



ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Программа «Фонд образовательный инноваций»

**«ЭКОНОМЕТРИЧЕСКОЕ
МОДЕЛИРОВАНИЕ В ПАКЕТЕ STATA:
СОЗДАНИЕ
ПРОГРАММНЫХ КОДОВ»**

(ПРАКТИКУМ ПО ЭКОНОМЕТРИКЕ-2)

Вакуленко Е.С. и Ратникова Т.А.

Кафедра математической экономики и эконометрики

Москва, 2011

ПРЕАМБУЛА

Практикум разработан для магистрантов факультета экономики всех направлений подготовки

Сведения об авторах:

- **Ратникова Т.А.** - к.ф.-м.н., доцент
кафедры математической экономики и эконометрики
- **Вакуленко Е.С.** - преподаватель и аспирант
кафедры математической экономики и эконометрики,

*При использовании материалов
практикума ссылка на разработчиков
обязательна*



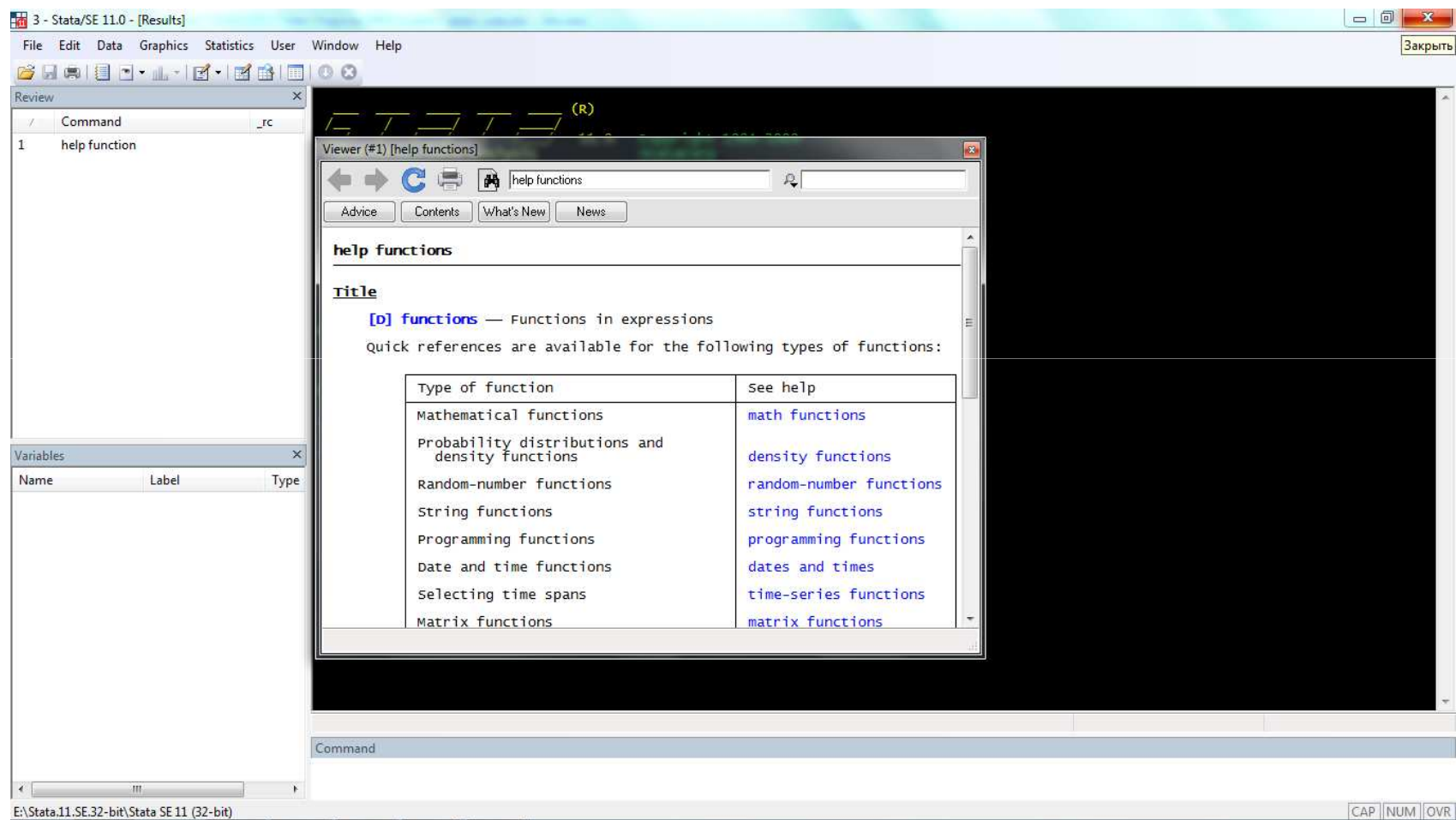


НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

ПРОГРАММА КУРСА

1. Введение: пакет статистической обработки данных STATA
2. Работа с массивами
3. Предварительный анализ данных
4. Генерирование случайных выборок
5. Оценивание линейных регрессионных моделей
6. Анализ структурной стабильности регрессионной модели
7. Квазимультиколлинеарность в регрессионных моделях
8. Ошибки спецификации и выбор оптимального набора регрессоров
9. Выбор функциональной формы регрессии
10. Обобщенная линейная регрессия: гетероскедастичность
11. Обобщенная линейная регрессия: автокорреляция
12. Эндогенность в регрессионных моделях
13. Анализ панельных данных
13. Метод максимального правдоподобия
14. Оценивание моделей бинарного и множественного выбора
15. Оценивание моделей по усеченным и цензурированным данным
16. Графическое представление результатов: вывод карт и графиков

ПАКЕТ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ STATA



ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ДАННЫХ

The screenshot displays the Stata/SE 11.0 software interface. The main window is the 'Do-file Editor - sem3', which contains a script for data analysis. The script includes the following commands and comments:

```
1 version 11.0
2 set more off
3 log using sem3.log, text replace
4
5 /* *****/
6 /* File Name: sem3.do */
7 /* Date: November, 2011 */
8 /* Authors: E. Vakulenko, T. Ratnikova */
9 /* Purpose: Summary statistics */
10 /* Input File: mgw.dta */
11 /* Output File: sem3.log */
12 /* *****/
13
14 /*Открытие файла с данными о стоимости канадских домов из учебника М.Вербика */
15 use "D:\housing.dta", clear
16
17 /*Вывод описательных статистик*/
18 summarize
19
20 /*Вывод детализированных описательных статистик*/
21 summarize, detail
22
23 /*Построение графиков*/
24 twoway dot price lotsize
25
26 /*Построение диаграмм рассеивания*/
27 twoway scatter price lotsize
28
29 /* Изучение распределений количественных показателей */
30 /* Построение гистограммы и сглаженной плотности распределения */
31 twoway hist price || kdensity price
32 twoway hist lotsize || kdensity lotsize
33
34 log close
```

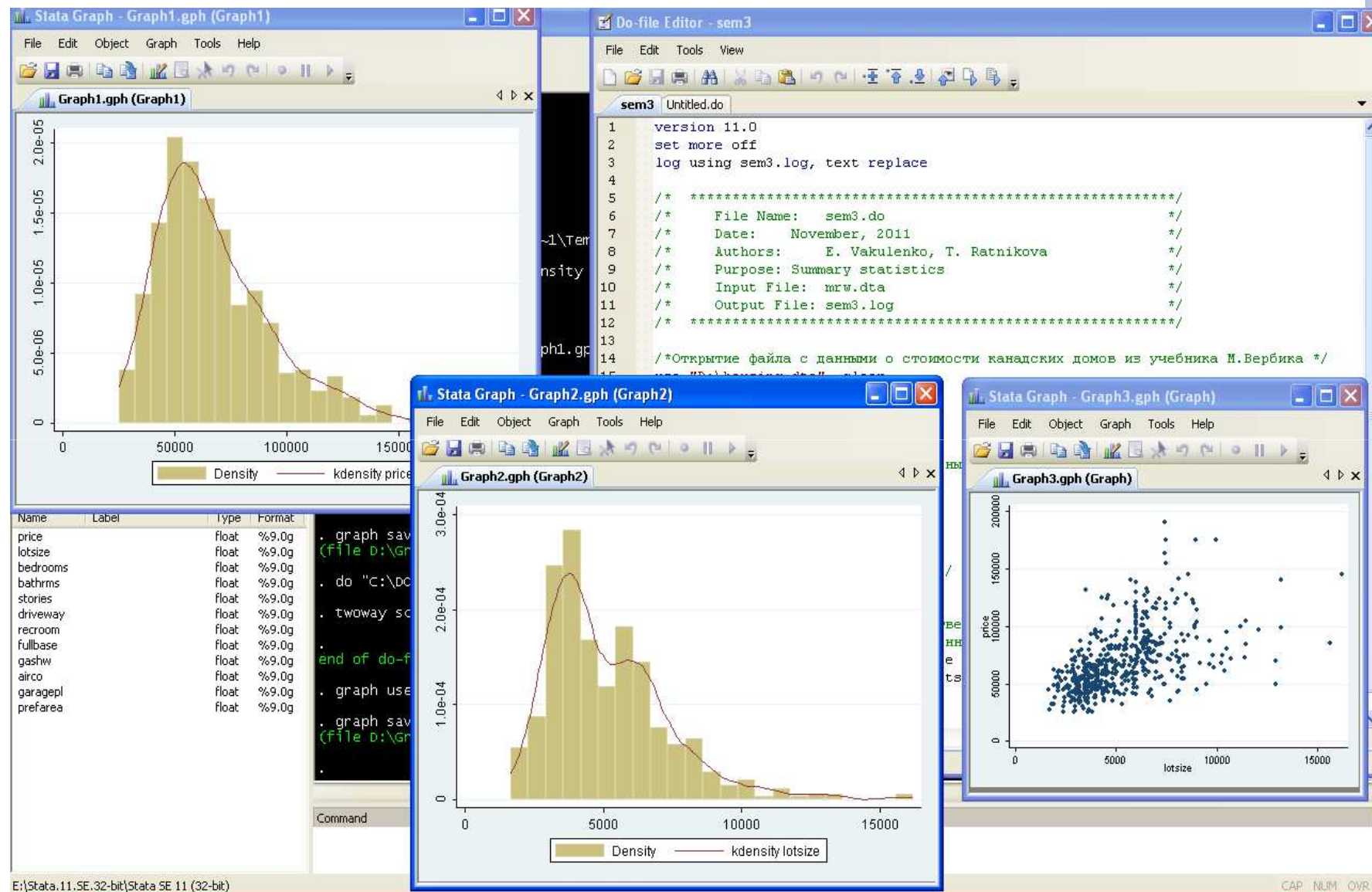
The 'Review' window on the left shows the command 'use "D:\housing.dta", clear' at line 1. The 'Variables' window at the bottom left lists the following variables:

Name	Label	Type	Format
price		float	%9.0g
lotsize		float	%9.0g
bedrooms		float	%9.0g
bathrms		float	%9.0g
stories		float	%9.0g
driveway		float	%9.0g
recroom		float	%9.0g
fullbase		float	%9.0g
gashw		float	%9.0g
airco		float	%9.0g
garagepl		float	%9.0g
prefarea		float	%9.0g

The status bar at the bottom indicates 'Line: 1, Col: 0' and 'CAP NUM OVR'.

Ратникова Т.А., Вакулenco Е.С., кафедра математической экономики и эконометрики НИУ-ВШЭ, Москва

ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ДАННЫХ



ГЕНЕРИРОВАНИЕ СЛУЧАЙНЫХ ВЕЛИЧИН

The image shows two windows from a Stata 11.0 environment. The left window is the Stata main interface, and the right window is a Do-file Editor.

Stata Main Interface:

- Review Window:** Shows a list of commands entered in the command window.
- Command Window:** Contains the following commands:

```
. drawnorm x1 x2, means(M) cov(C) n(150)
(obs 150)
. scatter x1 x2
. /*Генерирование стандартной нормал
. drawnorm epsilon
. hist epsilon
(bin=12, start=-2.5367177, width=.40481675)
. twoway kdensity epsilon
. /* Генерирование зависимой перемен
. generat y = 2 + 5*x1 - 3*x2 + epsilon
. summarize y x1 x2
```
- Output:** Displays summary statistics for variables y, x1, and x2:

Variable	obs	Mean	Std. Dev.	
y	150	-3.253932	6.119735	-1
x1	150	1.990482	.9715926	-.
x2	150	5.088496	1.058647	2
- Command Window:** Shows the following commands:

```
. correlate y x1 x2
(obs=150)
```
- Output:** Displays the correlation matrix:

	y	x1	x2
y	1.0000		
x1	0.8549	1.0000	
x2	-0.5948	-0.1212	1.0000
- Command Window:** Shows the command:

```
. end of do-file
```

Do-file Editor (Untitled1.do):

```
1 version 11.0
2 set more off
3 log using sem4.log, text replace
4
5 /* *****/
6 /* File Name: sem4.do */
7 /* Date: November, 2011 */
8 /* Authors: E. Vakulenko, T. Ratnikova */
9 /* Purpose: Regression data generation with random variables*/
10 /* Output File: sem4.log */
11 /* *****/
12
13 /*Создание вектора средних и ковариационной матрицы */
14 matrix M = (2 , 5)
15 matrix C = (1 , 0 \ 0 , 1)
16 set seed 100
17 drawnorm x1 x2, means(M) cov(C) n(150)
18 scatter x1 x2
19
20 /*Генерирование стандартной нормальной случайной величины */
21
22 drawnorm epsilon
23 hist epsilon
24 twoway kdensity epsilon
25
26 /* Генерирование зависимой переменной */
27
28 generat y = 2 + 5*x1 - 3*x2 + epsilon
29 summarize y x1 x2
30 correlate y x1 x2
31
32 clear
33 /*Генерирование случайных величин разных распределений*/
34 matrix A = (0 , 0, 0, 0, 0, 0, 0, 0, 0, 0)
35 matrix B = I(10) /*Создание единичной матрицы размера 10*/
36 drawnorm z1-z10, means(A) cov(B) n(200)
```

ОЦЕНИВАНИЕ РЕГРЕССИОННЫХ МОДЕЛЕЙ

The screenshot displays the Stata 11.0 interface with two windows open: the main Stata window and a Do-file Editor.

Stata Window:

- Command Window:** Shows the command `. reg lngdp85 lnngdelta lninv if intermed==1`.
- Review Window:** Lists commands from line 13 to 26, including `do "C:\Users\user\AppData\Lo...", reg lngdp85 lnngdelta lninv if ...`, and `shellout using "mrw.doc"`.
- Variables Window:** Lists variables: `obs`, `nonoil`, `intermed`, `oecd`, `gdp60`, `gdp85`, `gdpgrow`, `popgrow`, `inv`, `school`, `lngdp85`, `lngdp60`, `lngdpgrow`, `lnngdelta`.
- Results Window:** Displays regression results for the model `reg lngdp85 lnngdelta lninv if intermed==1`.

Source	SS	df	MS	Number of
Model	40.0886058	2	20.0443029	F(2,
Residual	26.847506	72	.372882028	Prob > F
Total	66.9361119	74	.904542052	R-squared

lnngdelta	Coef.	Std. Err.	t	P> t	[95% C
lnngdelta	-2.0172	.5338658	-3.78	0.000	-3.0814
lninv	1.317553	.1709425	7.71	0.000	.97678
_cons	5.345864	1.543075	3.46	0.001	2.2697

Results Window: Displays regression results for the model `. reg lngdp85 lnngdelta lninv if oecd==1`.

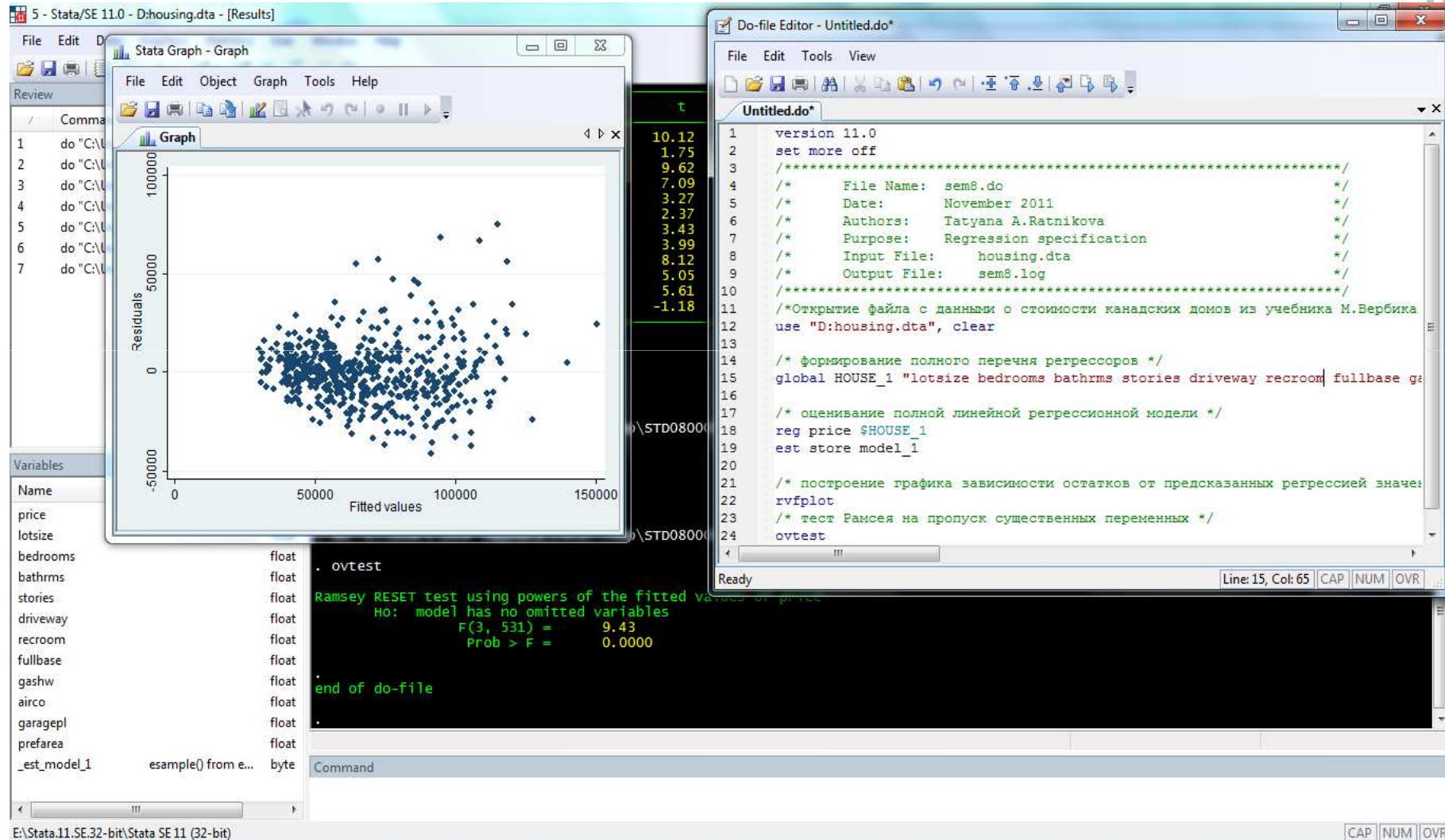
Source	SS	df	MS	Number of
Model	.320610261	2	.16030513	F(2,
Residual	2.70612185	19	.142427466	Prob > F
Total	3.02673211	21	.1441301	R-squared

lnngdelta	Coef.	Std. Err.	t	P> t	[95% C
lnngdelta	-.7419209	.8521945	-0.87	0.395	-2.5255
lninv	.4998895	.4338955	1.15	0.264	-.40826
_cons	8.020608	2.517888	3.19	0.005	2.7506
- Command Window:** Shows `. est store oecd` and `end of do-file`.

Do-file Editor Window: Shows the following code:

```
1 /* *****  
2 /* File Name: sem15.do  
3 /* Date: November, 2011  
4 /* Authors: E. Vakulenko, T. Ratnikova  
5 /* Purpose: Regression analysis  
6 /* *****  
7  
8 /*Задание для данных статьи */  
9 /*Mankiw G.N., David Romer, Weil D.N. A Contribution to th  
10 of Economic Growth // Quarterly Journal of Economics. 1992  
11  
12 /*Открытие файла с данными */  
13 use "D:\mrw.dta", clear  
14 set more off  
15 /*Оценивание модели Солсу для различных категорий стра  
16  
17 reg lngdp85 lnngdelta lninv if nonoil==1  
18 est store nonoil  
19 reg lngdp85 lnngdelta lninv if intermed==1  
20 est store intermed  
21 reg lngdp85 lnngdelta lninv if oecd==1  
22 est store oecd  
23  
24 /*Сводная таблица результатов*/  
25 est tab nonoil intermed oecd  
26  
27 /*Тестирование линейных гипотез*/  
28  
29 test lnngdelta + lninv = 0  
30  
31 /*Оценивание моделей с ограничениями на коэффициенты*/  
32  
33 constraint 1 lnngdelta=-lninv  
34 cnsreg lngdp85 lnngdelta lninv if nonoil==1, constraints(1  
35 cnsreg lngdp85 lnngdelta lninv if intermed==1, constraints
```


ОШИБКИ СПЕЦИФИКАЦИИ И ВЫБОР ОПТИМАЛЬНОГО НАБОРА РЕГРЕССОРОВ



ТЕСТИРОВАНИЕ ПРЕДПОСЫЛОК ТЕОРЕМЫ ГАУССА-МАРКОВА

The screenshot displays the Stata software interface with three main windows:

- Do-file Editor - Untitled.do*:** Contains Stata commands for regression and diagnostic testing:

```
48 /*Тесты на гетероскедастичность остатков*/  
49  
50 /* тест Бройша-Пагана на гетероскедастичность остатков */  
51 hettest  
52 /* тест Камеруна-Трайверди на гетероскедастичность и нормальн  
53 imtest  
54 /* ранговый тест на гетероскедастичность остатков */  
55 szroeter e_1  
56  
57 /* переход к другой функциональной форме количественных перемен  
58  
59 /* построение графиков остатков от отдельных регрессоров */  
60 rvppplot lotsize  
61 /* построение графиков квадрата остатков от отдельных регресс  
62 gen esq_2=e_1*e_1  
63 scatter esq_2 lotsize  
64 scatter esq_2 bedrooms  
65 scatter esq_2 bathrms  
66  
67 gen ae_2=abs(e_1)  
68 /* проведение тестов Глейзера на выявление источников гетерос  
69 foreach var of varlist $HOUSE_2 {  
70   qui reg ae_2 `var'  
71   est store ae2 `var'  
72 }  
73 est tab ae2_l1lotsize ae2_bedrooms ae2_bathrms ae2_stories ae2  
74  
75 tab airco  
76 tab prefarea
```
- Stata Graph - Graph:** A scatter plot showing the relationship between 'lotsize' (x-axis, 0 to 15000) and 'residual' (y-axis, -5000 to 10000). The plot shows a dense cluster of points around the origin with some outliers.
- Command Window:** Shows the execution of the do-file and the results of the regression and diagnostic tests:

```
1 . est store model_1  
2  
3 end of do-file  
4 . do "c:\Users\user\AppData\Local\Temp\STD070  
5  
6 . predict e_1, residual  
7  
8 end of do-file  
9  
10 . do "c:\Users\user\AppData\Local\Temp\STD070  
11  
12 /*Тесты на гетероскедастичность остатков*/  
13  
14 /* тест Бройша-Пагана на гетероскедастично  
15 . hettest  
16  
17 Breusch-Pagan / Cook-Weisberg test for heteroskedasticity  
18 Ho: Constant variance  
19 Variables: fitted values of price  
20  
21      chi2(1)      =    95.15  
22      Prob > chi2  =    0.0000  
23  
24 /* тест Камеруна-Трайверди на гетероскедастичность и нормальность остатков */  
25 . imtest  
26  
27 Cameron & Trivedi's decomposition of IM-test  
28  
29
```

Source	chi2	df	p
Heteroskedasticity	168.28	71	0.0000
Skewness	32.83	11	0.0006
Kurtosis	-4334.06	1	1.0000
Total	-4132.95	83	1.0000

ВЫВОД СВОДНЫХ ТАБЛИЦ РЕЗУЛЬТАТОВ РЕГРЕССИОННОГО АНАЛИЗА

The screenshot displays the Stata software interface. The main window shows the results of a regression analysis. The command used is `reg lngdp85 lnngdelta lninv if oecd==1`. The results are summarized in two tables.

Summary Statistics:

Source	SS	df	MS	Number of obs
Model	.320610261	2	.16030513	F(2, 19)
Residual	2.70612185	19	.142427466	Prob > F
Total	3.02673211	21	.1441301	R-squared
				Adj R-squared
				Root MSE

Regression Coefficients:

Variable	Coeff.	Std. Err.	t	P> t	[95% Conf. Int.]
lnngdelta	-.7419209	.8521945	-0.87	0.395	-2.525584
lninv	.4998895	.4338955	1.15	0.264	-.4082643
_cons	8.020608	2.517888	3.19	0.005	2.750607

The Do-file Editor window shows the following code:

```

1 /* ***** */
2 /* File Name: sem15.do */
3 /* Date: November, 2011 */
4 /* Authors: E. Vakulenko, T. Ratnikova */
5 /* Purpose: Regression analysis */
6 /* ***** */
7
8 /*Задание для данных статьи */
9 /*Mankiw G.N., David Romer, Weil D.N. A Contribution to
10 of Economic Growth // Quarterly Journal of Economics. 1!
11
12 /*Открытие файла с данными */
13 use "D:\mrw.dta", clear
14 set more off
15 /*Оценивание модели Солоу для различных категорий с
16
17 /*Представление результатов в виде сводных таблиц*/
18
19 reg lngdp85 lnngdelta lninv if nonoil==1
20 outreg2 using mrw.doc, replace word dec(3) ///
21 alpha(.01 , .05, .1) symbol(***, **, *) label
22
23 reg lngdp85 lnngdelta lninv if intermed==1
24 outreg2 using mrw.doc, append word dec(3) ///
25 alpha(.01 , .05, .1) symbol(***, **, *) label
26
27 reg lngdp85 lnngdelta lninv if oecd==1
28 outreg2 using mrw.doc, append word dec(3) ///
29 alpha(.01 , .05, .1) symbol(***, **, *) label
30
  
```

The Variables window lists the following variables:

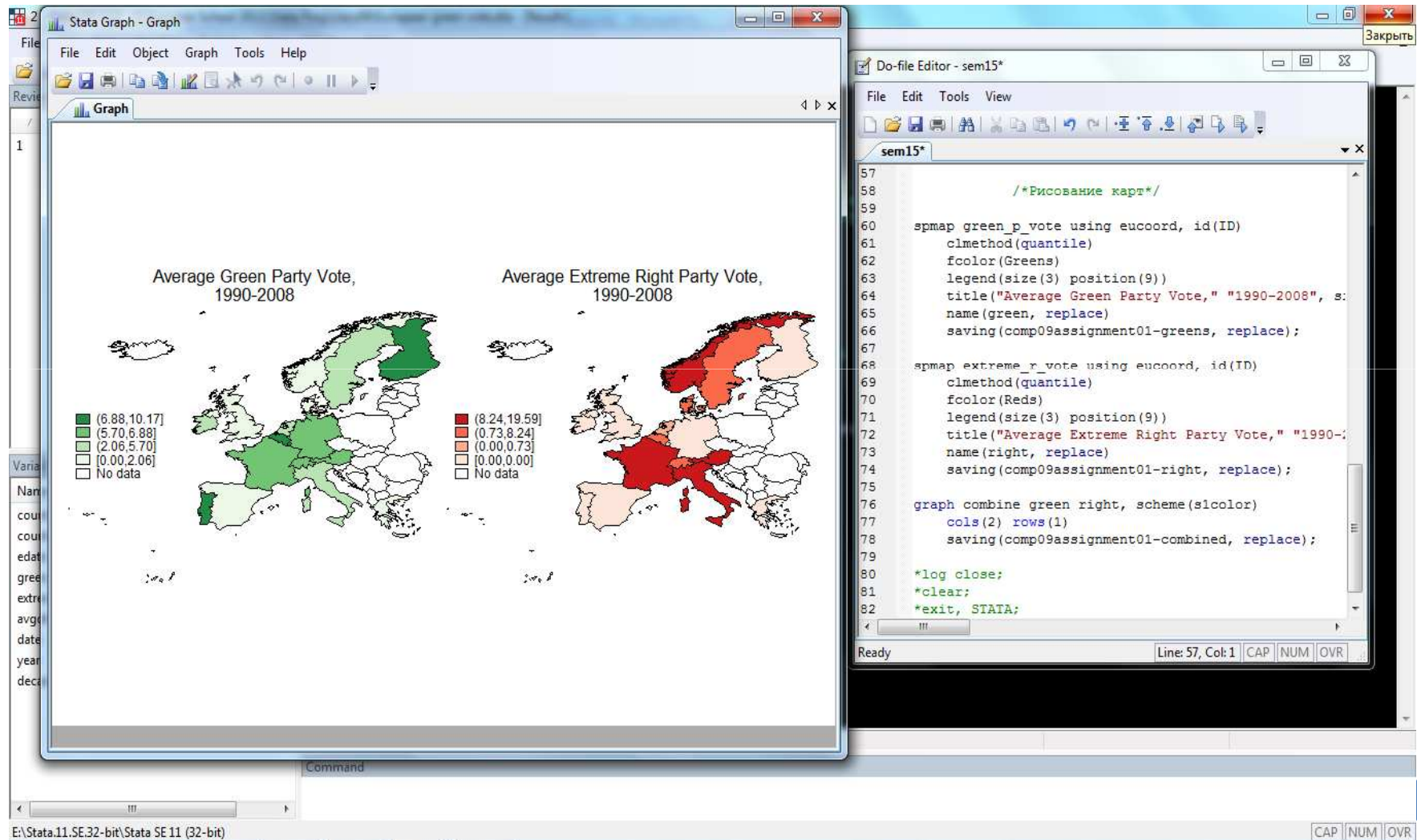
Name	Label	Type
obs		str
nonoil	dummy=1 for no...	by
intermed	dummy=1 for co...	by
oecd	dummy=1 for OE...	by
gdp60	gdp per adult, 1960	lor
gdp85	gdp per adult, 1985	int
gdpgrow	Average growth o...	flo
popgrow	Average growth o...	flo
inv	Average ratio of i...	flo
school	Percent of the wo...	flo
lngdp85		flo
lngdp60		flo
lngdpgrow		flo
lnngdelta		flo

The Do-file Editor window also shows a preview of the output table:

VARIABLES	(1)	(2)	(3)
	lngdp85	lngdp85	lngdp85
lnngdelta	-1.990*** (0.563)	-2.017*** (0.534)	-0.742 (0.852)
lninv	1.424*** (0.143)	1.318*** (0.171)	0.500 (0.434)
Constant	5.430*** (1.584)	5.346*** (1.543)	8.021*** (2.518)
Observations	98	75	22
R-squared	0.601	0.599	0.106

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ РЕЗУЛЬТАТОВ: ИЗОБРАЖЕНИЕ КАРТ



ЛИТЕРАТУРА

- Baum Christopher F. (2009). *An Introduction to Stata Programming*. Stata Press.
- Boehmke F. (2008). Interpreting and Presenting Regression Results. Presentation at the University of Kentucky. <http://myweb.uiowa.edu/fboehmke/Workshops/boehmke2008-02-29interactions.pdf>
- Cameron, A. Colin and Trivedi Pravin K. (2009). *Microeconometrics Using Stata*. Stata Press.
- Franzese, Robert J. and Cindy Kam (2007). *Modelling and Interpreting Interactive Hypotheses in Regression Analysis*. University of Michigan Press.
- Gould, William and William Sribney (2005). *Maximum Likelihood Estimation With Stata, 3rd Ed. College Station. Stata Press.*
- Long, J. Scott and Jeremy Freese (2005). *Regression Models for Categorical Dependent Variables Using Stata, 2nd Edition*. College Station, Texas. Stata Press.
- Mankiw G.N., Romer David, Weil D.N. (1992). *A Contribution to the Empirics of Economic Growth*. Quarterly Journal of Economics, 107, 2, 407-437.
- Mitchell, Michael N. (2008). *A Visual Guide to Stata Graphics, 2nd Edition*. College Station. Stata Press.
- Verbeek M. A . (2000). *Guide to Modern Econometrics*, Wiley.
- Берндт Э. (2005). *Практика эконометрики: классика и современность*. М.:ЮНИТИ-ДАНА, 863.
- Колеников С. (2003). *Прикладной эконометрической анализ в статистическом пакете Stata*.
- Ратникова Т.А. (2010). *Введение в эконометрический анализ панельных данных*. Москва: Издательский дом ГУ-ВШЭ.
- *Разработка учебно-методических материалов для преподавания курсов по применению количественного инструментария к решению социально-экономических задач* (2010). Исследовательский проект инициативной группы под руководством Елены Вакуленко. *Экономическая социология*. Т. 11. № 4.

