МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное
учреждение высшего образования

«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ
И РАДИОЭЛЕКТРОНИКИ» (ТУСУР)

Кафедра автоматизации обработки информации (АОИ)

УТВЕРЖДАЮ
Заведующий кафедрой АОИ

________________Ю.П. Ехлаков

«_____» ______________ 2017 г.

Методические указания к самостоятельным работам по дисциплине «Эконометрика»

Направление подготовки: **38.03.04 «Государственное и муниципальное управление»**

Форма обучения: **заочная**

Заочный и вечерний факультет (ЗиВФ)

Кафедра автоматизации обработки информации (АОИ)

Разработчик:

Старший преподаватель кафедры АОИ _____________ И.В. Потахова

2017

СОДЕРЖАНИЕ

1 Введение

Самостоятельная работа предусмотрена учебным планом. Цель самостоятельной работы студента в рамках курса «Эконометрика» — закрепление и расширение знаний, полученных во время проведения аудиторных занятий.

2 Содержание самостоятельной работы

1. Проработка лекционного материала осуществляется студентом с использованием конспекта лекций и рекомендуемых учебников. Цель — подготовка к восприятию очередной темы, рассматриваемой на лекции.

2. Подготовка к лабораторным работам. В соответствии с темой лабораторной работы студент должен изучить теоретический материал, подготовить решение задания к реализации на компьютере.

Темы лабораторных (соответственно, самостоятельных) работ:

1. Парная регрессия
2. Множественная регрессия
3. Различные аспекты множественной регрессии
4. Системы эконометрических уравнений

3. В рамках раздела «Изучение дополнительных тем курса» студент самостоятельно изучает дополнительные вопросы, связанные с построением и анализом моделей множественной регрессии, систем эконометрических уравнений .Для достижения этой цели сформулированы следующие домашние задания:

- Построение и анализ модели нелинейной парной регрессии
- Анализ случайных остатков в модели регрессии
- Построение модели регрессии в условиях гетероскедастичности
- Оценивание параметров структурной модели

3 Рекомендации по выполнению домашних заданий

Домашнее задание №1. Построение и анализ модели нелинейной парной регрессии

<u>Цель:</u> построение и исследование уравнения нелинейной регрессии.

Простая (парная) регрессия представляет собой регрессию между двумя переменными — y и x вида:

$$y = f(x) + \varepsilon, \tag{1.1}$$

где y — зависимая переменная (результативный признак);

x — независимая, или объясняющая, переменная (признак-фактор);

ε — случайная величина, характеризующая отклонения реального значения результативного признака от теоретического, найденного по уравнению регрессии $\hat{y} = f(x) + \varepsilon$.

Построение и анализ нелинейных моделей имеют свою специфику. В рамках данной лабораторной работы рассматриваются нелинейные модели, допускающие сведения их к линейному типу.

Расчетные соотношения.

1. ОЦЕНКА ПАРАМЕТРОВ НЕЛИНЕЙНОЙ РЕГРЕССИИ

1.1. Регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам.

- равносторонняя гипербола: $y = a + \frac{b}{x} + \varepsilon$. (1.2)
- полулогарифмическая: $y = a + b \cdot \ln(x) + \varepsilon$. (1.3)

При оценке параметров регрессий (1.1), (1.2) используется метод замены переменных. Суть его состоит в замене нелинейных объясняющих переменных, в результате чего нелинейные функции регрессии сводятся к линейным. К новой, преобразованной функции регрессии может быть применен обычный метод наименьших квадратов (МНК), рассмотренный в лабораторной работе №1.

Таблица 1. Возможные замены переменных

	Вид модели	Линеаризующие преобразования	Ограничения	Обратная замена переменных	
1	$\hat{y}_x = a + b \cdot \ln x$	$X = \ln x$	$x > 0$	$a = a$	$b = b$
2	$\hat{y}_x = a + b \cdot \frac{1}{x}$	$X = \frac{1}{x}$	$x \neq 0$	$a = a$	$b = b$

1.2. Регрессии, нелинейные относительно оцениваемых параметров, но линейных относительно включенных в анализ объясняющих переменных.

- логарифмическая модель (степенная): $y = a \cdot x^b + \varepsilon$ (1.4)
- показательная: $y = a \cdot b^x + \varepsilon$; (1.5)
- экспоненциальная: $y = e^{a + b \cdot x} + \varepsilon$ (1.6).

Оценка параметров регрессий (1.3), (1.4), (1,5) выполняется по следующему алгоритму:

1. уравнения приводятся к линейному виду с помощью логарифмирования и последующей замены переменных;
2. оцениваются параметры преобразованного уравнения с использованием метода наименьших квадратов;
3. выполняется обратная замена переменных и записывается исходное уравнение.

Таблица 2. Возможные замены переменных

	Вид модели	Линеаризующие преобразования	Ограничения	Обратная замена переменных	
1	$\hat{y}_x = a \cdot x^b$	$Y = \ln y,\ X = \ln x,$ $A = \ln a$	$x > 0,\ y > 0,$ $a > 0$	$a = e^A$	$b = b$
2	$\hat{y}_x = a \cdot b^x$	$Y = \ln y,\ B = \ln b,$ $A = \ln a$	$b > 0,\ y > 0,$ $a > 0$	$a = e^A$	$b = e^B$
3	$\hat{y}_x = e^{a + b \cdot x}$	$Y = \ln y$	$y > 0$	$a = a$	$b = b$

2. ИССЛЕДОВАНИЕ УРАВНЕНИЯ НЕЛИНЕЙНОЙ РЕГРЕССИИ

2.1. Оценка тесноты связи

Индекс корреляции. При нелинейной регрессии в качестве показателя тесноты связи выступает индекс корреляции. Его значение находится в границах $0 \le \rho_{xy} \le 1$.

$$\rho_{xy} = \sqrt{1 - \frac{\sigma^2_{ост}}{\sigma^2_y}}, \tag{1.7}$$

где

$$\sigma^2_y = \frac{1}{n}\sum(y-\bar{y})^2, \quad \sigma^2_{ост} = \frac{1}{n}\sum(y-\hat{y}_x)^2 \tag{1.8}$$

Индекс детерминации. Он характеризует долю дисперсии результативного признака *y*, объясняемую регрессией, в общей дисперсии результативного признака.

$$\rho^2{}_{xy} = \frac{\sigma^2_{объясн}}{\sigma^2_y} = 1 - \frac{\sigma^2_{ост}}{\sigma^2_y} \tag{1.9}$$

где

$$\sigma^2_y = \frac{1}{n}\cdot\sum_{i=1}^{n}(y_i-\bar{y})^2 = \overline{y^2} - \bar{y}^2,$$

$$\sigma^2_{объясн} = \frac{1}{n}\cdot\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2, \ \sigma^2_{ост} = \frac{1}{n}\cdot\sum_{i=1}^{n}(y_i-\hat{y}_i)^2, \tag{1.10}$$

Следует обратить внимание на то, что разности в соответствующих суммах $\sum(y-\bar{y})^2$, $\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$ и $\sum(y-\hat{y}_x)^2$ берутся не в преобразованных, а в исходных значениях результативного признака. Иначе говоря, при вычислении этих сумм следует использовать не преобразованные (линеаризованные) зависимости, а исходные нелинейные уравнения регрессии.

2.2. Оценка качества уравнения регрессии

Средняя ошибка аппроксимации. Средняя ошибка аппроксимации – среднее относительное отклонение расчетных значений $\hat{y}$ от фактических y.

$$\bar{A} = \frac{1}{n}\cdot\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_{x_i}}{y_i}\right|\cdot 100\% \tag{1.11}$$

Построенное уравнение регрессии считается удовлетворительным, если значение $\bar{A}$ не превышает 8–10 %.

Чем выше показатель детерминации или чем ниже средняя ошибка аппроксимации, тем лучше модель описывает исходные данные.

Коэффициент эластичности. Коэффициент эластичности. показывает, на сколько процентов измениться в среднем результат, если фактор изменится на 1%.

Формула для расчета коэффициента эластичности имеет вид:

$$Э = f'(x)\cdot\frac{x}{y}.$$

Так как для большинства функций коэффициент эластичности не является постоянной величиной, а зависит от соответствующего значения фактора x, то обычно рассчитывается средний коэффициент эластичности:

$$\overline{Э} = f'(x) \cdot \frac{\overline{x}}{\overline{y}}.$$

Таблица3. Формулы для расчета $\overline{Э}$ средних коэффициентов эластичности.

Вид модели, y	Первая производная, y'	Средний коэффициент эластичности, $\overline{Э}$
$y = a + \frac{b}{x} + \varepsilon$	$-\frac{b}{x^2}$	$-\frac{b}{a \cdot \overline{x} + b}$
$y = a \cdot x^b \cdot \varepsilon$	$y = a \cdot b \cdot x^{b-1}$	b
$y = a \cdot b^x \cdot \varepsilon$	$a \cdot \ln b \cdot b^x$	$\overline{x} \cdot \ln b$
$y = a + b \cdot \ln x + \varepsilon$	$\frac{b}{x}$	$\frac{b}{a + b \cdot \ln \overline{x}}$
$y = e^{a+b\cdot x} + \varepsilon$	$b \cdot e^{a+b\cdot x}$	$b \cdot \overline{x}$

2.3. Оценка значимости уравнения нелинейной регрессии.

$$F = \frac{\rho_{xy}^2}{1-\rho_{xy}^2} \cdot \frac{n-m-1}{m},$$

где ρ_{xy}^2 – индекс детерминации, n – число наблюдений, m – число параметров при переменной x. Фактическое значение F-критерия сравнивается с табличным при уровне значимости α и числе степеней свободы $k_2 = n-m-1$ (для остаточной суммы квадратов) и $k_1 = m$ (для факторной суммы квадратов). Вычисленное значение F-критерия признается достоверным, если оно больше табличного при заданном уровне значимости α. В этом случае делается вывод о существенности уравнения регрессии в целом.

3. ЗАДАНИЕ

1. На основании таблиц 4, 5 построить предложенные уравнения регрессий.
2. Вычислить индексы парной корреляции и индексы детерминации для каждого уравнения.
3. Проверить значимость уравнения регрессии в целом для каждой кривой выравнивания.
4. Определить лучшее уравнение на основе средней ошибки аппроксимации.
5. Определить лучшее уравнение на основе совместного анализа значений индекса детерминации и средней ошибки аппроксимации.
6. Вычислить средний коэффициент эластичности.
7. С помощью встроенной функции Excel построить линии тренда указанных функций, выведя соответствующее уравнение тренда и значение R^2. Сравнить с расчетными данными.
8. Сделать выводы по проделанной работе.

Таблица 4. Варианты заданий

Вариант	Графы табл. 4	Виды кривых аппроксимации				
		Гиперболическая	Полулогарифмическая	Степенная	Показательная	Экспоненциальная
1	**1, 2**	*		*		*
2	**2, 3**	*	*		*	
3	**3, 4**		*	*		*
4	**5, 6**	*		*	*	
5	**7, 8**		*		*	*
6	**9, 10**	*	*	*		
7	**1, 3**			*	*	*
8	**1, 4**		*	*	*	
9	**1, 5**	*			*	*
10	**1, 6**	*	*	*		
11	**2, 4**	*	*			*
12	**2, 6**	*			*	*
13	**2, 7**		*		*	*
14	**3, 6**	*		*		*
15	**3, 8**		*	*	*	

Таблица 5. Наличие предметов длительного пользования в домашних хозяйствах (шт)

Области и республики	Телевизоры	Видеомагнитофоны	Магнитофоны	Музыкальные центры	Персональные компьютеры	Холодильники	Стиральные машины	Пылесосы	Швейные машины	Легковые автомобили
	1	2	3	4	5	6	7	8	9	10
Белгородская область	113	39	65	10	1	103	93	77	58	26
Брянская область	124	37	47	5	2	99	72	64	50	18
Владимирская область	124	36	61	10	2	105	90	77	68	24
Воронежская область	122	36	44	13	2	102	96	66	60	25
Ивановская область	128	26	47	9	1	106	92	71	71	9
Калужская область	140	43	61	14	6	106	88	81	66	28
Костромская область	117	31	45	12	1	100	85	58	58	14
Курская область	113	40	47	15	3	100	78	66	51	28
Липецкая область	122	48	58	13	2	113	95	73	66	34
Московская область	139	64	57	27	14	106	87	81	62	22
Орловская	126	39	69	8	6	11	93	73	72	27

область										
Рязанская область	120	34	46	8	3	106	80	65	51	21
Смоленская область	125	39	55	24	6	115	93	66	49	21
Тамбовская область	118	37	59	8	1	108	99	74	65	23
Тверская область	122	35	52	8	4	102	87	64	65	12
Тульская область	133	54	58	15	5	102	93	79	36	16
Ярославская область	136	36	47	12	4	110	88	71	69	14
Республика Карелия	146	49	65	16	9	106	87	68	55	32
Республика Коми	148	58	59	23	5	11	92	78	69	34
Архангельская область	136	35	58	16	8	103	95	74	71	15
Вологодская область	138	34	51	10	3	104	95	64	60	19
Калининградская область	124	48	53	12	7	105	85	74	38	29
Ленинградская область	123	30	65	8	3	102	84	71	52	10
Мурманская область	149	59	63	29	8	107	92	87	74	21
Новгородская область	130	26	63	9	4	96	76	56	45	14
Псковская область	117	26	44	91	3	99	82	65	60	20
Краснодарский край	114	44	60	14	4	109	90	74	67	28
Ставропольский край	114	40	63	12	2	104	91	78	49	29
Астраханская область	126	54	55	11	5	116	87	76	66	29
Волгоградская область	109	41	55	8	1	106	93	74	70	23
Ростовская область	120	43	57	20	8	109	91	73	60	26
Республика Башкортостан	115	40	52	16	4	116	94	75	67	29
Республика Марий Эл	134	28	64	8	1	108	87	72	73	22
Республика Мордовия	130	33	52	10	1	109	89	77	58	18
Республика Татарстан	120	52	61	19	5	119	90	76	55	22
Удмуртская	123	32	53	7	4	111	97	69	74	26

республика										
Чувашская республика	128	31	60	11	0	105	85	76	78	19
Кировская область	144	27	60	7	2	120	109	74	86	22
Нижегородская область	125	36	52	6	3	114	101	81	75	21
Оренбургская область	124	47	57	17	6	119	105	82	66	36
Пензенская область	121	36	47	7	2	109	94	70	69	15
Пермская область	123	40	50	15	6	113	98	73	75	27
Самарская область	128	62	56	24	14	121	100	76	68	35
Саратовская область	118	38	51	12	4	124	87	65	62	27
Ульяновская область	116	37	51	9	4	109	96	77	67	25

Домашнее задание №2. Анализ случайных остатков в модели регрессии

Цель: научиться оценивать наличие эффекта гетероскедастичности.

Основные формулы и понятия:

<u>Тест Парка</u>

$$\ln e_i^2 = a + b \cdot \ln x_{ij} + v_i,$$

где x_{ij} $-$ *i-е значение о-го фактора*

v_i $-$ *случайный остаток*

Условие принятия гипотезы: $t_b > t_{\alpha,n-2}$

Если данное условие выполняется, то нулевая гипотеза о наличии гетероскедастичности будет принята при уровне значимости α.

<u>Тест ранговой корреляции Спирмена</u>

$$r_{x,e} = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot \left(n^2 - 1\right)}$$ — коэффициент ранговой корреляции Спирмена,

где x — одна из объясняющих переменных,

d_i — разность между рангом *i-го* наблюдения x и рангом модуля остатка *в i-м* наблюдении.

$$t_r = \frac{r_{x,e} \cdot \sqrt{n-1}}{\sqrt{1 - r_{x,e}^2}}$$ — статистика.

Если в модели регрессии имеется более одной объясняющей переменной, то проверка гипотезы может выполняться с использованием каждой из них.

Условие принятия гипотез: $t_r > t_{\alpha,n-2}$.

Если данное условие выполняется, то нулевая гипотеза об отсутствии гетероскедастичности отклоняется при уровне значимости α.

<u>Тест Голдфельда — Кванта</u>

В этом случае все наблюдения необходимо упорядочить по мере возрастания значений x. Затем построить регрессионную модель для первых ***k*** и последних ***k*** наблюдений. Соответственно обозначим через $SS_{ост}^{(1)}$ и $SS_{ост}^{(3)}$ необъясненную сумму квадратов отклонений в каждой регрессии. Тогда статистика имеет вид

$$F = \frac{SS_{ост}^{(3)}}{SS_{ост}^{(1)}}.$$

Если выполняется условие $F > F_\gamma(k-m-1, k-m-1)$, *то гипотеза об отсутствии гетероскедастичности отвергается.*

Для проведения теста ранговой корреляции Спирмена необходимо выполнить следующие действия:

1. Отсортировать данные в таблице по возрастанию значений *x*;
2. Придать каждому наблюдению ранг, для чего необходимо добавить новый столбец, в котором задать числа от 1 до *n*;
3. Вызвать из пакета анализа надстройку ***Регрессия***, указав в диалоговом окне опцию ***Остатки***. После выполнения данной надстройки появится дополнительная таблица, в которой содержатся номера наблюдений, прогнозы и остатки. Тот столбец таблицы, в котором находятся остатки, необходимо перенести к исходным данным. После выполнения этих действий наша таблица будет содержать четыре столбца: ранг наблюдения, упорядоченные значения регрессора *x*, значения *y* и значения остатков;
4. Отсортировать данные по возрастанию модулей остатков и добавить новый столбец рангов остатков, аналогичным образом задав значения от 1 до *n*;
5. В дополнительном столбце вычислить значения разности между двумя полученными рангами (это и будет значение d_i);
6. На основании формул подсчитать коэффициент ранговой корреляции и статистику;
7. Проверить гипотезу.

Вид таблицы для проведения теста ранговой корреляции Спирмена

Ранг по х	Ценах1(р.)	Спрос у (тыс. шт.)	Остатки	Ранг по остаткам	Разность рангов D_i	D_i* D_i
8	15,91р.	117,088	-0,34387	1	7	49
5	15,54р.	119,864	-0,39014	2	3	9
15	16,76р.	110,023	-0,84306	3	12	144
2	15,21р.	123,809	1,019821	4	-2	4
3	15,28р.	121,175	-1,11646	5	-2	4
9	15,92р.	116,17	-1,12322	6	3	9

10	15,95р.	118,344	1,257187	7	3	9
14	16,69р.	110,106	-1,31194	8	6	36
1	15,09р.	125,178	1,426776	9	-8	64
6	15,62р.	118,068	-1,5813	10	-4	16
11	16,31р.	116,201	1,847847	11	0	0
12	16,33р.	111,457	-2,67328	12	0	0
13	16,60р.	115,103	3,003645	13	0	0
4	15,49р.	116,914	-3,7319	14	-10	100
7	15,70р.	123,589	4,559903	15	-8	64
					Сумма	508

Следовательно, значение ранговой корреляции Спирмена будет равно

$$r_{x,e} = 1 - \frac{6 \cdot \sum_{i=1}^{n} D_i^2}{n \cdot \left(n^2 - 1\right)} = 1 - \frac{6 \cdot 508}{15 \cdot (225 - 1)} = 0.0928$$

А значение статистики будет $t = 0.0928 \cdot \sqrt{15 - 1} = 0.028$

Выбрав уровень значимости 5 %, получаем критическую точку $t_{0.05,13} = 2.16$. Данное значение получено формулой СТЬЮДРАСПОБР(0,05;13).

Поскольку условие $t < t_{\alpha,n-2}$ не выполняется, то гипотеза о наличии гетероскедастичности будет принята.

Для проверки подобной гипотезы на основании теста Гольдфельда — Кванта необходимо подобным образом отсортировать наблюдения по возрастанию значения *x*, а затем отдельно оценить каждую регрессионную модель для первой трети и для последней трети наблюдений. Просчитать соответствующую статистику и проверить гипотезу об отсутствии гетероскедастичности.

Задание для самостоятельной работы

Провести исследование табличных данных практической работы «множественная регрессия» на наличие гетероскедастичности, между значением ***y*** и каждым регрессором отдельно

a) Тестом Парка
b) Тестом ранговой корреляции Спирмена;
c) Тестом Гольдфельда — Кванта.

Сделать выводы.

Домашнее задание № 3. Построение модели регрессии в условиях гетероскедастичности

Цель: применение взвешенного метода наименьших квадратов с целью смягчения проблемы гетероскедастичности.

ВЗВЕШЕННЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ (ВМНК)

При нарушении предпосылок МНК о гомоскедастичности рекомендуется традиционный метод наименьших квадратов заменять взвешенным методом наименьших квадратов.

Применение обычного МНК к модели, в которой нарушена эта предпосылка, ведет к тому, что найденные параметры уравнения регрессии не будут эффективными оценками генеральных параметров. Кроме того, их дисперсии будут рассчитаны со смещением, что приведет к ложным выводам при оценке качества модели и при проведении прогнозирования по ней.

ВМНК используется для корректировки гетероскедастичности, за счет преобразования данных, позволяющего получать оценки, которые обладают не только свойством несмещенности, но и имеют меньшие выборочные дисперсии.

Пусть σ_{ε_i} — стандартное отклонение случайной ошибки ε_i в i–м наблюдении. В случае если σ_{ε_i} известно, гетероскедастичность можно корректировать, разделив каждое наблюдение на соответствующее ему значение σ_{ε_i}. Так для парной регрессии $y_i = a + b \cdot x_i + \varepsilon_i$ соответствующее преобразование данных будет иметь вид:

$$\frac{y_i}{\sigma_{\varepsilon_i}} = \frac{a}{\sigma_{\varepsilon_i}} + b \cdot \frac{x_i}{\sigma_{\varepsilon_i}} + \frac{\varepsilon_i}{\sigma_{\varepsilon_i}}. \qquad (4.6)$$

Тогда дисперсия остатков представляется в виде:

$$D\left(\frac{\varepsilon_i}{\sigma_{\varepsilon_i}}\right) = \frac{1}{\sigma^2_{\varepsilon_i}} \cdot \sigma^2(\varepsilon_i) = \frac{\sigma^2_{\varepsilon_i}}{\sigma^2_{\varepsilon_i}} = 1 \qquad (4.7)$$

В результате этой процедуры каждое наблюдение будет иметь случайную ошибку с единичной дисперсией. Следовательно, для преобразованной модели выполняется предпосылка МНК о гомоскедастичности дисперсии остатков, а оценки параметров регрессии, полученные по МНК будут наилучшими несмещенными оценками.

Применение вышеописанного метода в значительной степени ограничено тем, что на практике фактические значения σ_{ε_i} чаще всего неизвестны.

Рассмотрим практические решения этой проблемы (2 случая).

1. Применение ВМНК основано на преобразовании данных, в котором что фактические значения σ_{ε_i} заменяются оценкой $\hat{\sigma}_{\varepsilon}$

Алгоритм вычисления оценки $\hat{\sigma}_{\varepsilon}$ рассмотрим на примере:

№	x_1	x_2	y	№	x_1	x_2	y
1.	13	43	79	11	58	161	207
2.	28	56	110	12		108	152
3.	33	24	97	13.	69	86	199
4.	42	98	171	14	8	143	144
5.	12	176	204	15	60	42	140
6.	44	124	174	16	11	199	183
7.	36	130	184	17.	26	145	178
8.	33	291	311	18.	61	115	185
9.	34	141	206	19.	18	111	152
10	21	95	128	20.	30	192	204

По таблице исходных данных получена модель $\hat{Y} = 36.78 + 1.19 \cdot x_1 + 0.76 \cdot x_2$.

Упорядочим модули остатков данной модели и выделим группы, содержащие остатки, близкие по значениям модулей.

Остатки	*модуль*
-0,02531	0,025315
-0,22768	0,227681
2,639116	2,639116
-2,73419	2,734187
5,462125	5,462125
5,684687	5,684687
-5,97146	5,971459
-6,04654	6,046542
9,362728	9,362728
-9,51027	9,510271
9,643585	9,643585
-11,0514	11,05136
-11,9281	11,92808
13,61468	13,61468
-14,5311	14,53107
14,58852	14,58852
-18,2088	18,20879
19,08836	19,08836
-21,3305	21,33054
21,48151	21,48151

По каждой группе вычисляем оценку среднеквадратичного отклонения:

1 группа

-0,02531	0,000641
-0,22768	0,051839
2,639116	6,964934
-2,73419	7,475778
Дисперсия	14,49319/4
$\hat{\sigma}_{\varepsilon_1}$	**1,903496**

2 группа

5,462125	29,8348
5,684687	32,31567
-5,97146	35,65833
-6,04654	36,56067
9,362728	87,66067
-9,51027	90,44526
9,643585	92,99873
Дисперсия	405,4741/7
$\hat{\sigma}_{\varepsilon_2}$	**7,610839**

3 группа

-11,0514	122,1326
-11,9281	142,2791
13,61468	185,3594
-14,5311	211,152
14,58852	212,825
-18,2088	331,5602
19,08836	364,3655
-21,3305	454,9921
21,48151	461,4552
Дисперсия	2486,121/9
$\hat{\sigma}_{\varepsilon_3}$	**16,62034**

Преобразуем исходные данные, разделив на соответствующий коэффициент пропорциональности $\hat{\sigma}_{\varepsilon_j}$, j=1,2,3 и стром модель по преобразованным данным.

x_1/σ	x_2/σ	y/σ	x_1/σ	x_2/σ	y/σ
6,829538	22,59001	41,50258	3,4897	9,686926	12,45462
14,70977	29,41955	57,7884	12,08303	56,7377	79,85306
17,33652	12,60838	50,95886	4,15154	5,174383	11,97328
5,518445	12,87637	22,46796	0,481338	8,603916	8,664083
0,722007	10,58944	12,27412	31,52095	22,06466	73,54887
5,781228	16,29255	22,86213	0,66184	11,97328	11,01061
18,91257	68,29538	96,66423	13,65908	76,17562	93,51214
1,985519	17,50867	18,71201	3,670202	6,919233	11,13094
2,045686	8,483582	12,39445	2,365048	14,58446	19,97152
11,03233	49,90816	67,24468	1,805017	11,55211	12,27412

2. Применение ВМНК основано на предположении, что среднее значение остаточных величин равно нулю, а вот дисперсия их представлена в виде произведения некоторой величины K_i либо K_i^2 на постоянную величину σ^2:

$$\sigma_{\varepsilon_i}^2 = \sigma^2 \cdot K_i \quad \text{либо} \quad \sigma_{\varepsilon_i}^2 = \sigma^2 \cdot K_i^2$$

Например, рассматривается модель вида

$$y_i = a + b_1 \cdot x_{1_i} + b_2 \cdot x_{2_i} + \varepsilon_i,$$

для которой дисперсия остаточных величин оказалась пропорциональна K_i^2. Коэффициент пропорциональности K принимает различные значения для соответствующих i значений факторов x_1 и x_2.

Ввиду того, что $\sigma_{\varepsilon_i}^2 = \sigma^2 \cdot K_i^2$, для корректировки гетероскедастичности выполняется переход к уравнению с новым преобразованным переменным:

$$\frac{y_i}{K_i} = \frac{a}{K_i} + b_1 \cdot \frac{x_{1_i}}{K_i} + b_2 \cdot \frac{x_{2_i}}{K_i} + \frac{\varepsilon_i}{K_i}.$$

Параметры такой модели зависят от концепции, принятой для коэффициента пропорциональности K. В эконометрических исследованиях довольно часто выдвигается гипотеза, что остатки ε_i пропорциональны значениям какого-либо фактора. Так, если в уравнении

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \ldots + b_m \cdot x_m + \varepsilon$$

предположить, что $K_i = x_{1_i}$ и $\sigma_{\varepsilon_i}^2 = \sigma^2 \cdot x_{1_i}^2$, то ВМНК предполагает оценку параметров следующего трансформированного уравнения:

$$\frac{y_i}{x_{1_i}} = a \cdot \frac{1}{x_{1_i}} + b_1 + b_2 \cdot \frac{x_{2_i}}{x_{1_i}} + \ldots + b_m \cdot \frac{x_{m_i}}{x_{1i}} + \frac{\varepsilon_i}{x_{1_i}}.$$

Применение в этом случае ВМНК приводит к тому, что наблюдения с меньшими значениями преобразованных переменных $\frac{x}{K}$ имеют при определении параметров регрессии относительно больший вес, чем с первоначальными переменными. Вместе с тем, следует иметь в виду, что новые преобразованные переменные получают новое экономическое содержание и их регрессия имеет иной смысл, чем регрессия по исходным данным.

Задание для самостоятельной работы

2. Используя исходные данные предыдущей работы найти значения параметров регрессии, используя взвешенный метод наименьших квадратов, если

a) коэффициент пропорциональности K равен значению регрессора x_j

b) коэффициент пропорциональности K равен оценке $\hat{\sigma}_{\varepsilon}$

Тестом Гольдфельда — Кванта исследовать вновь полученную регрессионную

Домашнее задание №4. Оценивание параметров структурной модели.

Двухшаговый МНК основан на использовании, так называемых, «инструментальных» переменных и является универсальным методом. Как уже отмечалось, в системе одновременных уравнений нарушаются предпосылки о независимости факторов (выражаемых эндогенными переменными) и ошибок уравнений. Для преодоления этой трудности можно использовать замену эндогенных переменных y_i в правых частях уравнений модели на вспомогательные «инструментальные» переменные $\hat{y}_i$, которые были бы близки к исходным эндогенным переменным и при этом не зависели бы от ошибок уравнений. В качестве таких переменных предлагается использовать переменные, определяемые уравнениями приведенной формы модели

Согласно *двухшаговому МНК*, численные значения структурных параметров определяются в следующей последовательности:

1) Исходная система уравнений преобразуется в приведенную форму модели и определяются численные значения параметров δ_{ij} для каждого ее уравнения в отдельности с помощью традиционного МНК;

2) По полученным уравнениям приведенной формы находятся расчетные значения инструментальных переменных $\hat{y}_i$, соответствующих эндогенным переменным y_i для каждого наблюдения;

3) С помощью обычного МНК определяются параметры каждого структурного уравнения в отдельности, используя в качестве факторов фактические значения предопределенных переменных и полученные расчетные значения инструментальных переменных $\hat{y}_i$.

Более эффективным, но требующим существенно больших вычислительных затрат, является трехшаговый метод наименьших квадратов (ТМНК).

Он заключается в том, что двухшаговый метод наименьших квадратов применяется не к исходным уравнениям модели, а к уравнениям, преобразованным согласно обобщенному методу наименьших квадратов. Трехшаговый МНК является итерационной процедурой:

1) Параметры модели определяются обычным или двухшаговым МНК.

2) Вычисляются ошибки модели и определяется оценка корреляционной матрицы ошибок.

3) Уравнения преобразуются согласно обобщенному МНК.

4) Применяется двухшаговый МНК к преобразованным уравнениям и получается улучшенная модель (с улучшенными параметрами).

5) Процесс повторяется, начиная со второго шага, пока не будет достигнута заданная точность (либо превышено заданное количество итераций). Если случайные члены структурной модели не коррелируют, то трехшаговый метод сводится к двухшаговому.

Задание

1. Дана модифицированная модель Кейнса:

$$\begin{cases} C_t = a_1 + b_{11}Y_t + \varepsilon_1, \\ I_t = a_2 + b_{21}Y_t + b_{22}Y_{t-1} + \varepsilon_2, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где

C – потребление; Y – доход; I – инвестиции; G – государственные расходы;

t – текущий период; $t-1$ – предыдущий период.

годы	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Y	95,75	98,55	103,55	109	108,25	107,4	112,7	117,75	123,45	126,55	125,85	128,1	125,35	130,25	138,3	142,65	146,80	151,3	157,4	161,25
C	60,45	62,45	65,9	68,9	68,45	70	73,55	76,55	79,7	81,6	81,55	82,55	83,45	87,35	91,55	95,50	99	101,75	105,4	107,45
I	14,3	15,85	17,75	19,7	18,1	14,6	17,35	20	22,15	22,3	19,8	21	18	20	25,25	24,85	24,5	25	25,8	26,15

а) В предположении, что потребление зависит линейно от дохода (первое уравнение модели), оцените по МНК параметры a_1 и b_{11} функции потребления.

б) Оцените те же параметры по ВМНК и по ДМНК

в) Сравните полученные результаты. Сделайте выводы по качеству оценок.

Список литературы

Основная литература

1. Тихомиров, Николай Петрович. Эконометрика : учебник для вузов / Н. П. Тихомиров, Е. Ю. Дорохина . — М. : ЭКЗАМЕН, 2007 – 510[2] с. : ил., табл. (в библиотеке 11 экз.) (**Гриф**)

2. Яновский, Леонид Петрович. Введение в эконометрику : учебное пособие для вузов / Л. П. Яновский, А. Г. Буховец ; ред. Л. П. Яновский. - 2-е изд., доп. — М. : КноРус, 2009. - 254[2] с. : ил., табл. (в библиотеке 10 экз.)

3. Эконометрика : учебник для вузов / И. И. Елисеева [и др.] ; ред. И. И. Елисеева. - 2-е изд., перераб. и доп. - М. : Финансы и статистика, 2008. - 574[2] с. : ил., табл. (в библиотеке 5 экз.) (**Гриф**)

Дополнительная литература

1. Орлов, Александр Иванович. Эконометрика: Учебник для вузов/ А. И. Орлов. — 3-е изд., перераб и доп.. — М.: Экзамен, 2004. - 573[3] с.. (в библиотеке 1 экз.)
2. Практикум по эконометрике: Учебное пособие для вузов / Ирина Ильинична Елисеева, Светлана Владимировна Курышева, Нелли Михайловна Гордеенко и др; Ред. И. И. Елисеева. - М.: Финансы и статистика, 2001. - 192 с. (в библиотеке 2 экз.)
3. Бородич, Сергей Аркадьевич. Эконометрика: Учебное пособие для вузов. — Минск: Новое знание, 2001. - 408[8] с. : ил. (в библиотеке 4 экз.) (**Гриф**)
4. Кремер, Наум Шевелевич. Эконометрика: Учебник для вузов. — М.: ЮНИТИ-ДАНА, 2003. - 311 с. : ил. (в библиотеке 2 экз.) (**Гриф**)