Základy práce s ekonometrickým programom GRETL

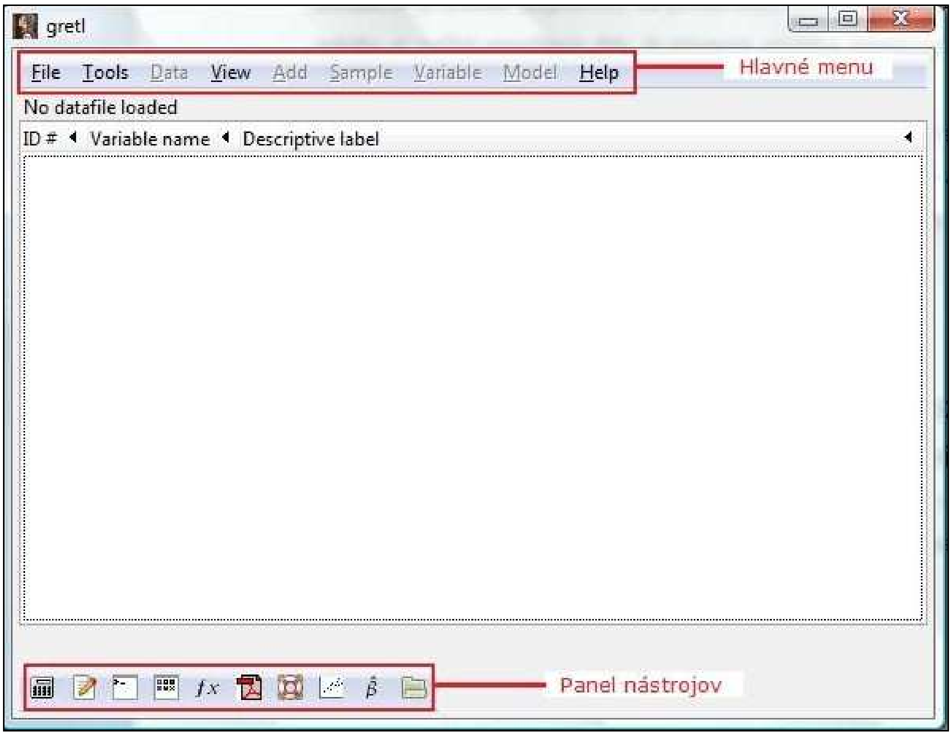
Martin Lukáčik, Viktor Slosiar

GRETL je voľne dostupný softvérový produkt so zameraním na štatistické metódy podporujúci ekonometrické analýzy[1]**.** Samotný názov je akronym pre *GNU Regression, Econometristic and Time-series Library*. Jeho vznik je výsledkom spoločného úsilia a vedeckej činnosti niekoľkých univerzít v Spojených štátoch amerických.

Používateľské rozhranie

Systém GRETL možno používať dvoma spôsobmi. Keďže snaha tvorcov systému od začiatku smerovala k priblíženiu ekonometrie širokej verejnosti, bolo okrem klasického príkazového riadku vytvorené aj grafické používateľské rozhranie (GUI – Graphical User Interface), ktoré je dnes pre väčšinu bežných používateľov prijateľnejšie. Po spustení programu sa objaví hlavné okno (*Obrázok 1*), kde je v hornej časti umiestnené *hlavné menu*, a v spodnej časti sa nachádza *panel nástrojov*. *Hlavné menu* obsahuje dve hlavné ponuky **`File`** a **`Tools`**. Ostatné sa sprístupnia v priebehu ďalších krokov. *Panel nástrojov* obsahuje praktické odkazy na najpoužívanejšie funkcie *Hlavného menu*, kalkulačku, návod na použitie a zoznam príkazov pre prácu v konzole.

Obrázok 1: Hlavné okno programu GRETL



[1] ROSENBLAD, A.: Journal of Statistical Software, Uppsala University, Marec 2008.

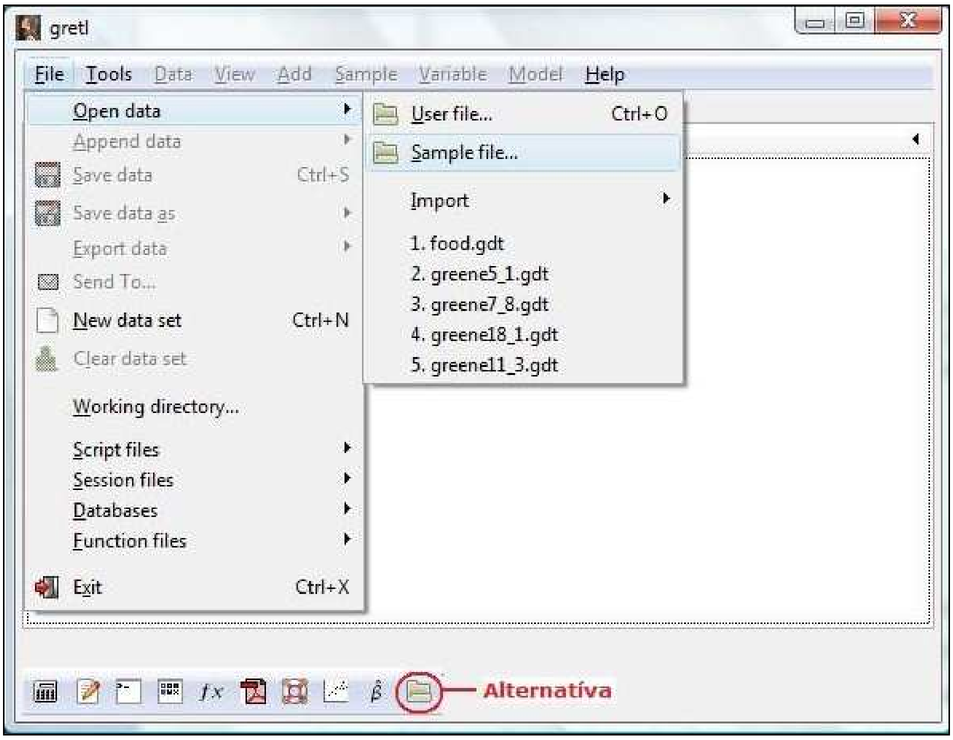
Popri práci v GUI je tu aj možnosť práce v rozhraní s príkazovým riadkom. Do toho sa dá dostať po kliknutí na **`Tools>Gretl console`** v *hlavnom menu***,** alebo cez *panel nástrojov* a tlačidlo „open gretl console“. V tomto móde sa pomocou príkazov vytvára syntax pre jednotlivé úlohy. Tento spôsob je oproti práci v GUI jednoduchší, ale vyžaduje oboznámenie sa s príkazmi. Ich úplne poznanie však nie je nevyhnutne potrebné, nakoľko je tu možnosť návratu do GUI, kde v *paneli nástrojov* tlačidlo „command reference“ obsahuje všetky príkazy aj s popisom. V ďalších častiach tejto práce sa budeme venovať používaniu programu GRETL iba v grafickom používateľskom prostredí.

Importovanie dát

Získavanie ekonometrických údajov a ich spracovanie do požadovaného formátu pre konkrétny softvér býva často komplikované. Existuje veľa rozličných programov, ktoré vyžadujú vlastné autorizované formáty údajov. V tomto ohľade však GRETL ponúka kompatibilitu a plnú funkčnosť. Je tak možné získať prístup k veľkému množstvu kvalitných údajov. Navyše, dáta zo systému GRETL je možné exportovať do ďalších programov v požadovanom formáte.

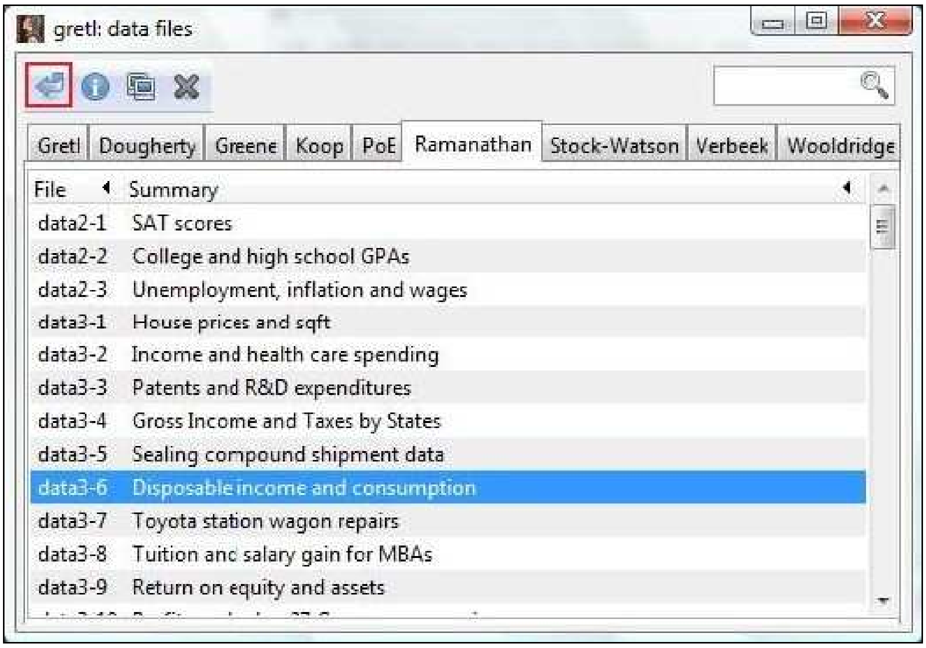
GRETL už po nainštalovaní základného balíka disponuje množstvom vzorových údajov. Otvoriť sa dajú z *hlavného menu* postupom krokov **`File>Open data>Sample file`**. alebo priamo z *panela nástrojov* (*Obrázok 2*).

Obrázok 2: Otváranie vzorových databáz



Tieto databázy možno ešte rozširovať doinštalovaním ďalších dátových súborov. Tie sú voľne stiahnuteľné z internetu – z oficiálnej stránky[2], alebo z ďalších špecializovaných stránok[3]. Pri inštalácii ich je potrebné uložiť do základného systémového balíka (napr. `c:\Program Files\gretl\data`). Po vyhľadaní potrebnej skupiny dát ich stačí označiť kurzorom a otvoriť tlačidlom „*open*“ v hornej časti okna. Vyberať možno napríklad z databáz údajov Ramanathan (2002), Greene (2003), Stock and Watson (2006) respektíve PoE (Hill - *Principles of Econometrics*, 3rd edition, 2008), ktorá bola doplnená zo stránky: http://www.learneconometrics.com/gretl.html. Po načítaní údajov s nimi možno priamo pracovať. Záložka **`Data`**, umiestnená v *hlavnom menu*, poskytuje veľký priestor na prispôsobenie štruktúry databázy podmienkam modelovania. Môžeme zobraziť, editovať, triediť hodnoty alebo pridávať pozorovania podľa potreby.

Obrázok 3: Zoznam databáz



Okrem vzorových databáz je tiež možné vytvoriť vlastnú priamo v programe GRETL (**`File>New data set>...`**), otvoriť databázu z počítača (**`File>Open data>User file`**), alebo importovať v príbuznom formáte – napr. ASCII, CVS, Excel – (**`File>Open data>Import`**).

To isté platí aj pri ukladaní databázy, je možné zvoliť formát v ktorom bude uložená podľa potreby používateľa (**`File>Save data as.. >> File>Export data>požadovaný formát`**)

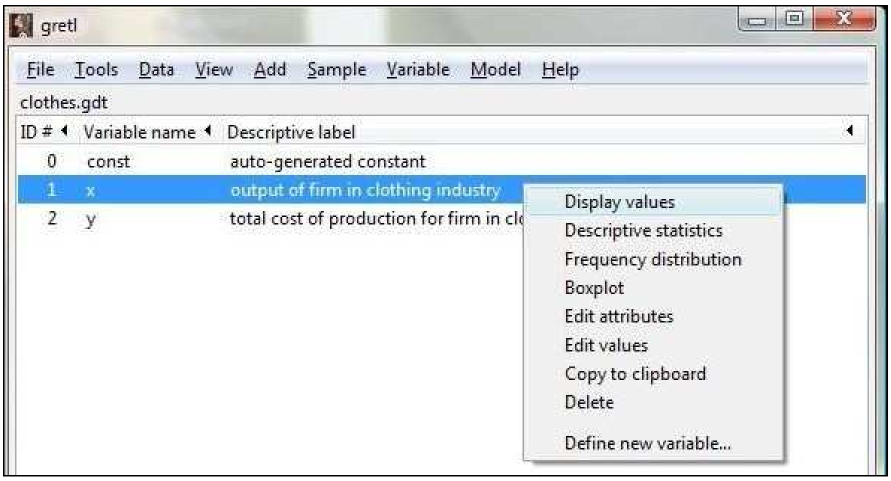
[2] http://gretl.sourceforge.net/gretl_data.html (2009-12-03)
[3] http://www.learneconometrics.com/gretl.html (2009-12-03)

Jednoduchý lineárny model s dvomi premennými

Prvým krokom analýzy je zabezpečenie vstupných údajov. Keďže GRETL ponúka veľké množstvo vzorových databáz, využijeme údaje súboru *clothes.gdt* zo zoznamu PoE (`File>Open data>Sample files>PoE>`*clothes*). Tento dátový súbor obsahuje údaje o *počte vyrobených kusov oblečenia podniku* (*output of firm in clothing industry*) a o *celkových nákladoch na výrobu oblečenia* (*total cost of production for firm in clothing industry*). Počet kusov oblečenia (v tis. ks) predstavuje nezávislú premennú *x*, celkové náklady (v tis. €) predstavujú závislú premennú *y*. Základná úloha môže preto znieť napríklad takto: „Ako sa zmenia celkové náklady podniku, ak dôjde k nárastu/poklesu výroby?“ Taktiež je zaujímavé zistiť aký silný je vzťah medzi premennými, či je tento vzťah priamo alebo nepriamo úmerný.

Niekedy je potrebné údaje prispôsobiť, aby vyhovovali podmienkam modelovania. To sa dá zabezpečiť kliknutím pravým tlačidlom na už vygenerovanú premennú v hlavnom okne. Zobrazené okno (*Obrázok 4*) ponúka možnosť zobrazenia hodnôt (*Display values*), ich úpravu (*Edit values*), meniť názov premennej (*Edit attributes*), pridávanie (*Define new varieble*), či zmazanie premennej (*Delete*). Dajú sa zobraziť popisné charakteristiky (*descriptive statistics*), čí graf *boxplot* pre vybranú premennú.

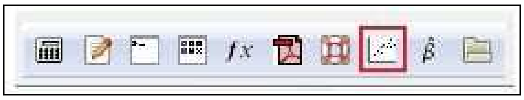
Obrázok 4: Menu na úpravu údajov



Grafické znázornenie dát

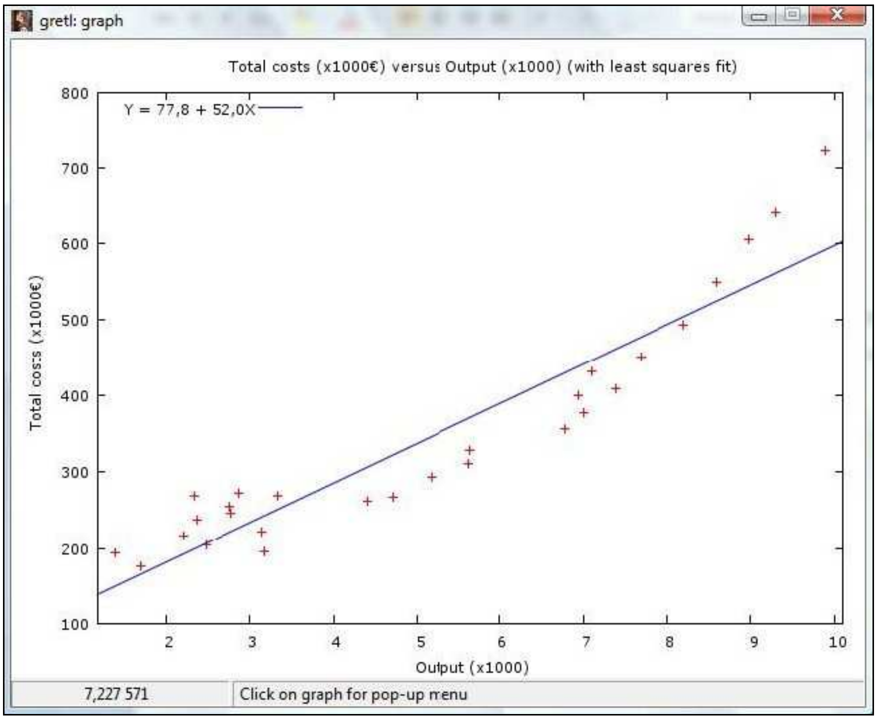
Na zobrazenie premenných *x,y* v súradnicovom grafe slúži tretie tlačidlo sprava na *paneli nástrojov* (*Obrázok 5*). Predtým je však potrebné označiť obe premenné v hlavnom okne. Po kliknutí na tlačidlo už stačí len zadefinovať osi pre zvolené premenné a graf sa vytvorí automaticky (*Obrázok 6*). Popis k osiam *x,y* sa pridáva v hlavnom okne po kliknutí pravým tlačidlom a výberom z menu „*Edit attributes*“ (Obrázok 4).

Obrázok 5: Tlačidlo vytvorenia grafu



Červené krížiky v grafe (*Obrázok 6*) znázorňujú jednotlivé bodové hodnoty. Modrá priamka je grafické znázornenie odhadu jednoduchého lineárneho modelu. Nazýva sa *vyrovnávajúca regresná priamka*. Priamka je lineárna a rastúca, to znamená, že s rastúcimi hodnotami premennej x rastú aj hodnoty premennej y. Zápis analytického tvaru rovnice $y = 77{,}8 + 52{,}0x$ sa nachádza v ľavom hornom rohu grafu a predstavuje odhadnutý tvar jednoduchého lineárneho modelu. (Pozor, obvykle zapisujeme rovnicu s indexmi.)

Obrázok 6: Grafické znázornenie premenných x a y

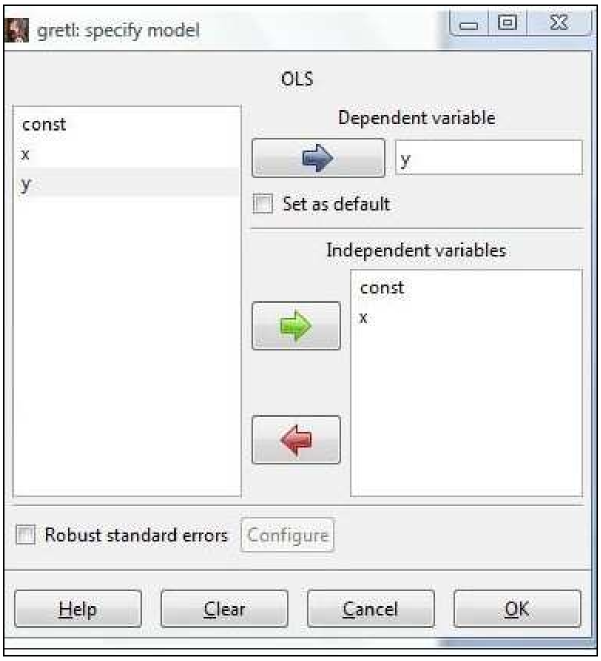


Odhad parametrov modelu

Odhad parametrov jednoduchého lineárneho modelu $y_i = \beta_0 + \beta_1 x_i + u_i$ v systéme GRETL sa vykoná tak, že sa označia obe premenné v hlavnom okne (premenné x a y). Ďalšie kroky sú nasledovné: *hlavné menu*`>Model>Ordinary Least Squares.`

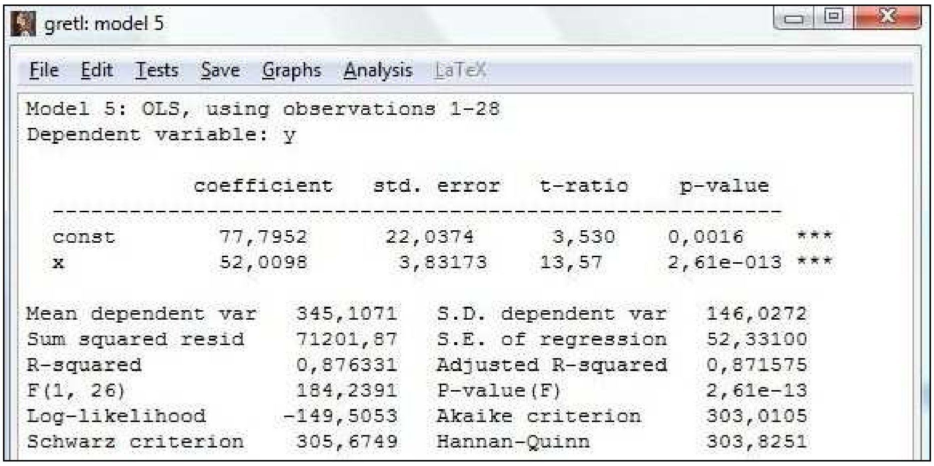
V okne, ktoré sa otvorí, je potrebné presne definovať závislú a nezávislú premennú (preniesť pomocou šípok z ľavej strany na pravú – *Obrázok* 7), potvrdiť tlačidlom OK a vidíme výsledok (*Obrázok* 8).

Obrázok 7: Špecifikácia modelu



Stĺpec „*coefficient*" označuje odhadované parametre. Prvá hodnota prislúchajúca riadku „*const*", je odhad parametra β_0 a nazýva sa *lokujúca konštanta*. Jej hodnota vyjadruje priemernú hodnotu premennej y za predpokladu, že premenná x = 0. V našom prípade sa dá interpretovať takto: „Za predpokladu, že sa nevyrobí ani jeden kus oblečenia, náklady na prevádzku podniku v priemere dosiahnu 77.795,2 €.

Obrázok 8: Výsledná tabuľka regresného odhadu




```
Model 5: OLS, using observations 1-28
Dependent variable: y

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const         77,7952      22,0374      3,530    0,0016    ***
  x             52,0098       3,83173    13,57     2,61e-013 ***

Mean dependent var   345,1071   S.D. dependent var   146,0272
Sum squared resid    71201,87   S.E. of regression   52,33100
R-squared            0,876331   Adjusted R-squared   0,871575
F(1, 26)             184,2391   P-value(F)           2,61e-13
Log-likelihood      -149,5053   Akaike criterion     303,0105
Schwarz criterion    305,6749   Hannan-Quinn         303,8251
```

Druhý parameter prislúchajúci riadku „x" je odhad parametra β_1 a nazýva sa *regresný koeficient*. Táto hodnota vyjadruje priemernú zmenu premennej y za predpokladu, že premenná x sa zmení o jednotku. Interpretácia je nasledovná: „So zvýšením výroby o tisíc kusov oblečenia celkové náklady podniku vzrastú v priemere o 52.009,8€."

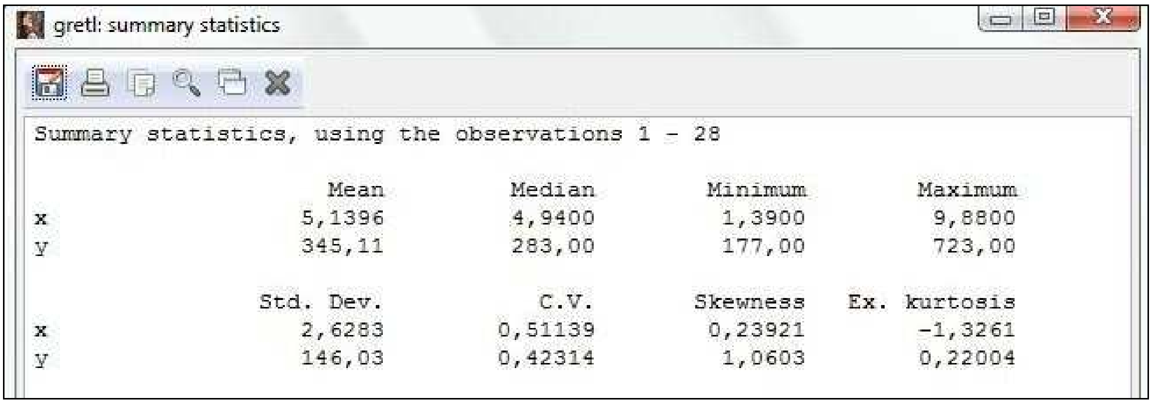
Výsledný vzťah pre odhad lineárneho regresného modelu má tvar:

$$\hat{y}_i = 77{,}7952 + 52{,}0098x_i$$

a je zhodný so vzťahom pre priamku v grafe na obrázku 6. Podľa tohto vzťahu sa môžu vytvárať prognózy. Napríklad, ak podnik vyrobí 4.500 (4,5 v tisícoch) kusov oblečenia, odhadované celkové náklady budú 311.839,3€ ($\hat{y}$ = 77,7952 + 52,0098*(4,5) = 311,8393).

Hodnota „*Mean depend var*“ je priemerná hodnota *y*. Ďalšie popisné štatistiky premenných sa získajú nasledovne: *hlavné menu*`>View>Summary statistics`. Zobrazí sa okno (*Obrázok 9*) s popisnými štatistikami pre jednotlivé premenné a zvolený počet pozorovaní („*Sample*“), kde vidíme priemer („*Mean*“), medián („*Median*“), minimum („*Minimum*“), maximum („*Maximum*“), štandardnú odchýlku („*Standard deviation*“), variačný koeficient („*C.V.*“), šikmosť („*Skewness*“) a špicatosť („*Kurtosis*“).

Obrázok 9: Popisné štatistiky premenných *x* a *y*


gretl: summary statistics

Summary statistics, using the observations 1 - 28

	Mean	Median	Minimum	Maximum
x	5,1396	4,9400	1,3900	9,8800
y	345,11	283,00	177,00	723,00

	Std. Dev.	C.V.	Skewness	Ex. kurtosis
x	2,6283	0,51139	0,23921	-1,3261
y	146,03	0,42314	1,0603	0,22004

Odhad intervalu spoľahlivosti

Pri odhadovaní parametrov metódou najmenších štvorcov nás zaujíma, nakoľko je tento odhad presný. Najrýchlejší spôsob ako to zistiť, je priamo porovnať hodnoty parametrov s odhadom ich rozptylov. Tento spôsob však nemusí odrážať reálne hodnoty, nakoľko je ovplyvnený výberom konkrétneho bodového odhadu. Aby sa odstránil vplyv náhody, vytvára sa intervalový odhad. Ten možno získať dvoma spôsobmi, a to buď vytvorením *konfidenčného intervalu* okolo hodnôt odhadu parametrov alebo *testovaním hypotéz* o parametroch β_0, β_1. Obe metódy sú rovnocenné.

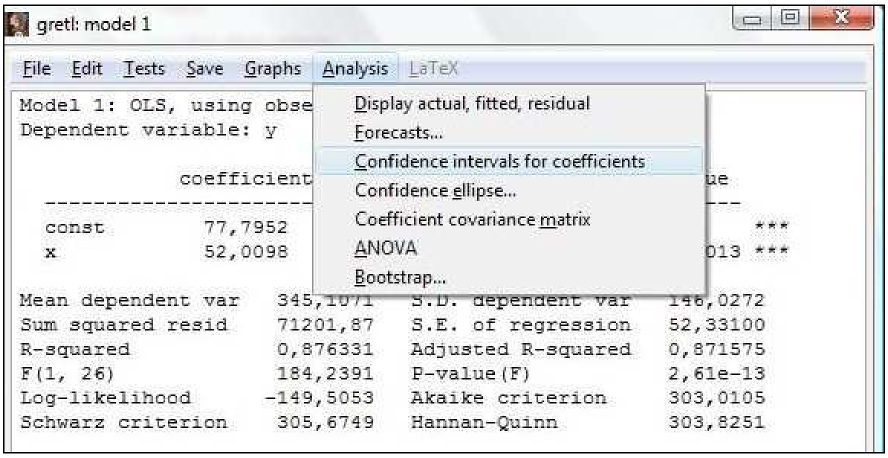
Vzťah na výpočet konfidenčného intervalu má tvar:

$$P\left\{\hat{\beta}_1 - t_{\alpha/2}\hat{\sigma}_{\beta_1} \le \beta_1 \le \hat{\beta}_1 + t_{\alpha/2}\hat{\sigma}_{\beta_1}\right\} = 1 - \alpha$$

kde $t_{\alpha/2}$ – kritická hodnota Studentovho rozdelenia (*t*) pre *n* - 2 stupne voľnosti, α je hladina významnosti, $\hat{\sigma}_{\beta_1}$ – odhad štandardnej odchýlky parametra β_1.

Na určenie hodnôt konfidenčného intervalu v systéme GRETL je potrebné mať odhadnutý jednoduchý lineárny model (*Obrázok* 8). V jeho výstupe na hornej lište cez ponuku „*Analysis*“ vyberieme „*Confidence intervals for coefficients*“ (*Obrázok* 10). Získame konfidenčný interval, kde sa dá cez tlačidlo **α** zmeniť hladina spoľahlivosti 1 – **α**.

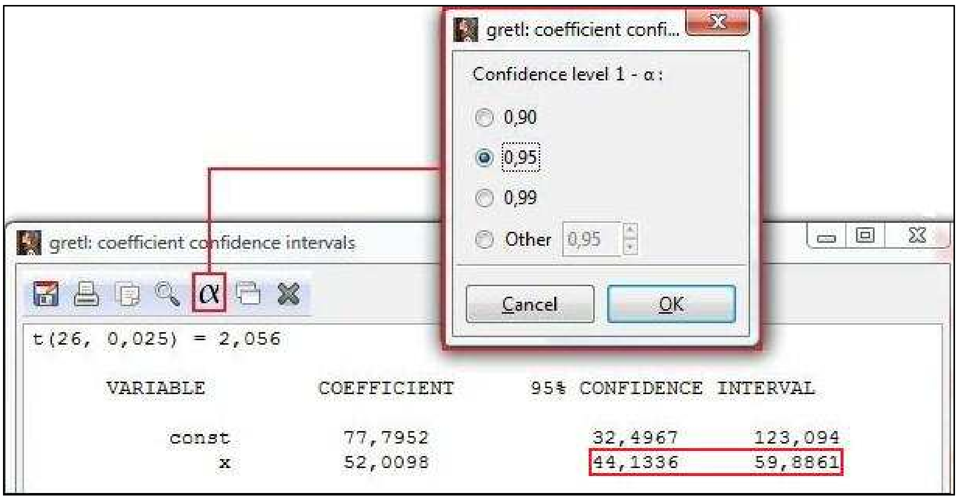
Obrázok 10: Spustenie výpočtu konfidenčného intervalu v odhadnutom modeli



Hodnoty výstupu (*Obrázok* 11) by sa v našom prípade interpretovali napr. takto: „V 95 prípadoch zo 100 po navýšení výroby oblečenia o 1.000 ks náklady na výrobu vzrastú o viac ako 44.133,6 € a menej ako 59.886,1 €.

$$P\{44{,}1336 \le \beta_1 \le 59{,}8861\} = 0{,}95$$

Obrázok 11: Zmena hladiny spoľahlivosti pri výpočte konfidenčného intervalu



Testovanie hypotéz

Testovaním hypotéz získavame dôležitý záver, či nami predpokladaný vzťah alebo celkový lineárny model je štatisticky významný. Základom je vytvorenie hypotéz, ktoré sa porovnávajú. Nulová hypotéza (H_0) sa porovnáva s alternatívnou (H_1). Podmienkou štatistickej významnosti je zamietnutie nulovej hypotézy (parameter sa rovná nule).

Testovanie hypotéz na vopred stanovenej hladine významnosti α je založené na vytvorení testovacej štatistiky, pre ktorú sú tabelované kritické hodnoty vhodných štatistických rozdelení. Hypotéza sa potom overuje porovnaním testovacej štatistiky s tabuľkovými kritickými hodnotami štatistík.

Príklady využitia testovania štatistickej významnosti na predchádzajúcej úlohe:

a) overenie štatistickej významnosti regresného koeficientu na hladine významnosti 0,05

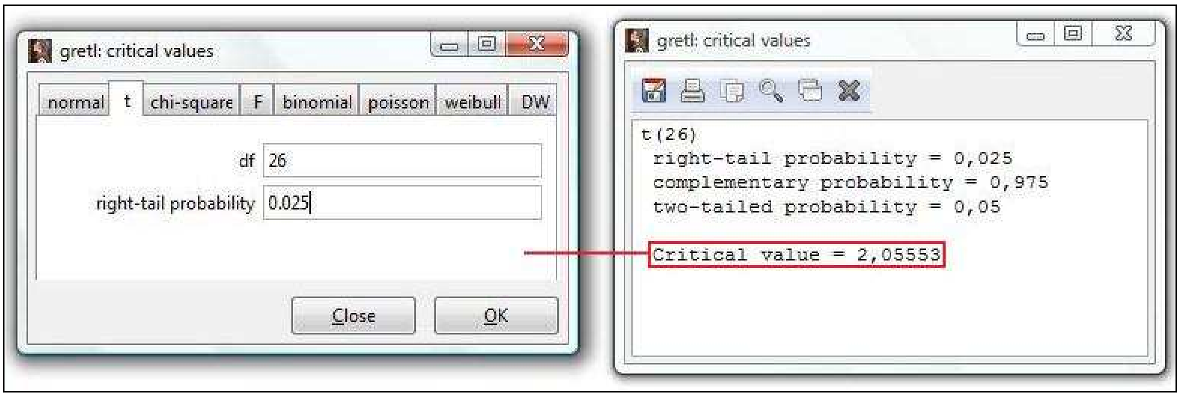
Formulácia hypotéz:

$$H_0\colon \beta_1 = 0$$

$$H_1\colon \beta_1 \neq 0$$

Kritické hodnoty sa v systéme GRETL dajú získať veľmi ľahko, a to postupom *hlavné menu*>**`Tools>Statistical tables`** (*Obrázok* 12). Potom stačí vybrať vhodné štatistické rozdelenie (v tomto prípade Studentovo rozdelenie), určiť počet stupňov voľnosti (*df*) a hodnotu α/2 (*right-tail probability*). Prípadnr iné údaje pre iné rozdelenia.

Obrázok 12: Určenie kritickej hodnoty testovanej štatistiky



Zamietnutie nulovej hypotézy sa vykoná potvrdením vzťahu

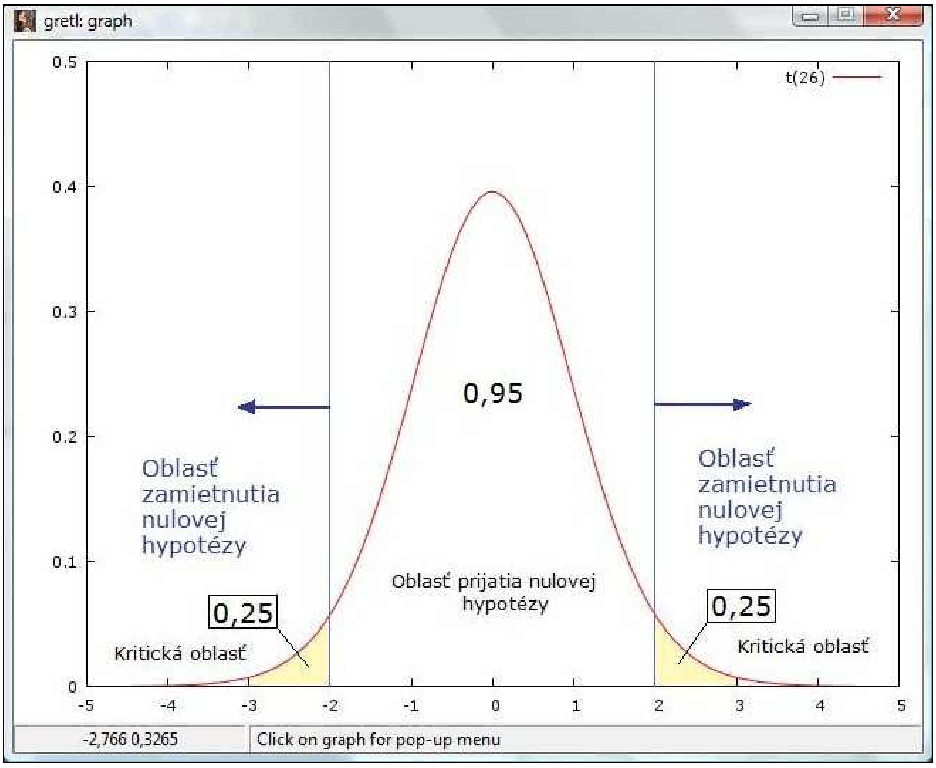
$$\left| \frac{\hat{\beta}_1 - 0}{\hat{\sigma}_{\beta_1}} \right| > t_{\alpha/2}$$

Po dosadení odhadu regresného koeficientu a odhadu jeho rozptylu spolu s testovacou štatistikou do vzťahu dostaneme nasledujúci záver:

$$\left| \frac{52{,}0098}{3{,}83173} \right| > 2{,}05553 \quad \rightarrow \quad 13{,}57345 > 2{,}05553$$

Keďže nerovnosť platí, zamietame nulovú hypotézu (hodnota 13,57345 sa nachádza v kritickej oblasti – *Obrázok* 13). Regresný koeficient je štatisticky významný – počet vyrobených kusov oblečenia ovplyvňuje výšku celkových nákladov podniku.

Obrázok 13: Graf Studentovho rozdelenia pre $t_{0,025}(26)$



b) overenie predpokladu, že na hladine významnosti 0,05 sa so zvýšením počtu vyrobených kusov oblečenia o 1.000 ks, v priemere zvýšia náklady o 45.000€.

Formulácia hypotéz:

$$H_0\colon \beta_1 = 45$$
$$H_1\colon \beta_1 \neq 45$$

Overenie hypotézy:

$$\left| \frac{\hat{\beta}_1 - 45}{\hat{\sigma}_{\beta_1}} \right| > t_{\alpha/2}$$

$$\left| \frac{52{,}0098 - 45}{3{,}83173} \right| > 2{,}05553 \quad \rightarrow \quad 5{,}7441 > 2{,}05553$$

Teda zamietame nulovú hypotézu H_0, že so zvýšením výroby o 1.000 ks oblečenia vzrastú celkové náklady v priemere o 45.000€.

c) overenie štatistickej významnosti modelu ako celku na hladine významnosti 0,1

Formulácia hypotéz:

$$H_0\colon \textit{model nie je štatisticky významný}$$
$$H_1\colon \textit{model je štatisticky významný}$$

Overenie hypotézy:

$$F > F_{0,1}(k, n - k - 1)$$

V tomto prípade sa štatistická významnosť overuje Fisherovym rozdelením. Kritickú hodnotu získame cez *hlavné menu* **>Tools>Statistical tables**. Zvolíme Fisherovo rozdelenie a po zadaní požadovaných hodnôt dostaneme $F_{0,1}(1, 26) = 2{,}90913$.

Vypočítaná hodnota F štatistiky je uvedená priamo vo výstupe odhadu lineárneho modelu a jej hodnota je v našom prípade $F = 184{,}2391$. Po dosadení do vzťahu dostaneme nerovnosť:

$$184{,}2391 > 2{,}90913$$

Nerovnosť (tentoraz bez absolútnej hodnoty) platí, zamietame nulovú hypotézu H_0, že regresný model nie je štatisticky významný.

Keďže je hladina významnosti α vždy daná vopred, ekonometrické programy poskytujú vo výstupe pravdepodobnostnú hodnotu (tzv. *p-value*). Je definovaná ako „najnižšia hodnota hladiny významnosti, na ktorej môžeme zamietnuť nulovú hypotézu na základe použitých výberových údajov.“[4] Nie je tak potrebné pri každom odhadovaní parametrov potrebné vyhľadávať hodnoty testovacích štatistík v tabuľkách.

V našej úlohe má *p-value* hodnotu $2{,}61e^{-13}$ pre parameter $\hat{\beta}_1$ a 0,0016 pre $\hat{\beta}_0$. V príkladoch a) a b) sme uvažovali na hladine významnosti $\alpha = 0{,}05$. Nulovú hypotézu H_0 môžeme zamietnuť, ak platí $p < 0{,}05$. Táto podmienka je splnená ($2{,}61e^{-13} < 0{,}05$) a prijímame alternatívnu hypotézu H_1. Analogicky v príklade c) so zadanou hladinou významnosti $\alpha = 0{,}01$ zamietame nulovú hypotézu H_0 ($2{,}61e^{-13} < 0{,}01$).

Koeficient determinácie

Koeficient determinácie vyjadruje akú časť variability závislej premennej vieme vysvetliť regresným modelom. Označuje sa R^2 a v systéme GRETL sa nachádza hneď vo viacerých výstupoch:

a) vo výstupe odhadu lineárneho modelu (*Obrázok* 14), kde je označený ako „R-squared“. Jeho hodnota je $R^2=0{,}876331$.
b) z analýzy rozptylu odhadnutého regresného modelu (ANOVA), ktorej výstup (*Obrázok* 15) získame, výberom záložky **Analysis>ANOVA** výstupe odhadu lineárneho modelu. Jeho hodnota je rovnaká $R^2=0{,}876331$.

Koeficient determinácie sa vypočíta nasledovne:

$$R^2 = \frac{\mathrm{ESS}}{\mathrm{TSS}} = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}}$$

[4] LUKÁČIKOVÁ, A.-LUKÁČIK, M.: Ekonometrické modelovanie s aplikáciami. Bratislava: EKONÓM, 2008, str. 70.

kde *ESS* – variabilita vysvetlená regresným modelom

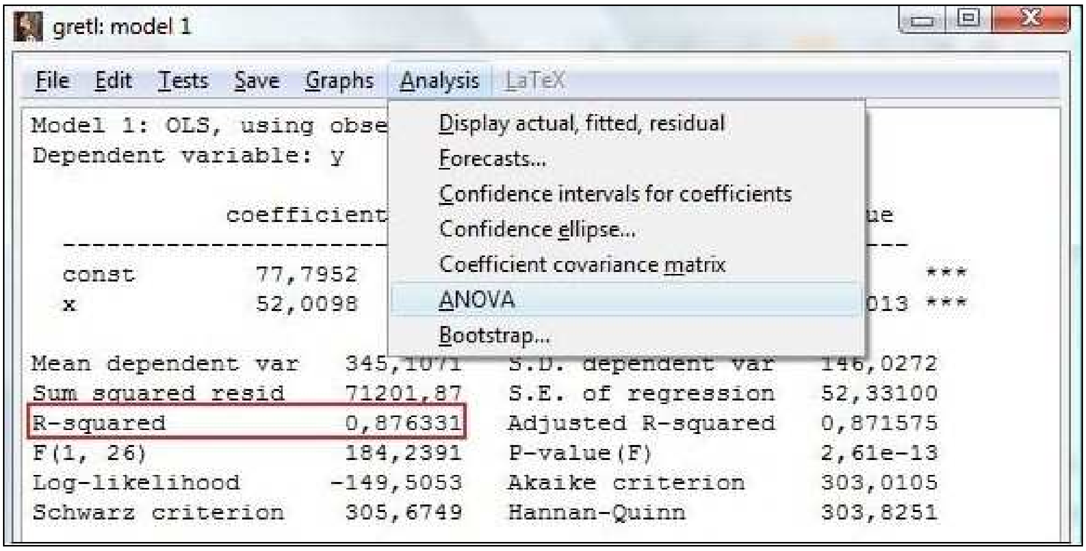
RSS – variabilita nevysvetlená regresným modelom (súčet štvorcov reziduálov)

TSS – celková variabilita závislej premennej (*TSS* = *RSS* + *ESS*)

Algebricky zapísané:

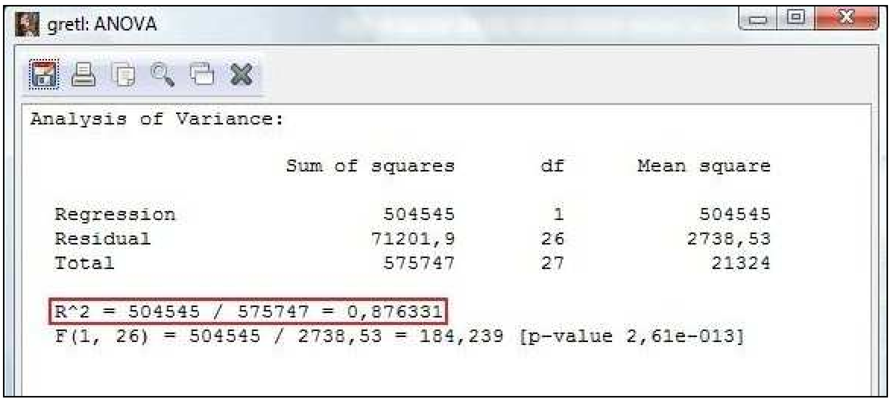
$$R^2 = \frac{504545}{504545 + 71202} = 0{,}876331$$

Obrázok 14: Koeficient determinácie a výber ANOVA vo výstupe odhadu modelu



Interpretácia koeficientu determinácie je v našej úlohe nasledovná: „Regresným modelom s nezávislou premennou *x* (počet kusov oblečenia) vieme vysvetliť 87,63% variability celkových nákladov podniku (závislá premenná *y*). Zvyšných 12,37% variability celkových nákladov spôsobujú činitele nezaradené do regresného modelu a náhodné vplyvy.

Obrázok 15: Koeficient determinácie R^2 z analýzy rozptylu ANOVA



Lineárny model s viacerými premennými

Odhad lineárneho modelu a ďalšie postupy sú podobné odhadu modelu s dvomi premennými. V lineárnom modeli s viacerými premennými sa vyžaduje podmienka, aby žiadna nezávislá premenná nebola lineárnou kombináciou iných nezávislých premenných.

Základy si ukážeme na údajoch súboru *andy.gdt* (*hlavné menu*>**File**>**Open data**>**Sample file**>**PoE**>***andy***). Tento súbor obsahuje tri premenné: mesačné zisky z predaja S (v tis. €), predajnú cenu produktu P (v €) a náklady na reklamu A (v tis. €). Závislou (vysvetľovanou) premennou y_i bude zisk z predaja, na ktorý vplýva predajná cena produktu (premenná x_{i1}) a náklady na reklamu (premenná x_{i2}). Lineárny model má tvar:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

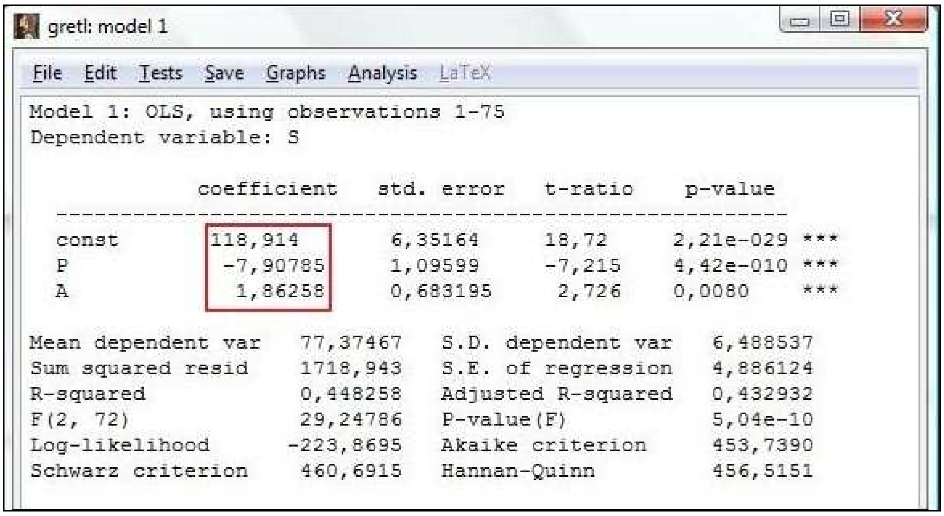
Na odhad parametrov použijeme metódu najmenších štvorcov a získame model:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}\,,$$

kde $\hat{y}_i$ je vysvetľovaná premenná a $\hat{\beta}_0, \hat{\beta}_1$ a $\hat{\beta}_2$ sú odhadnuté parametre.

V systéme GRETL sa odhad modelu s viacerými premennými líši oproti odhadu jednoduchého lineárneho modelu iba tým, že pri definovaní modelu je potrebné do časti pre nezávislé premenné presunúť všetky potrebné nezávislé premenné.

Obrázok 16: Výstup odhadu lineárneho modelu s tromi premennými


gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

```
Model 1: OLS, using observations 1-75
Dependent variable: S

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       118,914       6,35164      18,72     2,21e-029 ***
  P            -7,90785     1,09599      -7,215    4,42e-010 ***
  A             1,86258     0,683195      2,726    0,0080    ***

Mean dependent var   77,37467   S.D. dependent var   6,488537
Sum squared resid    1718,943   S.E. of regression   4,886124
R-squared            0,448258   Adjusted R-squared   0,432932
F(2, 72)             29,24786   P-value(F)           5,04e-10
Log-likelihood      -223,8695   Akaike criterion     453,7390
Schwarz criterion    460,6915   Hannan-Quinn         456,5151
```

Odhadnutý model má tvar:

$$\hat{s}_i = 118{,}914 - 7{,}90785 p_i + 1{,}86258 a_i$$

a vieme odhadnúť vplyv zmeny nezávislých premenných na závislú. Napríklad, s nárastom ceny produktu o jedno euro pri nezmenených nákladoch na reklamu, sa celkové mesačné zisky znížia v priemere o 7.907,85 €. Ak náklady na reklamu vzrastú o jednotku (1.000 €) pri nezmenenej cene, celkové mesačné zisky sa v priemere zvýšia o 1.862,58 €.

Môžeme takto modelovať rôzne situácie, napríklad vieme odhadnúť priemerný mesačný zisk za predpokladu, že cena produktu bude 5,5 € a mesačné náklady na reklamu budú 800 €:

$$\hat{s}_i = 118{,}914 - 7{,}90785(5{,}5) + 1{,}86258(0{,}8) = 76{,}910889$$

Mesačný zisk za hore uvedených predpokladov dosiahne hodnotu v priemere 76.910,89 €. Určenie konfidenčných intervalov je analogické ako pri jednoduchom lineárnom regresnom modeli.

Výstup regresie ponúka aj množstvo ďalších štatistických údajov. Štandardné odchýlky parametrov lineárneho modelu sú uvedené v stĺpci „std. error". Ďalšie údaje vieme získať analýzou rozptylu ANOVA, ich využitie je rovnaké ako bolo uvedené v druhej kapitole. Napríklad na určenie koeficientu determinácie R^2, ktorého hodnota $R^2 = 0{,}448258$ vyjadruje fakt, že pomocou premenných x_{i1} a x_{i2} vieme vysvetliť 44,83% variability mesačných ziskov (y_i).

Pridávaním nezávislých premenných do modelu dochádza k zníženiu hodnoty celkovej variability reziduálov (ESS) a k zvýšeniu hodnoty koeficientu determinácie R^2. Hodnota R^2-*adjusted* = 0,432932 koriguje takto vzniknutý rozdiel.

Testovanie združenej hypotézy (*F*-test)

Test združenej hypotézy o parametroch modelu sa používa, keď hypotéza obsahuje predpoklady s viacerými znakmi porovnania. Ak je hodnota F štatistiky väčšia ako tabuľková hodnota Fisherovho rozdelenia, zamietame nulovú hypotézu:

$$F = \frac{(RSS_R - RSS_U)/q}{RSS_U/(n-k-1)} > F(q, n-k-1) \qquad \rightarrow \qquad \text{zamietame } H_0$$

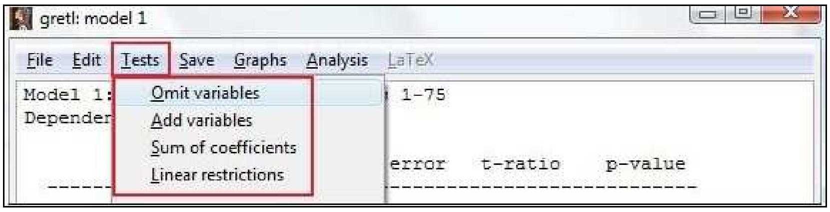
q – počet testovaných hypotéz (znakov porovnania). Indexy R a U označujú modely.

V našej modelovej úlohe sa budeme zaoberať testovaním hypotézy, či má cena produktu vplyv na celkový mesačný zisk. Nulová hypotéza H_0 bude predstavovať možnosť, že vplyv nemá, alternatívna hypotéza H_1 bude predstavovať možnosť, že má.

$$H_0: \beta_1 = 0 \quad \text{model } R \text{ má tvar: } \; y_i = \beta_0 + \beta_2 x_{i2} + u_i$$
$$H_1: \beta_1 \neq 0 \quad \text{model } U \text{ má tvar: } \; y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

V systéme GRETL sa takáto úloha rieši nasledovne: vo výstupe odhadu modelu vyberieme ponuku „Tests". Z nej sa testovania hypotéz týkajú prvé štyri možnosti (*Obrázok* 17) – a to vynechať premenné (*Omit variables*) v nulovej hypotéze H_0, pridať premenné (*Add variables*) v alternatívnej hypotéze H_1, testovanie súčtu koeficientov (*Sum of coefficients*) alebo testovanie ľubovoľného lineárneho obmedzenia (*Linear restrictions*).

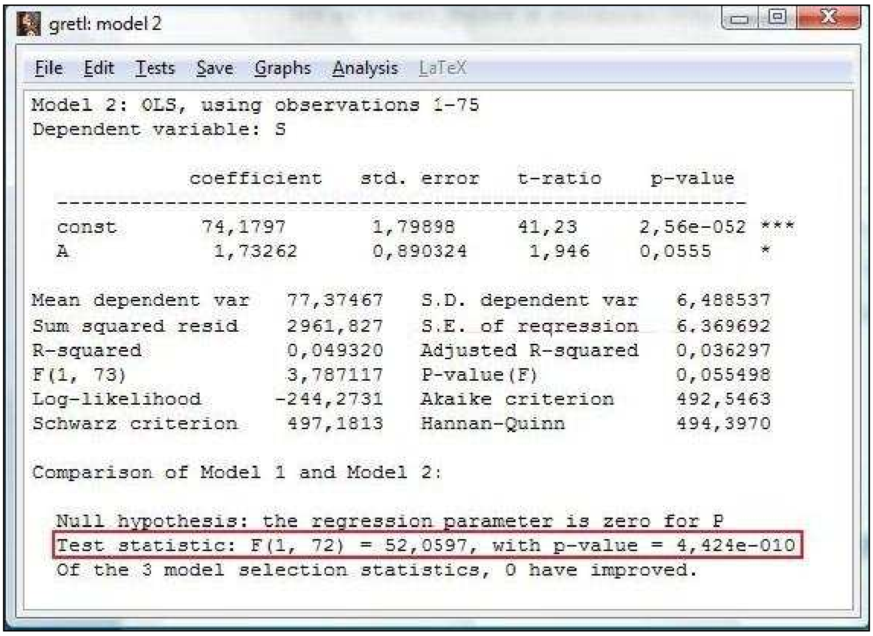
Obrázok 17: Možnosti testovania hypotéz vo výstupe odhadu modelu



V našom prípade chceme zistiť, či má na mesačný zisk vplyv cena produktu. Využijeme preto možnosť „*Omit variables*“. V zobrazenom dialógovom okne vyberieme premenné, ktoré chceme odstrániť z modelu (v našom prípade cenu *P)* t.j. x_{1i}o výbere možností odhadu necháme označenú možnosť „*Estimate reduced model*“ (príp. Waldov test ale v tomto prípade sa neodhadne redukovaný model) a potvrdíme tlačidlom OK. Zobrazí sa nám výstup modelu *R* (*Obrázok* 18), kde v spodnej časti sú vypočítané hodnoty pre $F(1, 72) = 52{,}0597$ a *p-value* = 4,424e^{-10}.

Rovnaký výsledok dostaneme aj použitím lineárneho ohraničenia „*Linear restrictions*“. V dialógovom okne lineárnej reštrikcie ho musíme špecifikovať. Chceme vylúčiť premennú *P* z modelu, preto sa jej parameter musí rovnať nule. V dialógovom okne sa to zapíše nasledovne: **b[2]=0**, kde 2 je poradové číslo parametra v modeli. Po potvrdení OK sa zobrazí výstup s rovnakými hodnotami *F* a *p-value* ako v prvom spôsobe.

Obrázok 18: Odhad modelu s ohraničením (model *R*)



```
Model 2: OLS, using observations 1-75
Dependent variable: S

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const        74,1797      1,79898      41,23     2,56e-052 ***
  A             1,73262     0,890324      1,946    0,0555    *

Mean dependent var    77,37467   S.D. dependent var    6,488537
Sum squared resid     2961,827   S.E. of regression    6,369692
R-squared             0,049320   Adjusted R-squared    0,036297
F(1, 73)              3,787117   P-value(F)            0,055498
Log-likelihood       -244,2731   Akaike criterion      492,5463
Schwarz criterion     497,1813   Hannan-Quinn          494,3970

Comparison of Model 1 and Model 2:

  Null hypothesis: the regression parameter is zero for P
  Test statistic: F(1, 72) = 52,0597, with p-value = 4,424e-010
  Of the 3 model selection statistics, 0 have improved.
```

$$F = \frac{(2961{,}827 - 1718{,}943)/1}{1718{,}943/(75-3)} > F(1,72)$$

52,05969 > 3,9739

Keďže nerovnosť platí, na hladine významnosti 0,05 zamietame nulovú hypotézu, že cena produktu nemá vplyv na celkový mesačný zisk.

Nesplnenie základných predpokladov modelu

Lineárny model musí spĺňať viaceré predpoklady, na základe ktorých je odhad modelu metódou najmenších štvorcov neskreslený a efektívny. Doteraz sme sa v texte stretli s podmienkou o lineárnej nezávislosti vysvetľujúcich premenných (nesplnenie tejto podmienky sa nazýva *multikolinearita*). Pri definícii náhodnej zložky u_i sa predpokladá, že priemerná hodnota náhodnej chyby je pre každú hodnotu nezávislej premennej x_{ik} nulová. Teda $E(u_i) = 0$; $i = 1,2,3,...,n$. Zjednodušene by sa dalo povedať, že vplyvy nezahrnuté do modelu pôsobia navzájom protichodne a rušia sa. V tejto časti sa budeme podrobnejšie zaoberať ďalšími tzv. základnými predpokladmi lineárneho regresného modelu.

Heteroskedasticita

Ďalším predpokladom lineárneho modelu je rovnaká hodnota rozptylov náhodnej zložky u_i pre rôzne hodnoty vysvetľujúcej premennej x_{ik}. Označuje sa ako podmienka o homoskedasticite lineárneho modelu a jej algebrický zápis je nasledovný:

$$\mathrm{var}\left(u_i | x_i\right) = E\left[u_i^2 | x_i\right] = \sigma^2 = \text{konštanta}$$

V prípade, že hodnoty rozptylov nie sú rovnaké, dochádza k nesplneniu homoskedasticity a nastáva jav označovaný ako *heteroskedasticita*[5]. V takýchto prípadoch sa výsledky odhadu modelu bez zohľadnenia tejto poruchy nedajú overiť, lebo štatistické testy sú neadekvátne. Boli preto vypracované postupy detekcie heteroskedasticity a metódy výpočtu želaných výstupov v takomto modeli bez problémov s indukčnými závermi. Heteroskedasticita je typická najmä *v súbore prierezových údajov*, kde je výrazná heterogenita štruktúry skúmaných údajov. Potom platí:

$$\mathrm{var}\left(u_i | x_i\right) = E\left[u_i^2 | x_i\right] = \sigma_i^2$$

To znamená, že rozptyl náhodnej zložky σ_i^2 je závislý od konkrétneho pozorovania *i*.

Testovanie heteroskedasticity

V niektorých prípadoch už samotná povaha problému napovedá, že sa v danom modeli sa bude vyskytovať heteroskedasticita. Sú to napríklad súbory obsahujúce údaje typu príjmy – výdavky. Z viacerých metód testovania vysvetlíme na Whitovom teste.

V dialógovom okne odhadnutého modelu vyberieme z ponuky v hornej lište možnosť **`Tests>Heteroskedasticity>White’s test`**. Ak si zvolíme príklad modelu s dvomi premennými, dostaneme výstup na obrázku 19.

---

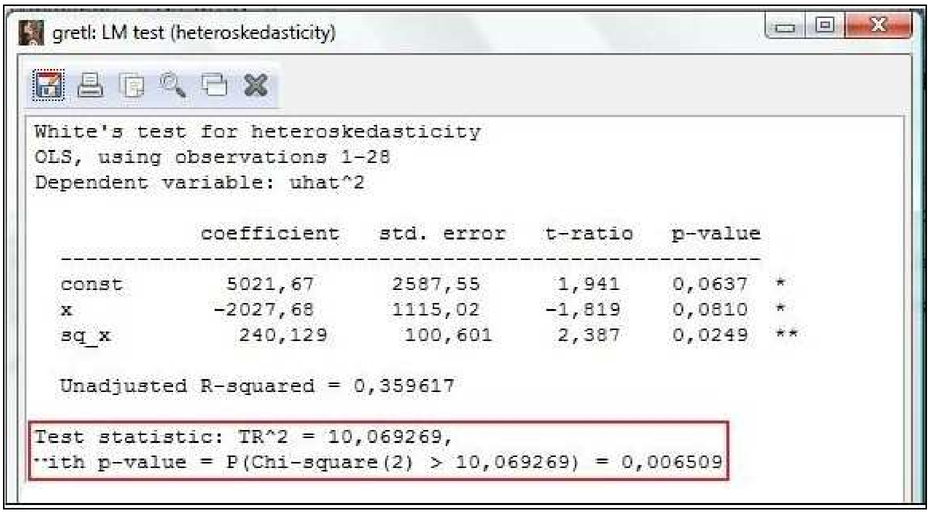
[5] LUKÁČIKOVÁ, A.-LUKÁČIK, M.: Ekonometrické modelovanie s aplikáciami. Bratislava: EKONÓM, 2008, str. 139.

Princíp Whitovho testu spočíva vo výpočte hodnoty štvorcov reziduálov u_i^2 pre každé pozorovanie a odhadnutí parametrov pomocnej regresie, pre ktorú sa vypočíta koeficient determinácie R^2. Testovacia štatistika má potom tvar $WH=nR^2$.

Samotný test je založený na porovnaní testovacej štatistiky *WH* s kritickými hodnotami χ^2-rozdelenia, v ktorom sa počet stupňov voľnosti *c* rovná počtu vysvetľujúcich premenných pomocnej regresie.

$$WH = nR^2 > \chi_c^2 \quad \rightarrow \quad zamietame\ H_0$$

Obrázok 19: Whiteov test heteroskedasticity



Výstup obsahuje údaje pomocnej regresie a hodnotu koeficientu determinácie R^2, pomocou ktorého je vypočítaná testovacia štatistika:

$$WH = 0{,}359617 \cdot 28 > \chi_2^2$$

$$10{,}069269 > 5{,}99146$$

Keďže nerovnosť platí, na hladine významnosti 0,05 zamietame nulovú hypotézu, že v modeli nie je prítomná heteroskedasticita.

Autokorelácia

Ďalším základným predpokladom lineárneho regresného modelu sú navzájom nekorelované náhodné zložky. To znamená, že náhodné zložky z dvoch ľubovoľných pozorovaní musia byť nezávislé. Algebricky zapísané:

$$\operatorname{cov}\left(u_t, u_{t+s} \mid x_t, x_{t+s}\right) = 0, \text{ pre } s > 0\,.$$

Nesplnenie tejto podmienky sa nazýva *autokorelácia*[6] a je typická pre modely *časových radov*.

[6] LUKÁČIKOVÁ, A.-LUKÁČIK, M.: Ekonometrické modelovanie s aplikáciami. Bratislava: EKONÓM, 2008, str. 205

Medzi hlavné príčiny výskytu autokorelácie v časových radoch ekonomických údajov patrí zotrvačnosť údajov (nevysvetlenie zmeny údajov závislej premennej vysvetľujúcimi premennými), nesprávna špecifikácia modelu a úprava údajov, ktorá predchádzajúca ich analýze.

Testovanie autokorelácie

Podobne ako pri testovaní heteroskedasticity, aj pri testovaní autokorelácie bolo vypracovaných viacero metód testovania. Najpoužívanejšou sa však stal *Durbin-Watsonov test*, aj keď je testom autokorelácie iba 1. rádu. Hypotézy majú tvar:

$$H_0: \rho_1 = 0$$

$$H_1: \rho_1 \neq 0$$

kde ρ je korelačný koeficient. Testovacia štatistika má tvar:

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

kde e_i sú reziduály získané z odhadu metódou najmenších štvorcov. Test pozostáva z porovnania testovacej štatistiky s tabuľkovou hodnotou. V nej na zvolenej hladine významnosti pre k nezávislých premenných a n pozorovaní dostaneme hornú kritickú hodnotu d_U a dolnú kritickú hodnotu d_L. O prijatí alebo zamietnutí danej hypotézy potom rozhodujú nasledujúce podmienky:

$$(0 < DW < d_L) \cup (4 - d_L < DW < 4) \quad \rightarrow \quad \textit{zamietame } H_0$$

$$d_U < DW < 4 - d_U \quad \rightarrow \quad \textit{prijímame } H_0$$

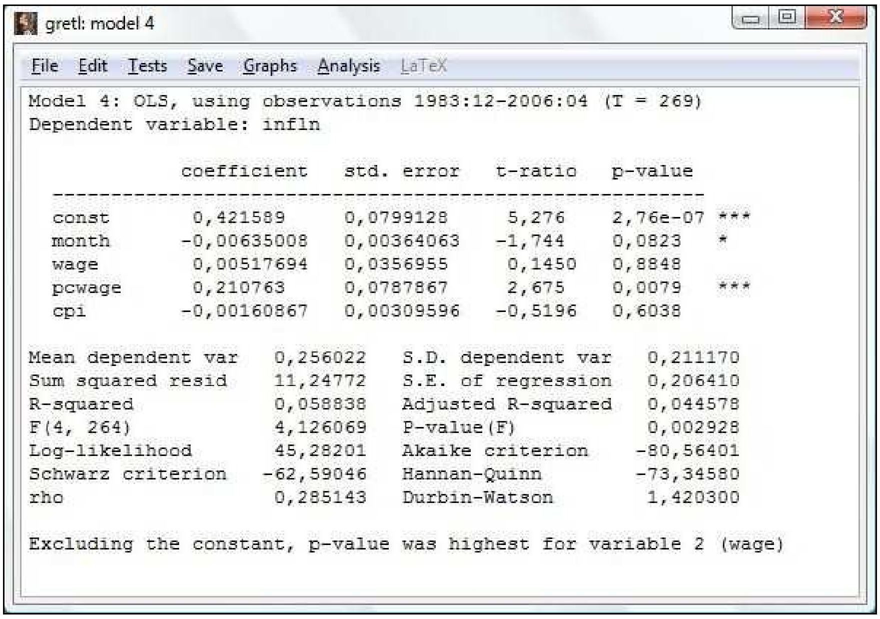
$$(d_L < DW < d_U) \cup (4 - d_U < DW < 4 - d_L) \quad \rightarrow \quad \textit{nevieme zistiť}$$

Ak na základe porovnania testovacej štatistiky a tabuľkových hodnôt zamietneme nulovú hypotézu H_0, v modeli je vysoko-pravdepodobná prítomnosť pozitívnej $(0 < DW < d_L)$ alebo negatívnej $(4 - d_L < DW < 4)$ autokorelácie 1. rádu.

V našom prípade pre ukážku využijeme súbor *inflation.gdt* (*hlavné menu*`>Open data>Sample file>PoE`). Ďalej je potrebné si prispôsobiť tieto údaje,lebo je dôležité definovať dátový súbor ako *časový rad*. Postup je nasledovný: *hlavné menu*`>Dataset structure`. Otvorí sa dialógové okno „*Dataset structure wizard – Structure of dataset*“, kde označíme možnosť časové rady „*Time series*“ a potvrdíme tlačidlom „*Forward*“. V ďalšom okne označíme frekvenciu sledovaných údajov (mesačná) a potvrdíme tlačidlom „*Forward*“. Nasleduje určenie počiatku sledovania v novom zobrazenom okne a posledný krok je potvrdenie celého nastavenia tlačidlom „*Apply*“.

Keď máme dáta upravené, môžeme pristúpiť k odhadnutiu lineárneho modelu a k samotnému testovaniu prítomnosti korelácie. Odhad modelu sa uskutoční analogicky ako už bolo prezentované (zoznam premenných modelu je zrejmý z obrázka 20).

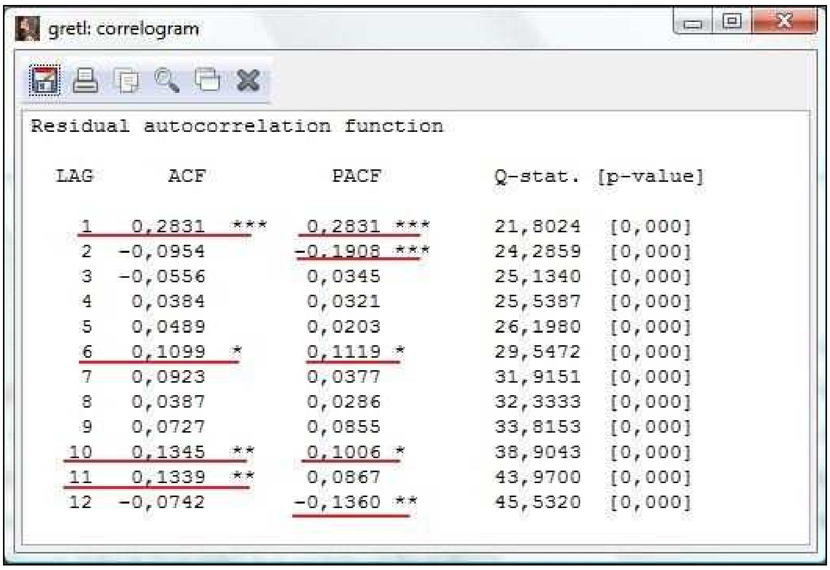
Obrázok 20: Výstup odhadu lineárneho modelu s časovými radmi




```
Model 4: OLS, using observations 1983:12-2006:04 (T = 269)
Dependent variable: infln

             coefficient    std. error    t-ratio   p-value
  ---------------------------------------------------------
  const       0,421589      0,0799128      5,276    2,76e-07 ***
  month      -0,00635008    0,00364063    -1,744    0,0823   *
  wage        0,00517694    0,0356955      0,1450   0,8848
  pcwage      0,210763      0,0787867      2,675    0,0079   ***
  cpi        -0,00160867    0,00309596    -0,5196   0,6038

Mean dependent var   0,256022   S.D. dependent var   0,211170
Sum squared resid    11,24772   S.E. of regression   0,206410
R-squared            0,058838   Adjusted R-squared   0,044578
F(4, 264)            4,126069   P-value(F)           0,002928
Log-likelihood       45,28201   Akaike criterion    -80,56401
Schwarz criterion   -62,59046   Hannan-Quinn        -73,34580
rho                  0,285143   Durbin-Watson        1,420300

Excluding the constant, p-value was highest for variable 2 (wage)
```

V odhadnutom modeli slúži na odhalenie prítomnosti autokorelácie *korelogram reziduálov*. Je to graf zaznamenávajúci korelácie medzi reziduálmi odhadnutého modelu. V okne výstupu odhadu lineárneho modelu (*Obrázok* 20) zvolíme „*Graphs*“ a vyberieme možnosť „*Residual correlogram*“. Určíme potrebný počet oneskorení reziduálov (*lag*) a potvrdíme tlačidlom OK. GRETL vytvorí výstup v tabuľkovej aj grafickej podobe.

Obrázok 21: Korelogram reziduálov v tabuľkovej forme.

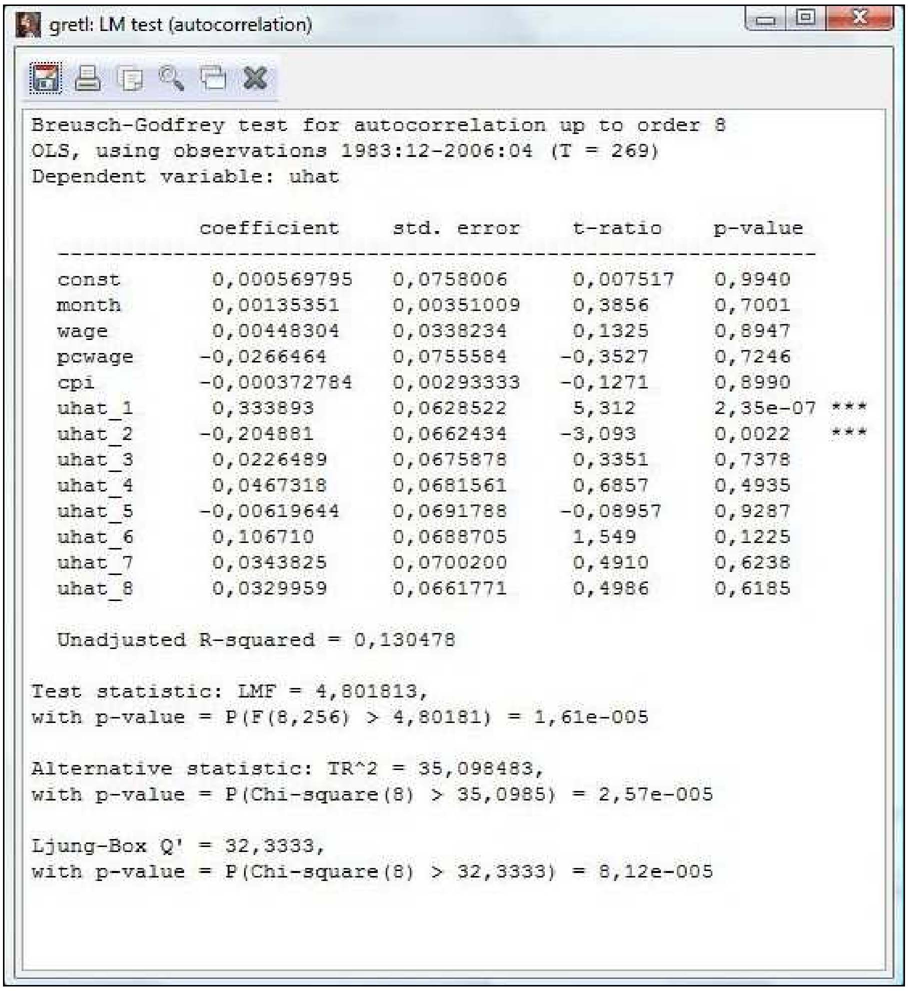



```
Residual autocorrelation function

  LAG      ACF          PACF         Q-stat. [p-value]

    1   0,2831  ***   0,2831 ***    21,8024  [0,000]
    2  -0,0954       -0,1908 ***    24,2859  [0,000]
    3  -0,0556        0,0345        25,1340  [0,000]
    4   0,0384        0,0321        25,5387  [0,000]
    5   0,0489        0,0203        26,1980  [0,000]
    6   0,1099  *     0,1119 *      29,5472  [0,000]
    7   0,0923        0,0377        31,9151  [0,000]
    8   0,0387        0,0286        32,3333  [0,000]
    9   0,0727        0,0855        33,8153  [0,000]
   10   0,1345  **    0,1006 *      38,9043  [0,000]
   11   0,1339  **    0,0867        43,9700  [0,000]
   12  -0,0742       -0,1360 **     45,5320  [0,000]
```

Hodnota ACF označuje hodnotu výberovej autokorelačnej funkcie a hodnota PACF hodnotu výberovej parciálnej autokorelačnej funkcie. Na grafe sú hodnoty potvrdzujúce autokoreláciu tie, ktoré prekračujú hranice naznačenej priamky. Sú to tie isté, ktoré v tabuľkovom korelograme označujú aspoň 2 hviezdičky.

Testovanie autokorelácie v odhadnutom lineárnom modeli je možné aj testom Lagrangeových multiplikátorov – v ponuke „*Tests*“ odhadu lineárneho regresného modelu vyberieme možnosť „*Autocorrelation*“. Určíme počet oneskorení (*lag*) a potvrdíme.

Obrázok 22: Test Lagrangeových multiplikátorov


gretl: LM test (autocorrelation)

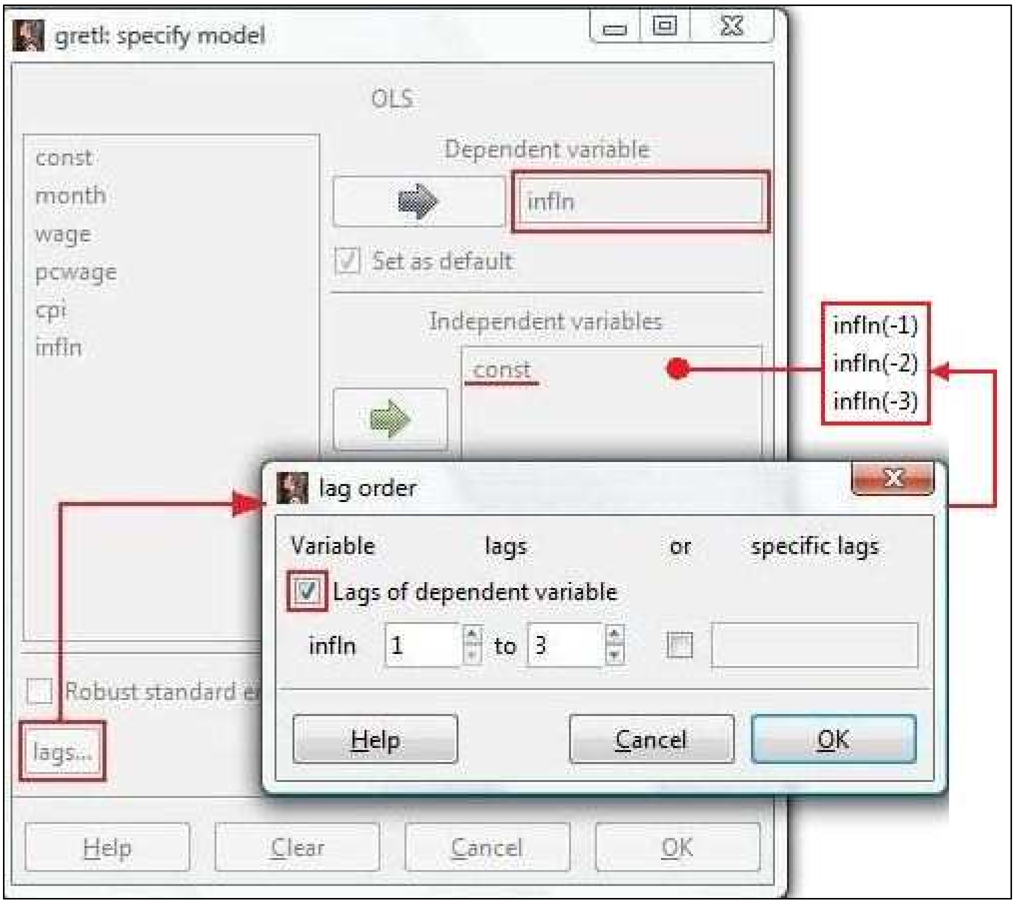
```
Breusch-Godfrey test for autocorrelation up to order 8
OLS, using observations 1983:12-2006:04 (T = 269)
Dependent variable: uhat

             coefficient    std. error    t-ratio    p-value
  -----------------------------------------------------------
  const       0,000569795   0,0758006     0,007517   0,9940
  month       0,00135351    0,00351009    0,3856     0,7001
  wage        0,00448304    0,0338234     0,1325     0,8947
  pcwage     -0,0266464     0,0755584    -0,3527     0,7246
  cpi        -0,000372784   0,00293333   -0,1271     0,8990
  uhat_1      0,333893      0,0628522     5,312      2,35e-07 ***
  uhat_2     -0,204881      0,0662434    -3,093      0,0022   ***
  uhat_3      0,0226489     0,0675878     0,3351     0,7378
  uhat_4      0,0467318     0,0681561     0,6857     0,4935
  uhat_5     -0,00619644    0,0691788    -0,08957    0,9287
  uhat_6      0,106710      0,0688705     1,549      0,1225
  uhat_7      0,0343825     0,0700200     0,4910     0,6238
  uhat_8      0,0329959     0,0661771     0,4986     0,6185

  Unadjusted R-squared = 0,130478

Test statistic: LMF = 4,801813,
with p-value = P(F(8,256) > 4,80181) = 1,61e-005

Alternative statistic: TR^2 = 35,098483,
with p-value = P(Chi-square(8) > 35,0985) = 2,57e-005

Ljung-Box Q' = 32,3333,
with p-value = P(Chi-square(8) > 32,3333) = 8,12e-005
```

Vo výstupe je vypočítaných viacero testovacích štatistík, ktoré po porovnaní s tabuľkovou hodnotou rozdelení. Všetky na hladine významnosti 0,05 potvrdzujú prítomnosť autokorelácie v modeli.

Prognóza hodnôt závislej premennej

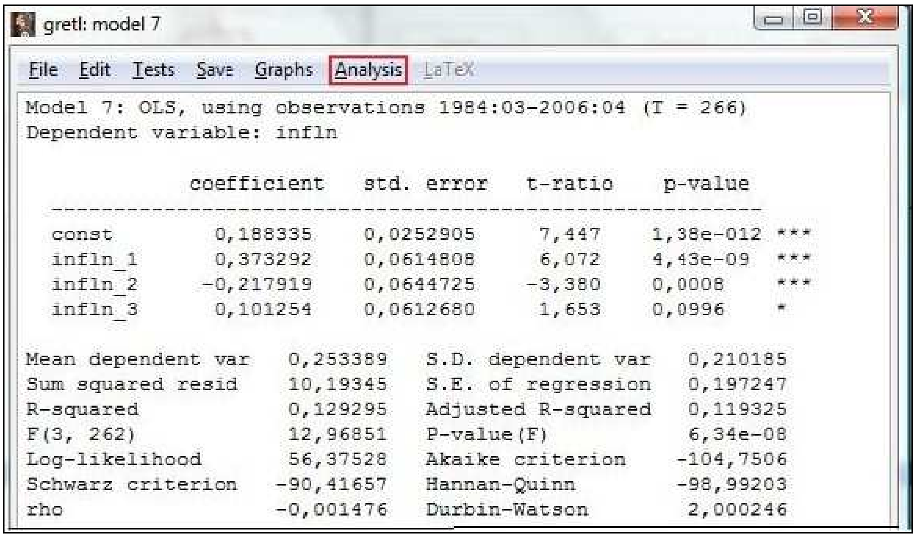
Prognózovanie a modelovanie nasledujúceho vývoja skúmaných javov je jednou zo základných aplikácií ekonometrického modelu. V našom prípade chceme odhadnúť, aké hodnoty bude nadobúdať inflácia v nasledujúcich troch mesiacoch. Postupujeme ako pri odhade lineárneho modelu (*hlavné menu*>**`Model>Ordinary Least Squares`**), pričom zadáme závislú premennú, ktorej sa týka prognóza. Nezávislými premennými v prvom modeli budú iba oneskorené premenné. Keďže sme na začiatku údaje špecifikovali ako rad, v spodnej časti okna sa pridalo tlačidlo „*lags*...“. Po kliknutí naň v dialógovom okne potvrdíme, že odhadujeme závislú premennú *infln* a určíme počet jej oneskorení (*Obrázok* 23). Po potvrdení sa do okna špecifikácie modelu pridajú tri premenné (*infln*(-1), *infln*(-2), *infln*(-3)), pomocou ktorých budeme prognózovať. Potvrdíme tlačidlom OK a máme odhad modelu pre premennú *infln* spolu s oneskorenými premennými.

Obrázok 23: Doplnenie oneskorených premenných do modelu



Po tejto jednoduchšej prognóze si ukážeme, že prognózovať sa dá aj v modeloch s vysvetľujúcimi premennými, kde nie sú iba oneskorené závislé premenné, ale vtedy musíme najskôr predĺžiť rozsah údajov. Následne musíme získať prognózu vysvetľujúcich premenných a doplniť ju medzi údaje. Ďalší postup je zhodný s tým, ktorý uvádzame.

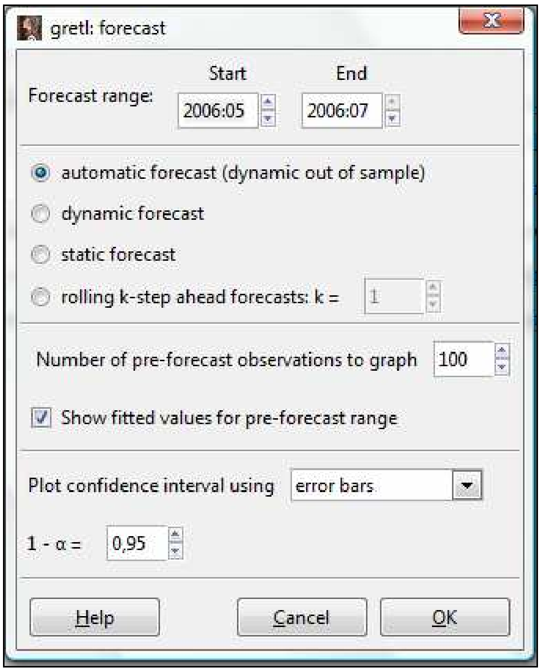
Obrázok 24: Odhad lineárneho regresného modelu s oneskorenými premennými



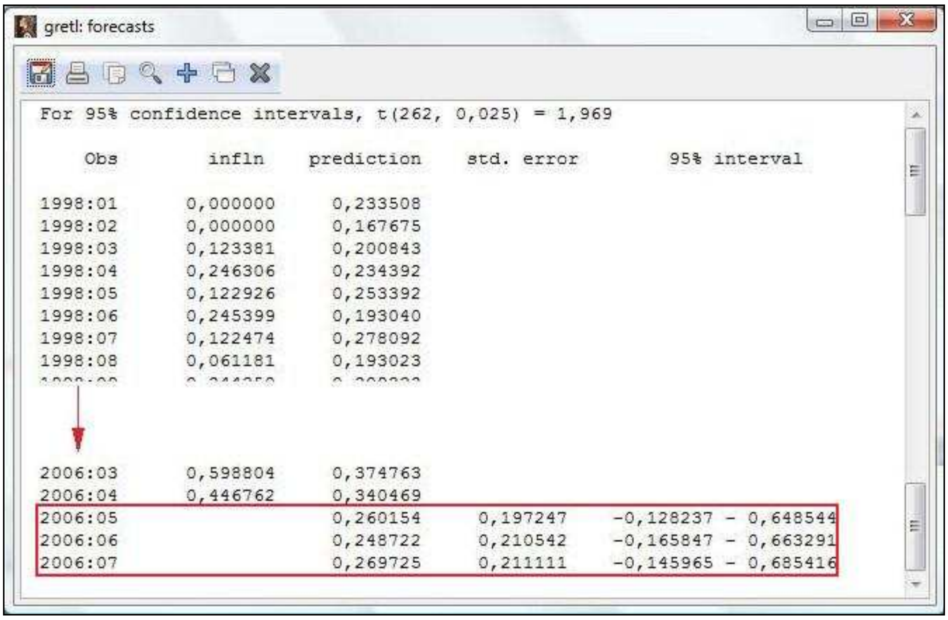

```
Model 7: OLS, using observations 1984:03-2006:04 (T = 266)
Dependent variable: infln

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       0,188335     0,0252905     7,447    1,38e-012 ***
  infln_1     0,373292     0,0614808     6,072    4,43e-09  ***
  infln_2    -0,217919     0,0644725    -3,380    0,0008    ***
  infln_3     0,101254     0,0612680     1,653    0,0996    *

Mean dependent var   0,253389   S.D. dependent var   0,210185
Sum squared resid    10,19345   S.E. of regression   0,197247
R-squared            0,129295   Adjusted R-squared   0,119325
F(3, 262)            12,96851   P-value(F)           6,34e-08
Log-likelihood       56,37528   Akaike criterion    -104,7506
Schwarz criterion   -90,41657   Hannan-Quinn        -98,99203
rho                 -0,001476   Durbin-Watson        2,000246
```

Po odhade modelu môžeme pristúpiť k samotnej prognóze: v hornej lište ponuky vyberieme záložku „*Analysis*“ a v nej možnosť „*Forecasts*..“. Otvorí sa nám okno, v ktorom sa určí rozsah odhadu – posledná hodnota inflácie je z 2006:04 a my potrebujeme prognózu na nasledujúce tri mesiace takže do okna „*Forrecast range*“ zadáme hodnotu 2006:05 a 2006:07. Ďalej označíme, že vytvárame automatickú predpoveď, zadáme počet pozorovaní zobrazených na grafe (100) a na záver určíme spoľahlivosť odhadu (0,95%).

Obrázok 25: Špecifikácia prognózy



Obrázok 26: Prognóza inflácie na mesiace 05/2006 až 07/2006



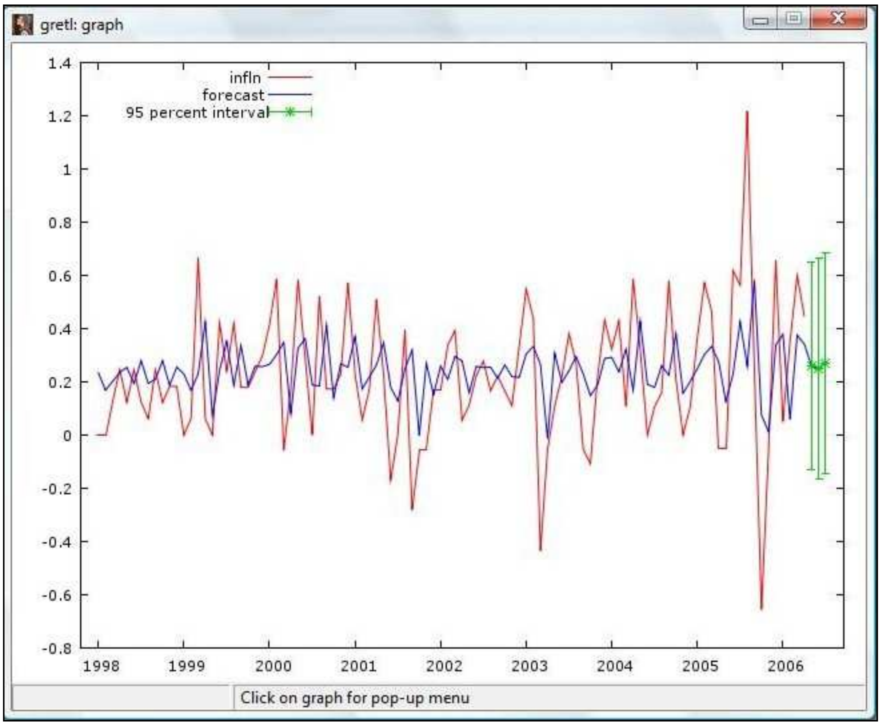
gretl: forecasts

For 95% confidence intervals, t(262, 0,025) = 1,969

Obs	infln	prediction	std. error	95% interval
1998:01	0,000000	0,233508		
1998:02	0,000000	0,167675		
1998:03	0,123381	0,200843		
1998:04	0,246306	0,234392		
1998:05	0,122926	0,253392		
1998:06	0,245399	0,193040		
1998:07	0,122474	0,278092		
1998:08	0,061181	0,193023		
2006:03	0,598804	0,374763		
2006:04	0,446762	0,340469		
2006:05		0,260154	0,197247	-0,128237 - 0,648544
2006:06		0,248722	0,210542	-0,165847 - 0,663291
2006:07		0,269725	0,211111	-0,145965 - 0,685416

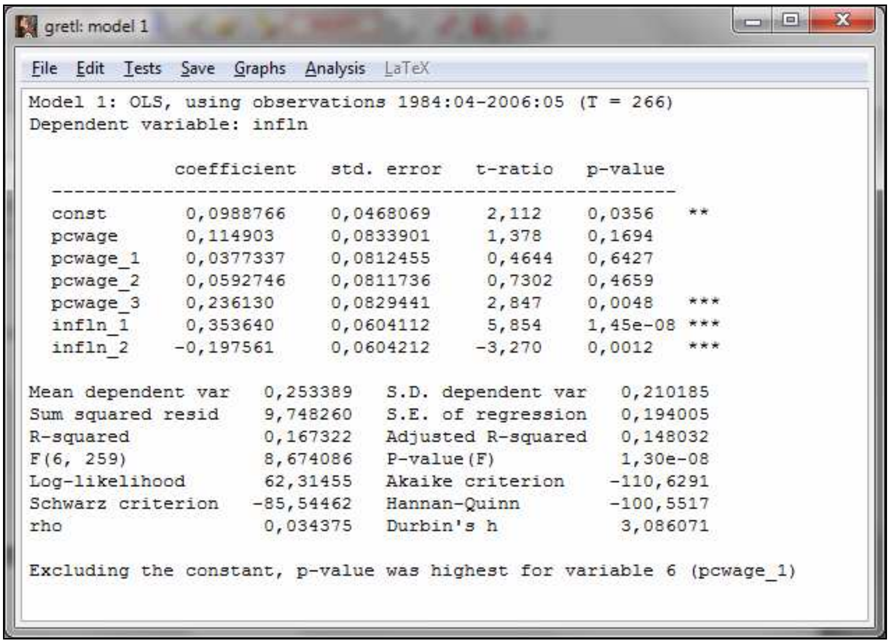
Prognóza získaná takýmto jednoduchým autoregresným modelom nie je v tomto prípade príliš vhodná, aj keď by mohla prípadne vyjsť. Tento záver potvrdzuje aj graf.

Obrázok 27: Graf vývoja inflácie v rokoch 1998 – 2006 s pridanou prognózou



Prv ako vytvoríme prognózu pomocou modelu s vysvetľujúcimi premennými, kde nie sú iba oneskorené závislé premenné, nájdeme vhodný model. Nám sa pozdával ako vhodný autoregresný model s rozloženým oneskorením na obrázku 28.

Obrázok 28: Model pre prognózu s vysvetľujúcimi premennými



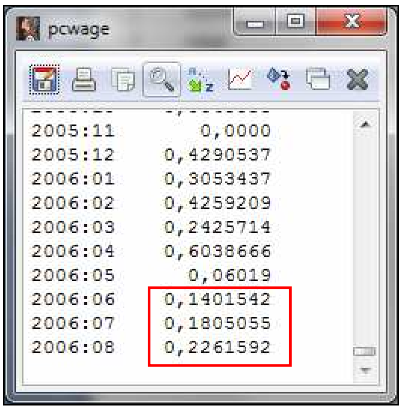

```
Model 1: OLS, using observations 1984:04-2006:05 (T = 266)
Dependent variable: infln

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        0,0988766    0,0468069     2,112   0,0356    **
  pcwage       0,114903     0,0833901     1,378   0,1694
  pcwage_1     0,0377337    0,0812455     0,4644  0,6427
  pcwage_2     0,0592746    0,0811736     0,7302  0,4659
  pcwage_3     0,236130     0,0829441     2,847   0,0048    ***
  infln_1      0,353640     0,0604112     5,854   1,45e-08  ***
  infln_2     -0,197561     0,0604212    -3,270   0,0012    ***

Mean dependent var   0,253389   S.D. dependent var   0,210185
Sum squared resid    9,748260   S.E. of regression   0,194005
R-squared            0,167322   Adjusted R-squared   0,148032
F(6, 259)            8,674086   P-value(F)           1,30e-08
Log-likelihood       62,31455   Akaike criterion    -110,6291
Schwarz criterion   -85,54462   Hannan-Quinn        -100,5517
rho                  0,034375   Durbin's h           3,086071

Excluding the constant, p-value was highest for variable 6 (pcwage_1)
```

Najskôr musíme predĺžiť rozsah údajov pomocou ponuky *hlavné menu*>**Data>Add observations…** zobrazenej na obrázku 29.

Obrázok 29: Rozšírenie počtu pozorovaní o tri, po august 2006

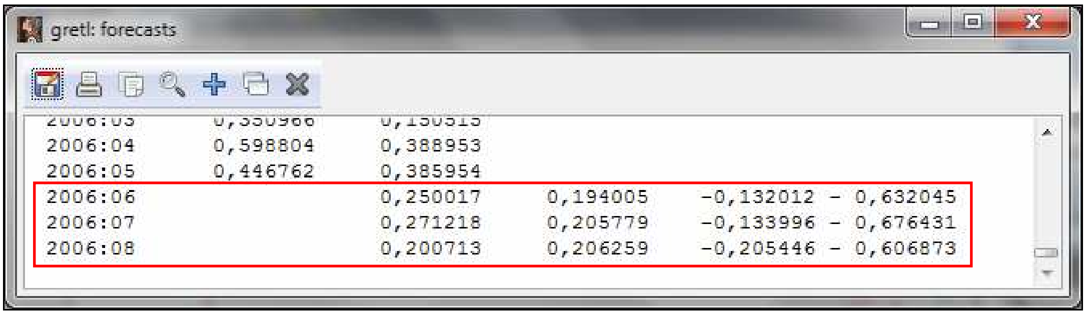
Následne doplníme rad hodnôt vysvetľujúcich premenných o prognózované hodnoty pre obdobia o ktoré bol rozšírený dátový súbor (*Obrázok* 30).

Obrázok 30: Vysvetľujúca premenná doplnená o 3 údaje potrebné pre prognózu



Ďalší postup sme už realizovali, teda špecifikujeme prognózu podľa obrázku 25. Výsledok prognózy závisiaci od hodnôt z obrázka 30 vidíme zobrazený na obrázku 31.

Obrázok 31: Výsledok prognózy



Literatúra

[1] ADKINS, L.C.: Using gretl for Principles of Econometrics 3rd Ed., 2009.

[2] COTRELL, A.: Gretl User's Guide, 2009.

[3] LUKÁČIKOVÁ, A. – LUKÁČIK, M.: Ekonometrické modelovanie s aplikáciami. Bratislava: EKONÓM, 2008.

[4] ROSENBLAD, A.: Journal of Statistical Software, Uppsala University, 2008.