2.039 + 0.441 \ln RCapInc_t - 0.381 \ln RChicken_t + 0.021 \ln RBeef_t + u_t$$

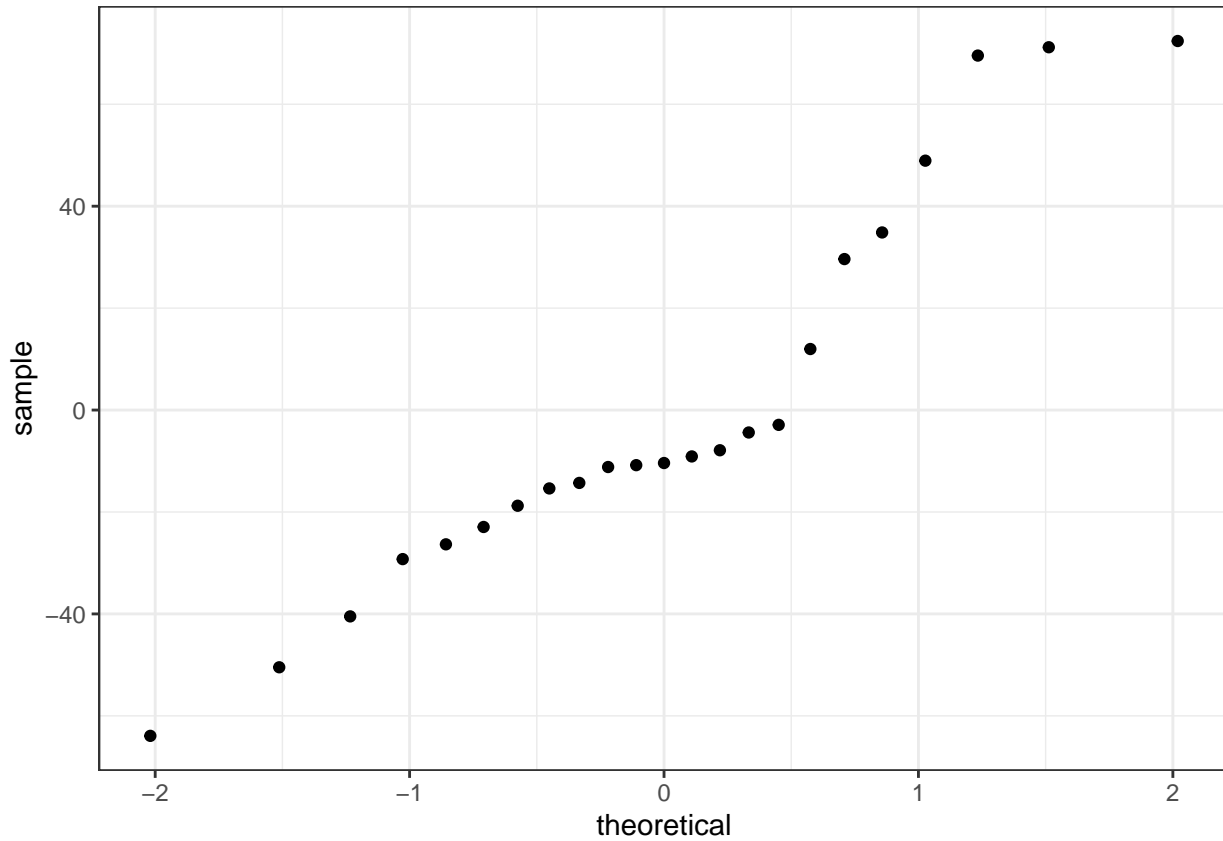
From the perspective of  $R^2 = 0.98$ , the model explains most of the variation of the dependent variable. From table 5, we can figure that only the price of beef is not significant for this model. (*significance level: 0.05*) The null hypothesis is  $\gamma_2 = \gamma_3 = \gamma_4 = 0$ . From the outcome of the F-test, the p-value is much lower than  $\alpha = 0.05$ , and thus we reject the null hypothesis. This confirms that our model makes sense to some extent.

A relative increase in per capita income will, according to this model, result in a relative increase in the consumption of chicken. An increase in the price of beef will also increase the consumption of chicken.

After estimating the model, the price of the beef has a much smaller effect on the consumption of chicken compared to the income per capita and price of chicken. We can see that the real per capita income has a positive effect on the consumption of chicken which was not expected, and that the real price of chicken has a positive effect which was expected. The price of beef has a positive effect on consumption, which was not expected, but was also not significant.

#### 4.3.2 Tests for model 3





According to the statistics theory, if the model were normal, we should have that the sample quantiles coincide with the theoretical quantiles of the normal distribution (dots should be along a diagonal line). However, this is **not** the case.

The model seems to fulfill the normality assumption according to the QQ-plot.

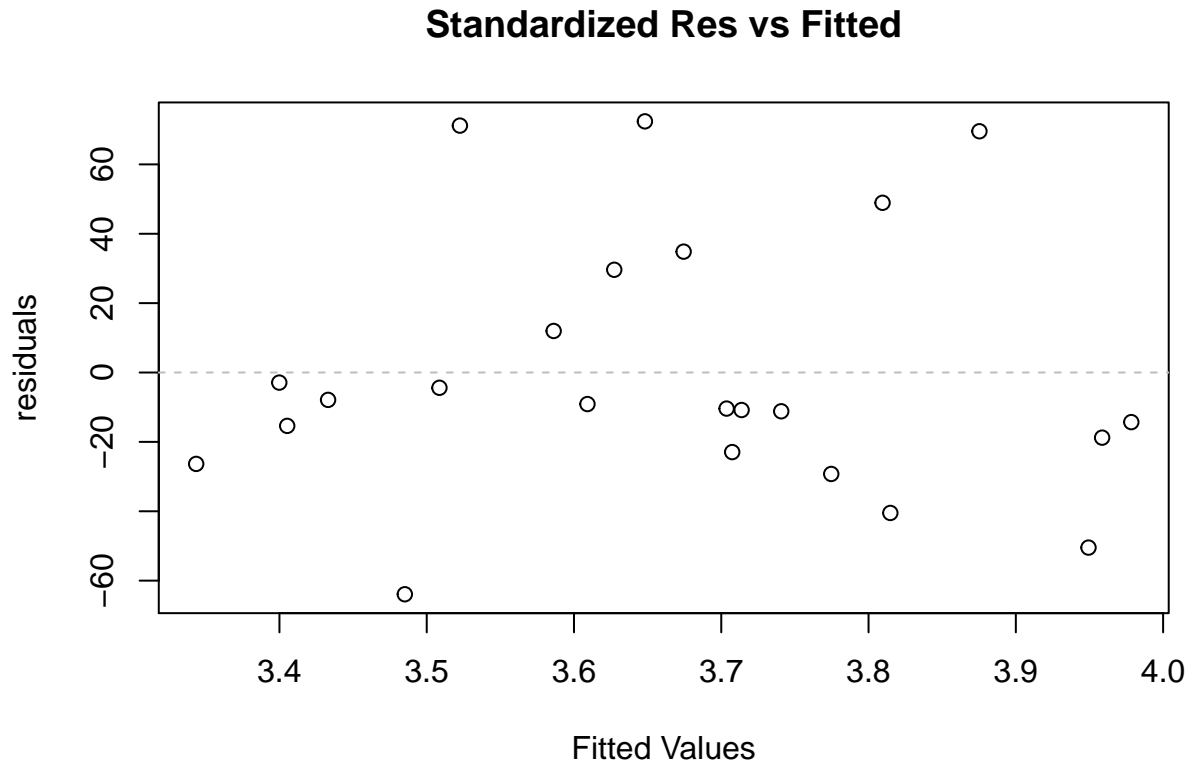
To make it more clear, we perform a JB-test.

Table 12: Jarque-Bera test for normality

Model	JB	p_value
1	1.676	0.176

Since the p-value is bigger than 0.05, we can not reject  $H_0$ .





We don't see any issue with the residuals for model 3 - the residuals are not following any kind of pattern which could indicate heteroscedasticity.

Table 13: studentized Breusch-Pagan test

	Model	BP	df	p_value
BP	3	0.953	3	0.813

The Breusch-Pagan test, or BP-test, tests for heteroscedasticity. If the p-value  $< 0.05$  then the error term of the model is heteroscedastic. Since the p-value is higher than 0.05, the error term of the model is not heteroscedastic.

Table 14: Variance Inflation Factors for the Model

	VIF
log(RCapInc)	24.25
log(RChicken)	7.79
log(RBeef)	33.17

Since  $VIF > 10$ , we do observe a risk of multicollinearity.

## 4.4 Model 4

### 4.4.1 Estimation of Model 4

Our fourth model is:

$$\ln CapChicken_t = \theta_1 + \theta_2 \ln RCapInc_t + \theta_3 \ln RChicken_t + \theta_4 \ln RPork_t + \theta_5 \ln RBeef_t + u_t$$

In this model the per capita consumption of chicken is a function of the per capita income of the consumer, plus the price of chicken, plus the price of pork, plus the price of beef. Here we can assume that an increase in the price of chicken and beef and a decrease in the price of pork will lead to less consumption of chicken due to consumers buying more pork. We can also connect level of income to what kind of meat consumers might purchase. If the base level of income is quite low then an increase to a certain point of income will lead to an increase in the amount of chicken that consumers purchase. However, a bigger increase of income will lead to a decrease in the amount of chicken that consumers purchase because the consumer will be able to buy more beef.

If the income decreases we assume that the consumer will buy more pork instead, since we assume it to be generally cheaper than chicken. We assume income to have a positive effect on chicken consumption to a certain point, so,  $(\theta_2 > 0)$  if  $RCapInc_t < k$ , and  $(\theta_2 < 0)$  if  $RCapInc_t > k$ , holding all other variables constant. We also assume that the price of chicken should have a negative effect on consumption  $(\theta_3 < 0)$ , and pork price should have a positive effect on chicken consumption  $(\theta_4 > 0)$ , holding all other variables constant. The price of beef should have a negative effect on chicken consumption  $(\theta_5 < 0)$ , holding all other variables constant.

Table 15: Regression Results for Model 4

	<i>Dependent variable:</i>
	log(CapChicken)
log(RCapInc)	0.343*** (0.083)
log(RChicken)	-0.505*** (0.111)
log(RPork)	0.149 (0.100)
log(RBeef)	0.091 (0.101)
Constant	2.190*** (0.156)
Observations	23
R <sup>2</sup>	0.982
Adjusted R <sup>2</sup>	0.978
Residual Std. Error	0.028 (df = 18)
F Statistic	249.928*** (df = 4; 18)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Estimate of Model 4:

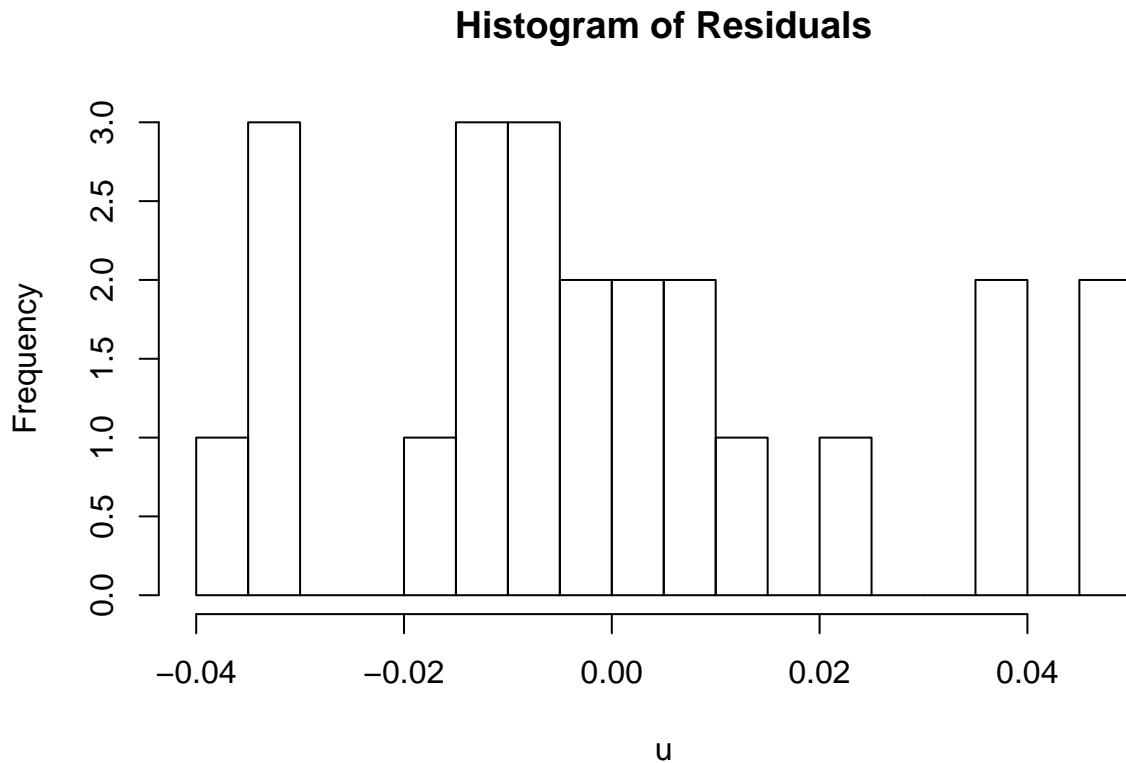
$$\ln CapChicken_t = 2.190 + 0.343 \ln RCapInc_t - 0.505 \ln RChicken_t + 0.149 \ln RPork_t + 0.091 \ln RBeeft_t + u_t$$

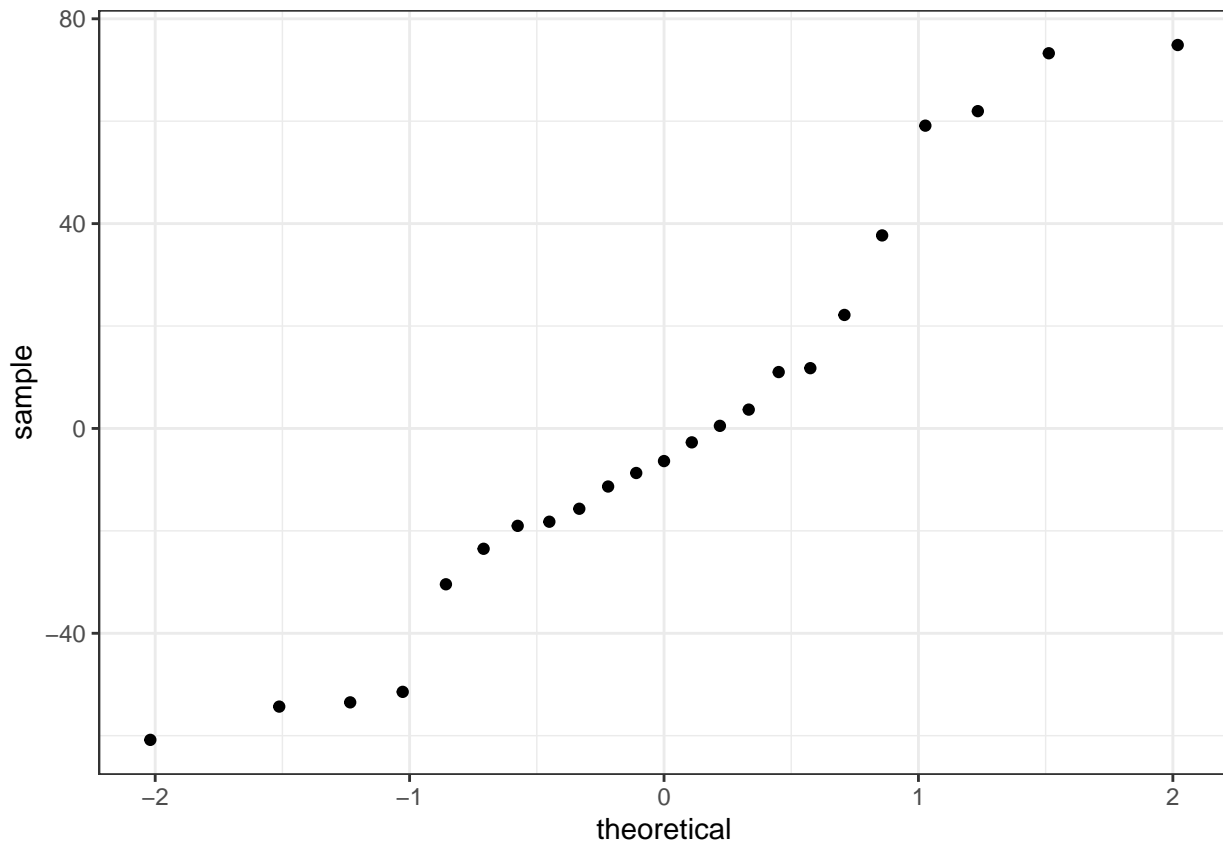
From the perspective of  $R^2 = 0.982$ , the model explains 98.2% of the variation of the dependent variable. From the table on the previous page, we can observe that the variables that are significant for this model are the income, the price of chicken. (*significance level: 0.05*) The null hypothesis is  $\theta_2 = \theta_3 = \theta_4 = \theta_5 = 0$ . From the outcome of the F-test the p-value is much lower than  $\alpha = 0.05$ , and thus we reject the null hypothesis.

A relative increase in per capita income, price of pork and beef will, according to this model, result in a relative increase in the consumption of chicken. The effect of the price of beef was not expected. An increase in the price of chicken will decrease the consumption of chicken, holding all other variables constant.

If we look at our model, the price of the beef has a much smaller effect on the consumption of chicken compared to the income per capita. We can see that the real per capita income, and the price pork and beef all have a positive effect on the consumption of chicken and that the real price of chicken has a negative effect.

#### 4.4.2 Tests for model 4





According to the statistics theory, if the model were normal, we should have that the sample quantiles coincide with the theoretical quantiles of the normal distribution (dots should be along a diagonal line). However, this is **not** the case.

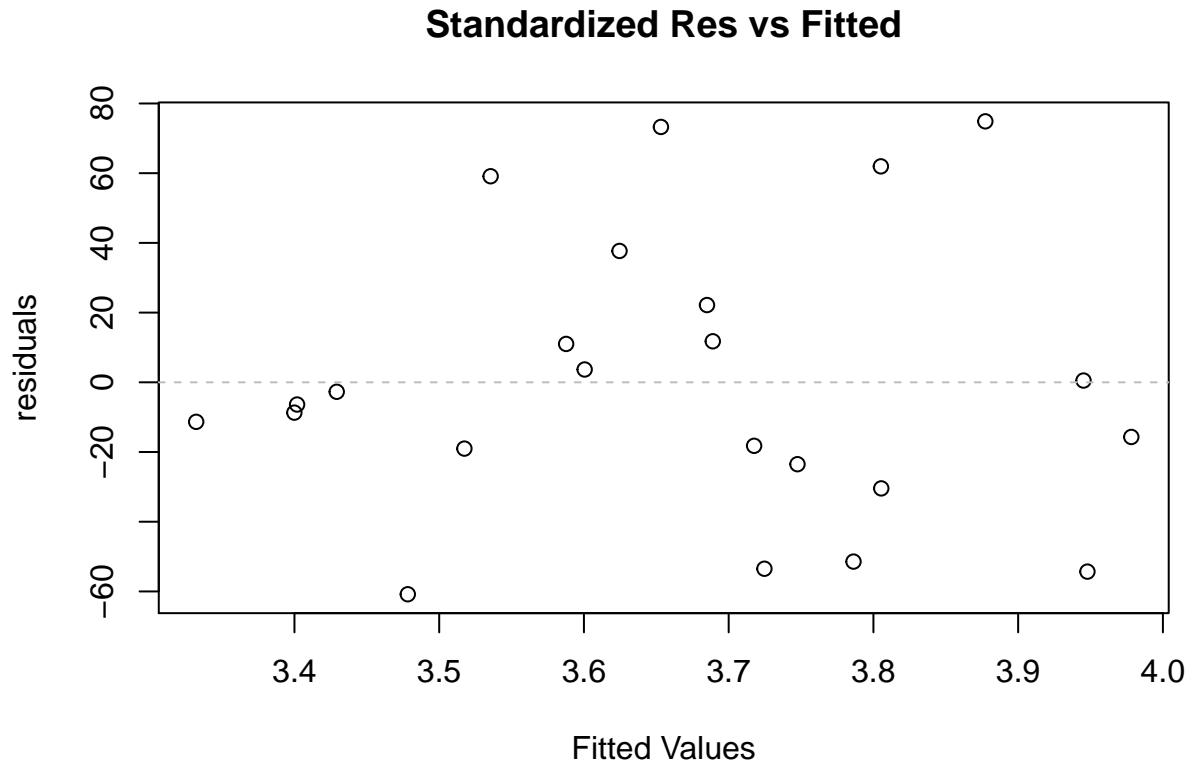
The model seems to fulfill the normality assumption according to the QQ-plot, but the histogram is not approximately normally distributed.

To make it more clear, we perform a JB-test.

Table 16: Jarque-Bera test for normality

Model	JB	p_value
1	1.676	0.176

Since the p-value is bigger than 0.05, we can not reject  $H_0$ .



We don't see any issue with the residuals for model 4 - the residuals are not following any kind of pattern which could indicate heteroscedasticity.

Table 17: studentized Breusch-Pagan test

	Model	BP	df	p_value
BP	4	5.977	4	0.201

The Breusch-Pagan test, or BP-test, tests for heteroscedasticity. If the p-value  $< 0.05$  then the error term of the model is heteroscedastic. Since the p-value is higher than 0.05, the error term of the model is not heteroscedastic.

Table 18: Variance Inflation Factors for the Model

	VIF
log(RCapInc)	65.11
log(RChicken)	17.49
log(RPork)	41.43
log(RBeef)	42.31

Since  $VIF > 10$ , we do observe a risk of multicollinearity.

## 4.5 Model 5

### 4.5.1 Estimation of Model 5

Our fifth and final model is:

$$\ln CapChicken_t = \beta_1 + \beta_2 \ln RCapInc_t + \beta_3 \ln RChicken_t + \beta_4 \ln RSub_t + u_t$$

In this model the per capita consumption of chicken is a function of the per capita income of the consumer, plus the price of chicken, plus the price of the related substitute for chicken. Here we can assume that an increase in the price of chicken and beef and a decrease in the price of related substitute for chicken will lead to less consumption of chicken due to consumers buying more pork. We can also connect level of income to what kind of meat consumers might purchase.

If the base level of income is quite low then an increase to a certain point of income will lead to an increase in the amount of chicken that consumers purchase. However, a bigger increase of income will lead to a decrease in the amount of chicken that consumers purchase because the consumer will be able to buy food more expensive than chicken, holding all other variables constant. If the income decreases we assume that the consumer will buy something more pork instead, since we assume it to be generally cheaper than chicken, holding all other variables constant. We assume income to have a positive effect on chicken consumption to a certain point, so,  $(\beta_2 > 0)$  if  $RCapInc_t < k$ , and  $(\beta_2 < 0)$  if  $RCapInc_t > k$ , holding all other variables constant. We also assume that the price of chicken should have a negative effect on consumption  $(\beta_3 < 0)$ , and price for related substitute for chicken should have a positive effect on chicken consumption  $(\beta_4 > 0)$ , holding all other variables constant.

Table 19: Regression Results for Model 5

	<i>Dependent variable:</i>
	log(CapChicken)
log(RCapInc)	0.481*** (0.068)
log(RChicken)	-0.351*** (0.079)
log(RSub)	-0.061 (0.130)
Constant	2.030*** (0.119)
Observations	23
R <sup>2</sup>	0.980
Adjusted R <sup>2</sup>	0.977
Residual Std. Error	0.028 (df = 19)
F Statistic	315.206*** (df = 3; 19)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Estimate of Model 5:

$$\ln CapChicken_t = 2.030 + 0.481 \ln RCapInc_t - 0.351 \ln RChicken_t - 0.061 \ln RSub_t + u_t$$

From the perspective of  $R^2 = 0.98$ , the model explains 98% of the variation of the dependent variable. Above, we can see that the significant variables for this model are the income and the price of chicken. (*significance level: 0.05*)

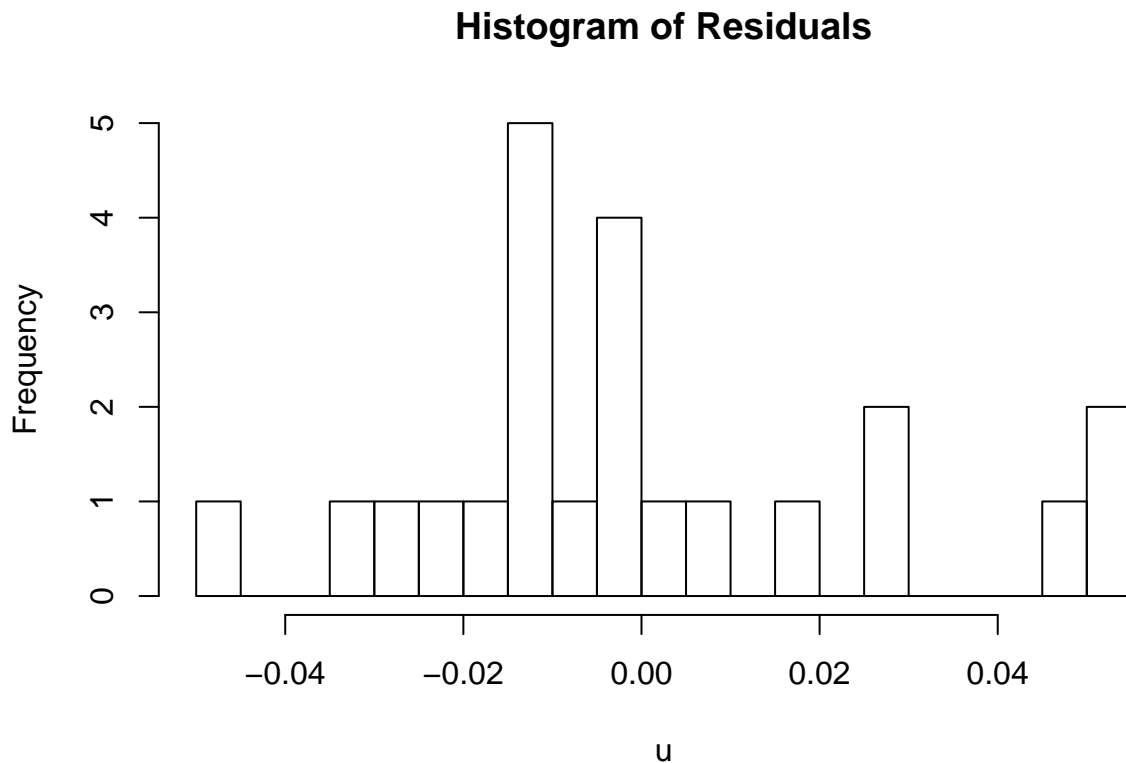
The null hypothesis is  $\theta_2 = \theta_3 = \theta_4 = 0$ .

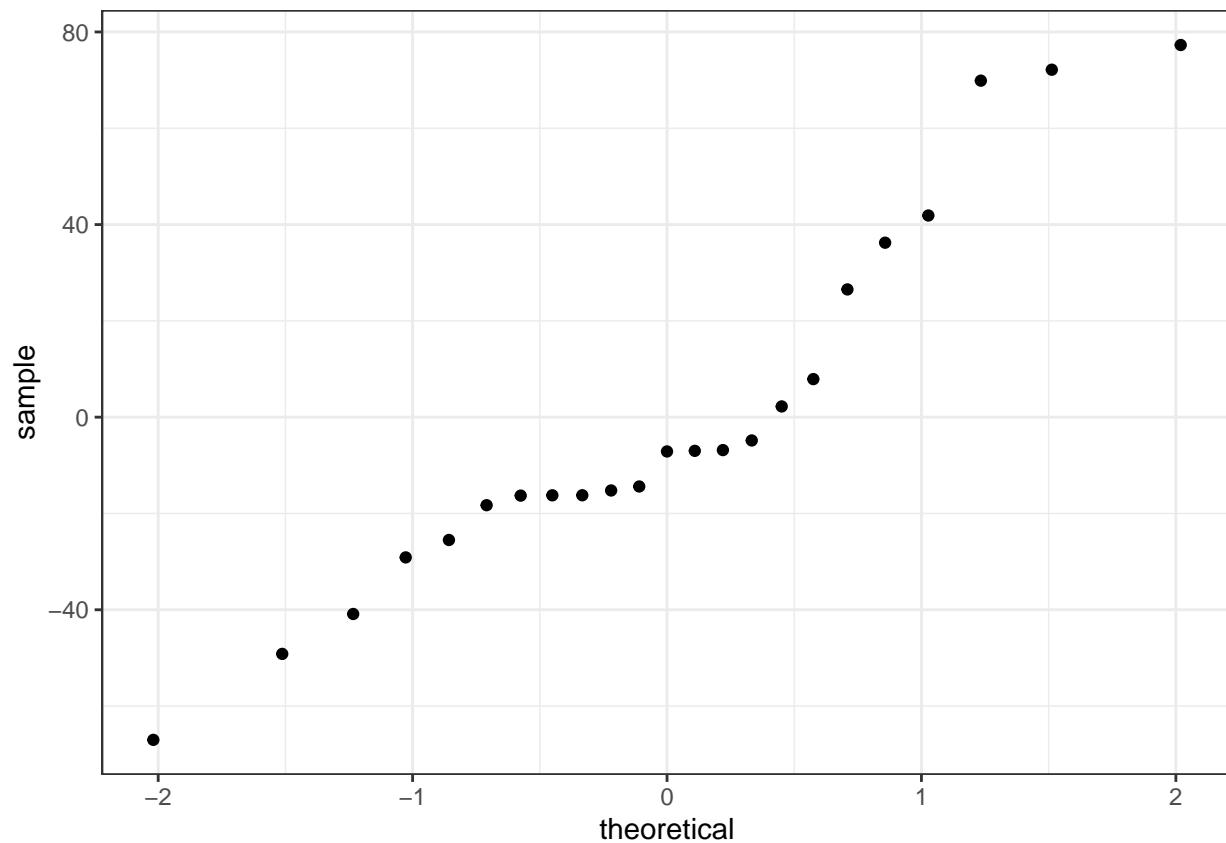
From the outcome of the F-test the p-value is much lower than  $\alpha = 0.05$ , and thus we reject the null hypothesis.

A relative increase in per capita income will, according to this model, result in a relative increase in the consumption of chicken. An increase in the price of chicken will decrease the consumption of chicken, holding all other variables constant.

If we look at our model, the price of the related substitute for chicken has a much smaller effect on the consumption of chicken compared to the income per capita and the price for chicken. However, the negative effect of the real prices of chicken was what we expected.

#### 4.5.2 Tests for model 5





According to the statistics theory, if the model were normal, we should have that the sample quantiles coincide with the theoretical quantiles of the normal distribution (dots should be along a diagonal line). However, this is **not** the case.

The model seems to fulfill the normality assumption according to the QQ-plot.

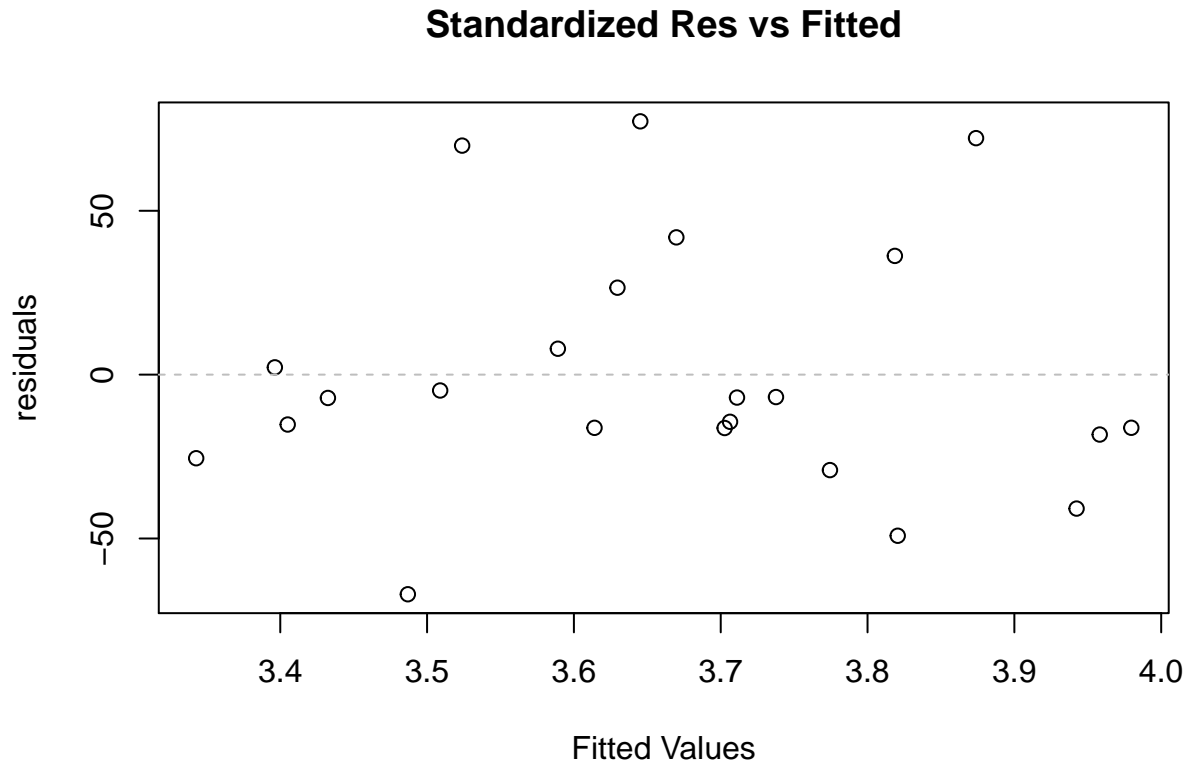
To make it more clear, we perform a JB-test.

Table 20: Jarque-Bera test for normality

Model	JB	p_value
1	1.676	0.176

Since the p-value is bigger than 0.05, we can not reject  $H_0$ .





We don't see any issue with the residuals for model 5 - the residuals are not following any kind of pattern which could indicate heteroscedasticity.

Table 21: studentized Breusch-Pagan test

	Model	BP	df	p_value
BP	5	0.5	3	0.919

The Breusch-Pagan test, or BP-test, tests for heteroscedasticity. If the p-value  $< 0.05$  then the error term of the model is heteroscedastic. Since the p-value is higher than 0.05, the error term of the model is not heteroscedastic.

Table 22: Variance Inflation Factors for the Model

	VIF
log(RCapInc)	41.39
log(RChicken)	8.50
log(RSub)	57.89

Since  $VIF > 10$ , we do observe a risk of multicollinearity.

## 5 Task 1b

Table 23: Comparison among 5 models

	<i>Dependent variable:</i>				
	log(CapChicken)				
	(1)	(2)	(3)	(4)	(5)
log(RCapInc)	0.452*** (0.025)	0.406*** (0.045)	0.441*** (0.052)	0.343*** (0.083)	0.481*** (0.068)
log(RChicken)	-0.372*** (0.063)	-0.439*** (0.083)	-0.381*** (0.076)	-0.505*** (0.111)	-0.351*** (0.079)
log(RPork)		0.107 (0.088)		0.149 (0.100)	
log(RBeef)			0.021 (0.092)	0.091 (0.101)	
log(RSub)					-0.061 (0.130)
Constant	2.033*** (0.116)	2.125*** (0.138)	2.039*** (0.122)	2.190*** (0.156)	2.030*** (0.119)
Observations	23	23	23	23	23
R <sup>2</sup>	0.980	0.982	0.980	0.982	0.980
Adjusted R <sup>2</sup>	0.978	0.979	0.977	0.978	0.977
Residual Std. Error	0.028 (df = 20)	0.027 (df = 19)	0.028 (df = 19)	0.028 (df = 18)	0.028 (df = 19)
F Statistic	491.868***	336.181***	312.419***	249.928***	315.206***

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 24: studentized Breusch-Pagan test

Model	BP	df	p_value
1	1.005	2	0.605
2	4.565	3	0.207
3	0.953	3	0.813
4	5.977	4	0.201
5	0.500	3	0.919

Table 25: Jarque-Bera test for normality

Model	JB	p_value
1	1.676	0.176
2	1.069	0.392
3	1.675	0.199
4	1.078	0.395
5	1.648	0.208

For the BP and JB tests, all the models pass both tests, which means they all fulfill the assumptions of normality and heteroscedasticity.

## 5.1 Comparison of BIC and AIC for all five models.

Now we are going to investigate all models and compare their BIC(Bayesian information criterion) and AIC(Akaike information criterion). The best value for both of these is the lowest value.

Table 26: AIC and BIC for 5 Models

Model	df	AIC	BIC
1	4	-94.78	-90.24
2	5	-94.50	-88.82
3	5	-92.84	-87.17
4	6	-93.52	-86.71
5	5	-93.04	-87.37

For the AIC we can see that model 1 is the best, since -94.77735 is the lowest value. For BIC, model 1 wins again, since -90.23538 is the lowest value.

## 5.2 Choosing the best model

When we have estimated all of the five models, we need to choose the one we deem the most appropriate to use. We have a couple of different criteria to look at:

- 1) Adjusted  $R^2$ . This shows how much of the variation of the dependant variable is explained by the regressors, adjusted for the number of regressors.
- 2) AIC. We want this value to be as low as possible.
- 3) BIC. We want this value to be as low as possible.
- 4) multicollinearity. Here we want our model to have as low a value as possible.

In terms of criteria 1 model 2 is our winner, since it has the highest adjusted  $R^2$  at 97.9%, as seen in table 13. However, all the models are fairly close to each other and none stand out significantly from the rest. For criteria 2 model 1 wins and for criteria 3, model 1 wins again, since it has the lowest values. Finally, model 1 is the only model that doesn't have a problem with multicollinearity. Another factor in favor of model 1 is that the variables in model 1, Real Capita and Price of Chicken (which are shared with the other models), are all significant, while none of the variables that are unique to the other models, such as price of beef or pork, are significant, as can be seen in table 13. Thus we cannot confirm that these other variables actually contribute to explaining the per capita consumption of chicken. The difference in  $R^2$  between model 1 and model 2 is also very small.

As we can see in our table above, the estimate of the Intercept is 2.032. Without taking account of any other factors, the consumption of chicken per capita on itself is important. The estimate of the coefficient for the per capita income is 0.451. It is smaller than 1 but still significant. An increase in the income will affect positively the chicken consumption, holding all other variables constant. The estimate of the coefficient for the real price of chicken is  $-0.372$ . A not so large increase in the price of chicken would affect negatively the chicken consumption, holding all other variables constant.

### 5.3 T-test

We perform a t-test for the parameters  $\alpha_2$  and  $\alpha_3$ .

#### 1. Hypotheses

$$H_{0,i} : \alpha_i = 0$$

where  $i = 2, 3$

$$H_{1,i} : \alpha_i \neq 0 \quad \text{where } i = 2, 3$$

#### 2. Significance level

$$\alpha = 0.05$$

#### 3. Estimators

$$\hat{\alpha}_2 = 0.406$$

$$\hat{\alpha}_3 = -0.439$$

$$\hat{\alpha}_4 = 0.107$$

#### 4. Assumptions

$$\mu_t \sim NID(0, \sigma^2)$$

$$Cov(\log(RCapInc_t), \mu_t) = 0$$

$$Cov(\log(RChicken_t), \mu_t) = 0$$

No exact multicollinearity

#### 5. Test statistic:

$$t_{obs,i} = \frac{\hat{\alpha}_i - \alpha^{H_0}}{\hat{\sigma}_{\hat{\alpha}_i}} \sim t_{n-k}$$

$$i = 1, \dots, 4$$

$k$  : no. of parameters in the model,  $k = 3$

$n$  : no. of observations,  $n = 23$

#### 6. Rejection rule and figure:

Reject the null hypothesis if  $t_{obs,i} > t_{n-k, \alpha/2}$  or if  $t_{obs,i} < -t_{n-k, \alpha/2}$  where  $i = 2, 3$ .

#### 7. Calculations:

At  $H_{0,2}$  we get  $t_{obs,2} = 18.284 > 2.085 = t_{20, 0.025}$  At  $H_{0,3}$  we get  $-2.093 < t_{obs,3} = -5.865 < -2.085$

Hence, we reject  $H_{0,2}$  and  $H_{0,3}$

## 8. Conclusions:

On a 5% significance level, we reject  $H_0 = \alpha_2 = 0$  and  $H_0 = \alpha_3 = 0$ . Thus, we reject the null hypothesis that the regressors RCapInc and RChicken do not affect the variation of the per capita consumption of chicken.

## 5.4 F-test

We perform a F-test to test if the model explains any of the variation in the dependent variable.

### 1. Hypotheses

$$H_0 : \alpha_2 = \alpha_3 = 0$$

$$H_0 : \text{At least one of } \alpha_i \neq 0 \quad \text{where } i = 2, 3$$

.

### 2. Significance level

$$\alpha = 0.05$$

### 3. Estimators

$$\hat{\alpha}_2 = 0.451$$

$$\hat{\alpha}_3 = -0.372$$

$$R_R^2 = 0$$

$$R_{UR}^2 = 0.98$$

### 4. Assumptions

$$\mu_t \sim NID(0, \sigma^2)$$

$$Cov(\log(RCapInc_t), \mu_t) = 0$$

$$Cov(\log(RChicken_t), \mu_t) = 0$$

No exact multicollinearity

### 5. Test statistic:

$$F = \frac{(R_{UR}^2 - R_R^2)/m}{(1 - R_{UR}^2)/(n-k)} \sim F_{m, (n-k)}$$

$m$  : no. of restrictions ,  $m = 2$

$k$  : no. of parameters in the model,  $k = 3$

$n$  : no. of observations,  $n = 23$

### 6. Rejection rule and figure:

Reject the null hypothesis if  $F_{obs} > F_{m, n-k, \alpha}$ .

### 7. Calculations:

$$F_{obs} = 491.868 > 3.492 = F_{3, 19, 0.05}$$

Hence, we reject the null hypothesis.

#### **8. Conclusions:**

On the 5% significance level, we reject  $H_0: \alpha_2 = \alpha_3 = 0$ . Thus we reject the hypothesis that the model does not explain any variation of the consumption of chicken.



## 6 Task 1c

Now that we have chosen our model, we will direct our attention to the fourth model and investigate any potential statistical problems with it. After that we are going to compare the model specification of model 4 with the model specification of model 5. We want to see which model is the best from a “specification” point of view, and what the reasons for this are.

We can see that model 4 has the highest amount of variables out of all our models. This might make the model less robust than we want it to be, due to potential risk of multicollinearity: When one predictive variable can be understood in terms other predictive variables. For example, if a high degree of covariance is detected between the various predicting variables, which in turn means that the predicting variables can be explained in terms of each other.

To this end we are going to calculate the variance inflation factor (VIF) for our variables.

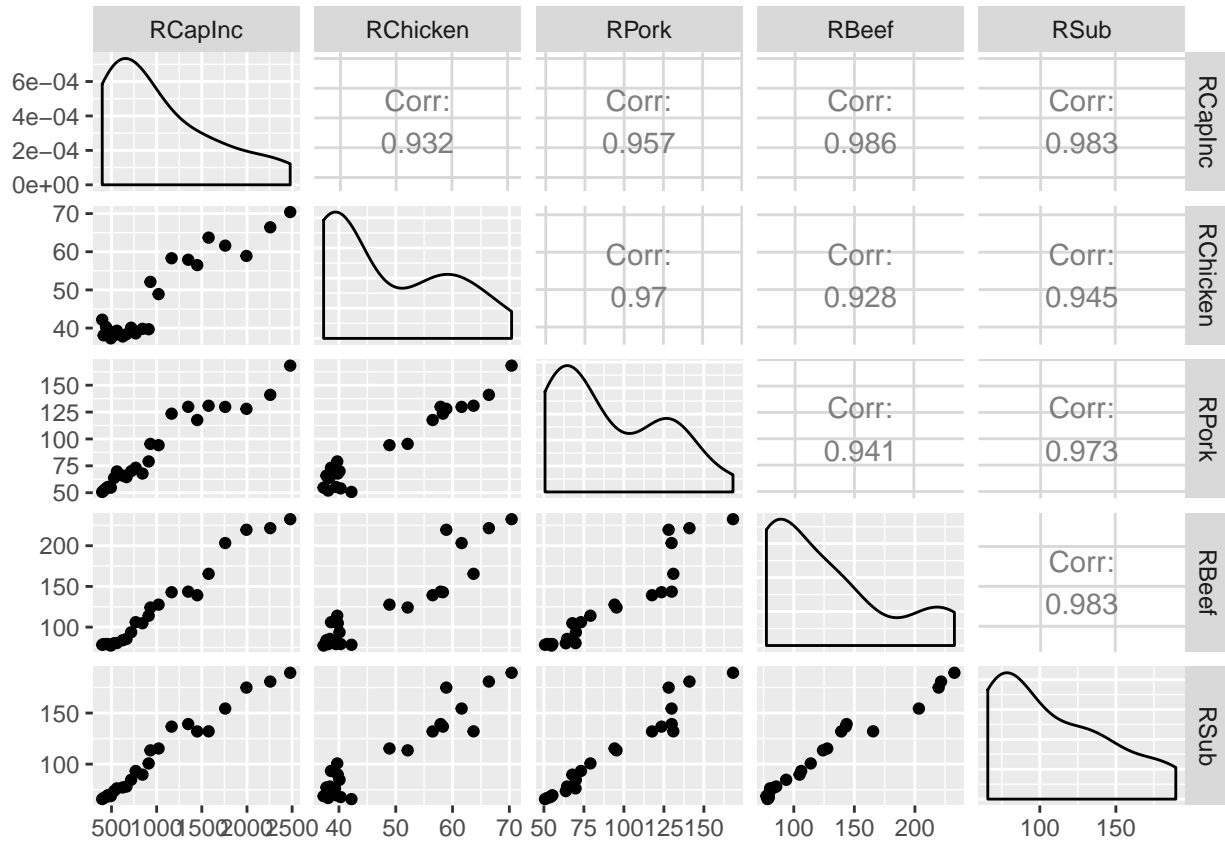
We know that if the VIF is  $> 10$ , we have a problem with multicollinearity.

Table 27: Variance Inflation Factors for Model 4

	VIF
log(RCapInc)	65.11
log(RChicken)	17.49
log(RPork)	41.43
log(RBeef)	42.31

As we can see from the above calculations, we have a problem with multicollinearity, due to all values being above 10.

This means that trying to determine the effect of one variable will be difficult. We can observe that the variable for per capita income, RCapInc, has a value fairly close to 10. This means it is not as strongly correlated with the other variables as the variables representing meat prices are. The price of beef especially seems to have a high variance inflation factor. This could mean some variables risk being redundant.



Plot: This plot is a distribution and correlation plot for all the explanatory variables. The upper part of the matrix shows the correlation between the variables listed along the horizontal axis and the vertical axis. For example, the correlation between RChicken and RCapInc is 0.932. The diagonal part of the matrix shows the distribution for each explanatory variable in a graph format. The lower part of the matrix shows the spread of the error terms of the correlated variables.

Our variables are all correlated and explain each other, as can be seen in our model above. All of our variables have a correlation with each other of at least 92.8%. Our highest correlation is between the price of beef and the real per capita income, at 98,6%. Since all the variables are highly correlated, it will be difficult to determine the impact of one variable on its own, due to the variables all explaining each other.

One way to prevent multicolliniarty could be decreasing the amount of variables we have to work with. To do this we can redefine several variables into one new variable. This is exactly what's been done in model 5 - here, the variables of price of pork and price of beef have been combined into price of chicken substitutes. The idea is that we want the predictive variables to have low correlation with each other, but high correlation with the dependant variable.

Table 28: Variance Inflation Factors for Model 5

	VIF
log(RCapInc)	41.39
log(RChicken)	8.50
log(RSub)	57.89

We can see that model 5 also has some colliniarity, but that the effect of the new composite variable is lower compared to model 4. The new redefined variables of RSub seems to have a higher multicollinearity than RPork and RBeef each had, but the VIF of real income has decreased. The VIF of the price of chicken has decreased heavily. Overall, there is not that big of a difference except for the price of chicken which has a VIF

smaller than 10. It should thus be easier to determine the effects of the price of chicken in model 5 compared to model 4. This in turn, might make model 5 preferable to model 4 - in model 4 all explanatory variables have a problem with multicollinearity, whereas in model 5 one variable does not.

Table 29: Comparison between Model 4 and Model 5

	<i>Dependent variable:</i>	
	log(CapChicken)	
	(1)	(2)
log(RCapInc)	0.343*** (0.083)	0.481*** (0.068)
log(RChicken)	-0.505*** (0.111)	-0.351*** (0.079)
log(RPork)	0.149 (0.100)	
log(RBeef)	0.091 (0.101)	
log(RSub)		-0.061 (0.130)
Constant	2.190*** (0.156)	2.030*** (0.119)
Observations	23	23
R <sup>2</sup>	0.982	0.980
Adjusted R <sup>2</sup>	0.978	0.977
Residual Std. Error	0.028 (df = 18)	0.028 (df = 19)
F Statistic	249.928*** (df = 4; 18)	315.206*** (df = 3; 19)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

And according to the comparison table, model 4 has a slightly higher value of *adjusted R<sup>2</sup>* compared to the model 5. The explanatory variable RSub which is the composite price of chicken substitutes in model 5 is **not significant**. In fact, none of the variables other than the variables from model 1 are significant. Thus we can't really say that there is a huge difference between the models, but model 4 still has a slightly higher *adjusted R<sup>2</sup>*. Model 5 does have less of a problem with multicollinearity, which does make it more preferable, since the other differences are almost minimal.

## 7 Conclusion

For this homework assignment, we have studied the consumption of chicken in the US between the years 1960 and 1982. We looked at the per capita consumption of chicken, and tried to explain it using variables such as per capita income and various data sets about the real price of goods, such as chicken, pork and beef. We have studied five different models and compared how well they explained our data. We found that the most basic model, the first model, fulfilled most of our criterion. It was therefore chosen as the most appropriate model for explaining the data.

We were surprised that the most basic model was the best one. However, we noticed that the added variables to the other models were not significant. The first model was also the only model that did not have a problem with multicollinearity. It made sense in the end to choose it as our best model.

In the final part of the assignment, we looked at the problem of multicollinearity. This is when variables can be explained in terms of each other. We compared our fourth model, which had the most variables, with our fifth model, which had a composite variable. We wanted to measure multicollinearity due to the fact that a high multicollinearity can make it difficult to measure the effect of one single explanatory variable on the dependant variable. We found that our composite model had less of a problem with multicollinearity than the model with lots of variables, and thus judged our composite model, our fifth model, to be the more preferable of the two.