

**Econometrics:
estimating a model
using Statistica**

© Artur Marques, 2003

Table of Contents

1 – Introduction	3
2 – Estimating the model – Trimming	4
3 – Estimating the model – Residuals.....	11
4 – Final Model	17
5 – References	18

1 – Introduction

This document describes how I used Statistica, from <http://www.statsoft.com>, to estimate a IS based Model, as requested in the terms of one assignment, using raw data provided as an excel spreadsheet. Check the final references for URLs that link to the assignment, to the spreadsheet and to the Statistica workbook that built the solution.

The model was built step-by-step and every step was recorded as a component of the Statistica workbook.

The final Statistica file holds 75 tables, 7 graphs and 3 raw data spreadsheets. You don't have to have *any* previous Statistica experience to be able to read the workbook; however, if you download and open it, you'll be able to understand in greater detail the steps taken towards the solution.

The workbook is neatly organized as a tree with branches corresponding to requests/suggestions of the original assignment. This document is a "digest" of the workbook, illustrated with selected snapshots and explaining the why and the how of the relevant steps.

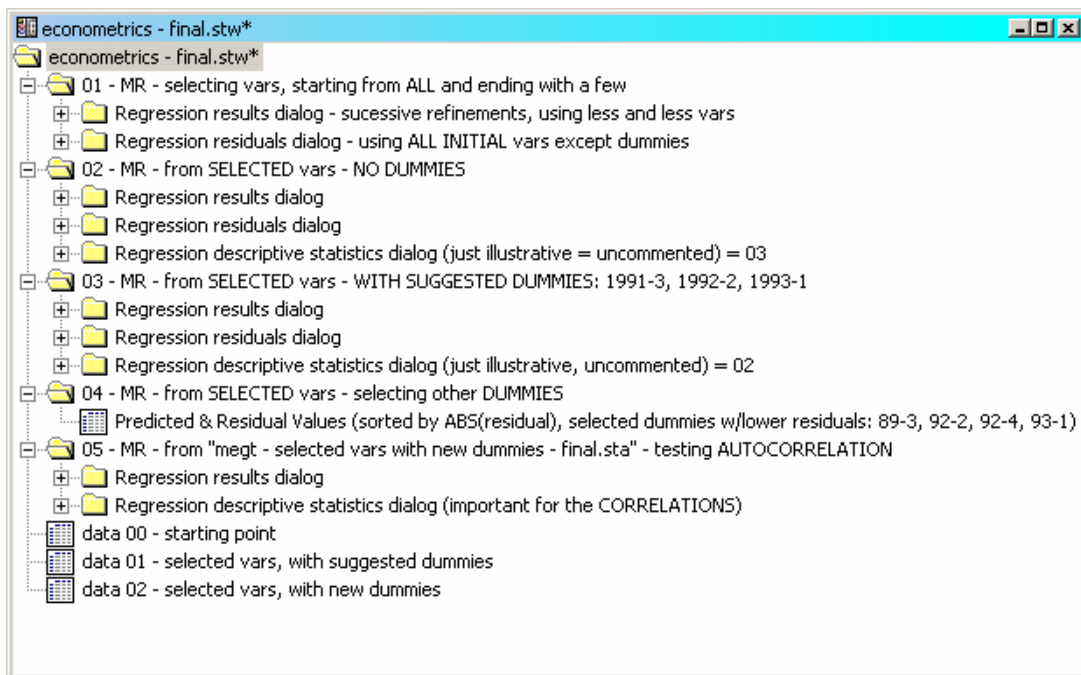


Figure 1 – "Statistica workbook, organized as a tree"

2 – Estimating the model – Trimming

The model to estimate is a model of aggregate output, involving a dependent variable and a number of independent others, along the lines

$$\Delta LY_t = \alpha_0 + \sum_{i=1}^{k_1} \alpha_i \Delta LY_{t-i} + \sum_{i=0}^{k_2} \beta_i \Delta LG_{t-i} + \sum_{i=0}^{k_3} \chi_i \Delta RR10Y_{t-i} + \sum_{i=0}^{k_4} \delta_i \Delta LEER_{t-i} + \sum_{i=0}^{k_5} \phi_i \Delta LY^*_{t-i} + \sum_{i=0}^{k_6} \varphi_i \Delta LPIM_{t-i} + \varepsilon_t$$

where:

- “delta” is a difference operator, meaning that the variables it affects are expressed in rates of changes. In Statistica, my “delta” variables have names starting with a capital D;
- a capital L denotes natural logs;
- Y represents the GDP (Gross Domestic Product);
- PIM represents the import prices;
- EER represents the effective exchange rate;
- G represents government consumption;
- Y_STAR represents Y*, the foreign output;
- RR10Y represents the 10-year real interest rate.

All these concepts were translated to Statistica variables; check the next snapshot. Notice that there are three variables (D91q3, D92q2 and D93q1) commented as “boolean”, in the sense that they just signal their corresponding year-quarter case, being false or zero for all the remaining cases. These variables will be useful at a later stage as “dummies” for the multiple regression, but they were ignored in the first regressions and are only shown right now, because they were already present as columns on the spreadsheet first used with Statistica.

G, Y_STAR and PIM are considered to have a positive (+) effect on GDP, while RR10Y and EER have a negative effect (-).

	Name	Type	MD code	Length	Long Name (label or formula)
1	DLY	Double	-9999		difference operator (D) of LOG (L) for GDP
2	DLPIM	Double	-9999		DL for import prices (PIM)
3	DLEER	Double	-9999		DL for Effective Exchange Rate
4	DLG	Double	-9999		DL for Government Consumption
5	DLY_STAR	Double	-9999		DL for Foreign Output
6	DRR10Y	Double	-9999		DL for Real 10-year Interest Rate
7	D91q3	Double	-9999		boolean for 1991, 3rd quarter case
8	D92q2	Double	-9999		boolean for 1992, 2nd quarter case
9	D93q1	Double	-9999		boolean for 1993, 1st quarter case

Figure 2 – “Initial variables and their short descriptions”

	1 DLY	2 DLPIM	3 DLEER	4 DLG	5 DLY STAR	6 DRR10Y	7 D91q3	8 D92q2	9 D93q1
1985-2	0.004721	-0.00668	0.002682	0.025628	0.00983529269	0.000196	0	0	0
1985-3	0.006815	-0.02207	0.022455	0.025874	0.0116568416	0.007819	0	0	0
1985-4	0.011849	-0.01081	0.032034	0.01834	0.00878089872	-0.00821	0	0	0
1986-1	-0.00281	-0.02121	0.051556	0.009134	0.0080916115	-0.00122	0	0	0
1986-2	0.012743	-0.02107	-0.00496	0.026379	0.00624864176	-0.00975	0	0	0
1986-3	0.003189	-0.01337	0.032174	0.01511	0.00879400746	-0.00066	0	0	0
1986-4	0.012593	0.02201	0.042903	0.012898	0.0063267354	-0.00169	0	0	0
1987-1	-0.00563	0.003568	0.034231	0.001025	0.00838451925	-0.00312	0	0	0
1987-2	0.011728	0.000928	-0.01817	0.023451	0.00951441647	0.003054	0	0	0
1987-3	0.00685	0.006509	-0.00504	0.015006	0.0116826543	0.009483	0	0	0
1987-4	0.01973	-0.00613	0.017695	0.024194	0.0182553838	-0.00044	0	0	0
1988-1	0.009727	-0.00212	-0.00443	0.014358	0.00892256723	-0.00518	0	0	0
1988-2	0.001559	0.012115	-0.02143	0.007211	0.0112555143	-0.00292	0	0	0
1988-3	0.008542	0.018714	-0.03495	0.015261	0.00987482371	0.001589	0	0	0
1988-4	0.018418	0.001861	0.0116	0.014753	0.0113611986	-0.00179	0	0	0
1989-1	0.010728	0.010814	-0.01751	0.015994	0.0120744701	0.002184	0	0	0
1989-2	0.005127	0.023774	0.002196	0.01508	0.00384145942	0.001544	0	0	0
1989-3	-0.00028	0.006593	0.011959	0.014177	0.00883287459	0.005449	0	0	0
1989-4	0.020821	0.009987	0.045666	0.032955	0.0066696586	0.001958	0	0	0
1990-1	0.015621	0.0037	0.04604	0.031994	0.00887154622	0.006201	0	0	0
1990-2	-0.00106	0.007264	0.011989	0.020961	0.00872264066	0.002899	0	0	0
1990-3	0.004504	-0.00599	-0.00167	0.02478	0.000256924434	0.000315	0	0	0
1990-4	0.009711	0.004075	0.012263	0.018873	-0.00565464835	-0.00249	0	0	0
1991-1	0.00796	0.004513	-0.00524	0.027507	-0.00131968127	-0.0013	0	0	0
1991-2	0.00197	0.004186	-0.0583	0.028736	0.00520794611	-0.00452	0	0	0
1991-3	0.000722	0.003161	-0.00159	0.022239	0.00128654356	-0.00213	1	0	0
1991-4	0.009056	-0.00124	0.029269	0.017944	0.00479631719	0.001868	0	0	0
1992-1	0.014495	-0.00163	0.006009	0.022601	0.00784680551	-0.006	0	0	0
1992-2	-0.00709	0.002429	0.003811	0.009569	0.00489971031	0.001716	0	1	0
1992-3	-0.00181	-0.00385	0.043763	0.018334	0.00552415924	0.009003	0	0	0
1992-4	-0.00239	0.009498	-0.01412	0.017738	0.00950270629	-0.00612	0	0	0
1993-1	-0.00731	0.017579	-0.02953	0.014081	0.0025673074	-0.01186	0	0	1
1993-2	0.000164	-0.006	-0.0209	0.021513	0.00373244368	-0.0008	0	0	0
1993-3	0.004551	0.003699	-0.04659	0.010454	0.00376582837	-0.00779	0	0	0
1993-4	0.003824	0.003721	-0.0019	0.003916	0.0109755023	-0.00504	0	0	0
1994-1	0.008796	-0.00719	-0.01549	0.003528	0.00886543331	-0.00401	0	0	0
1994-2	0.004557	0.004838	0.018056	0.008033	0.00984264494	0.011994	0	0	0
1994-3	0.007238	0.001825	0.026417	0.01266	0.00723135519	0.009979	0	0	0
1994-4	0.008323	0.003215	0.00062	0.011858	0.00945727397	0.005436	0	0	0
1995-1	0.00496	0.004439	0.015772	0.011754	0.00239022228	-0.00322	0	0	0
1995-2	0.006722	0.000829	0.010646	0.010713	0.00322856034	-0.00306	0	0	0
1995-3	0.001499	0.009046	0.012196	0.010534	0.00773893306	-0.00014	0	0	0
1995-4	0.001828	0.006603	0.015933	0.006405	0.00884621055	-0.00392	0	0	0
1996-1	0.002207	0.002259	-0.00438	0.009266	0.00799808296	-0.00954	0	0	0
1996-2	0.006682	-0.00113	-0.01546	0.005924	0.0130913416	-0.00003	0	0	0
1996-3	0.005271	-0.00614	0.008442	0.005884	0.0030163747	0.003361	0	0	0
1996-4	0.002029	-0.00481	-0.01636	0.004183	0.0117295003	-0.01008	0	0	0
1997-1	0.003401	-0.00006	-0.04191	-0.00062	0.0149549269	-0.00603	0	0	0
1997-2	0.010961	-0.00558	-0.03136	0.008647	0.00270443885	0.003357	0	0	0
1997-3	0.006246	-0.0062	-0.04145	-0.00418	0.00870049937	-0.00366	0	0	0
1997-4	0.010312	-0.00042	0.025131	-0.00465	0.00694211447	-0.00098	0	0	0
1998-1	0.008811	-0.0106	-0.02238	0.01482	0.00934611895	-0.00514	0	0	0
1998-2	0.004095	-0.00147	0.02209	0.009949	0.00470469237	-0.00461	0	0	0
1998-3	0.005151	-0.00368	0.018558	0.007957	0.00468686225	0.000305	0	0	0
1998-4	0.001269	-0.01957	0.009662	0.002761	0.0113921668	-0.00301	0	0	0
1999-1	0.009003	-0.00656	-0.035	0.010901	0.00620012245	-0.00372	0	0	0
1999-2	0.006236	0.002804	-0.03929	0.003374	0.00701055351	-0.00262	0	0	0
1999-3	0.010822	-0.00118	-0.01681	0.008711	0.00863250288	0.013186	0	0	0
1999-4	0.009763	-0.00123	-0.02897	0.00845	0.0104963812	0.001296	0	0	0
2000-1	0.0089	0.0051	-0.03389	0.008286	0.00992896569	-0.00112	0	0	0
2000-2	0.008763	0.000801	-0.03316	0.007891	0.0105168971	-0.00121	0	0	0
2000-3	0.003954	0.006482	-0.01474	0.007821	0.00166156987	0.000493	0	0	0
2000-4	0.005869	0.005207	-0.02687	0.007782	0.00541950422	-0.00066	0	0	0
2001-1	0.005511	0.003606	0.063535	0.006976	0.00529380966	-0.00345	0	0	0
2001-2	0.000685	-0.00052	-0.03315	0.002996	-0.00185174301	-0.00468	0	0	0
2001-3	0.001434	-0.00951	0.00988	0.001998	-0.00301646909	0.010006	0	0	0
2001-4	-0.00095	-0.00692	0.003366	0.003992	-0.000687658653	-0.00155	0	0	0
2002-1	0.001798	0.000712	0.003671	0.003992	0.00176936845	-0.00404	0	0	0

Figure 3 – “the raw data, imported into Statistica as the spreadsheet “data 00 - starting point”

Once the data is imported, it is possible to start estimating the model, by multiple linear regressions (MR), using DLY as the dependent variable and DLPIM, DLEER, DLG, DLY_STAR and DRR10Y as the independent ones.

At this point, the assignment suggests to “start with lags 0 to 4 for the regressors and DLY”, so I inserted 4 new variables for each regressor and 4 new variables for DLY, representing the suggested forward lags. Each forward lag was named like +1, +2, +3 and +4, denoting the number of trimesters it lags forward, relatively to its corresponding original time series.

So, the original spreadsheet grew to include the variables DLY, DLY(+1), DLY(+2), DLY(+3), DLY(+4), DLPIM, DLPIM(+1), DLPIM(+2)... and so on.

The current equation should be:

$$DLY = coef1 + coef_DLY * DLY + coef_DLY1 * DLY(+1) + coef_DLY2 * DLY(+2) + coef_DLY3 * DLY(+3) + coef_DLY4 * DLY(+4) + \dots + coef_DRR10Y * DRR10Y + coef_DRR10Y1 * DRR10Y(+1) + coef_DRR10Y2 * DRR10Y(+2) + coef_DRR10Y3 * DRR10Y(+3) + coef_DRR10Y4 * DRR10Y(+4)$$

Doing a multiple regression now, would compute the outputs “1.1” of the workbook, which are a “summary statistics” and a “regression summary for DLY”. But not all of the current variables are worth keeping for the estimation of the model and, following the assignment’s suggestions, I would trim them.

The trim process involved 26 steps, labeled 1.1 to 1.26 in the workbook, each having two outputs: the summary statistics and the regression summary.

I paid attention to Statistica’s “regression summary”, where

- “**B**” represents the regression coefficient; that is, the number that multiplies the regressor in order to better estimate the dependent variable (DLY). B is the weight of the independent variable for the estimation.
- “**Std. Err. Of B**” represents the standard error of the regression coefficient. The lower this number, the better, representing a lower statistical noise.
- “**t(degrees-of-freedom)**” is the t-statistics, a test to the null hypothesis. See p-level, below.
- **p-level** is such that the lower its score, the stronger the evidence towards rejecting the null hypothesis in favor of the alternative, meaning the stronger the acceptance of the computed B. Statistica automatically highlights computations with a low p-level. The user can set the value below which such highlighting happens; the default value is 5%. So, the higher the p-level, the less one can believe that the observed relation between variables in the sample is a reliable indicator of the relation between the respective variables in the population. The p-level represents the probability of error that is involved in accepting the observed result as valid, that is, as “representative of the population.”

In Statistica’s “summary statistics” I paid attention to **R²** – the proportion of variance in the dependent variable that can be explained by the independent variable(s), so that when **R²=1** there is a perfect regression.

Step 1.1, the statistics and the regression for immediately after inserting the lags are documented in the following pictures.

Summary Statistics; DV: DLY	
Statistic	Value
Multiple R	0.861518
Multiple R ²	0.742214
Adjusted R ²	0.522337
F(29,34)	3.375593
p	0.000419
Std.Err. of Estimate	0.003982

Figure 4 – “Summary statistics for step 1.1, just after inserting the lags”

Regression Summary for Dependent Variable: DLY (megt_econ4 - trying lag - forward.sta)
R= .86151825 R²= .74221369 Adjusted R²= .52233713
F(29,34)=3.3756 p<.00042 Std.Error of estimate: .00398

N=64	Beta	Std.Err. of Beta	B	Std.Err. of B	t(34)	p-level
Intercept			-0.004253	0.002316	-1.83668	0.075012
DLY (+4)	0.343981	0.135979	0.343866	0.135934	2.52966	0.016220
DLY (+3)	-0.112110	0.165578	-0.111359	0.164469	-0.67708	0.502936
DLY (+2)	0.057923	0.166352	0.057267	0.164470	0.34819	0.729841
DLY (+1)	0.141115	0.164243	0.139207	0.162023	0.85918	0.396258
DLPIM (+4)	0.001886	0.118547	0.001184	0.074427	0.01591	0.987398
DLPIM (+3)	0.091385	0.135805	0.057661	0.085688	0.67292	0.505549
DLPIM (+2)	0.036679	0.128410	0.024120	0.084444	0.28564	0.776887
DLPIM (+1)	-0.209128	0.134594	-0.138623	0.089218	-1.55376	0.129501
DLPIM	-0.159458	0.127923	-0.111729	0.089632	-1.24652	0.221093
DLEER (+4)	0.287483	0.128565	0.061346	0.027435	2.23609	0.032017
DLEER (+3)	-0.068693	0.132857	-0.014490	0.028024	-0.51704	0.608472
DLEER (+2)	-0.366097	0.131400	-0.077572	0.027842	-2.78613	0.008661
DLEER (+1)	-0.181109	0.123238	-0.038823	0.026417	-1.46959	0.150865
DLEER	0.110089	0.111062	0.024357	0.024573	0.99124	0.328571
DLG (+4)	-0.153818	0.198843	-0.105265	0.136078	-0.77357	0.444533
DLG (+3)	0.208045	0.252325	0.143328	0.173834	0.82451	0.415396
DLG (+2)	-0.100459	0.228632	-0.069699	0.158627	-0.43939	0.663158
DLG (+1)	-0.274830	0.219317	-0.189942	0.151575	-1.25312	0.218712
DLG	0.635576	0.174564	0.436387	0.119856	3.64093	0.000894
DLY_STAR (+4)	0.117496	0.120657	0.170808	0.175403	0.97380	0.337033
DLY_STAR (+3)	0.121680	0.121337	0.170347	0.169867	1.00282	0.323030
DLY_STAR (+2)	0.030250	0.119496	0.040817	0.161240	0.25314	0.801681
DLY_STAR (+1)	0.174259	0.127386	0.229768	0.167965	1.36796	0.180300
DLY_STAR	0.072012	0.121318	0.094037	0.158423	0.59358	0.556723
DRR10Y (+4)	-0.203302	0.122277	-0.224857	0.135241	-1.66263	0.105579
DRR10Y (+3)	0.026649	0.131193	0.029353	0.144507	0.20313	0.840246
DRR10Y (+2)	-0.040110	0.122006	-0.043645	0.132758	-0.32876	0.744357
DRR10Y (+1)	-0.042411	0.122368	-0.046886	0.135279	-0.34659	0.731037
DRR10Y	0.134286	0.130353	0.147989	0.143655	1.03017	0.310202

Figure 5 – “Regression summary for step 1.1”

Notice how Statistica highlighted the cases where p-level was $\leq 5\%$ in order to show them as “significant”.

Less noticeable is the fact that many of the computed Bs have the wrong sign: negative instead of positive; or positive, instead of negative. Those coefficients will be affecting variables so that they will have an opposite to the expected contribution to the dependent DLY.

In order to trim to model, I focused precisely in (1) detecting the wrong signs and (2) eliminating the cases with the higher p-levels. The trim process took 26 steps. For each step, the corresponding statistical and regression summaries were kept and formatted in order to highlight exactly what the trimming was.

My "algorithm" works on Statistica's regression summary sheets, built from data sheets, via multiple regression.

//instruction 1 was responsible for steps [1.1 to 1.15]

```
1) while (there are independent variables with an opposite sign for B) do
begin
    1.1) format with a yellow background the vars with a B of opposite sign;
    1.2) format in bold-italic the row with the higher absolute B coefficient;
    1.3) eliminate the variable with the bold-italic row from the data
spreadsheet;
    1.4) repeat the multiple regression process for the new data spreadsheet;
end
```

//instruction 2 was responsible for steps [1.16 to 1.26]

```
2) while (there are independent variables with p-levels >5%)
begin
    2.1) format with a yellow background the var with the highest p-level;
    2.2) format in bold-italic the row with the var with the highest p-level;
    2.3) eliminate the variable with the bold-italic row from the spreadsheet;
    2.4) repeat the multiple regression process for the new spreadsheet;
end
```

To make work easier, there were some sheets where I highlighted with a green background the variables with the correct sign.

On step 1.26 I decided to stop the trimming process, although there was still a variable – DRR10Y(+4) – with a p-level of 7.7%, that the algorithm would eliminate. That variable DRR10Y(+4) was kept because I think that the real interest rate is too determinant for investment, hence for DLY, for not having it represented, as it would have been the case, should I have decided to strictly follow the algorithm.

The decision was even more difficult, because I had already lost all the DLPIM variables. Does this mean that I find import prices of not such a great importance for GDP? No, the reason for letting DLPIMs go, was their consistent wrong signs:

DLPIM is selected for elimination in step 1.4, due to its wrong sign;
DLPIM(+1) is selected for elimination in step 1.6, due to its wrong sign;
DLPIM(+3) is selected for elimination in step 1.12, due to its wrong sign;
DLPIM(+4) is selected for elimination in step 1.13, due to its wrong sign.

Regression Summary for Dependent Variable: DLY (megt_econ4 -
R= .75667264 R²= .57255348 Adjusted R²= .40157487
F(18,45)=3.3487 p<.00051 Std.Error of estimate: .00446

N=64	Beta	Std.Err. of Beta	B	Std.Err. of B	t(45)	p-level
Intercept			-0.005263	0.002004	-2.62618	0.011760
DLY (+4)	0.244034	0.113101	0.243953	0.113064	2.15766	0.036330
DLY (+3)	-0.048901	0.113622	-0.048573	0.112861	-0.43038	0.668971
DLY (+2)	-0.076832	0.110910	-0.075962	0.109655	-0.69274	0.492032
DLY (+1)	0.061017	0.121013	0.060193	0.119377	0.50422	0.616566
DLPIM (+4)	-0.052085	0.112124	-0.032701	0.070395	-0.46454	0.644504
DLPIM (+2)	-0.015965	0.116521	-0.010499	0.076625	-0.13702	0.891629
DLEER (+3)	-0.078543	0.121987	-0.016568	0.025731	-0.64386	0.522932
DLEER (+2)	-0.344478	0.116924	-0.072991	0.024775	-2.94617	0.005081
DLEER (+1)	-0.114087	0.111111	-0.024456	0.023818	-1.02679	0.310008
DLEER	0.092756	0.107468	0.020522	0.023777	0.86311	0.392655
DLG	0.510260	0.128530	0.350345	0.088249	3.96997	0.000256
DLY_STAR (+4)	0.105484	0.114104	0.153345	0.165877	0.92445	0.360181
DLY_STAR (+3)	0.201528	0.122511	0.282132	0.171511	1.64498	0.106943
DLY_STAR (+2)	0.023304	0.124141	0.031445	0.167507	0.18772	0.851939
DLY_STAR (+1)	0.157478	0.130817	0.207642	0.172487	1.20381	0.234958
DLY_STAR	0.071233	0.119676	0.093020	0.156279	0.59522	0.554680
DRR10Y (+4)	-0.205789	0.126295	-0.227607	0.139685	-1.62943	0.110203
DRR10Y (+2)	-0.102577	0.116621	-0.111617	0.126899	-0.87958	0.383762

Figure 6 – “DLPIM(+4), the last of the DLPIMs is eliminated on step 1.13; “wrong” sign variables are highlighted with a yellow background. DLPIM(+4) is also on bold because of its higher B”

Regression Summary for Dependent Variable: DLY (megt_econ4 -
R= .75058564 R²= .56337881 Adjusted R²= .42693469
F(15,48)=4.1290 p<.00009 Std.Error of estimate: .00436

N=64	Beta	Std.Err. of Beta	B	Std.Err. of B	t(48)	p-level
Intercept			-0.005481	0.001918	-2.85850	0.006279
DLY (+4)	0.269289	0.107033	0.269199	0.106997	2.51595	0.015270
DLY (+3)	-0.025708	0.107901	-0.025536	0.107178	-0.23826	0.812696
DLY (+2)	-0.082265	0.104063	-0.081334	0.102886	-0.79053	0.433107
DLY (+1)	0.081277	0.115531	0.080178	0.113970	0.70350	0.485142
DLEER (+3)	-0.072097	0.115571	-0.015208	0.024378	-0.62383	0.535693
DLEER (+2)	-0.317321	0.109463	-0.067237	0.023194	-2.89889	0.005631
DLEER (+1)	-0.104296	0.107697	-0.022357	0.023086	-0.96843	0.337686
DLG	0.499099	0.115568	0.342681	0.079349	4.31867	0.000078
DLY_STAR (+4)	0.103254	0.108913	0.150104	0.158331	0.94804	0.347856
DLY_STAR (+3)	0.202514	0.118793	0.283512	0.166306	1.70476	0.094705
DLY_STAR (+2)	0.013170	0.121001	0.017771	0.163270	0.10884	0.913780
DLY_STAR (+1)	0.154789	0.126255	0.204097	0.166473	1.22601	0.226177
DLY_STAR	0.067898	0.116743	0.088664	0.152450	0.58160	0.563559
DRR10Y (+4)	-0.229080	0.121065	-0.253367	0.133900	-1.89221	0.064503
DRR10Y (+2)	-0.124934	0.110456	-0.135944	0.120191	-1.13107	0.263646

Figure 7 – “On step 1.16, there were NO more “wrong” sign independent variables. The algorithm went on trimming, based on the higher p-levels”

Regression Summary for Dependent Variable: DLY (megt_econ4 -
R= .71790895 R²= .51539327 Adjusted R²= .46438203
F(6,57)=10.104 p<.00000 Std.Error of estimate: .00422

N=64	Beta	Std.Err. of Beta	B	Std.Err. of B	t(57)	p-level
Intercept			-0.003826	0.001541	-2.48316	0.015988
DLY (+4)	0.258301	0.099886	0.258215	0.099853	2.58595	0.012291
DLEER (+2)	-0.410362	0.095092	-0.086951	0.020149	-4.31544	0.000064
DLG	0.432660	0.101151	0.297064	0.069450	4.27736	0.000073
DLY_STAR (+3)	0.217469	0.103002	0.304449	0.144199	2.11131	0.039143
DLY_STAR (+1)	0.225511	0.103366	0.297346	0.136293	2.18167	0.033272
DRR10Y (+4)	-0.182990	0.101728	-0.202390	0.112513	-1.79882	0.077342

Figure 8 – “On step 1.25, there was still one var with a p-level >5%, but I kept her”

Statistic	Summary Value
Multiple R	0.71791
Multiple R ²	0.51539
Adjusted R ²	0.46438
F(6,57)	10.10352
p	0.00000
Std.Err. of Estimate	0.00422

Figure 9 – “Summary statistics for step 1.25”

Step 1.26 is actually 1.25 replicated, but without the color formatting that the algorithm was doing. In other words, after the trimming process, DLY was dependent on DLY(+4), DLEER(+2), DLG, DLY_STAR(+3), DLY_STAR(+1) and DRR10Y(+4).

All the trimming process is illustrated in branch “**01 - MR - selecting vars, starting from ALL and ending with a few**” of the Statistica workbook.

3 – Estimating the model – Residuals

The assignment suggests to look at the residuals and try to assess the need for dummies, hinting the cases 1991-3, 1992-2 and 1993-1.

This assessment corresponds in the workbook to branches 02, 03 and 04:

- 02 - MR - from selected vars – without dummies;
- 03 - MR - from selected vars – with suggested dummies: 91-3, 92-2, 93-1;
- 04 - MR - from selected vars – selecting other dummies.

Figure 11 – “Predicted and residual values, after trimming” illustrates the predicted and the observed values for each case, for each of the variables. The difference between the observed and the predicted values is a “residual”. The greater the residual, the less accurate was the estimation. The table was ordered by descending order of absolute residual, so that the greater abs(residual) would come first. The abs(residual) is an extra column that Statistica does not output by default, but that is easy to add using the built-in functions.

After sorting, case 1993-1 show up as king of the absolute residual, while 1992-2 is well placed on the same ranking; however, 1991-3 is modestly placed. By using dummies one will be ignoring the residuals.

Before dummies, the summary statistics is, of course, the same shown on Figure 9 – “Summary statistics for step 1.25”, since the multiple regression – based on the spreadsheet “data 01 - selected vars, with suggested dummies” is not yet using the dummies as regressors, as the following picture illustrates.

Regression Summary for Dependent Variable: DLY (megt - selecte						
R= .71790895 R ² = .51539327 Adjusted R ² = .46438203						
F(6,57)=10.104 p<.00000 Std.Error of estimate: .00422						
N=64	Beta	Std.Err. of Beta	B	Std.Err. of B	t(57)	p-level
Intercept			-0.003826	0.001541	-2.48316	0.015988
DLY (+4)	0.258301	0.099886	0.258215	0.099853	2.58595	0.012291
DLEER (+2)	-0.410362	0.095092	-0.086951	0.020149	-4.31544	0.000064
DLG	0.432660	0.101151	0.297064	0.069450	4.27736	0.000073
DLY_STAR (+3)	0.217469	0.103002	0.304449	0.144199	2.11131	0.039143
DLY_STAR (+1)	0.225511	0.103366	0.297346	0.136293	2.18167	0.033272
DRR10Y (+4)	-0.182990	0.101728	-0.202390	0.112513	-1.79882	0.077342

Figure 10 – “Regression before any dummies”

Predicted & Residual Values (meqt - selected vars - final.sta)										
Dependent variable: DLY										
Case No.	Observed Value	Predicted Value	Residual	Standard Pred. v	Standard Residual	Std.Err. Pred.Val	Mahalanobis Distance	Deleted Residual	Cook's Distance	ABS(residual) =Abs(v3)
1993-1	-0.007313	0.005827	-0.013139	0.00760	-3.11638	0.001540	7.41945	-0.015162	0.246431	0.01313938
Minimum	-0.007313	-0.006736	-0.013139	-3.02998	-3.11638	0.000696	0.73324	-0.015162	0.000000	0.01313938
1992-4	-0.002393	0.007104	-0.009497	0.31636	-2.25244	0.000696	0.73324	-0.009763	0.020884	0.00949681
1989-3	-0.000282	0.008393	-0.008675	0.62815	-2.05757	0.001000	2.55723	-0.009192	0.038171	0.008675214
2001-3	0.001434	-0.006736	0.008171	-3.02998	1.93790	0.002025	13.54375	0.010620	0.208995	0.008170636
Maximum	0.020821	0.017192	0.008171	2.75571	1.93790	0.002073	14.24175	0.010620	0.246431	0.008170636
1992-2	-0.007094	0.000620	-0.007714	-1.25122	-1.82962	0.001218	4.27076	-0.008416	0.047481	0.007714101
1992-3	-0.001812	0.004633	-0.006446	-0.26096	-1.52876	0.000984	2.44438	-0.006817	0.020323	0.006445588
1992-1	0.014495	0.008357	0.006139	0.61936	1.45594	0.000931	2.08450	0.006453	0.016301	0.006138575
1990-2	-0.001063	0.004768	-0.005832	-0.24825	-1.38320	0.001193	4.06138	-0.006340	0.025868	0.005831896
1995-2	0.006722	0.000964	0.005758	-1.16818	1.36569	0.001485	6.82904	0.006573	0.043065	0.005758076
1990-1	0.015621	0.010119	0.005501	1.04549	1.30481	0.001431	6.27449	0.006218	0.035797	0.005501365
1987-4	0.019730	0.014562	0.005168	2.11964	1.22583	0.001332	5.30816	0.005742	0.026464	0.005168377
1988-4	0.018418	0.013257	0.005161	1.80406	1.22414	0.001560	7.64232	0.005980	0.039353	0.005161247
1999-2	0.006236	0.001597	0.004639	-1.01501	1.10025	0.000958	2.26934	-0.004892	0.009931	0.004638908
1986-2	0.012743	0.008359	0.004384	0.61987	1.03980	0.001442	6.38304	0.004965	0.023163	0.004384032
2002-1	0.001798	-0.002146	0.003944	-1.92000	0.93544	0.001639	8.53557	0.004646	0.026213	0.003944021
2000-3	0.003954	0.007892	-0.003938	0.50707	-0.93406	0.002073	14.24175	-0.005193	0.052384	0.003938222
1993-2	0.000164	0.004059	-0.003894	-0.41988	-0.92361	0.001825	10.82048	-0.004792	0.034580	0.003894163
1988-2	0.001559	0.005398	-0.003838	-0.09612	-0.91041	0.001151	3.70966	-0.004148	0.010300	0.003838486
1997-2	0.010961	0.007262	0.003699	0.35468	0.87722	0.001671	8.90659	0.004387	0.024287	0.003698868
1989-4	0.020821	0.017192	0.003629	2.75571	0.86065	0.001822	10.78297	0.004462	0.029887	0.003628697
1989-1	0.010728	0.014329	-0.003601	2.06332	-0.85397	0.001281	4.82849	-0.003967	0.011666	0.003600548
2001-4	-0.000950	0.002606	-0.003556	-0.77119	-0.84344	0.001520	7.20670	-0.004088	0.017458	0.003556141
1987-3	0.006850	0.003368	0.003482	-0.58685	0.82594	0.000978	2.40372	0.003680	0.005854	0.00348237
1991-3	0.000722	0.004162	-0.003440	-0.39478	-0.81597	0.001892	11.70067	-0.004308	0.030026	0.003440339
1994-4	0.008323	0.004984	0.003339	-0.19615	0.79206	0.000920	2.01769	0.003507	0.004709	0.003339495
2001-2	0.000685	0.003988	-0.003303	-0.43687	-0.78348	0.001344	5.41885	-0.003677	0.011044	0.003303333
1997-4	0.010312	0.007225	0.003088	0.34563	0.73231	0.001850	11.14538	0.003824	0.022623	0.003087577
1999-1	0.009003	0.005934	0.003070	0.03349	0.72806	0.001121	3.46758	0.003303	0.006196	0.003069681
1994-3	0.007238	0.010303	-0.003065	1.08984	-0.72685	0.001151	3.71419	-0.003312	0.006573	0.003064581
1996-2	0.006682	0.003637	0.003045	-0.52185	0.72228	0.000830	1.45808	0.003188	0.003127	0.003045318
1987-1	-0.005632	-0.003014	-0.002618	-2.13002	-0.62089	0.001388	5.83912	-0.002936	0.007502	0.002617795
1997-3	0.006246	0.003631	0.002615	-0.52336	0.62028	0.001603	8.12385	0.003057	0.010859	0.002615229
1994-1	0.008796	0.006184	0.002612	0.09405	0.61944	0.001824	10.80669	0.003213	0.015527	0.002611692
1986-4	0.012593	0.010237	0.002356	1.07393	0.55880	0.001135	3.57893	0.002540	0.003756	0.002356051
1986-3	0.003189	0.000888	0.002301	-1.18641	0.54571	0.001405	6.01314	0.002588	0.005980	0.002300853
1999-4	0.009763	0.007492	0.002270	0.41036	0.53846	0.001027	2.75509	0.002414	0.002779	0.002270259
1999-3	0.010822	0.008626	0.002196	0.68451	0.52081	0.001028	2.76222	0.002335	0.002605	0.002195845
1993-4	0.003824	0.001677	0.002147	-0.99581	0.50921	0.001264	4.67600	0.002359	0.004018	0.002146948
1989-2	0.005127	0.007234	-0.002108	0.34799	-0.49988	0.001037	2.82787	-0.002243	0.002447	0.002107612
1987-2	0.011728	0.009845	0.001883	0.97916	0.44669	0.001735	9.68619	0.002267	0.006997	0.001883331
1994-2	0.004557	0.002713	0.001844	-0.74525	0.43741	0.000946	2.18962	0.001942	0.001527	0.001844211
1997-1	0.003401	0.005230	-0.001829	-0.13675	-0.43368	0.001557	7.60860	-0.002117	0.004914	0.001828504
1990-3	0.004504	0.002980	0.001524	-0.68057	0.36143	0.001600	8.09089	0.001780	0.003669	0.001523867
2001-1	0.005511	0.006866	-0.001355	0.25897	-0.32134	0.000867	1.68255	-0.001415	0.000681	0.001354859
1988-1	0.009727	0.008381	0.001346	0.62511	0.31931	0.001854	11.19931	0.001669	0.004329	0.001346274
1996-4	0.002029	0.003357	-0.001328	-0.58948	-0.31502	0.001073	3.09915	-0.001420	0.001051	0.001328202
1988-3	0.008542	0.009847	-0.001304	0.97959	-0.30937	0.001845	11.08042	-0.001613	0.004006	0.001304385
1995-1	0.004960	0.006261	-0.001301	0.11263	-0.30850	0.000964	2.31229	-0.001373	0.000792	0.001300724
2000-4	0.005869	0.007144	-0.001275	0.32622	-0.30248	0.001193	4.05702	-0.001386	0.001236	0.001275344
1990-4	0.009711	0.008495	0.001215	0.65288	0.28827	0.001635	8.48686	0.001430	0.002472	0.001215425
1996-1	0.002207	0.003411	-0.001205	-0.57638	-0.28575	0.000924	2.04272	-0.001266	0.000619	0.001204809
1998-3	0.005151	0.006349	-0.001198	0.13388	-0.28412	0.000815	1.36700	-0.001244	0.000464	0.00119793
1993-3	0.004551	0.003560	0.000992	-0.54048	0.23519	0.001603	8.12413	0.001159	0.001561	0.0009916211
1998-4	0.001269	0.002174	-0.000905	-0.87558	-0.21455	0.001193	4.05566	-0.000983	0.000622	0.0009045997
1995-4	0.001828	0.001229	0.000599	-1.10408	0.14216	0.001496	6.95016	0.000686	0.000476	0.0005993636
1991-4	0.009056	0.009566	-0.000510	0.91169	-0.12098	0.001857	11.24280	-0.000633	0.000625	0.0005100965
1991-2	0.001970	0.002469	-0.000498	-0.80436	-0.11819	0.001878	11.51242	-0.000622	0.000616	0.000498303
2000-1	0.008900	0.008430	0.000469	0.63718	0.11135	0.000923	2.03298	0.000493	0.000094	0.0004694583
1998-2	0.004095	0.004523	-0.000427	-0.30764	-0.10136	0.001106	3.35077	-0.000459	0.000116	0.0004273616
1998-1	0.008811	0.009166	-0.000355	0.81506	-0.08427	0.001257	4.61534	-0.000390	0.000109	0.0003553042
1991-1	0.007960	0.008244	-0.000284	0.59197	-0.06732	0.002021	13.48435	-0.000368	0.000251	0.0002838187
1995-3	0.001499	0.001620	-0.000122	-1.00944	-0.02886	0.001309	5.08674	-0.000135	0.000014	0.0001216948
1996-3	0.005271	0.005304	-0.000033	-0.11878	-0.00784	0.001103	3.32915	-0.000035	0.000001	0.0000330559
Median	0.005391	0.005880	-0.000014	0.02055	-0.00323	0.001295	4.95762	-0.000015	0.0005917	0.00001362409
2000-2	0.008763	0.008757	0.000006	0.71623	0.00138	0.000878	1.75048	0.000006	0.000000	0.000005807728
Mean	0.005795	0.005795	-0.000000	0.000000	-0.000000	0.001347	5.90625	0.000040	0.018623	0.00000000007275958

Figure 11 – “Predicted and residual values, after trimming and before any dummies”

Predicted & Residual Values (meqt - selected vars - final sta)										
Dependent variable: DLY										
Case No.	Observed Value	Predicted Value	Residual	Standard Pred. v.	Standard Residual	Std.Err. Pred.Val	Mahalanobis Distance	Deleted Residual	Cook's Distance	ABS(t)= A
1992-4	-0.002393	0.007498	-0.009891	0.36272	-2.74082	0.000605	0.78392	-0.010176	0.022320	
Minimum	-0.007313	-0.007313	-0.009891	-2.79293	-2.74082	0.000605	0.78392	-0.010176	0.000000	
1989-3	-0.000282	0.007524	-0.007806	0.36830	-2.16306	0.000882	2.78005	-0.008302	0.031624	
1992-3	-0.001812	0.005147	-0.006959	-0.13811	-1.92853	0.000870	2.67618	-0.007389	0.024359	
1990-2	-0.001063	0.005768	-0.006831	-0.00584	-1.89300	0.001042	4.27074	-0.007453	0.035579	
2001-3	0.001434	-0.005221	0.006655	-2.34720	1.84420	0.001771	14.18592	0.008766	0.142087	
Maximum	0.020821	0.017619	0.006655	2.51930	1.84420	0.003609	62.01563	0.008766	0.142087	
1995-2	0.006722	0.000461	0.006261	-1.13657	1.73495	0.001285	7.00086	0.007170	0.050031	
1992-1	0.014495	0.009086	0.005409	0.70120	1.49895	0.000832	2.36102	0.005713	0.013307	
2001-2	0.000685	0.005317	-0.004632	-0.10181	-1.28366	0.001212	6.12674	-0.005222	0.023633	
1987-4	0.019730	0.015190	0.004540	2.00179	1.25807	0.001160	5.52580	0.005063	0.020342	
1986-2	0.012743	0.008375	0.004368	0.54976	1.21029	0.001238	6.43567	0.004951	0.022166	
1988-2	0.001559	0.005918	-0.004358	0.02607	-1.20775	0.001011	3.96116	-0.004730	0.013485	
1997-4	0.010312	0.006045	0.004267	0.05328	1.18243	0.001667	12.46097	0.005425	0.048227	
1997-3	0.006246	0.002041	0.004205	-0.79995	1.16527	0.001422	8.80418	0.004979	0.029573	
1988-4	0.018418	0.014240	0.004178	1.79935	1.15770	0.001362	7.99047	0.004872	0.025964	
1990-1	0.015621	0.011461	0.004159	1.20731	1.15252	0.001302	7.21570	0.004781	0.022851	
1987-1	-0.005632	-0.002153	-0.003479	-1.69363	-0.96396	0.001214	6.14524	-0.003923	0.013371	
1989-1	0.010728	0.014064	-0.003336	1.76184	-0.92437	0.001103	4.90475	-0.003680	0.009720	
1999-3	0.010822	0.007500	0.003322	0.36316	0.92069	0.000914	3.05744	0.003550	0.006209	
1999-2	0.006236	0.002949	0.003287	-0.60642	0.91088	0.000870	2.67516	0.003490	0.005432	
1989-4	0.020821	0.017619	0.003202	2.51930	0.88735	0.001577	11.04694	0.003958	0.022974	
2000-3	0.003954	0.007115	-0.003161	0.28121	-0.87590	0.001786	14.45316	-0.004187	0.032983	
2001-4	-0.000950	0.002152	-0.003103	-0.77616	-0.85985	0.001306	7.26734	-0.003571	0.012823	
1996-1	0.002207	0.005221	-0.003015	-0.12232	-0.83536	0.000894	2.88024	-0.003212	0.004858	
1994-3	0.007238	0.010157	-0.002919	0.92945	-0.80896	0.000997	3.82829	-0.003161	0.005860	
1993-3	0.004551	0.001689	0.002863	-0.87501	0.79335	0.001431	8.92266	0.003397	0.013936	
1993-2	0.000164	0.002956	-0.002791	-0.60499	-0.77352	0.001606	11.49833	-0.003481	0.018438	
1994-4	0.008323	0.005610	0.002713	-0.03938	0.75182	0.000808	2.17304	0.002856	0.003140	
1994-1	0.008796	0.006307	0.002489	0.10906	0.68968	0.001582	11.12651	0.003081	0.014014	
1999-4	0.009763	0.007291	0.002472	0.31861	0.68505	0.000895	2.89300	0.002634	0.003280	
1995-1	0.004960	0.007395	-0.002435	0.34085	-0.67466	0.000874	2.70687	-0.002586	0.003009	
1989-2	0.005127	0.007559	-0.002433	0.37589	-0.67407	0.000893	2.87322	-0.002591	0.003157	
1987-3	0.006850	0.004723	0.002128	-0.22850	0.58959	0.000887	2.82462	0.002265	0.002381	
1997-1	0.003401	0.005435	-0.002034	-0.07665	-0.56375	0.001392	8.38543	-0.002390	0.006523	
1997-2	0.010961	0.008974	0.001987	0.67735	0.55048	0.001510	10.04742	0.002408	0.007798	
1996-2	0.006682	0.004744	0.001938	-0.22400	0.53713	0.000751	1.74272	0.002026	0.001364	
1998-4	0.001269	0.003079	-0.001810	-0.57873	-0.50149	0.001059	4.43788	-0.001980	0.002591	
2002-1	0.001798	0.000043	0.001755	-1.22554	0.48631	0.001505	9.96693	0.002124	0.006023	
1993-4	0.003824	0.002151	0.001673	-0.77645	0.46347	0.001110	4.98101	0.001847	0.002482	
1998-2	0.004095	0.005714	-0.001618	-0.01740	-0.44838	0.000979	3.64944	-0.001747	0.001723	
1986-3	0.003189	0.001766	0.001423	-0.85852	0.39440	0.001217	6.17870	0.001606	0.002251	
1988-1	0.009727	0.008344	0.001383	0.54308	0.38319	0.001594	11.30273	0.001718	0.004419	
1998-3	0.005151	0.006494	-0.001343	0.14887	-0.37213	0.000698	1.37306	-0.001395	0.000559	
1991-4	0.009056	0.010392	-0.001337	0.97952	-0.37039	0.001684	12.74261	-0.001709	0.004887	
1995-4	0.001828	0.003006	-0.001178	-0.59426	-0.32641	0.001362	7.98934	-0.001374	0.002064	
1986-4	0.012593	0.011417	0.001176	1.19785	0.32583	0.001006	3.91131	0.001275	0.000970	
2001-1	0.005511	0.006623	-0.001112	0.17647	-0.30814	0.000759	1.80410	-0.001163	0.000460	
1991-2	0.001970	0.002998	-0.001028	-0.59592	-0.28492	0.001679	12.65856	-0.001312	0.002864	
1996-4	0.002029	0.003030	-0.001001	-0.58924	-0.27732	0.000927	3.17338	-0.001071	0.000582	
1994-2	0.004557	0.003638	0.000919	-0.45962	0.25468	0.000867	2.65589	0.000975	0.000422	0
1998-1	0.008811	0.009693	-0.000882	0.83050	-0.24444	0.001126	5.14482	-0.000977	0.000713	
1999-1	0.009003	0.008150	0.000854	0.50165	0.23661	0.001065	4.50748	0.000935	0.000586	0
1990-3	0.004504	0.003734	0.000771	-0.43929	0.21360	0.001393	8.40873	0.000906	0.000940	
1988-3	0.008542	0.007926	0.000616	0.45397	0.17083	0.001653	12.23637	0.000780	0.000981	0
2000-1	0.008900	0.009347	-0.000447	0.75686	-0.12398	0.000814	2.21846	-0.000471	0.000087	0
2000-4	0.005869	0.006277	-0.000408	0.10276	-0.11318	0.001041	4.25550	-0.000445	0.000127	0
1990-4	0.009711	0.009371	0.000340	0.76187	0.09422	0.001423	8.81571	0.000403	0.000194	0
1996-3	0.005271	0.005593	-0.000322	-0.04305	-0.08931	0.000949	3.36912	-0.000346	0.000064	0
1991-1	0.007960	0.007640	0.000319	0.39313	0.08852	0.001739	13.63972	0.000416	0.000308	0
1995-3	0.001499	0.001371	0.000128	-0.94270	0.03540	0.001122	5.10155	0.000141	0.000015	0
1987-2	0.011728	0.011845	-0.000117	1.28906	-0.03239	0.001551	10.64722	-0.000143	0.000029	0
2000-2	0.008763	0.008777	-0.000014	0.63530	-0.00376	0.000752	1.75177	-0.000014	0.000000	0.0
Mean	0.005795	0.005795	0.000000	0.00000	-0.00000	0.001295	8.85938	0.000071	0.012347	0.000000
1993-1	-0.007313	-0.007313	0.000000	-2.79293	0.00000	0.003609	62.01563			
1991-3	0.000722	0.000722	0.000000	-1.08093	0.00000	0.003609	62.01563			
Median	0.005391	0.005981	0.000000	0.03967	0.00000	0.001143	5.33531	0.000141	0.004858	
1992-2	-0.007094	-0.007094	0.000000	-2.74627	0.00000	0.003609	62.01563			

Figure 12 – “Predicted and residuals, using the suggested dummies”

When the multiple regression uses the dummies as independent variables for the estimation of DLY, it makes the corresponding cases 100% accurate; that is, with zero residual. In fact, the dummy is just a coded vector that implies that mathematical effect. Notice on Figure 12 – “Predicted and residuals, using the suggested dummies” how the dummies’ cases come last, since they have null residual.

The regression corresponding to Figure 12 – “Predicted and residuals, using the suggested dummies” is shown next.

Regression Summary for Dependent Variable: DLY (megt - selecte						
R= .81466368 R ² = .66367691 Adjusted R ² = .60762306						
F(9,54)=11.840 p<.00000 Std.Error of estimate: .00361						
N=64	Beta	Std.Err. of Beta	B	Std.Err. of B	t(54)	p-level
Intercept			-0.002702	0.001373	-1.96752	0.054269
DLY (+4)	0.338425	0.087902	0.338313	0.087873	3.85004	0.000315
DLEER (+2)	-0.308971	0.084684	-0.065468	0.017944	-3.64853	0.000595
DLG	0.408102	0.087432	0.280203	0.060031	4.66762	0.000021
DLY_STAR (+3)	0.073762	0.098996	0.103264	0.138591	0.74510	0.459441
DLY_STAR (+1)	0.284130	0.089568	0.374638	0.118100	3.17221	0.002496
DRR10Y (+4)	-0.224765	0.088188	-0.248594	0.097538	-2.54870	0.013688
D91q3	-0.129381	0.088049	-0.005963	0.004058	-1.46942	0.147521
D92q2	-0.217404	0.082323	-0.010020	0.003794	-2.64086	0.010792
D93q1	-0.350481	0.084435	-0.016153	0.003891	-4.15089	0.000118

Figure 13 – “Regression after the suggested dummies”

Statistic	Value
Multiple R	0.81466
Multiple R ²	0.66368
Adjusted R ²	0.60762
F(9,54)	11.83999
p	0.00000
Std.Err. of Estimate	0.00361

Figure 14 – “Summary statistics after the suggested dummies”

From Figure 9 – “Summary statistics for step 1.25” to Figure 14 – “Summary statistics after the suggested dummies”, R increased from 0.71791 to 0.81466 and R² increased from 0.51539 to 0.66368. As mentioned before, R² represents the proportion of variance in the dependent variable that can be explained by the independent variables, so the model apparently became more representative.

On the other hand, the regression now has three variables with p-levels above 5%, including the one for the dependent DLY; the others are DLY_STAR(+3) with a terrible p-level of nearly 46% and the dummy D91q3, also with a very low B. This suggests that D91q3 may not be a smart choice for regressor.

If dummies are a way of forgetting residuals, then they should be used for cases with significant residuals that may be safely ignored... when can this be? For special cases, where particular events occurred.

The workbook’s branch 04 does nothing more than including the “predicted and residual values” just after trimming, using no dummies, sorted by descending

order of abs(residual), thus listing the highest abs(residual) cases on the first rows. The top five abs(residual) cases are 1993-1, 1992-4, 1989-3, 2001-3, and 1992-2. Which of these could safely be used as dummies? For which ones could I forget the significant residual, improve the model's estimation of DLY, yet not compromising its faithfulness?

The option for dummies was:

- D93q1 representing 1993-1, because 1993-1 hosts *both* the lowest DLY and the lowest DRR10Y, making it unique. Since the lower the DRR10Y, the better for DLY, this period could be seen as a "turning point", when the government decided it could not let DLY decelerate further and would boost investment incentives to the maximum, lowering the effective exchange rate to its minimum.
- D92q4 representing 1992-4, because 1992-4 is immediately before 1993-1, is the second highest abs(residual) and the 5th lowest DLY among only 9 cases with negative DLYs.
- D89q3 representing 1989-3, because it is the 3rd highest residual and has nearly exactly the same DLG as 1993-1. It also scores almost the same DRR10Y(+4) as 1992-4, the trimester before 1993-1, meaning it featured a somewhat similar context.
- D92q2 representing 1992-2, because it is the 5th highest residual and the 2nd lowest DLY.

The final workbook branch "05 - MR with new dummies" includes the regression results after using the aforementioned dummies, based on the spreadsheet "data 02 - selected vars, with new dummies".

Summary :	
Statistic	Value
Multiple R	0.85564
Multiple R ²	0.73212
Adjusted R ²	0.68158
F(10,53)	14.48512
p	0.00000
Std.Err. of Estimate	0.00325

Figure 15 – "Summary statistics, after using the new dummies"

Regression Summary for Dependent Variable: DLY (megt - selecte
R= .85564123 R²= .73212191 Adjusted R²= .68157887
F(10,53)=14.485 p<.00000 Std.Error of estimate: .00325

N=64	Beta	Std.Err. of Beta	B	Std.Err. of B	t(53)	p-level
Intercept			-0.003222	0.001208	-2.66791	0.010103
DLY (+4)	0.341801	0.079193	0.341687	0.079166	4.31607	0.000070
DLEER (+2)	-0.321178	0.076546	-0.068054	0.016219	-4.19589	0.000104
DLG	0.422353	0.078649	0.289988	0.054000	5.37011	0.000002
DLY_STAR (+3)	0.171510	0.082431	0.240107	0.115400	2.08065	0.042315
DLY_STAR (+1)	0.237006	0.080770	0.312503	0.106498	2.93434	0.004931
DRR10Y (+4)	-0.229310	0.079162	-0.253621	0.087554	-2.89673	0.005469
D89q3	-0.191089	0.072702	-0.008807	0.003351	-2.62839	0.011199
D92q2	-0.200394	0.073916	-0.009236	0.003407	-2.71111	0.009018
D92q4	-0.223191	0.071556	-0.010286	0.003298	-3.11912	0.002932
D93q1	-0.343190	0.076004	-0.015817	0.003503	-4.51542	0.000036

Figure 16 – “Regression summary, after using the new dummies”

The new regression has all p-levels below 5% and also shows an improvement in R², contrary to what was observed before, when the improvement in R² came at the expense of p-levels.

From Figure 14 – “Summary statistics after the suggested dummies” to Figure 15 – “Summary statistics, after using the new dummies”, R increased from 0.81466 to 0.85564 and R² increased from 0.66368 to 0.73212.

4 – Final Model

Hence, the final model is

$$\begin{aligned} DLY = & -0.003222 + 0.341687 * DLY(+4) + -0.068054 * DLEER(+2) + 0.289988 * DLG \\ & + 0.240107 * DLY_STAR(+3) + 0.312503 * DLY_STAR(+1) + - \\ & 0.253621 * DRR10Y(+4) + -0.008807 * D89q3 + -0.009236 * D92q2 + - \\ & 0.010286 * D92q4 + -0.015817 * D93q1 \end{aligned}$$

The Statistica workbook computed the Durbin-Watson statistic at least once for every branch of steps. The Durbin-Watson d value increased from 1.602502 right after trimming and before dummies, to 1.945934 when using the suggested dummies and, finally, to 2.161233. The serial correlation measured by this statistic was always decreasing.

Serial correlation is the correlation of adjacent residuals, also known as autocorrelation. Because the Durbin-Watson statistic is useful for evaluating the presence or absence of a serial correlation of residuals (i.e., whether or not residuals for adjacent cases are correlated, indicating that the observations or cases in the data file are not independent), I believe that the model became healthier in terms of autocorrelation as I worked on it.

Durbin-Watson d (meg and serial correlation o		
	Durbin- Watson d	Serial Corr.
Estimate	1.602502	0.184420

Figure 17 – “Durbin-Watson statistic after trimming and before dummies”

Durbin-Watson d (meg and serial correlation c		
	Durbin- Watson d	Serial Corr.
Estimate	1.945934	0.011330

Figure 18 – “Durbin-Watson statistic after trimming, using the suggested dummies”

Durbin-Watson d (meg and serial correlation of		
	Durbin- Watson d	Serial Corr.
Estimate	2.161233	-0.098557

Figure 19 – “Durbin-Watson statistic, using final dummies”

5 – References

Introductory Econometrics (2003), by Jeffrey Wooldridge

<http://www.statsoft.com/> - Statistica website

<http://www.statsoft.com/textbook/stathome.html> - A free textbook on statistics

<http://arturmarques.com/megt/econometrics/a2.pdf> - the original assignment

<http://arturmarques.com/megt/econometrics/a2.xls> - the raw data

<http://arturmarques.com/megt/econometrics/a2.stw> - the Statistica workbook