

Ilustrační příklad odhadu LRM v SW Gretl

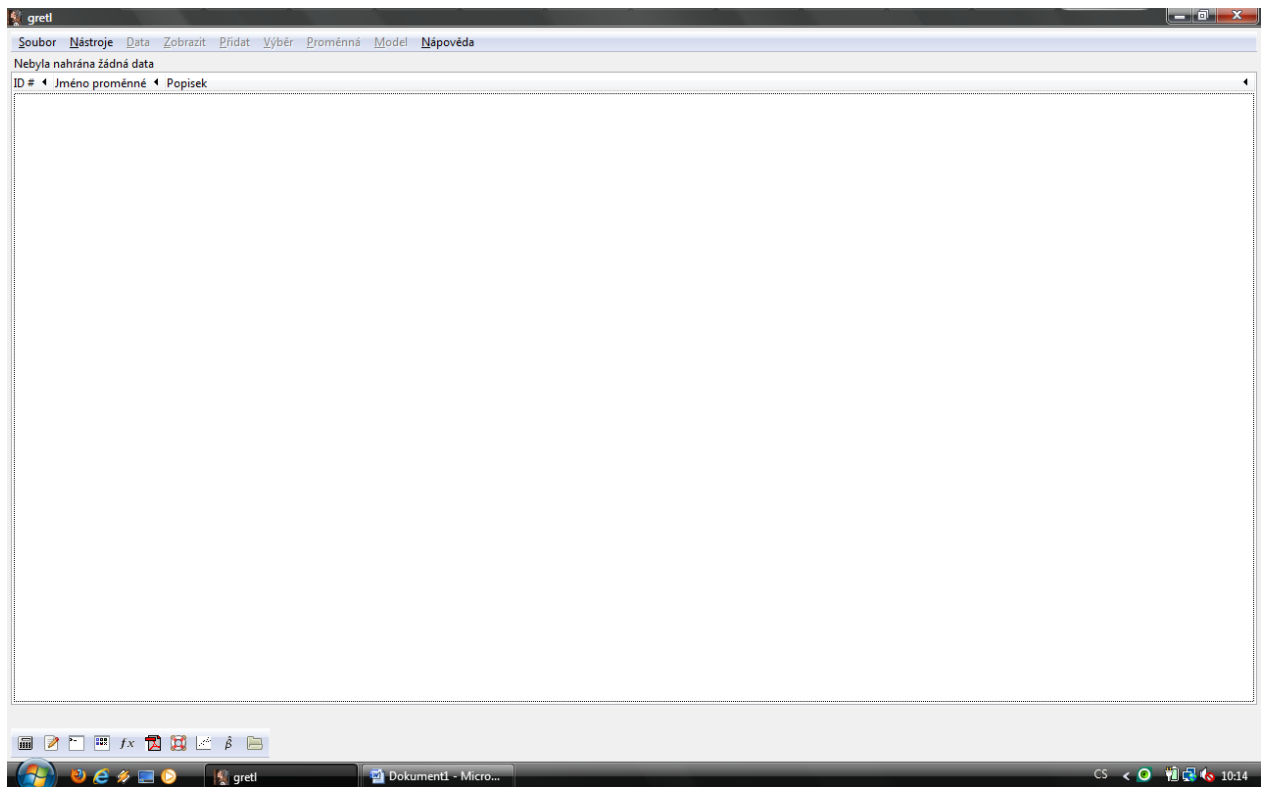


Podkladové údaje
Korelační matice
Odhad lineárního regresního modelu (LRM)
Verifikace modelu

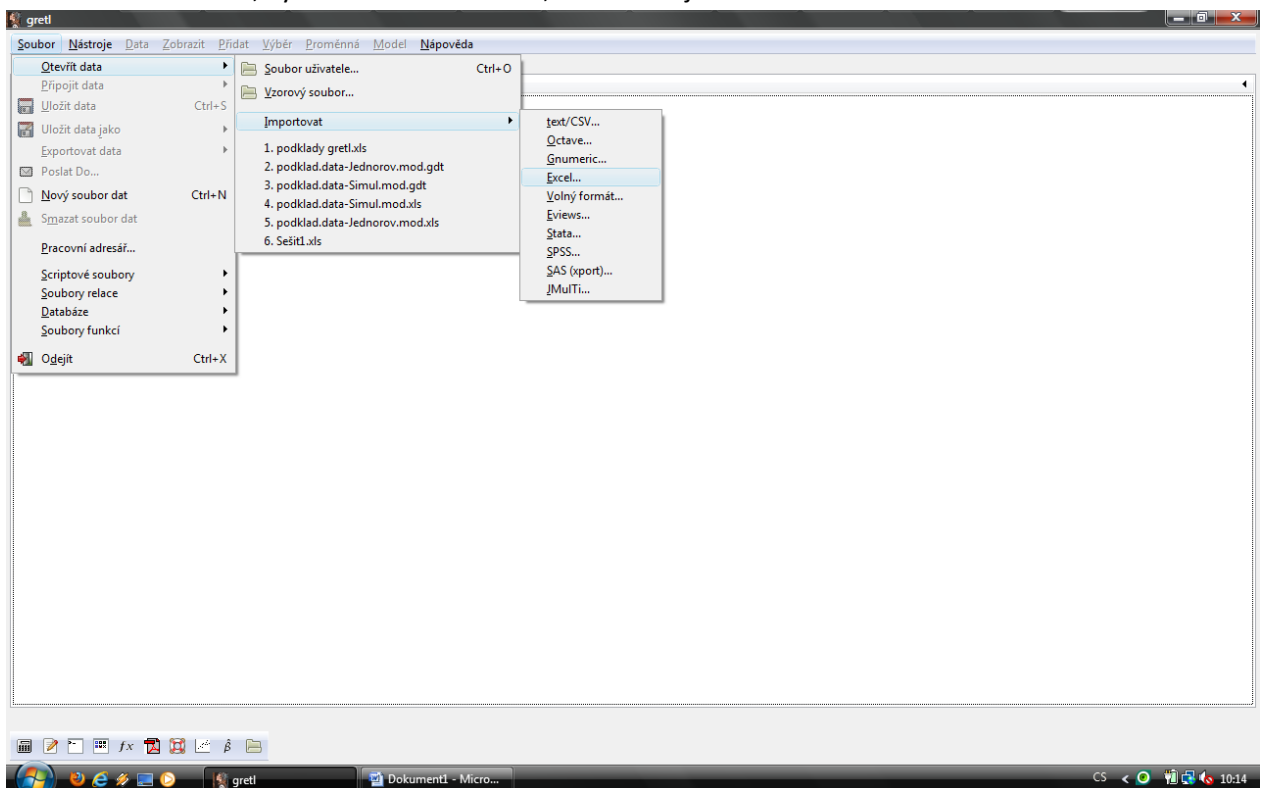
PEF ČZU Praha
Určeno pro posluchače
předmětu Ekonometrie

Needitovaná studijní pomůcka
MM2011

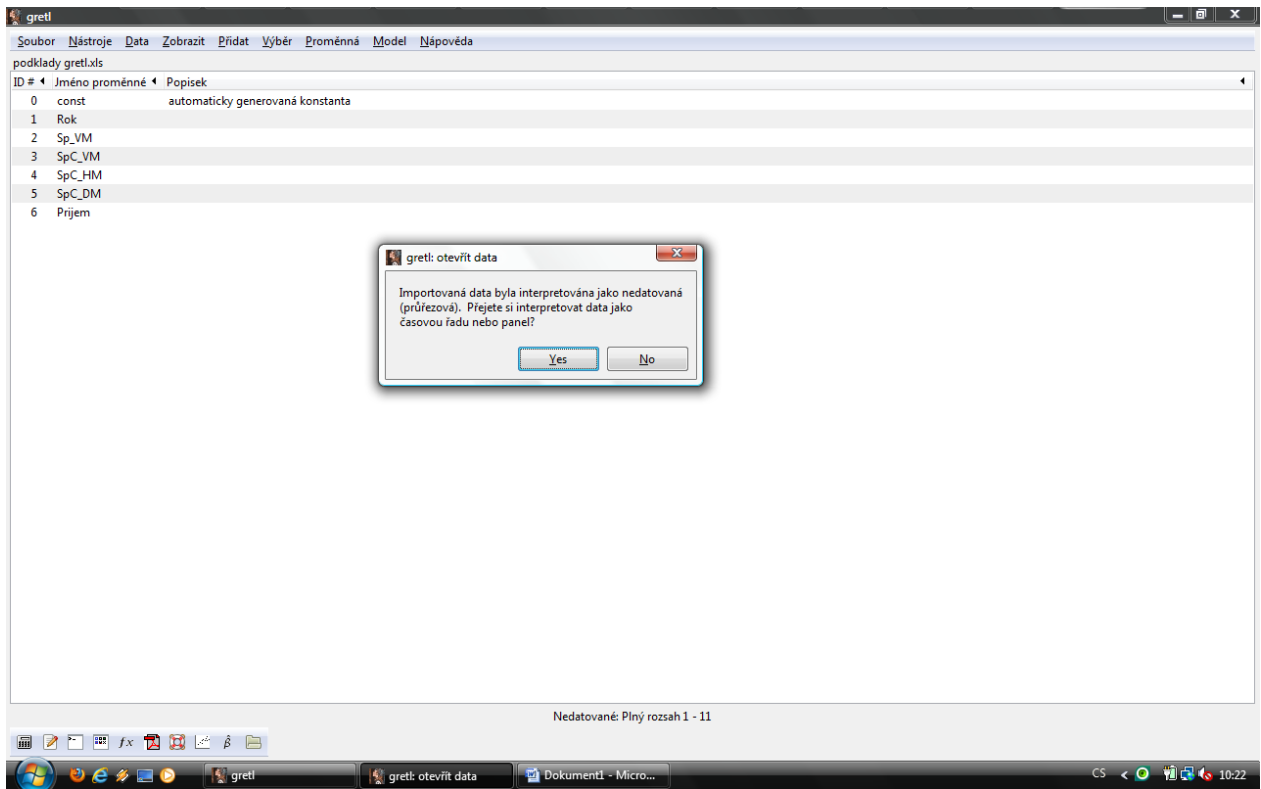
Úvodní obrazovka Gretlu po jeho instalaci a spuštění



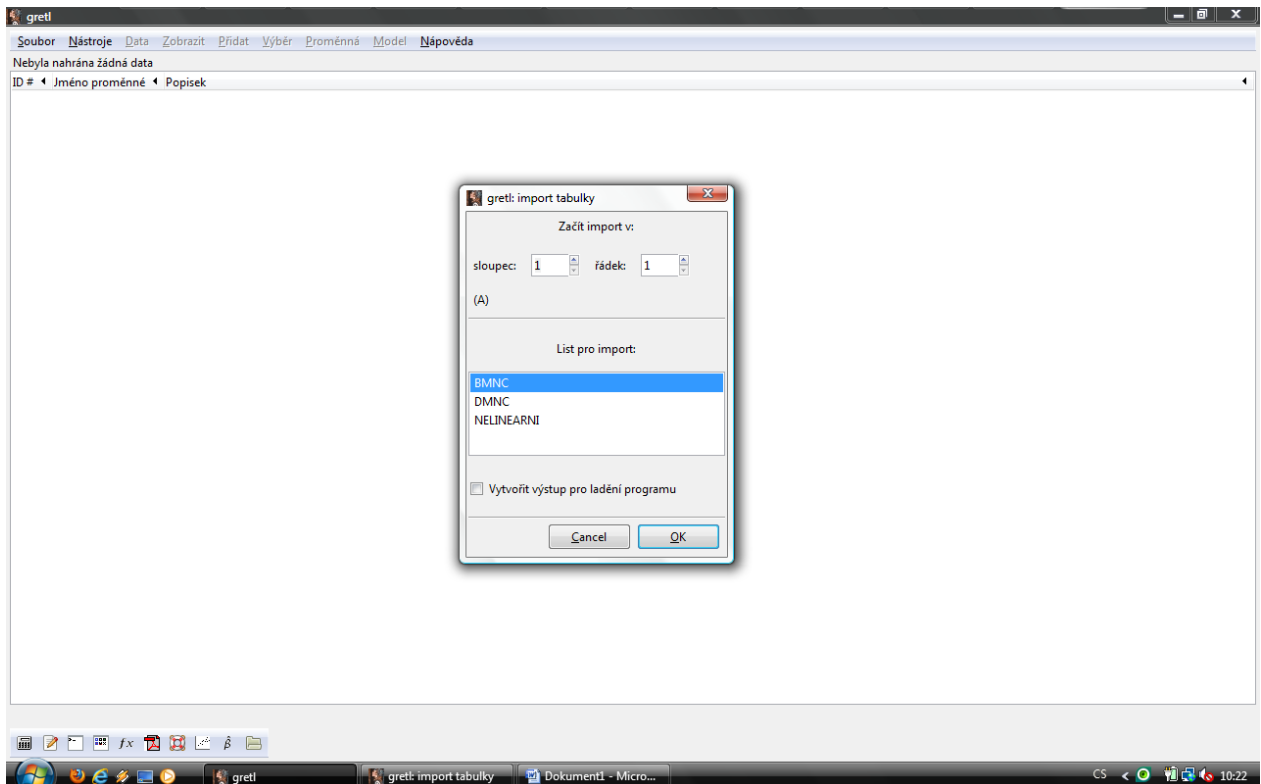
Nahrání dat – krok 1, výběr formátu souboru, ve kterém jsou data uložena



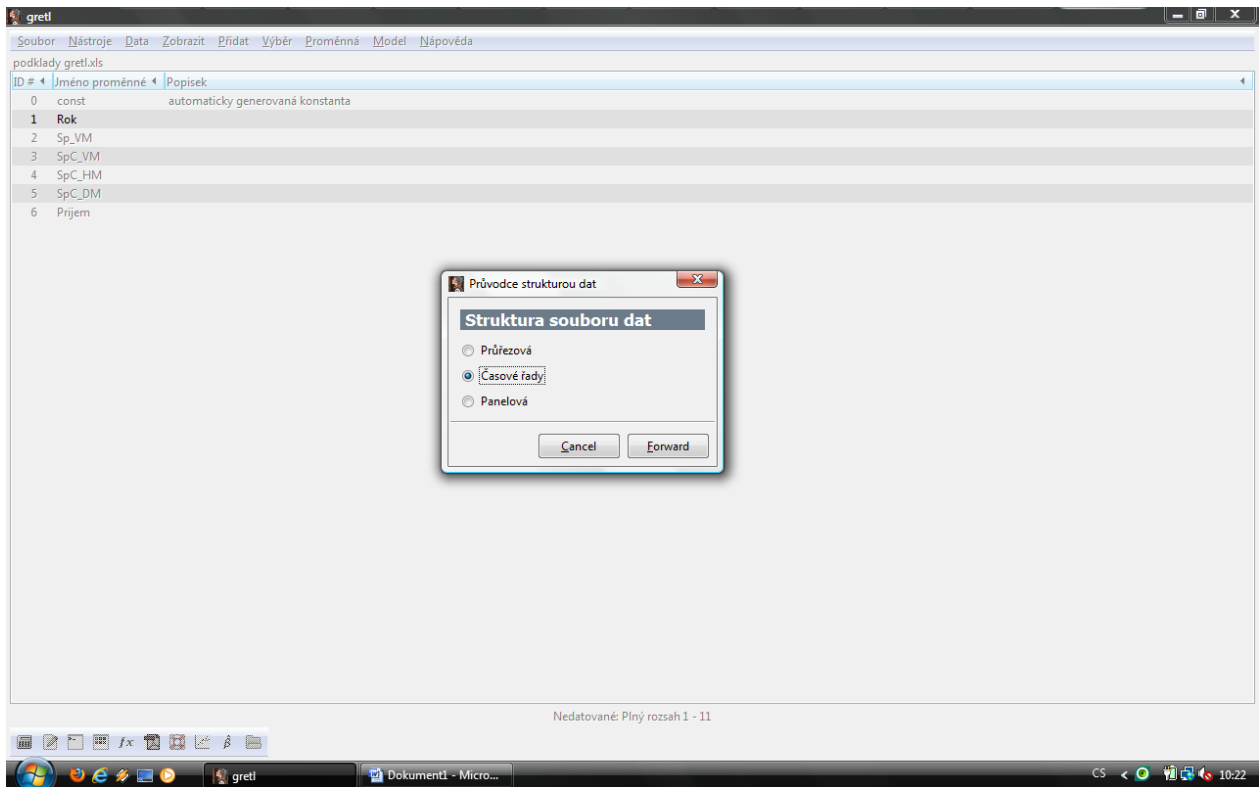
Nahrání dat – krok 2, dotaz na typ podkladových údajů, v našem případě volíme ANO (YES) - data jsou ve formě časových řad



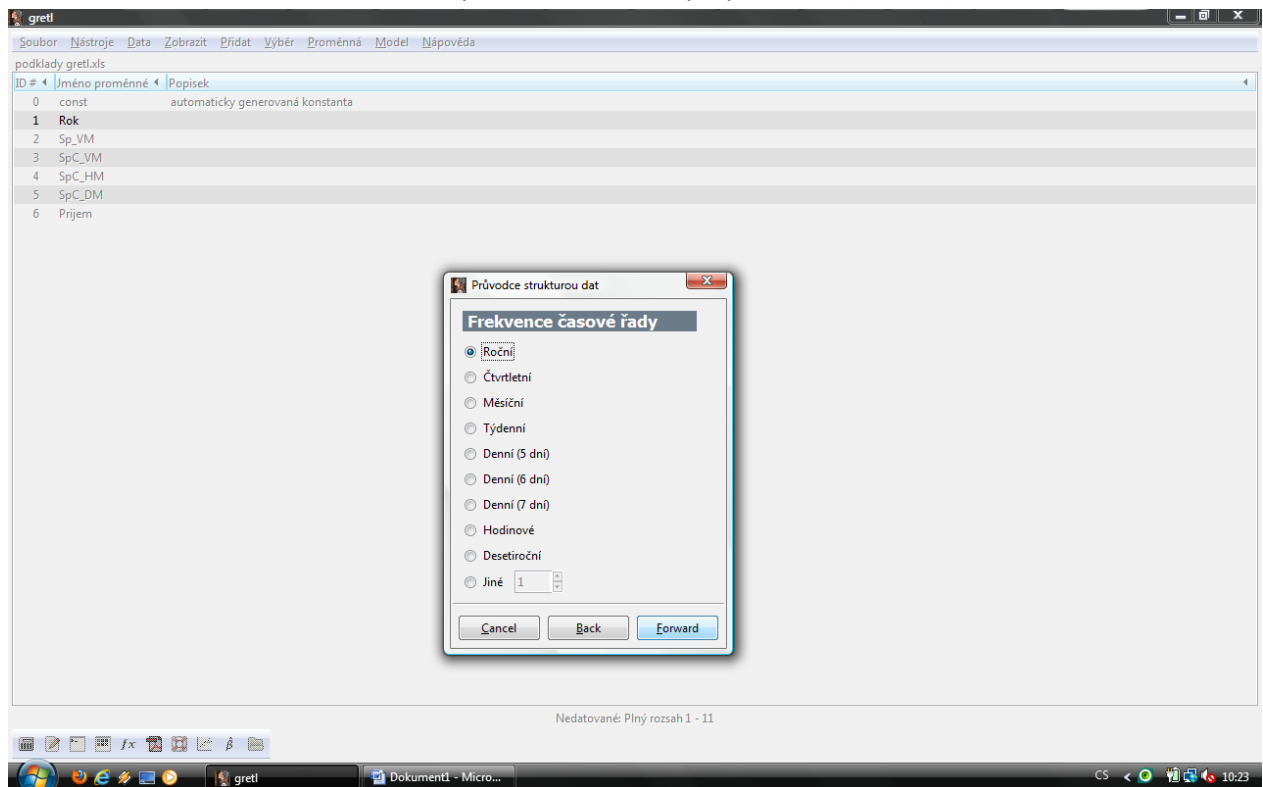
Nahrání dat – krok 3, volba listu z excelovského souboru na kterém jsou data a potvrzení (OK)



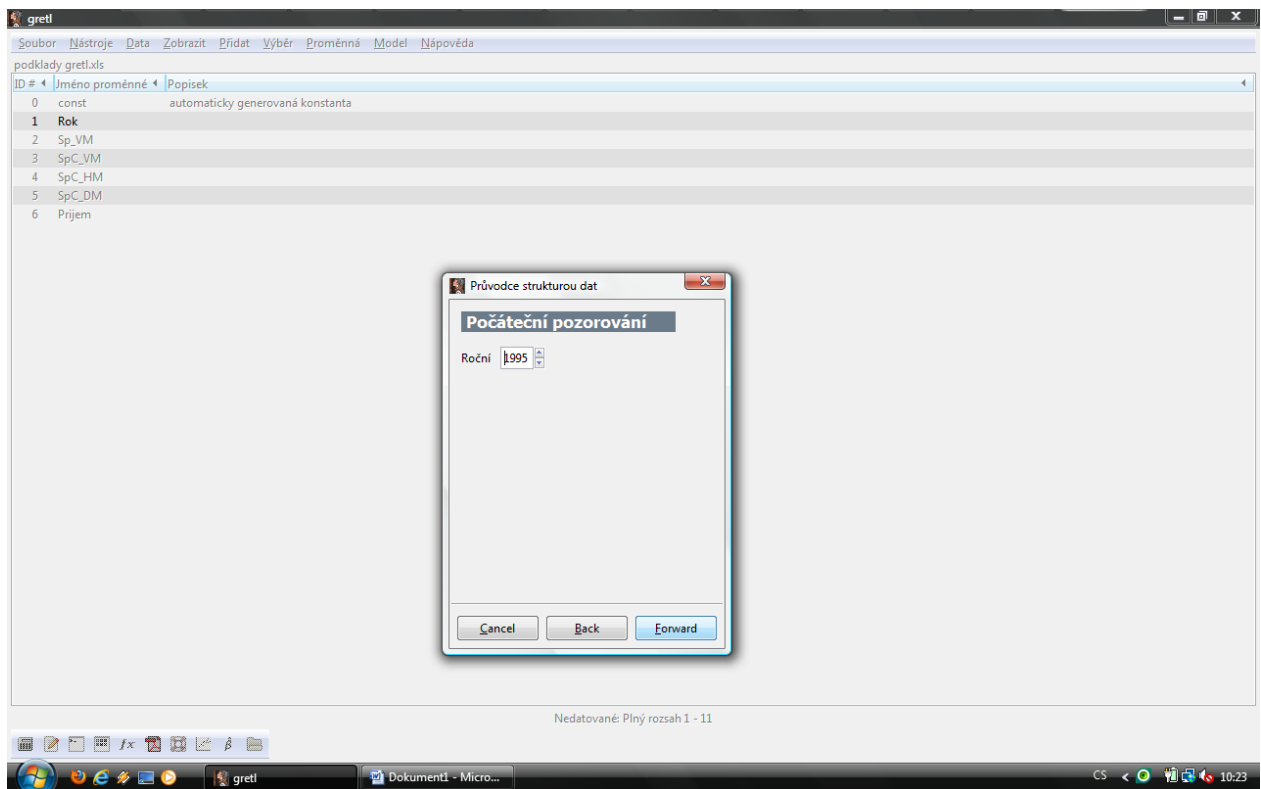
Nahrání dat – krok 4, dotaz na typ datových údajů, volíme časové řady a potvrzujeme krok dopředu (Forward)



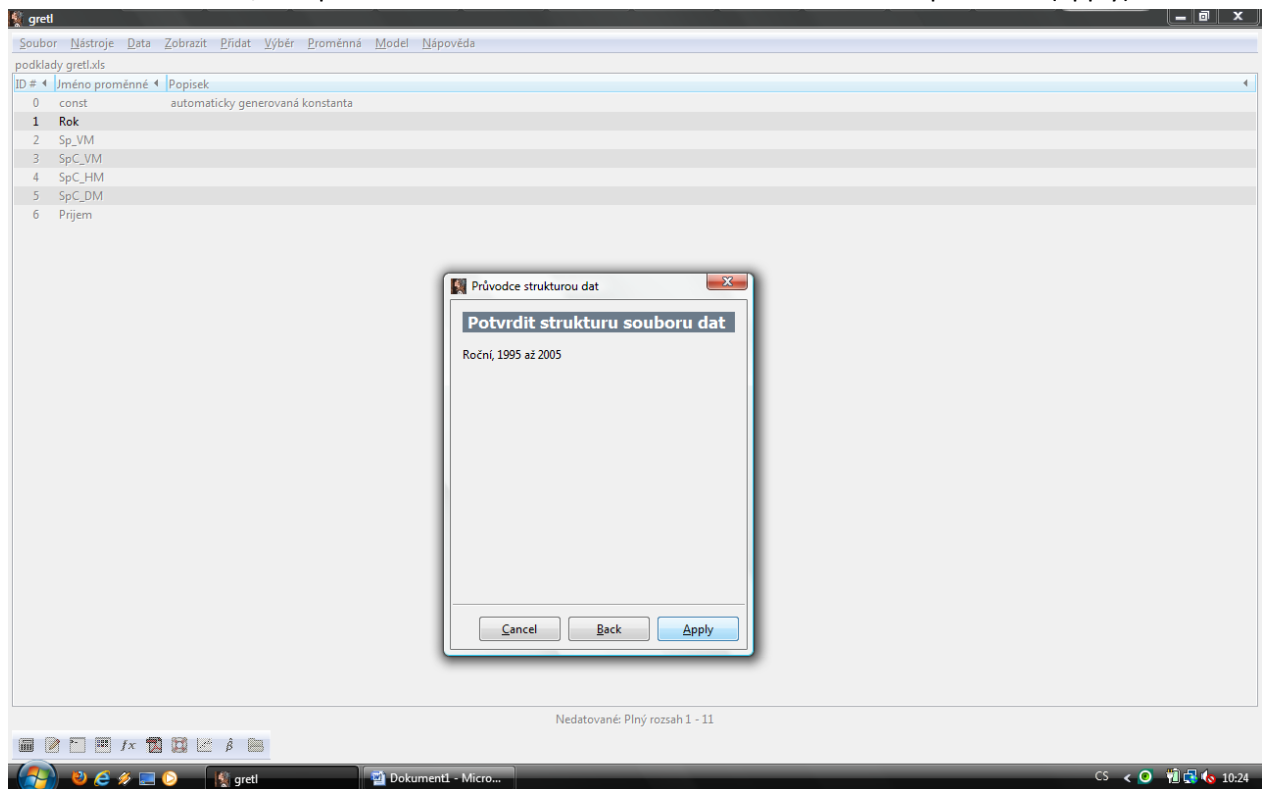
Nahrání dat – krok 5, volba frekvence použité časové řady + potvrzení (forward)



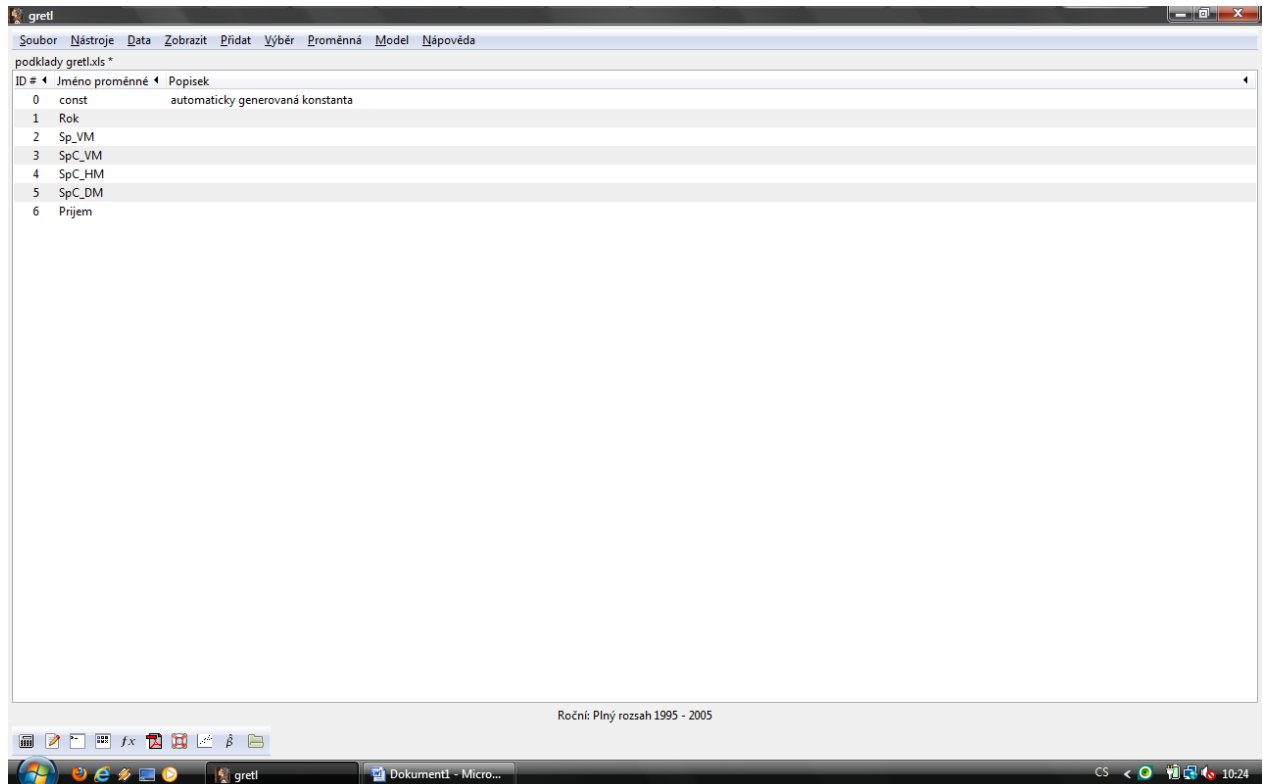
Nahrání dat – krok 6, volba počátečního období pozorování + potvrzení (forward)



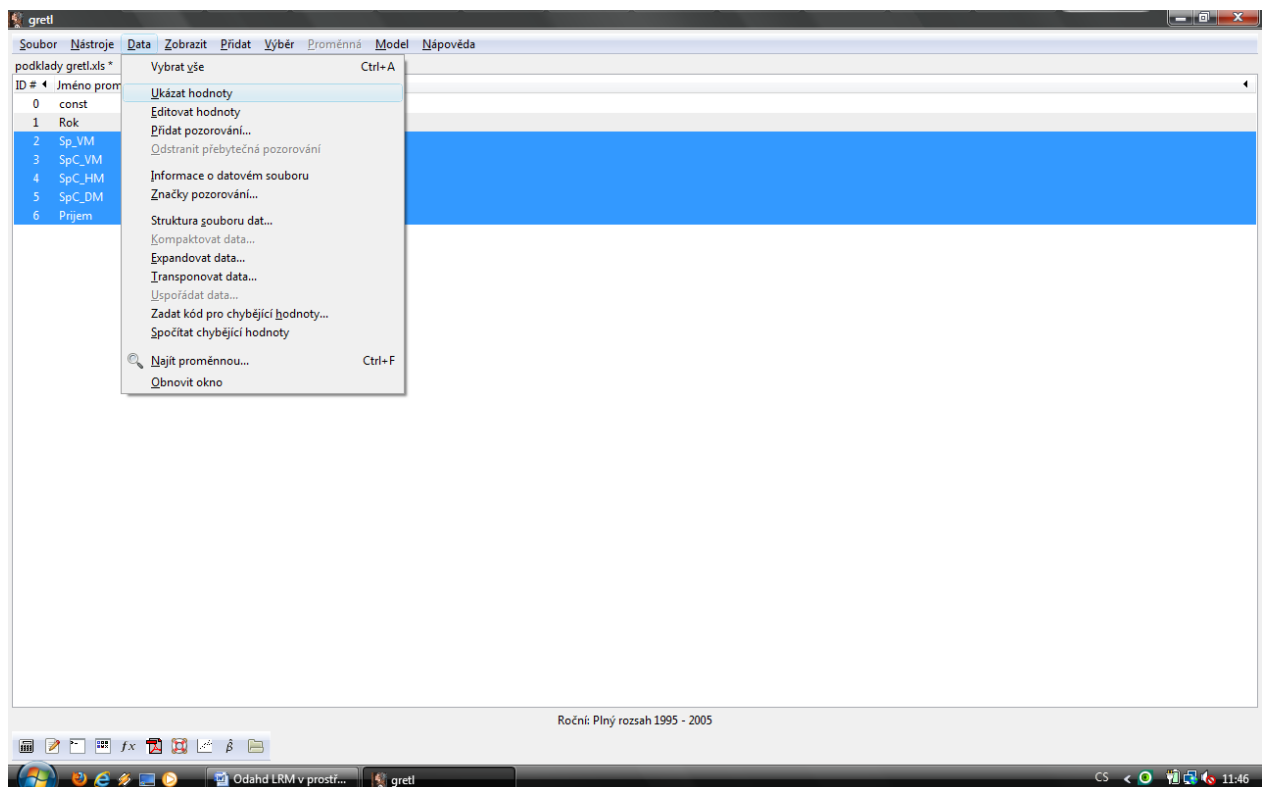
Nahrání dat – krok 7, rekapitulace začátku a konce načteného intervalu dat + potvrzení (apply)



Nahrání dat – krok 8, konečný stav základní obrazovky zobrazující nahrané proměnné



Možnost kontroly nahraných podkladových údajů – označíme vybrané proměnné a postupujeme přes kontextovou nabídku



Výsledné okno zobrazující nahraná data pro značené proměnné

The screenshot shows the gretl software interface. The main window displays a list of variables under the heading "podklady gretl.xls *". The variables are:

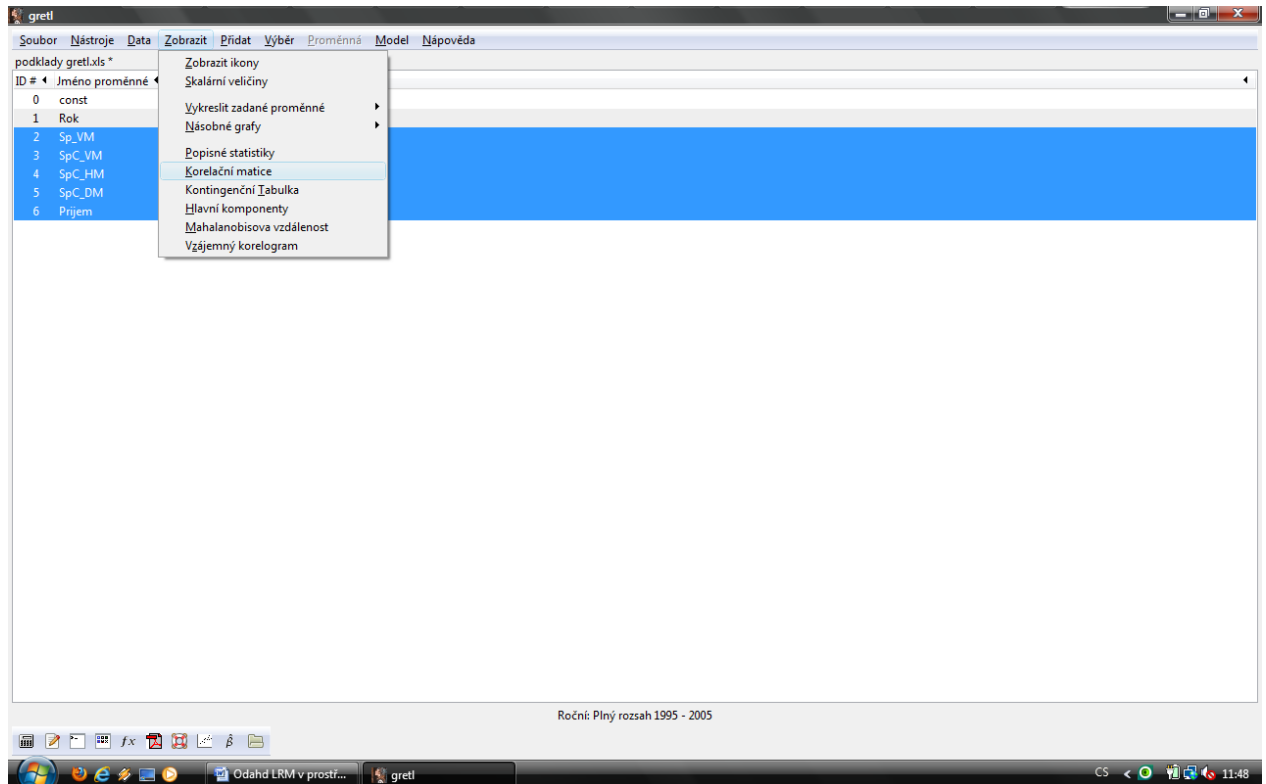
- 0 const (automaticky generovaná konstanta)
- 1 Rok
- 2 Sp_VM
- 3 SpC_VM
- 4 SpC_HM
- 5 SpC_DM
- 6 Prijem

A smaller window titled "gretl: ukázat data" is overlaid on the main window, displaying a table of data for the years 1995 to 2005. The table has the following columns: Sp_VM, SpC_VM, SpC_HM, SpC_DM, and Prijem.

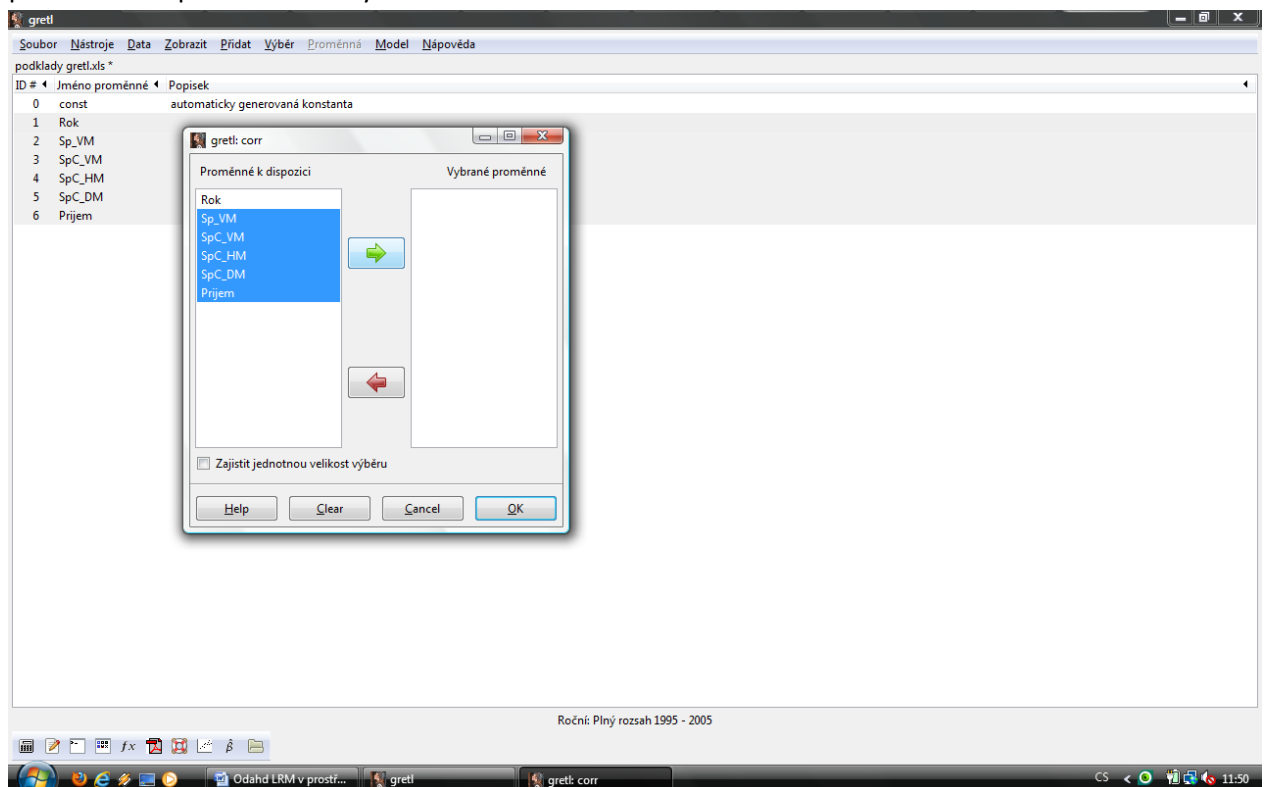
	Sp_VM	SpC_VM	SpC_HM	SpC_DM	Prijem
1995	8,04	84,20	94,81	52,32	55,5780
1996	8,87	90,42	102,12	62,77	64,1140
1997	8,74	92,11	104,82	70,64	70,9680
1998	10,36	86,39	110,16	73,31	77,9420
1999	9,78	80,47	107,80	56,51	80,7710
2000	8,94	90,04	111,53	61,83	83,4220
2001	9,05	101,66	112,56	71,28	90,1670
2002	9,55	89,84	112,99	62,40	93,1530
2003	10,14	82,74	108,02	60,67	98,1020
2004	9,97	85,36	112,84	62,55	102,2170
2005	11,18	85,30	117,73	62,73	116,5735

The status bar at the bottom of the gretl window indicates "Roční: Plný rozsah 1995 - 2005". The Windows taskbar at the bottom shows the system tray with the time 11:47.

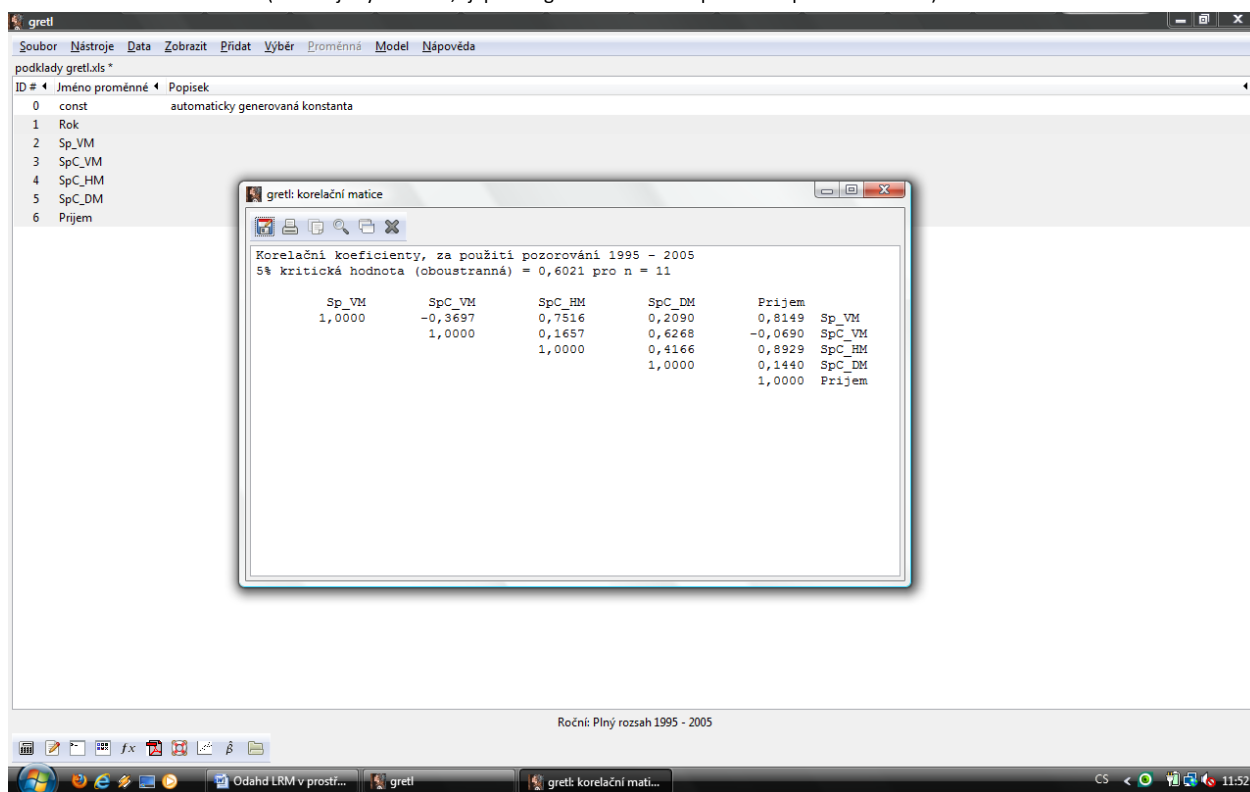
Zobrazení korelační matice přes kontextovou nabídku



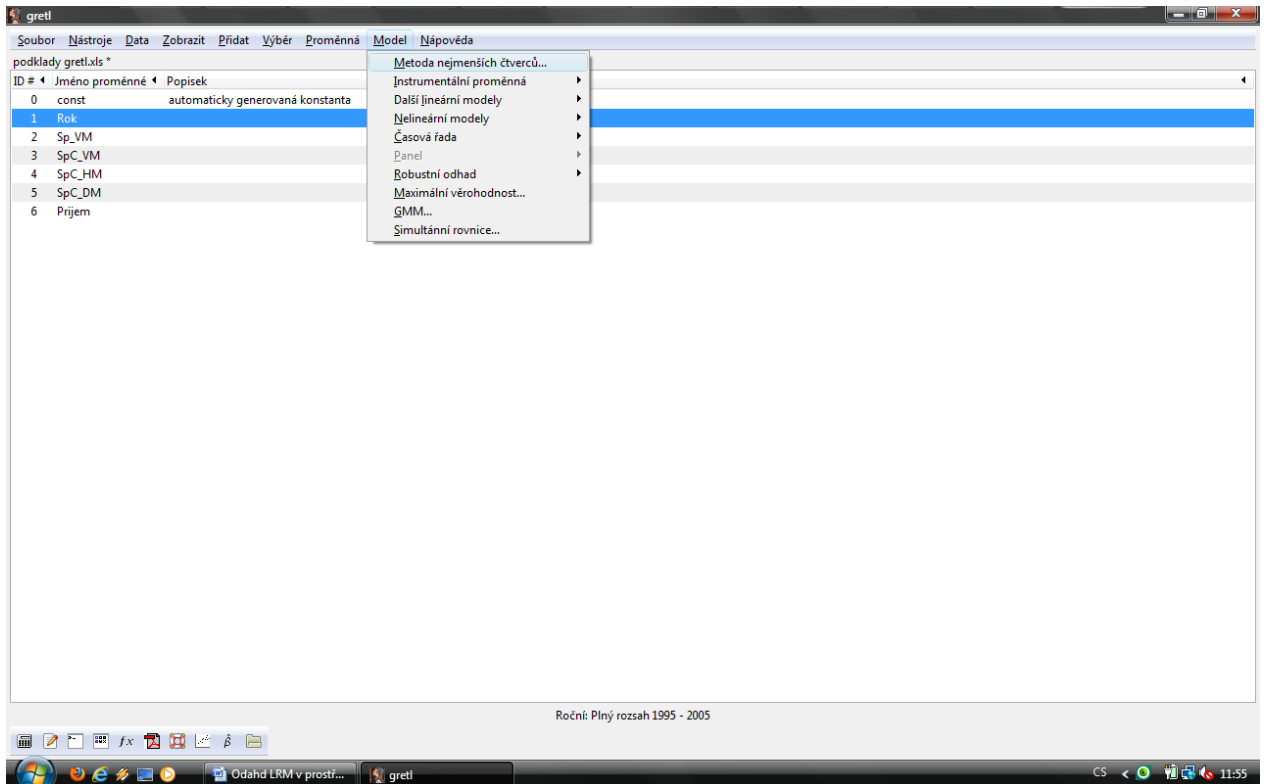
Výběr proměnných zahrnutých do korelační matice – pomocí zelené šipky vkládáme označené proměnné do pravého okna výběru



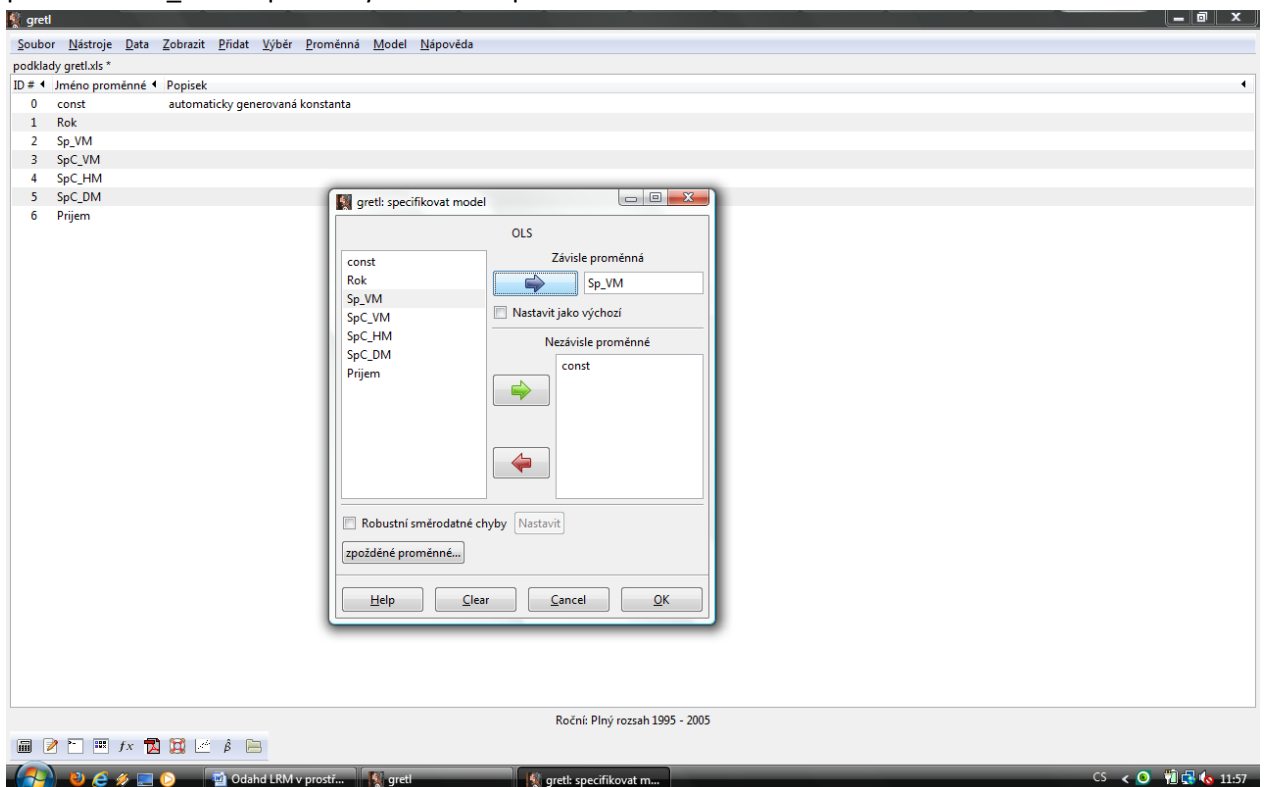
Výsledné zobrazení korelační matice – po potvrzení předchozí volby (OK) je zobrazena naddiagonální část korelační matice (matice je symetrická, tj. poddiagonální část lze explicitně doplnit zrcadlením)



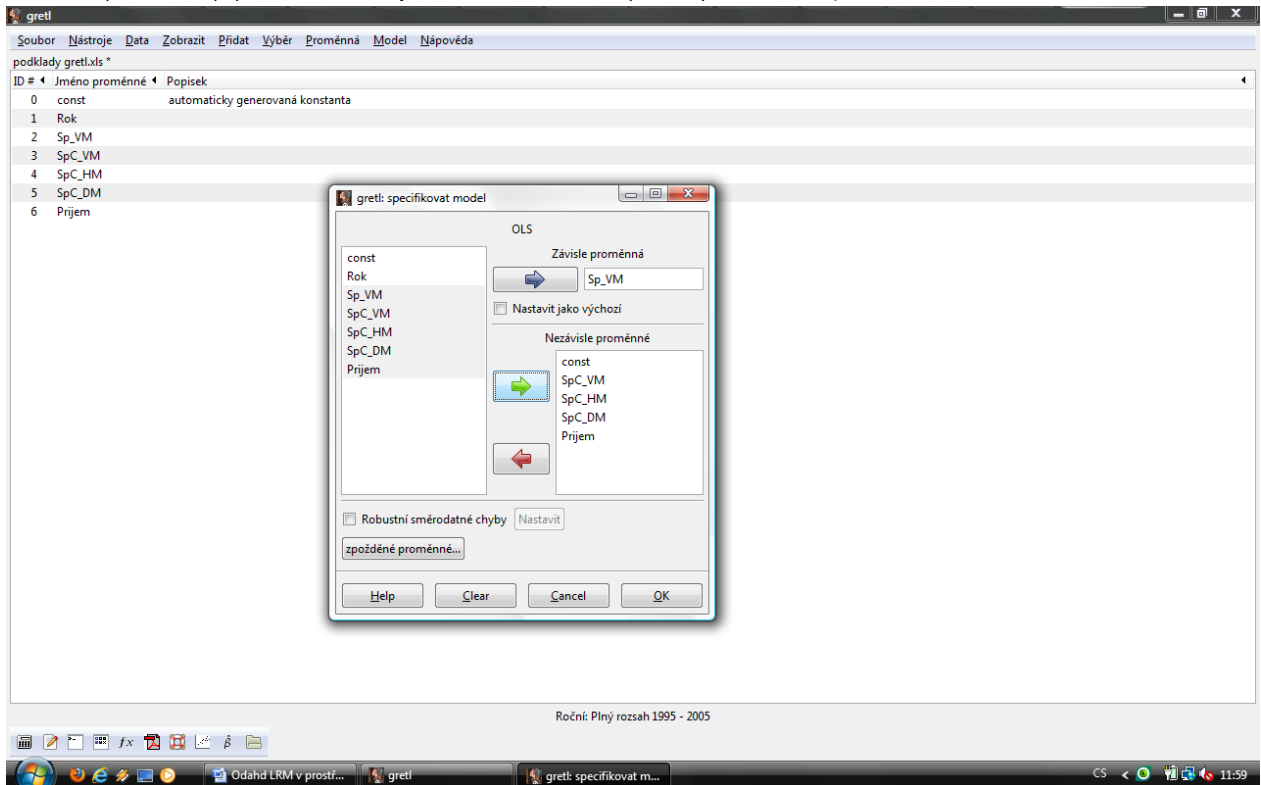
Odhad modelu metodou nejmenších čtverců – v kontextové nabídce volíme MNČ



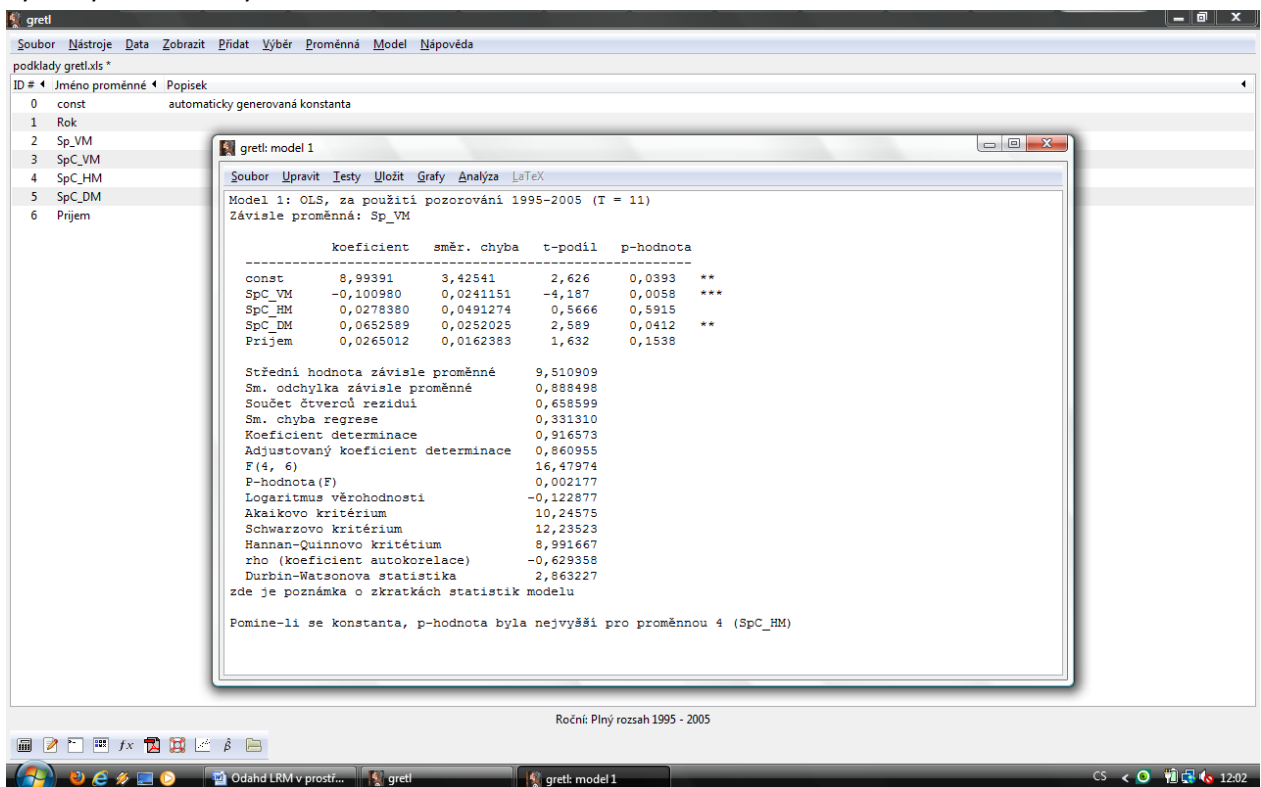
Výběr a rozdělení proměnných na vysvětlovanou (závislou) a vysvětlující (nezávislé), zde výběr proměnné SP_VM do pozice vysvětlované proměnné



Výběr a rozdělení proměnných, zde výběr dalších označených proměnných do pozice vysvětlujících proměnných; (Pozn.: Gretl automaticky ke každému výběru vysvětlujících vkládá jednotkový vektor (konstantu), kterou již nemusíme do modelu přidávat, resp. pokud chceme, lze ji označením a červenou šipkou z výběru odstranit)



Výsledný odhad – po potvrzení volby proměnných (OK) je již zobrazen kompletní odhad, včetně vybraných statistických vlastností odhadu



Verifikace modelu

gretl: model 1

Model 1: OLS, za použití pozorování 1995-2005 (T = 11)
Závisle proměnná: Sp_VM

	koeficient	směr. chyba	t-podíl	p-hodnota
const	8,99391	3,42541	2,626	0,0393 **
SpC_VM	-0,100980	0,0241151	-4,187	0,0058 ***
SpC_HM	0,0278380	0,0491274	0,5666	0,5915
SpC_DM	0,0652589	0,0252025	2,589	0,0412 **
Prijem	0,0265012	0,0162383	1,632	0,1538

Střední hodnota závisle proměnné 9,510909
 Sm. odchylka závisle proměnné 0,888498
 Součet čtverců reziduí 0,658599
 Sm. chyba regrese 0,331310
 Koeficient determinace 0,916573
 Adjustovaný koeficient determinace 0,860955
 F(4, 6) 16,47974
 P-hodnota(F) 0,002177
 Logaritmus věrohodnosti -0,122877
 Akaiikovo kritérium 10,24575
 Schwarzovo kritérium 12,23523
 Hannan-Quinnovo kritérium 8,991667
 rho (koeficient autokorelace) -0,629358
 Durbin-Watsonova statistika 2,863227

zde je poznámka o Zkratkách statistik modelu
 Pomine-li se konstanta, p-hodnota byla nejvyšší pro proměnnou 4 (SpC_HM)

K základní statistické verifikaci lze využít výstupů předchozího odhadu → p-hodnoty a koeficientu determinace.

P-hodnota informuje o hladině významnosti „ α “ na níž je zamítána nulová hypotéza (H_0) o statistické nevýznamnosti parametru (v Gretlu automaticky nastaveno $\alpha=0,05$). Obecně je-li *p-hodnota* menší než zvolená α , zamítáme nulovou hypotézu (H_0) o statistické nevýznamnosti parametru, tj. analyzovaný parametr je statisticky významný na dané hladině významnosti.

(1-(*p-hodnota*)) = pravděpodobnost statistické významnosti zkoumaného parametru)

(Pozn.: Zjednodušeně o průkaznosti parametru rovněž informují zobrazené hvězdičky za tabulkou odhadu. Čím více je zobrazeno hvězdiček (maximum je 3), tím vyšší je pravděpodobnost statistické významnosti parametru (* $\alpha=0,1$, ** $\alpha=0,05$, *** $\alpha=0,01$). Pokud nejsou u daného parametru žádné, tak parametr není při $\alpha=0,05$ statisticky významný.)

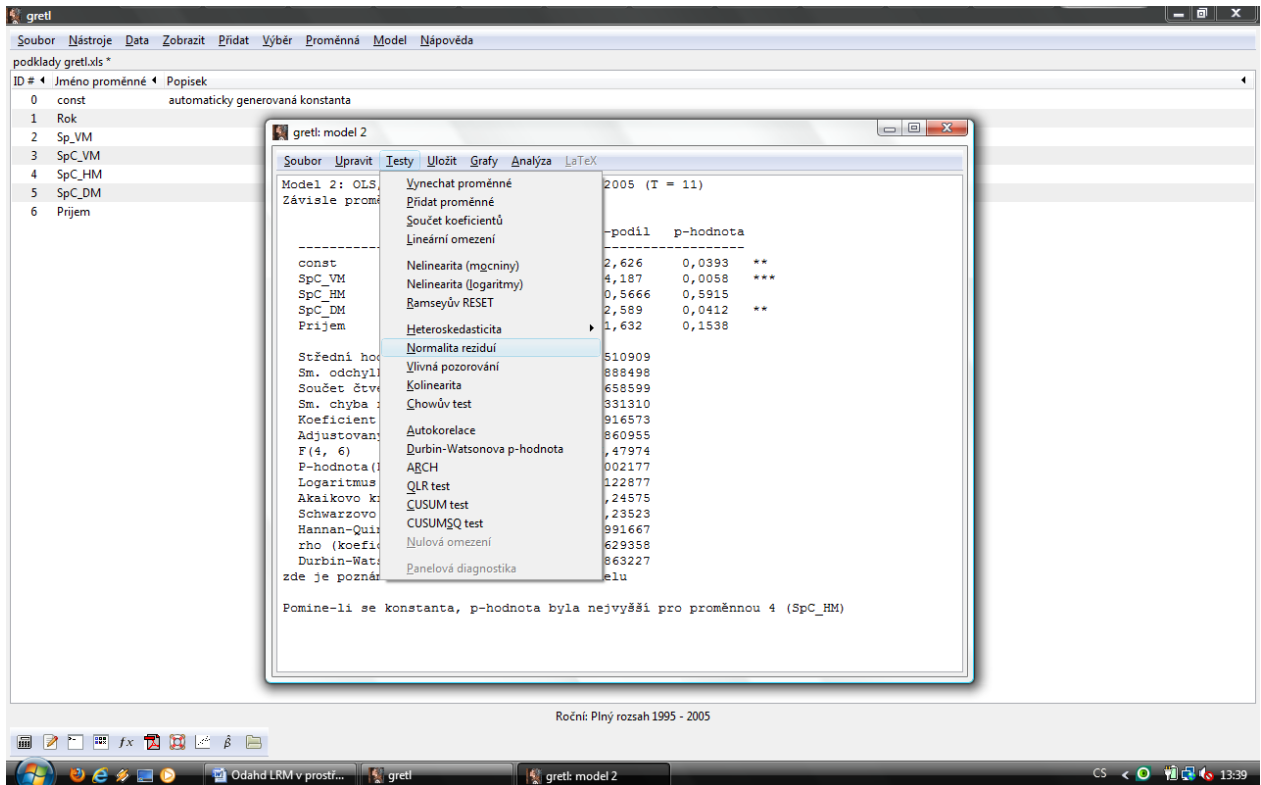
Koeficient determinace (R^2), resp. korigovaný (adjustovaný) koeficient determinace informuje o těsnosti závislosti. Výslednou hodnotu R^2 lze interpretovat v procentickém vyjádření, přičemž udává, z kolika procent jsou změny ve vysvětlované proměnné, závislé na změnách vysvětlujících proměnných.

Mezi další hodnocené ukazatele z prvotního odhadu lze doporučit Durbin-Watsonovu statistiku (DW), která slouží k testování přítomnosti autokorelace reziduí. Postup je následující:

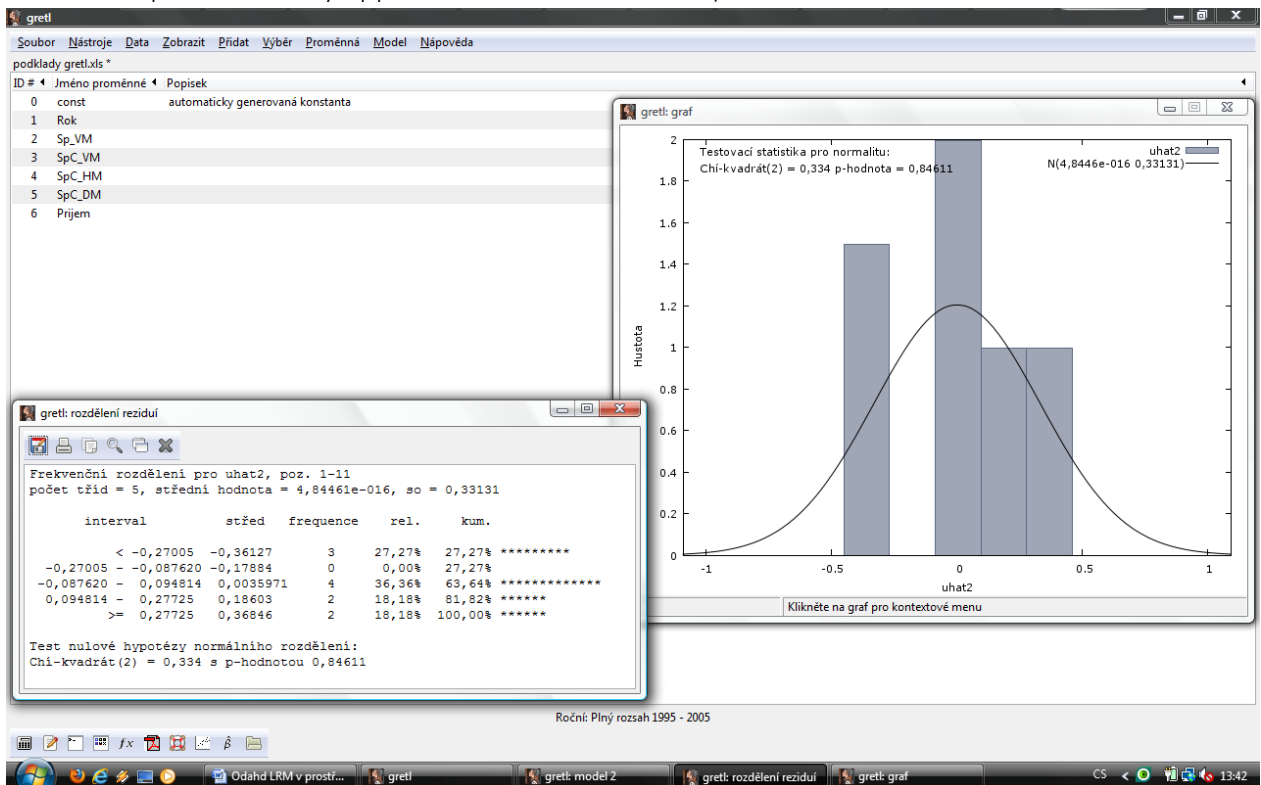
1. Statistika má střední hodnotu $E(d) = 2$ a nachází se v intervalu $<0;4>$,
2. Z příslušných tabulek stanovíme podle počtu stupňů volnosti tabulkové hodnoty d_b (dolní mez) a d_H (horní mez),
3. porovnáme naši hodnotu DW statistiky s následujícími intervaly a na základě její zařazení vyhodnotíme autokorelaci:
 - a. Interval $<0;d_b>$ označuje pozitivní autokorelaci
 - b. V intervalu $<d_b;d_H>$ nelze spolehlivě rozhodnout, zda se jedná o korelaci, či nikoliv = tzv. šedé pásmo
 - c. Interval $<d_H;2>$ poukazuje na statisticky nevýznamnou pozitivní autokorelaci
 - d. Interval $<2;4-d_H>$ poukazuje na statisticky nevýznamnou negativní autokorelaci
 - e. V intervalu $<4-d_H;4-d_b>$ opět nelze spolehlivě rozhodnout, zda se jedná o korelaci, či nikoliv = šedé pásmo
 - f. Interval $<4-d_b;4>$ poukazuje na statisticky významnou negativní autokorelaci

Mezi další postupy ekonometrické verifikace lze zařadit výběr následujících přídatných testů, které ověřují dodržení základních předpokladů LRM.

Test normality reziduí (Jargue–Bera test) – jeho výběr z kontextové nabídky odhadu



Výstup zobrazeného testu normality reziduí (Pozn.: zde provádí program řadu výpočtů a zobrazení, proto je nutné zachovat trpělivost a chvíli na výstup počkat – otázka několika desítek vteřin)

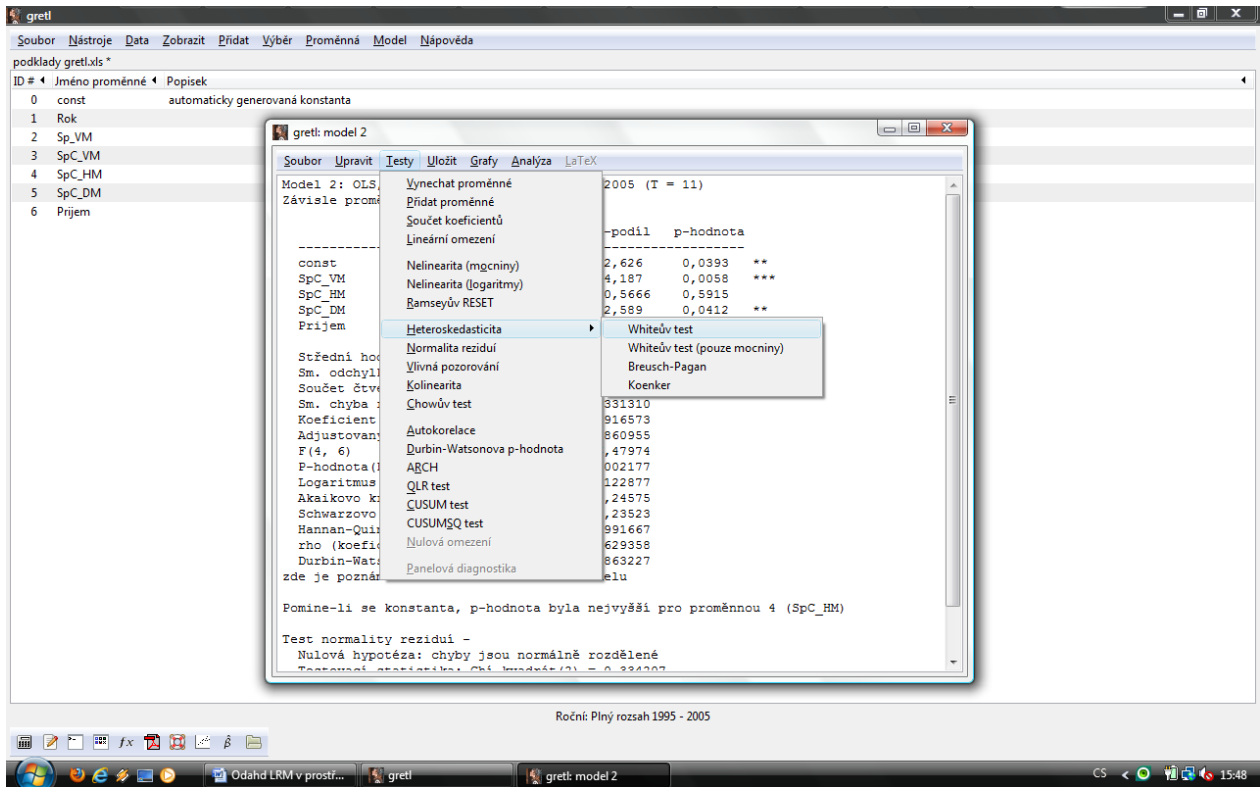


Vyhodnocení provedeného testu normality je pravděpodobně nejsnazší odvodit z průběhu grafu předpokládaného normálního rozdělení v porovnání se skutečným rozdělením reziduí a analýzou p-hodnoty Chí-kvadrát testu.

H_0 : Rezidua mají normální rozdělení, tj. nulovou střední hodnotu a konstantní rozptyl

p-hodnota $\text{vypočtená} = 0,84611 > \text{zvolené } \alpha = 0,05 \rightarrow H_0 \text{ nelze zamítnout} \rightarrow \text{normalita reziduí}$

Test heteroskedasticity (White test) – jeho výběr z kontextové nabídky



Pozn.: Alternativně lze samozřejmě zvolit i Breusch-Pagan test s analogickým vyhodnocením.

Výstup zvoleného testu (White test)

gretl: LM test (heteroskedasticita)

Whiteův test heteroskedasticity
 OLS, za použití pozorování 1995–2005 (T = 11)
 Závisele proměnná: uhat^2

	koeficient	směr. chyba	t-podíl	p-hodnota
const	-63,3264	63,9165	-0,9908	0,4262
SpC_VM	0,345603	0,409385	0,8442	0,4874
SpC_HM	1,18496	1,26559	0,9363	0,4480
SpC_DM	-0,387734	0,605693	-0,6401	0,5876
Prijem	-0,104414	0,0869121	-1,201	0,3526
sq_SpC_VM	-0,00179546	0,00212838	-0,8436	0,4877
sq_SpC_HM	-0,00545818	0,00585143	-0,9328	0,4494
sq_SpC_DM	0,00286773	0,00451613	0,6350	0,5904
sq_Prijem	0,000617781	0,000534902	1,155	0,3675

Varování: matice dat je téměř singulární!
 Neadjustovaný koeficient determinace = 0,818386
 Testovací statistika: $TR^2 = 9,002243$,
 s p-hodnotou = $P(\text{Chi-kvadrát}(8) > 9,002243) = 0,342107$

Pomíne-li se konstanta, p-hodnota byla nejvyšší pro proměnnou 4 (SpC_HM)

Test normality reziduí -
 Nulová hypotéza: chyby jsou normálně rozdělené
 Testovací statistika: Chi-kvadrát(2) = 0,342107

Vyhodnocení lze provést opět na základě odvozené p-hodnoty.

H_0 : Homoskedasticita (tj. konstantní rozptyl rezidua)

P-hodnota = 0,342107 > $\alpha = 0,05 \rightarrow H_0$ nelze zamítnout \rightarrow potvrzení homoskedasticity

Test autokorelace (Breusch – Godfrey test); alternativa (rozšíření či ověření) k DW testu – výběr z kontextu

gretl: model 2

Model 2: OLS
 Závisele proměnná: uhat^2
 2005 (T = 11)

		-podíl	p-hodnota
const	Nelinearita (mecniny)	2,626	0,0393 **
SpC_VM	Nelinearita (logaritmy)	4,187	0,0058 ***
SpC_HM	Ramseyův RESET	0,5666	0,5915
SpC_DM	Heteroskedasticita	2,589	0,0412 **
Prijem	Heteroskedasticita	1,632	0,1538
Střední hodnota	Normalita reziduí	510909	
Sm. odchylka	Vlivná pozorování	888498	
Součet čtverců	Kolinearita	658599	
Sm. chyba	Chowův test	31310	
Koeficient	Autokorelace	860955	
Adjustovaný R	Durbin-Watsonova p-hodnota	4,47974	
F(4, 6)	ARCH	002177	
P-hodnota (F)	QLR test	122877	
Logaritmus	CUSUM test	2,24575	
Akaikevo kritérium	CUSUMSQ test	2,23523	
Schwarzovo kritérium	Nulová omezení	991667	
Hannan-Quinnovo kritérium	Panelová diagnostika	629358	
rho (koeficient)		863227	
Durbin-Watsonův test		0,1538	

Varování: matice dat je téměř singulární!
 Pomíne-li se konstanta, p-hodnota byla nejvyšší pro proměnnou 4 (SpC_HM)

Test normality reziduí -
 Nulová hypotéza: chyby jsou normálně rozdělené
 Testovací statistika: Chi-kvadrát(2) = 0,342107

Test autokorelace – volba zpoždění (Pozn.: testujeme zda u_t je závislé na $u_{(t-1)}$, proto volíme zpoždění „1“)

The screenshot shows the gretl interface with a regression model window open. The model is OLS for the period 1995-2005 (T=11) with dependent variable Sp_VM. A dialog box titled 'gretl: autokorelace' is open, showing 'Řád zpoždění pro test:' set to 1. The regression results table is as follows:

	koeficient	směr. chyba	t-podíl	p-hodnota
const	8,99391	3,42541	2,626	0,0393 **
SpC_VM	-0,100980	0,0241151	-4,187	0,0058 ***
SpC_HM	0,0278380	0,0491274	0,5666	0,5915
SpC_DM	0,0652589	0,0252025	2,589	0,0412 **
Prijem	0,0265012	0,0162383	1,632	0,1538

Below the table, the text indicates that the constant term's p-value is the lowest for variable 4 (SpC_HM). The test normality of residuals is also shown.

Test autokorelace – výstup

The screenshot shows the gretl interface with the 'gretl: LM test (autokorelace)' window open. The test is for the first lag of the residuals from the OLS model. The output is as follows:

	koeficient	směr. chyba	t-podíl	p-hodnota
const	0,191294	2,80595	0,06817	0,9483
SpC_VM	-1,64047e-05	0,0197425	-0,0008309	0,9994
SpC_HM	0,00986173	0,0405242	0,2434	0,8174
SpC_DM	-0,0107119	0,0213246	-0,5023	0,6368
Prijem	-0,00717992	0,0137758	-0,5212	0,6245
uhat_1	-0,850025	0,427578	-1,988	0,1035

The adjusted coefficient of determination is 0,441474. The Breusch-Godfrey test statistics are: LMF = 3,952138, p-hodnotou = P(F(1,5) > 3,95214) = 0,104. Alternative statistics: TR^2 = 4,856216, p-hodnotou = P(Chi-kvadrát(1) > 4,85622) = 0,0275. Ljung-Box Q' = 3,8573, p-hodnotou = P(Chi-kvadrát(1) > 3,8573) = 0,0495.

Vyhodnocení:

H_0 : Nepřítomnost autokorelace reziduí (časové řady jsou stacionární)

P-hodnota = 0,104 > $\alpha = 0,05 \Rightarrow H_0$ nelze zamítnout \Rightarrow nepřítomnost autokorelace prvního řádu