# Solution to Task T7.

Residual analysis.

Beach umbrella rental

To estimate model (1), click

Model --> Ordinary Least Squares ...

<u>F</u> ile <u>E</u> dit <u>T</u> ests	<u>S</u> ave <u>G</u> raphs <u>A</u> n	alysis <u>L</u> a	TeX				
Model 1: OLS, Dependent vari	using observa able: U	ations :	2013-05-0	5:2013-09-	29 (T	= 22)	
c	oefficient	std. e:	rror t-	ratio p-	value		
const	21.3194	51.43	74 0	.4145 0.	6837		
P	0.456259	3.36	725 0	.1355 0.	8938		
Т	12.0924	2.42	373 4	.989 0.	0001	***	
WW	-9.33935	9.81	757 -0	.9513 0.	3548		
time	-1.20704	1.21	759 -0	.9913 0.	3354		
Mean dependent	var 381.2	727 S	.D. depen	dent var	60.60	0110	
Sum squared re	sid 6920.	784 S	.E. of re	gression	20.17	7684	
R-squared	0.9102	262 A	djusted R	-squared	0.889	9148	
F(4, 17)	43.110	025 P	-value(F)		1.100	e-08	
Log-likelihood	-94.480	031 A	kaike cri	terion	198.9	9606	
Schwarz criter	ion 204.41	158 H	annan-Qui	nn	200.2	2457	
rho	0.129	518 D	urbin-Wat:	son	1.683	3612	

To save the residuals, go up to the menu bar of the estimation results window and click

## Save --> Residuals

The name given to the series of residuals by default is uhat1. You may change this name in the dialog box. Let's call them uhatumbrella. This new variable is listed in the main window of Gretl.

	🛐 greti	
	<u>File T</u> ools <u>D</u> ata <u>V</u> iew <u>A</u> dd <u>S</u> ample <u>V</u> ariable <u>M</u> odel <u>H</u> elp	
	umbrellas.gdt *	
	ID #  Variable name Descriptive label	4
	0 const	
	1 U Number of rented umbrellas	
	2 P Daily rental price (in euros)	
greti: variable attributes	3 T weekly average temperature (in degrees Celsius)	
	# 4 W =1 if windy week; = 2 if non windy week	
Name of variable: uhatumbrella	6 time time trend variable	
Description:	7 uhatumbrella residual from model 1	
residual from model 1		
	Weekly: Full range 2013-05-05 - 2013-09-29	
<u>Cancel</u>	🖩 🖉 🛅 fx 🔁 🙀 🖉 🛱 🗎 🗁 🗁	

To obtain the descriptive statistics of the residuals, highlight the variable *uhatumbrella*, right-click and select the option *Summary statistics* from the pulldown menu.

gretl: summary stats: uhatumbre	lla	
3 4 G Q 1X 🗞 G	*	
Summary statistics, usi	ng the observations 2013- mbrella' (22 valid observ	05-05 - 2013-09-29 ations)
		,
Mean	2.3254e-014	
Median	2.5884	
Minimum	-44.333	
Maximum	22.441	
Standard deviation	18.154	
c.v.	7.8067e+014	
Skewness	-1.0634	
Ex. kurtosis	0.51566	
5% percentile	-43.495	
95% percentile	22.040	
Jos perochorre		
Interquartile range	22.007	

To obtain the boxplot of the residuals, highlight the variable *uhatumbrella*, go up to the Menu Bar and click

```
Variable --> Boxplot
```

Then, mark Simple boxplot.



To estimate the density function of the residuals, highlight the variable *uhatumbrella*, go up to the Menu Bar and click

Variable --> Estimated density plot ...

In this example, there are not enough data to estimate the density function. Let's plot the frequency distribution clicking

Variable --> Frequency distribution ...



To graph the residuals against time, go up to the menu bar of the estimation results window and click

Graphs --> Residual plot --> Against time



Conclusions:

- The mean of the residuals is almost zero, because it is one of the properties of the Sample Regression Function.
- In general, to reach some conclusion about the distribution of the residuals we should perform some tests, for instance, normality tests.
- The time series plot of the residuals shows clusters of positive residuals followed by clusters of negative residuals. This result suggests that the error term might be autocorrelated. It would be necessary to test this hypothesis using the Durbin-Watson or Breusch-Godfrey tests.

## Registered vehicles

The estimation results of model (2) are shown in the window below.

💐 gretl: model 1					- 0 <b>- X</b>
<u>F</u> ile <u>E</u> dit <u>T</u> ests <u>S</u> ave <u>(</u>	<u>G</u> raphs <u>A</u> nalysis	<u>L</u> aTeX			8
Model 1: OLS, using Dependent variable:	observation RV	13 2004	1:01-2012:12	(T = 108)	
coeffi	cient std.	erroi	r t-ratio	p-value	
const 5826.2	2 274.	528	21.22	1.39e-039	***
TPTR 23.7	94157 /. 1944 11	1173	-1.160	0.2466	**
time -25.3	465 4.	61027	-5.498	2.76e-07	***
Mean dependent var	3893.463	S.D.	dependent va	ar 1255.1	15
Sum squared resid	62155981	S.E.	of regression	on 773.08	06
R-squared	0.631250	Adjus	sted R-square	ed 0.6206	13
F(3, 104)	59.34457	P-val	lue(F)	1.93e-2	22
Log-likelihood	-869.4488	Akai)	ce criterion	1746.8	98
Schwarz criterion	1757.626	Hanna	an-Quinn	1751.2	48
rho	0.215736	Durbi	in-Watson	1.5557	01
Excluding the const	ant, p-value	e was h	nighest for v	variable 3	(BOP)

To save the residuals, go up to the menu bar of the estimation results window and click

#### Save --> Residuals

Let's call the residuals of this model *uhat1vehicle*.

	🛐 gret	l.								
	Eile	<u>T</u> ools <u>D</u> ata	View	Add	Sample	Variable	Model	Help		
	vehicle	s.adt *	_	_		-	_			
	ID # 4	Variable nar	ne 🖣 De	escriptiv	/e label					
	0	const								
	1	RV	N	lumber	of register	red vehicle	s in the Ba	asque Count	iry	
	2	RVSP	N	lumber	of register	red vehicle	s in Spain			
	3	BOP	В	rent oil	price (in d	Iollars per b	oarrel)			
	4	IPI	In	dustria	l Producti	on Index in	Spain (ra	w data)		
	5	IPIR	a	nnual v	ariation ra	te of the Sp	oanish IPI			
	6	IPIBC	Ir	dustria	l Producti	on Index in	the Basq	ue Country	(seasonally a	djusted data)
	7	IPIBCR	a	nnual v	ariation ra	te of the Ba	asque IPI			
	8	crisis	=	1 if cris	is period	(obs>=200	8:10 && a	obs<=2012:1	2); 0 otherwi	se
	9	PIVE	=	1 if pla	n PIVE is i	n force (ob	s>=2008:	11 && obs<	=2010:06); 0	otherwise
	10	time	ti	me tren	d variable					
	11	dm1	=	1 if mo	onth = 1, 0	otherwise				
	12	dm2	=	1 if mo	onth = 2, 0	otherwise				
	13	dm3	=	1 if mo	onth = 3, 0	otherwise				
	14	dm4	=	1 if mo	onth = 4, 0	otherwise				
	15	dm5	=	1 if mo	onth = 5, 0	otherwise				
	16	dтб	=	1 if mo	onth = 6, 0	otherwise				
	17	dm7	=	1 if mo	onth = 7, 0	otherwise				
	18	dm8	=	1 if mo	onth = 8, 0	otherwise				
	19	dm9	=	1 if mo	onth = 9, 0	otherwise				
ittributes	20	dm10	=	1 if mo	onth = 10,	0 otherwise	e			
: uhat1vehicle	21	dm11	=	1 if mo	onth = 11,	0 otherwise	e			
	22	dm12	=	1 if mo	onth = 12,	0 otherwise	2			
	23	uhativehic	e re	esidual f	rom mod	el 1				
nodel 1					Monthly:	Full range	2004:01 -	2012:12		
<u>Cancel</u>		2 - 10	fx 📘	1	<u>⊮</u> β		3			

To obtain the descriptive statistics of the residuals, highlight the variable *uhat1vehicle*, go up to the Menu Bar and click

Variable --> Summary statistics

🙀 gretl: summary stats: uhat1vehicle		
🖬 占 🕞 🔍 📧 🗞 🗁 🗴	8	
Summary statistics, usin for the variable 'uhatly	g the observations ehicle' (108 valio	2004:01 - 2012:12 observations)
Mean	9.6844e-014	
Median	-101.12	
Minimum	-2409.6	
Maximum	1890.7	
Standard deviation	762.17	
c.v.	7.8700e+015	
Skewness	0.18836	
Ex. kurtosis	0.42147	
5% percentile	-1197.7	
95% percentile	1507.0	
Interguartile range	1009.7	
Missing obs.	0	

To obtain the boxplot of the residuals, highlight the variable *uhat1vehicle* and click Variable --> Boxplot in the Menu Bar. Then mark Simple boxplot.

To plot the estimated density function, highlight the variable *uhat1matriculacion*, go up to the Menu Bar and click

```
Variable --> Estimated density plot ...
```



To obtain the time series plot of the residuals, go up to the menu bar of the estimation results window and click

Graphs --> Residual plot --> Against time



Conclusions:

- The mean of the residuals is almost zero, because it is one of the properties of the Sample Regression Function.
- In general, to reach some conclusion about the distribution of the residuals we should perform some tests, for instance, normality tests.
- The time series plot of the residuals shows clusters of positive residuals followed by clusters of negative residuals. This result suggests that the error term might be autocorrelated. It would be necessary to test this hypothesis using the Durbin-Watson or Breusch-Godfrey tests.

Two characteristics of these residuals should be noted:

- A change of level in the residuals is observed from 2010 onwards. This change could be explained by some factors that have not been included in the model. If this were the case, the model (2) would be omitting relevant variables.
- The residuals show a cyclical behaviour of period one year, which is called seasonality. Since this cyclical pattern has not been included in the specification of the model, it appears in the residuals. Seasonal dummies should be included in the model and their statistical significance tested. If seasonality were a significant factor, the model (2) would be omitting a relevant variable.

# Wages

The results of estimating model (3) are shown in the figure below.

📓 gretl: model	1						X
<u>F</u> ile <u>E</u> dit <u>T</u> e	sts <u>S</u> ave <u>G</u> rap	ohs <u>A</u> nalysi	s <u>L</u> aTeX				
Model 1: 0 Dependent	LS, using o variable: W	bservatio	ns 1-5;	26			*
	coeffici	ent std	l. erro:	r t-ratio	p-value		
const	-1.56794	0.7	24551	-2.164	0.0309	**	
ED	0.57150	5 0.0	493373	11.58	9.09e-028	***	
EX	0.02539	59 0.0	115694	2.195	0.0286	**	
Т	0.14100	5 0.0	211617	6.663	6.83e-011	***	=
F	-1.81085	0.2	64825	-6.838	2.26e-011	***	
Mean depen	dent var	5.896103	s.D.	dependent v	var 3.6930	86	
Sum square	d resid	4557.308	S.E.	of regress:	ion 2.9575	72	
R-squared		0.363541	Adju	sted R-squa	red 0.3586	55	
F(4, 521)		74.39801	P-va	lue(F)	7.30e-	50	
Log-likeli	hood -	1314.228	Akail	ke criterio	n 2638.4	55	
Schwarz cr	iterion	2659.782	Hanna	an-Quinn	2646.8	05	-

To save the residuals, click Save --> Residuals in the estimation results window. Let's call these residuals *uhat1wage*.

	🛐 greti	
	<u>File T</u> ools <u>D</u> ata <u>V</u> iew <u>A</u> dd <u>S</u> ample <u>V</u> ariable <u>M</u> odel <u>H</u> elp	
	wages1.gdt *	
	ID #  Variable name Descriptive label	•
	0 const	
	1 W average hourly earnings (in dollars)	
	2 ED years of education	
	3 EX years potential experience	
	4 T years with current employer	
	5 NW =1 if nonwhite	
	6 F =1 if female	
	7 M =1 if married	
gretl: variable attributes	8 ND number of dependents	
Name of variable: ubat1wage	9 C =1 if live in SMSA	
Description	10 uhat1wage residual from model 1	
residual from model 1	Undated: Full range 1 - 526	
Cancel QK	🖩 🖻 🎦 🎟 f× 🔁 🔛 🖆 🔒 🔚 🗁	

To obtain the descriptive statistics, go up to the Menu Bar and click

Variable --> Summary statistics

gretl: summary stats: uhat1wage	
🖥 占 🖟 🔍 📧 😽 🕞 🕽	8
Summary statistics, usin	g the observations 1 - 526
for the variable 'uhat1w	age' (526 valid observations)
Mean	1.2605e-015
Median	-0.42285
Minimum	-7.7675
Maximum	14.008
Standard deviation	2.9463
c.v.	2.3374e+015
Skewness	1.5889
Ex. kurtosis	4.7488
5% percentile	-3.7089
95% percentile	5.2973
Interquartile range	2.8664
Missing obs.	0

To obtain the boxplot of the residuals, highlight the variable *uhat1wage* and click Variable --> Boxplot in the Menu Bar. Then mark Simple boxplot.

To plot the estimated density function, select the variable uhat1wate, go up to the Menu Bar and click

Variable --> Estimated density plot ...



To obtain the plot of the residuals by observation or against a regressor, click in the estimation results window

#### Graphs --> Residual plot

and select the option:





Conclusions:

- The mean of the residuals is almost zero, because it is one of the properties of the Sample Regression Function.
- In general, to reach some conclusion about the distribution of the residuals we should perform some tests, for instance, normality tests.
- The plot of the residuals by observation is not of much interest because cross-section data are not ordered.
- The residuals plot against experience and tenure show a quite homogeneous distribution.
- The residuals plots against education suggest that the higher the level of education, the higher the variability in the residuals. This fact suggests that the error term might be heteroskedastic and this heteroskedasticity may be caused by the variable education. The homoskedasticity hypothesis should be tested using the White, Breusch-Pagan or Goldfed-Quandt tests.

Holiday cottages in Bilbao

The results of estimating model (4) appear in the figure below.

📓 gretl: model 1						x
<u>F</u> ile <u>E</u> dit <u>T</u> ests	<u>S</u> ave <u>G</u> raphs	<u>A</u> nalysis <u>L</u> aTeX				
Model 1: OLS, Dependent var	, using obse riable: RP	rvations 1-7	5			
	coefficient	std. erro	r t-ratio	p-value		
const	38.4321	7.22899	5.316	1.14e-06	***	
NR	2.26766	1.20082	1.888	0.0630	*	
BP	1.49558	1.09746	1.363	0.1772		
Mean depender	nt var 56.	13893 S.D.	dependent v	ar 14.98	446	
Sum squared :	resid 152	63.15 S.E.	of regressi	on 14.55	982	
R-squared	0.0	81392 Adju	sted R-squar	ed 0.055	875	
F(2, 72)	3.1	89724 P-va	lue(F)	0.047	064	
Log-likelihoo	od -305	.7595 Akai	ke criterion	617.5	189	
Schwarz crite	erion 624	.4714 Hann	an-Quinn	620.2	950	

To save the residuals, click Save --> Residuals in the estimation results window and then call these residuals uhat1cottage.

	🙀 greti	
	<u>File T</u> ools <u>D</u> ata <u>V</u> iew <u>A</u> dd <u>S</u> ample <u>V</u> ariable <u>M</u> odel <u>H</u> elp	
	cottages.gdt *	
	ID # 4 Variable name 4 Descriptive label	4
	0 const	
	1 RP average room price (in euros)	
	2 NR number of rooms	
	3 BP breakfast price (in euros)	
	5 LOC location	
	6 NP proximity to a natural park	
	T LK proximity to a lake or a reservoir	
	BE proximity to a beach	
gretl: variable attributes	B 9 Q Tourism Quality Q certificate	
Name of variable: uhat1cottage	⊞ 10 C Commitment to quality tourism award	
Description	30 uhat1cottage residual from model 1	
Description:		
residual from model 1	Undated: Full range 1 - 75	
<u>Cancel</u>	🖩 🖉 🎦 🗰 fx 🔁 🙀 🖉 🗳 🔒 🚍 🗖	

To obtain the descriptive statistics, go up to the Menu Bar and click

```
Variable --> Summary statistics
```

gretl: summary stats: uhat1cotta	ge 📃 🗖 🗖
🖥 占 🕞 🔍 🖿 🍫 🕒	×
Summary statistics, usi	ng the observations 1 - 75
for the variable 'uhat1	cottage' (75 valid observations)
Mean	9 2844e-015
Median	-3.2329
Minimum	-18.818
Maximum	70.495
Standard deviation	14.362
c.v.	1.5469e+015
Skewness	2.3610
Ex. kurtosis	7.7028
5% percentile	-17.437
95% percentile	29.135
	11.504
Interquartile range	

To obtain the boxplot of the residuals, highlight the variable *uhat1cottage* and click Variable --> Boxplot in the Menu Bar. Then mark *Simple boxplot*.

To plot the estimated density function, select the variable *uhat1cottage* using the cursor and click Variable --> Estimated density plot ... in the Menu Bar.



To plot the residuals by observation or against a regressor, click

### Graphs --> Residual plot

in the menu bar of estimation results window and select the option:



Conclusions:

- The mean of the residuals is almost zero, because it is one of the properties of the Sample Regression Function.
- In general, to reach some conclusion about the distribution of the residuals we should perform some tests, for instance, normality tests.
- The plot of the residuals by observation is not of much interest because cross-section data are not ordered.
- The distribution of the residuals is quite homogeneous in the residuals plot against the regressor number of rooms.
- The residual plot against price of breakfast shows that the higher the price, the higher the variability in the residuals. This fact suggests that the variance of the error term might depend on the price of breakfast. The homoskedasticity hypothesis should be tested using the White, Breusch-Pagan or Goldfed-Quandt tests.