Solution to Exercise E5.

The Multiple Regression Model. Estimation.

Exercise E5.1. Beach umbrella rental

Part I. Simple Linear Regression Model.

a. Regression model: $U_t = \beta_1 + \beta_2 T_t + u_t \quad t = 1, \ldots, 22$

Model 1: OLS, using observations 2013-05-05–2013-09-29 ($T = 22$)
Dependent variable: U

	Coefficient	Std. Error	t-ratio	p-value
const	27.0692	26.9160	1.0057	0.3266
T	11.4595	0.860078	13.3238	0.0000

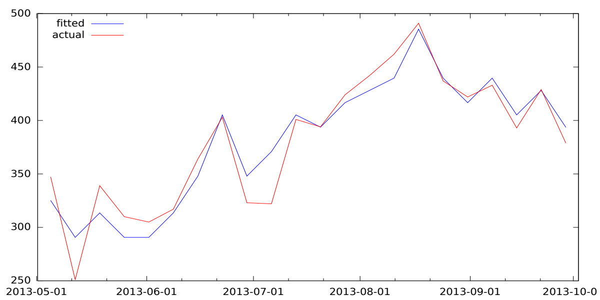
Mean dependent var	381.2727	S.D. dependent var	60.60110
Sum squared resid	7808.908	S.E. of regression	19.75969
R^2	0.898747	Adjusted R^2	0.893684
$F(1, 20)$	177.5241	P-value(F)	2.09e–11
Log-likelihood	−95.80841	Akaike criterion	195.6168
Schwarz criterion	197.7989	Hannan–Quinn	196.1308
$\hat{\rho}$	0.127938	Durbin–Watson	1.661439

b. SRF: $\hat{U}_t = 27.0692 + 11.4595\, T_t$

c. The increase in the number of beach umbrellas rented when the temperature increases by 1°C is estimated at 11.4595 umbrellas. The estimate has the expected sign, because the higher the temperature is, the more umbrellas are rented.

d. 89.8747% of the sample variation in the number of umbrellas rented is explained by the variation in temperature.

e. The fit seems adequate: the long-term behaviour of the number of rented umbrellas is properly reflected.



g. `Add --> Define new variable ...` *total=sum(U)*

8388 umbrellas were rented in this period.

h. Save the fitted values by clicking `Save --> Fitted values` in the menu bar of the estimation results window.

   Highlight the variables U and $yhat1$, right-click and select the *Summary Statistics* option from the pulldown menu.

   The sample mean of the number of rented umbrellas is 381.27, which coincides with the sample mean of the fitted number of rented umbrellas because it is one of the properties of the Sample Regression Function.

i. Highlight the variable $yhat1$ using the cursor, right-click and select the *Display values* option from the pulldown menu. The number of umbrellas rented in the first week of August is estimated at 428.1526.

j. Save the residuals by clicking `Save --> Residuals` in the menu bar of the estimation results window. Highlight the variable $uhat1$, right-click and select the *Display values* option. The estimation error works out at -14.77403 umbrellas, that is, the number of rented umbrellas is overestimated. This error is called residual and it comes from two sources:

   - the estimation error derived from estimating the coefficients of the model and
   - the fact that the error term is unobservable and unpredictable.

k. If the average temperature over a week is 26°C, the estimated number of beach umbrellas rented in the week is 325.0162.

l. If the average temperature rises by 2°C from one week to the next, the estimated change in the number of rented umbrellas is $2 \times \hat{\beta}_2 = 22.919$.

Part II. General Linear Regression Model.

a. Regression model: $U_t = \beta_1 + \beta_2 T_t + \beta_3 P_t + v_t \quad t = 1, \ldots, 22$

Model 2: OLS, using observations 2013-05-05–2013-09-29 ($T = 22$)
Dependent variable: U

	Coefficient	Std. Error	t-ratio	p-value
const	21.6998	40.7472	0.5325	0.6005
T	11.7963	2.07726	5.6788	0.0000
P	−0.591532	3.30317	−0.1791	0.8598

Mean dependent var	381.2727	S.D. dependent var	60.60110
Sum squared resid	7795.750	S.E. of regression	20.25593
R^2	0.898917	Adjusted R^2	0.888277
$F(2,19)$	84.48230	P-value(F)	3.50e–10
Log-likelihood	−95.78985	Akaike criterion	197.5797
Schwarz criterion	200.8528	Hannan–Quinn	198.3508
$\hat{\rho}$	0.152956	Durbin–Watson	1.616466

b. SRF: $\hat{U}_t = 21.6998 + 11.7963\, T_t - 0.5915\, P_t \quad t = 1, \ldots, 22$

c. Interpretation of the estimated coefficients:

$\hat{\beta}_2$: It is estimated that the number of rented umbrellas increases by 11.7963 when the temperature increases by 1°C, holding the price fixed. This figure has the expected sign, because the warmer it gets the more umbrellas are rented.

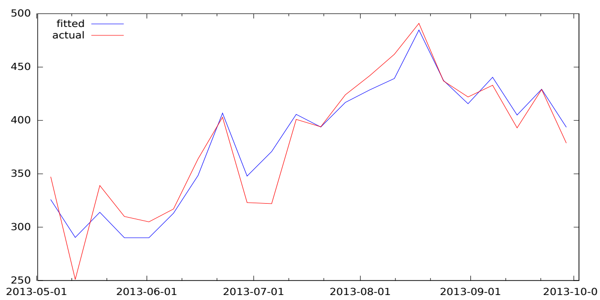
$\hat{\beta}_3$: It is estimated that the number of rented umbrellas falls by 0.591532 when the price increases by €1, holding the temperature fixed. This figure has the expected sign, because the more expensive it is to rent umbrellas, the fewer are rented. Given the estimate obtained, a price increase of €2 would be required to bring down the number of umbrellas rented by one unit.

d. This model contains one more explanatory variable, the average price of renting a beach umbrella.

e. No, because although these are estimates of the coefficients for the same explanatory variable they come from two different models.

f. 89.8917% of the variation in the number of rented umbrellas in the sample can be explained by the temperature and price variations. The value of the coefficient of determination is higher than the one obtained for the previous model because model (2) contains one more explanatory variable. This does not mean that this model is better specified: the significance of the additional variable (price) would have to be analysed.

g. There are not great differences between this actual-fitted values plot and the one obtained for the previous model. It seems that the inclusion of the variable price in the model might not have a relevant influence in the results.



h. First, save the fitted values and the residuals for model (2). Then, highlight the variables U, T, P, $yhat2$ and $uhat2$, right-click and select the option *Summary statistics*. The results obtained are:

Summary Statistics, using the observations 2013-05-05–2013-09-29

Variable	Mean	Median	Minimum	Maximum
U	381.273	393.500	251.000	491.000
P	8.52273	9.00000	4.50000	15.0000
T	30.9091	32.5000	23.0000	40.0000
yhat2	381.273	399.462	290.058	484.681
uhat2	0.000000	2.04077	−48.8580	25.0534

Variable	Std. Dev.	C.V.	Skewness	Ex. Kurtosis
U	60.6011	0.158944	−0.252001	−0.696409
P	3.15277	0.369925	0.495965	−0.466571
T	5.01340	0.162198	−0.267677	−1.06131
yhat2	57.4567	0.150697	−0.282628	−1.06230
uhat2	19.2672	3.38953e+014	−0.975379	0.519199

Variable	5% perc.	95% perc.	IQ range	Missing obs.
U	259.100	486.650	107.250	0
P	4.50000	15.0000	4.25000	0
T	23.0000	39.4000	9.25000	0
yhat2	290.058	478.047	106.011	0
uhat2	−47.4324	24.7049	23.7001	0

As indicated by the properties of the SRF, the sample means of the dependent and fitted variables match, and the sample mean of the residuals is zero. The variable that shows the greatest variability is the beach umbrella variable, and that which shows the least is the price variable. The sample mean of the price is €8.52273 and the sample mean of the temperature is 30.9091°C.

i. If the average temperature in a given week were 39°C, the estimated number of umbrellas rented in that week would be $\hat{U}_t = 481.7555 - 0.5915\, P_t$.

   And if the average charge per day were €13, the estimated number of beach umbrellas rented in that week would be $\hat{U}_t = 474.066$.

j. If the family firm decided to charge the same amount throughout the season, then the third column in the data matrix X would be constant. This means that the X matrix would not have full rank, and it would not be possible to estimate all the coefficients of the model individually. This problem is known as perfect collinearity.

Part III. General Linear Regression Model.

a. Regression model: $U_t = \beta_1 + \beta_2\, T_t + \beta_3\, P_t + \beta_4\, WW_t + v_t \quad t = 1, \ldots, 22$

Model 3: OLS, using observations 2013-05-05–2013-09-29 ($T = 22$)
Dependent variable: U

	Coefficient	Std. Error	t-ratio	p-value
const	44.5693	45.7575	0.9740	0.3429
T	11.0471	2.18129	5.0645	0.0001
P	−0.0524909	3.32631	−0.0158	0.9876
WW	−10.5245	9.73983	−1.0806	0.2942

Mean dependent var	381.2727	S.D. dependent var	60.60110
Sum squared resid	7320.861	S.E. of regression	20.16716
R^2	0.905075	Adjusted R^2	0.889254
$F(3, 18)$	57.20762	P-value(F)	2.11e–09
Log-likelihood	−95.09850	Akaike criterion	198.1970
Schwarz criterion	202.5612	Hannan–Quinn	199.2251
$\hat{\rho}$	0.107465	Durbin–Watson	1.685968

SRF: $\hat{U}_t = 44.5693 + 11.0471\, T_t - 0.0524909\, P_t - 10.5245\, WW_t \quad t = 1, \ldots, 22$

b. The explanatory variable *windy week* has been added. This is a qualitative variable. It is introduced into the model by means of the dummy variable WW, which takes the value 1 when the observation comes from a windy week and 0 if it is from a non windy week.

c. `Add --> Define new variable ...` *total=sum(U*WW)*

The estimated number of umbrellas rented in a windy week is 3131.

`Add --> Define new variable ...` *total=sum(U)-sum(U*WW)*

The estimated number of umbrellas rented in a non windy week is 5257.

d. It is estimated that the difference between the number of beach umbrellas rented in a non windy week and in a windy week is 10.5245, holding the remaining characteristics (temperature and price) constant.

e. It is estimated that the number of beach umbrellas rented falls by 0.05249 when the price increases by €1 holding the remaining explanatory variables constant.

f. It is estimated that the number of beach umbrellas rented when the price is €7 and the average temperature for the week is 30°C is

$$\hat{U}_t = 375.61487 - 10.5245\ WW_t.$$

Windy week: $\hat{U}_t = 365.09037$ umbrellas.

Non windy week: $\hat{U}_t = 375.61487$ umbrellas.

Exercise E5.2 Holiday cottages

Model A

a. Regression model: $RP_i = \alpha_1 + \alpha_2\, NR_i + \alpha_3\, BP_i + u_i \quad i = 1, 2, \ldots, 75$

Model 1: OLS, using observations 1–75
Dependent variable: RP

	Coefficient	Std. Error	t-ratio	p-value
const	38.4321	7.22899	5.3164	0.0000
NR	2.26766	1.20082	1.8884	0.0630
BP	1.49558	1.09746	1.3628	0.1772

Mean dependent var	56.13893	S.D. dependent var	14.98446
Sum squared resid	15263.15	S.E. of regression	14.55982
R^2	0.081392	Adjusted R^2	0.055875
$F(2, 72)$	3.189724	P-value(F)	0.047064
Log-likelihood	−305.7595	Akaike criterion	617.5189
Schwarz criterion	624.4714	Hannan–Quinn	620.2950

SRF: $\widehat{RP}_i = 38.4321 + 2.26766\, NR_i + 1.49558\, BP_i \quad i = 1, 2, \ldots, 75$

b. Interpretation of the estimated coefficients:

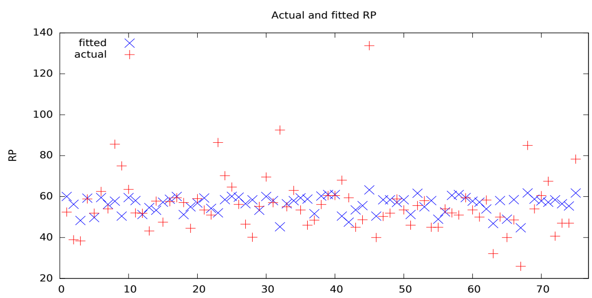
$\hat{\alpha}_2$: The estimated variation in the price of a room when the holiday cottage has an additional room and the price of breakfast remains fixed is €2.26766.

$\hat{\alpha}_3$: The estimated variation in the price of the room when the price of breakfast increases by €1 holding the number of bedrooms fixed is €1.49558.

c. The estimated price of a room when breakfast is included in the price and the holiday cottage has 10 bedrooms is €61.1087.

d. If breakfast is included in the room price the estimated variation in price between a holiday cottage with 15 bedrooms and one with 10 bedrooms is $\hat{\alpha}_2 \times 5 =$ €11.3383.

e. The fit is quite poor: only the average level of the series is reflected.



f. No. Given the value of the coefficient of determination (0.081392) the fit is quite poor.

Model B

a. Regression model:

$$RP_i = \lambda_1 + \lambda_2\, NR_i + \lambda_3\, BP_i + \lambda_4\, WIFIF_i + \lambda_5\, WIFIP_i + \lambda_6\, LOCC_i + u_i$$

Model 2: OLS, using observations 1–75
Dependent variable: RP

	Coefficient	Std. Error	t-ratio	p-value
const	40.5761	7.39661	5.4858	0.0000
NR	1.94192	1.21303	1.6009	0.1140
BP	0.559911	1.21918	0.4593	0.6475
WIFIF	6.98544	3.65362	1.9119	0.0600
WIFIP	−5.75696	12.0827	−0.4765	0.6352
LOCC	2.11170	5.43209	0.3887	0.6987

Mean dependent var	56.13893	S.D. dependent var	14.98446
Sum squared resid	14429.98	S.E. of regression	14.46133
R^2	0.131536	Adjusted R^2	0.068604
$F(5, 69)$	2.090130	P-value(F)	0.077001
Log-likelihood	−303.6545	Akaike criterion	619.3089
Schwarz criterion	633.2138	Hannan–Quinn	624.8610

SRF:

$$\begin{aligned}\widehat{RP}_i &= 40.5761 + 1.94192\, NR_i + 0.559911\, BP_i + 6.98544\, WIFIF_i - \\ &- 5.75696\, WIFIP_i + 2.11170\, LOCC_i\end{aligned}$$

b. Interpretation of the estimated coefficients.

$\hat{\lambda}_4$: The estimated difference in the price of a room between a holiday cottage that offers free WiFi access and one that does not offer WiFi access is €6.98544, holding the remaining characteristics constant. The sign is expected to be positive, as what is on offer is a new service free.

$\hat{\lambda}_5$: The estimated difference in the price of a room between a holiday cottage that does not offer WiFi access and one that offers it for an additional fee is €5.75696, holding the remaining characteristics constant. The sign is not expected to be positive, as what is on offer is the possibility of opting for a service.

c. The estimated price of a room when the holiday cottage has 6 rooms, offers WiFi access and the price of breakfast is €3 is:

$$\widehat{RP}_i = 53.907353 + 6.98544\, WIFIF_i - 5.75696\, WIFIP_i + 2.11170\, LOCC_i \text{ euros.}$$

Free WiFi access: $\widehat{RP}_i = 60.892793 + 2.11170\, LOCC_i$ euros.

WiFi access costs €2: $\widehat{RP}_i = 48.150393 + 2.11170\, LOCC_i$ euros.

d. The estimated price for the first cottage in the sample is €62.20859 while the actual price is €52.430. The fitted value does not match the actual value. This difference, called residual, is due to the estimation error derived from estimating the coefficients of the model and to the fact that the disturbance is unpredictable.

Model C

a. Regression model:

$$RP_i = \beta_1 + \beta_2\, NR_i + \beta_3\, BP_i + \beta_4\, WIFIF_i + \beta_5\, NPR_i + \beta_6\, BER_i + \beta_7\, LKR_i + u_i$$

Six explanatory variables are included in the model.

The differences between this regression model and the previous one are:

- This model contains three more explanatory variables: *proximity to a natural park, proximity to a lake or a reservoir* and *proximity to a beach.* Since only the dummy variables NPR, LKR and BER have been used to represent these explanatory variables, the model only differentiates between the holiday cottages located less than 1 km from the service and the ones further away.
- The qualitative explanatory variable WiFi has a different number of categories: while it had 3 categories in the previous model, it only has two in this one: free WiFi and no WiFi/paid WiFi.

Model 3: OLS, using observations 1–75
Dependent variable: RP

	Coefficient	Std. Error	t-ratio	p-value
const	33.9408	6.72069	5.0502	0.0000
NR	1.36156	1.11199	1.2244	0.2250
BP	1.62738	1.07599	1.5124	0.1351
WIFIF	9.02246	3.32099	2.7168	0.0084
NPR	3.33934	3.93959	0.8476	0.3996
BER	16.1587	4.47207	3.6133	0.0006
LKR	12.0185	7.81149	1.5386	0.1285

Mean dependent var	56.13893	S.D. dependent var	14.98446
Sum squared resid	11550.44	S.E. of regression	13.03301
R^2	0.304840	Adjusted R^2	0.243503
$F(6,68)$	4.969876	P-value(F)	0.000285
Log-likelihood	−295.3075	Akaike criterion	604.6151
Schwarz criterion	620.8375	Hannan–Quinn	611.0925

SRF:

$$\begin{aligned}\widehat{RP}_i &= 33.9408 + 1.36156\, NR_i + 1.62738\, BP_i + 9.02246\, WIFIF_i + \\ &+ 3.33934\, NPR_i + 16.1587\, BER_i + 12.0185\, LKR_i\end{aligned}$$

b. Interpretation of the estimated coefficients.

$\hat{\beta}_5$: The estimated difference in the price of a room between a holiday cottage located less than 1 km from a natural park and one further away is €3.33934, holding the rest of the characteristics constant. The sign is positive as expected since the proximity to a natural park should increase the price.

$\hat{\beta}_6$: The estimated difference in the price of a room between a holiday cottage located less than 1 km from a beach and one further away is €16.1587, holding the rest of the characteristics constant. The sign is positive as expected since the proximity to a beach should increase the price.

$\hat{\beta}_7$: The estimated difference in the price of a room between a holiday cottage located located less than 1 km from a lake or a reservoir and one further away is €12.0185, holding the rest of the characteristics constant. The sign is positive as expected since the proximity to a lake should increase the price.

c. Yes, the fit has improved. The coefficient of determination of this model is 0.304840, three times the coefficient of determination of the previous model. Nevertheless, the significance of the explanatory variables would have to be analysed.

Exercise E5.3 Soy milk

Part I. Data file organization.

To give the data set a time series structure click `Data --> Dataset structure` and choose the following options:

Structure of data set: Time series

Time series frequency: Monthly

Starting observation: 1990:01

Confirm data set structure.

To change the name and characteristics of the variables, highlight the variable of interest, right-click and select the *Edit attributes* option from the pulldown menu.

Save all the changes in the file `soymilk-sales.gdt`

Part II. $S = f(P)$

Regression model: $S_t = \gamma_1 + \gamma_2 P_t + u_t \qquad t = 1990:1, \ldots, 2012:6$

a. Descriptive statistics of sales.

Summary Statistics, using the observations 1990:01–2012:06
for the variable S (270 valid observations)

Mean	Median	Minimum	Maximum
120.848	119.963	76.2000	171.506
Std. Dev.	C.V.	Skewness	Ex. kurtosis
16.8791	0.139673	0.228360	0.729406
5% perc.	95% perc.	IQ Range	Missing obs.
89.3864	152.127	19.1710	0

Range: 76.2 - 171.51. Sample mean of sales: 120.85 thousands of euros.

b. The correlation coefficient between sales and price is: corr(S, P) = -0.60412271.

c. Estimation results:

Model 1: OLS, using observations 1990:01–2012:06 ($T = 270$)
Dependent variable: S

	Coefficient	Std. Error	t-ratio	p-value
const	204.889	6.82123	30.0370	0.0000
P	−0.846206	0.0681840	−12.4106	0.0000

Mean dependent var	120.8475	S.D. dependent var	16.87912
Sum squared resid	48668.71	S.E. of regression	13.47589
R^2	0.364964	Adjusted R^2	0.362595
$F(1, 268)$	154.0235	P-value(F)	3.01e–28
Log-likelihood	−1084.353	Akaike criterion	2172.707
Schwarz criterion	2179.903	Hannan–Quinn	2175.597
$\hat{\rho}$	0.994798	Durbin–Watson	0.042141

SRF: $\hat{S}_t = 204.889 - 0.846206\, P_t$

d. $\hat{\gamma}_2$. It is estimated that sales will fall by €846.206 if the price of soy milk increases by 1 one euro cent. The negative sign of this estimate was expected since the higher the price the fewer sales there are.

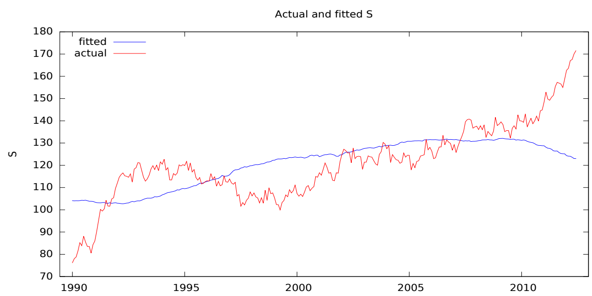
e. It is estimated that sales will drop by $30 \times \hat{\gamma}_2 = 25.38618$ thousands of euros if the price of soy milk increases by 30 euro cents.

f. 36.4964% of the sample variation in sales is explained by the variations in price.

g. $\hat{\sigma}^2 = \dfrac{SSR}{T-k} = \dfrac{48668.71}{270-2} = 181.597.$

h. $\widehat{Var}(\hat{\gamma}_2^{OLS}) = 0.0681840^2 = 0.00464905.$

i. The fit is quite poor: it does not even reflect the long-term behaviour of sales.



Part III. $S = f(P, AE)$

Regression model:

$$S_t = \beta_1 + \beta_2\, P_t + \beta_3\, AE_t + \beta_4\, AE_t^2 + u_t \quad t = 1990:1, \ldots, 2012:6$$

a. The model includes two explanatory variables: price and advertising expenditures. This model is different from the previous one because it contains one more explanatory variable: advertising expenditures. Note that sales are a quadratic function of advertising expenditures.

b. Yes, the regression model is linear in the coefficients. The model is not linear in the variables because the relationship between sales and expenditures is quadratic, but this fact does not affect the assumptions of the Multiple Regression Model.

c. Correlation matrix.

Correlation coefficients, using the observations 1990:01–2012:06
5% critical value (two-tailed) = 0.1194 for n = 270

S	P	AE	
1.0000	−0.6041	0.1614	S
	1.0000	−0.7444	P
		1.0000	AE

The simple correlation coefficients have the expected sign. Sales are proportional to advertising expenditure (more advertising means more sales) and inversely proportional to prices (the higher the price the fewer sales there are).

d. Estimation results

Model 2: OLS, using observations 1990:01–2012:06 ($T = 270$)
Dependent variable: S

	Coefficient	Std. Error	t-ratio	p-value
const	57.5860	43.7095	1.3175	0.1888
P	−1.38907	0.0807449	−17.2032	0.0000
AE	3.77879	0.610062	6.1941	0.0000
sq_AE	−0.0166627	0.00229460	−7.2617	0.0000

Mean dependent var	120.8475	S.D. dependent var	16.87912
Sum squared resid	28697.27	S.E. of regression	10.38674
R^2	0.625554	Adjusted R^2	0.621331
$F(3, 266)$	148.1277	P-value(F)	1.89e–56
Log-likelihood	−1013.042	Akaike criterion	2034.083
Schwarz criterion	2048.477	Hannan–Quinn	2039.863
$\hat{\rho}$	0.965596	Durbin–Watson	0.084094

SRF: $\hat{S}_t = 57.5860 - 1.38907\, P_t + 3.77879\, AE_t - 0.0166627\, AE_t^2$

e. The estimated variation in sales if advertising expenditures increase by €100, holding the price of soy milk constant, is

$$(3.77879 - 2 \times 0.0166627 \times AE_t) \quad \text{thousands of euros.}$$

This effect is not constant over the whole sample because it depends on the level of expenditures at each moment in time.

If advertising expenditures are €1500, the estimated variation in sales would be

$(3.77879 - 2 \times 0.0166627 \times 15 = 3.278909)$ thousands of euros $=$ sales are estimated to increase by €3278.909.

If advertising expenditures are €15000, the estimated variation in sales would be

$(3.77879 - 2 \times 0.0166627 \times 150 = -1.22002)$ thousands of euros $=$ sales are estimated to fall by €1220.002.

f. The estimated variation in sales if the price increases by 1 euro cent, holding advertising expenditures constant is $\hat{\beta}_2$ thousands of euros. This variation is constant throughout the sample.

If the price increases by half a euro, the estimated variation in sales would be

$50\hat{\beta}_2 = -69.4535$ thousands of euros $=$ sales are estimated to decrease by €69453.5.

This variation does not depend on the price of soy milk: whether the price is 123 or 80 euro cents, the estimated decrease in sales would be the same: 69.4535 thousands of euros.

g. First, save the fitted values for this model.

Estimated sales for December 1990 total 103.4355 thousands of euros. The difference between this estimated value and the actual value is -19.0716 thousands of euros. The OLS residual is negative meaning that sales for December 1990 have been overestimated.

h. Point prediction.

$\hat{S}_{2012:7} = 57.5860 - 1.38907\, P_{2012:7} + 3.77879\, AE_{2012:7} - 0.0166627\, AE^2_{2012:7}$

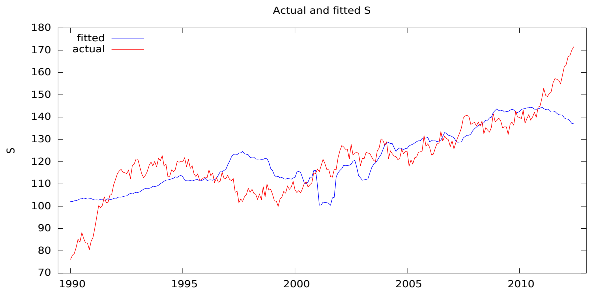
$\hat{S}_{2012:7} = 57.5860 - 1.38907 \times 123 + 3.77879 \times 146 - 0.0166627 \times 146^2 = 83.2516$ thousands of euros.

i. Covariance matrix of the OLS estimator.

Coefficient covariance matrix

const	P	AE	sq_AE	
1910.5	−1.7979	−25.831	0.094797	const
	0.0065197	0.014197	−4.1443e–005	P
		0.37218	−0.0013941	AE
			5.2652e–006	sq_AE

j. The fit is slightly better than the fit of the previous model, but the long-term behaviour of the sales is not yet properly reflected.



Part IV. Trend

a. Regression model:

$$S_t = \alpha_1 + \alpha_2\, P_t + \alpha_3\, AE_t + \alpha_4\, AE_t^2 + \alpha_5\, time + u_t \quad t = 1990:1, \ldots, 2012:6$$

Model 3: OLS, using observations 1990:01–2012:06 ($T = 270$)
Dependent variable: S

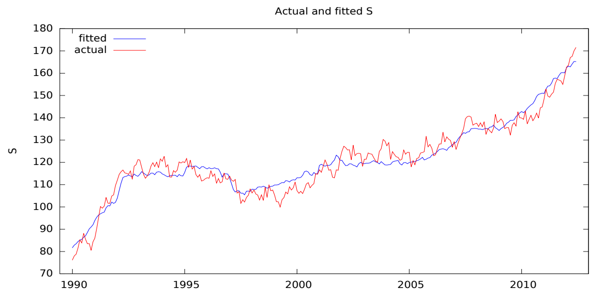
	Coefficient	Std. Error	t-ratio	p-value
const	−385.413	23.1683	−16.6353	0.0000
P	2.64647	0.124963	21.1780	0.0000
AE	2.17358	0.270360	8.0396	0.0000
sq_AE	−0.00633613	0.00104691	−6.0522	0.0000
time	0.495534	0.0147224	33.6585	0.0000

Mean dependent var	120.8475	S.D. dependent var	16.87912
Sum squared resid	5440.154	S.E. of regression	4.530881
R^2	0.929016	Adjusted R^2	0.927945
$F(4, 265)$	867.0608	P-value(F)	7.5e–151
Log-likelihood	−788.5374	Akaike criterion	1587.075
Schwarz criterion	1605.067	Hannan–Quinn	1594.300
$\hat{\rho}$	0.813395	Durbin–Watson	0.372147

SRF:

$$\hat{S}_t = -385.413 + 2.64647\, P_t + 2.17358\, AE_t - 0.00633613\, AE_t^2 + 0.495534\, time$$

b. The fit is much better than the one of the previous model. The long-term behaviour of the sales is properly reflected. However, some fluctuations in the series have yet to be explained.



c. The estimated annual variation rate is €495.534, holding price and advertising expenditures constant.

d. Yes, the graph of the adjusted series suggests that the trend variable provides information that is relevant for determining soy milk sales. A significance test should have to be performed to confirm this.