Федеральное государственное автономное образовательное учреждение высшего образования "Национальный исследовательский университет "Высшая школа экономики"

Факультет экономических наук
Департамент прикладной экономики

Программа дисциплины Эконометрика (продвинутый уровень)

для направления 38.04.01 «Экономика» подготовки магистра
для магистерской программы «Мировая экономика»

Авторы программы: к.ф.-м.н., доцент Ратникова Т.А. (taratnikova@yandex.ru),
д.э.н., профессор Пересецкий А.А. (aperesetsky@hse.ru),
к.ф.-м.н., профессор Катышев П.К., (pkatyshev@hse.ru),
к.э.н., старший преподаватель Семерикова Е.В. (esemerikova@hse.ru).

Одобрена на заседании департамента прикладной экономики
«___»_____ 2018г.
Руководитель департамента
Авдашева С.Б, _____ [подпись]

Утверждена Академическим советом образовательной программы
«___»_____ 2018 г., № протокола_____

Академический руководитель образовательной программы
Григорьев Л.М. _____ [подпись]

Москва, 2018

Аннотация

Курс «Эконометрика – продвинутый уровень» рассчитан на студентов 1-го года обучения по магистерской программе «Мировая экономика» по специализациям: «Регулирование внешнеэкономических отношений», «Регулирование энергетических и сырьевых отраслей в России и в мире», «Внешнеторговая деятельность компаний», «Международные валютно-финансовые отношения», и представляет собой одну из базовых дисциплин фундаментального экономического образования. Материал курса предназначен для использования в дисциплинах, связанных с количественным анализом реальных экономических явлений. Программа рассчитана на студентов, прослушавших курс математического анализа, включающий дифференциальное и интегральное исчисление, а также курсы линейной алгебры, методов оптимальных решений, экономической статистики, теории вероятностей и математической статистики, эконометрики-1.

Учебная задача курса: дать магистрантам представления о теоретических основах современных эконометрических методов анализа данных, показать как можно более широкий спектр инструментов анализа данных, описывающих экономические процессы, ознакомить с наиболее важными эконометрическими подходами и научить корректному использованию инструментов на практике при работе со специализированной эконометрической программой STATA.

Тематический план учебной дисциплины

Название темы	Всего часов по дисциплине	Аудиторные часы		Самостоятельная работа
		Лекции	Сем. и практ. занятия	
1.**Методология эконометрического исследования.**	8	2	2	4
2. **Классическая линейная регрессионная модель.**	36	8	8	20
3. **Регрессионный анализ при нарушении условий теоремы Гаусса-Маркова и предположения о нормальности.**	36	8	8	20
4. **Оценка максимального правдоподобия**	24	4	4	16
5. **Модели с дискретными зависимыми переменными.**	32	6	6	20
6. **Инструментальные переменные в линейной модели.**	32	6	6	20
7. **Одномерные модели временных рядов.**	20	4	4	12
8. **Модели анализа панельных данных.**	28	4	4	20
9. **Модели анализа эффекта воздействия**	12	2	2	8
Итого	228	44	44	140

Базовые учебники

M. Verbeek. A Guide to Modern Econometrics. 3rd ed. Wiley. 2008.

Я. Магнус, П. Катышев, А. Пересецкий. Эконометрика. Начальный курс. 8-е изд. 2007.

П. Катышев, Я. Магнус, А. Пересецкий, С. Головань. Сборник задач к начальному курсу эконометрики. Москва, Дело. 2007. 4–е дополненное и переработанное издание, 2007.

Формы контроля знаний студентов

Предполагается посещение студентами лекций и семинарских занятий. Выполнение домашних работ, «проекта» зачета и экзамена и активность на семинарах.

Методика формирования результирующей оценки

Итоговая оценка складывается из следующих компонент (каждая оценивается по 100-балльной системе):

- 10% — оценок за домашние задания и квизы (письменные 15-мин. опросы на лекциях),
- 10% — оценка за зачет,
- 20% — оценка за «проект» (большое домашнее задание с реальными данными),
- 60% — оценка за финальный письменный экзамен.

К итоговой сумме добавляются баллы за активность на семинарах (т.е. теоретически итоговая оценка может быть более 100 баллов). В связи с принятой в НИУ ВШЭ процедурой оценки по 100-балльной системе конвертируются в оценки по 10-балльной системе. Пороги заранее не фиксированы.

При пересдачах оценивание происходит с использованием аналогичной процедуры, но пороги не обязательно совпадают с порогами на основном экзамене, т.к. сложность экзаменов может отличаться.

Содержание программы

1. Введение.

1.1. Методология эконометрического исследования на примере простой макроэкономической модели.

1.2. Обзор основных разделов и методов эконометрики.

1.3. Повторение теории вероятностей и математической статистики.

2. Классическая линейная регрессионная модель.

2.1. Оценивание параметров линейной регрессионной модели. Преимущества и недостатки различных методов оценивания. Метод наименьших квадратов (МНК). Матричная система обозначений. Операторы-проекторы и их свойства. Геометрическая интерпретация МНК. Свойства оценок метода наименьших квадратов при отсутствии предположения о случайном характере ошибок. Роль гипотезы о включении в регрессию свободного члена. Анализ вариации зависимой переменной в регрессии. Коэффициент множественной детерминации и его свойства. Коэффициент множественной детерминации, скорректированный на степени свободы.

2.2. Классическая линейная регрессия в предположении о случайном характере ошибок. Статистические характеристики ошибок, остатков, МНК-оценок параметров. Теорема Гаусса-Маркова. Несмещенная оценка дисперсии ошибок.

2.3. Гипотеза о нормальном распределении случайной ошибки. Законы распределения оценок регрессионных параметров. *t*-статистика для оценок коэффициентов регрессии, доверительные интервалы для теоретических значений коэффициентов и прогнозного значения зависимой переменной, доверительный интервал для дисперсии ошибки. *F*-статистика для линейной комбинации коэффициентов. Статистическая проверка общей линейной гипотезы о коэффициентах регрессии.

2.4. Проверка гипотезы о наличии структурных изменений. Тест Чоу. Использование фиктивных переменных для учета структурных изменений.

3. Регрессионный анализ при нарушении условий теоремы Гаусса-Маркова или предположения о нормальности.

3.1. Мультиколлинеарность. Внешние признаки, методы диагностики, методы устранения. Метод главных компонент.

3.2. Ошибки спецификации. Виды ошибок спецификации и их последствия. Диагностика ошибок спецификации. Выбор «оптимального» набора регрессоров и функциональной формы регрессионной зависимости.

3.3. Гетероскедастичность случайного возмущения и ее причины. Внешние признаки, методы диагностики. Свойства оценок МНК при наличии гетероскедастичности. Оценивание моделей с гетероскедастичными ошибками. Взвешенный метод наименьших квадратов. Доступный обобщенный метод наименьших квадратов.

3.4. Автокорреляция в линейных моделях. Внешние признаки, тесты.

4. Оценка максимального правдоподобия.

4.1. Оценка максимального правдоподобия (ML). Концепция, общие свойства оценок ML.

4.2. Оценка максимального правдоподобия параметров линейной регрессионной модели. Критерии для тестирования гипотез в линейной модели. Тесты Вальда (W), множителей Лагранжа (LM), отношения правдоподобия (LR) для тестирования общих ограничений в классической регрессионной модели. QML, понятие об обобщенном методе моментов (GMM).

5. Модели с дискретными зависимыми переменными.

5.1. Дискретные зависимые переменные: номинальные, ранжированные, количественные. Модели бинарного выбора. Probit и Logit модели. Интерпретация коэффициентов в моделях бинарного выбора. Предельные эффекты. Критерии качества моделей. Ошибки спецификации в моделях бинарного выбора.

5.2. Модели множественного выбора. Упорядоченный и неупорядоченный выбор.

5.3. Модели с урезанными и цензурированными выборками. Tobit модель. Модель Хекмана. Пример Mroze: занятость женщин..

5.4. Регрессия с переключением (switch regression model).

5.5. Модели времени жизни: от чего зависит длительность пребывания в состоянии безработицы?

6. Проблема эндогенности.

6.1. Неприменимость МНК в случае коррелированности регрессоров и случайной ошибки. Автокорреляция в динамических авторегрессионных моделях. Ошибки измерений переменных. Одновременность. Пропущенные регрессоры.

6.2. Инструментальные переменные. Где искать инструменты? Пригодность и уместность инструментов. Тест Хаусмана, его интерпретация. Тест Саргана.

6.3. Приложение: оценивание отдачи от образования.

7. Одномерные временные ряды

7.1 Понятие случайного (стохастического) процесса. Временной ряд, как дискретный случайный процесс. Слабо и сильно стационарные случайные процессы. Характеристики случайных процессов (математическое ожидание, дисперсия, автоковариационная и автокорреляционная функции). Теорема Вольда. Оператор лага.

7.2 Стационарность случайных процессов. Модели скользящего среднего MA(q). Условие обратимости. Модели авторегрессии AR(p). Уравнения Юла-Уокера. Условие стационарности. Модели авторегрессии-скользящего среднего ARMA (p,q). Автокорреляционная и частная автокорреляционная функции.

7.3 Оценивание коэффициентов авторегрессионных моделей. Оценивание коэффициентов процессов ARMA (p). Качество подгонки моделей временных рядов. Информационные критерии Акаике (AIC) и Шварца (BIC). Подход Бокса-Дженкинса к идентификации моделей стационарных временных рядов.

8. Модели анализа панельных данных.

8.1. Преимущества использования панельных данных. Трудности, возникающие при работе с панельными данными.

8.2. Понятие о модели однокомпонентной ошибки или модели со специфическим индивидуальным эффектом. Спецификация модели. Детерминированный и случайный индивидуальный эффект. Операторы «Between» и «Within». Виды оценок. Сравнительный анализ оценок.

8.3. Тестирование спецификации в моделях панельных данных. Тест Хаусмана. Тест на наличие случайного индивидуального эффекта. Тест на наличие детерминированного индивидуального эффекта.

9. Модели анализа эффекта воздействия

9.1 Средний эффект воздействия ATE. Средний эффект воздействия на подвергшихся воздействию. Смещение выборки из-за отбора.

9.2 Оценка среднего эффекта воздействия с помощью регрессии. Инструментальные переменные. Мэтчинг. Мера склонности.

Дополнительная литература

Greene W.H. Econometric Analysis, 5th ed., Prentice Hall. 2003.

Johnston J. And Dinardo J. (1997), Econometrics Methods, 4th edition, McGrow-Hill.

Wooldridge J.M. Econometric Analysis of Cross-Section and Panel Data. MIT Press, 2002.

Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. –М.: «ЮНИТИ», 1998.

Ратникова Т.А. Анализ панельных данных в пакете STATA. Методические указания к компьютерному практикуму по курсу «Эконометрический анализ панельных данных». М.: ГУ-ВШЭ, 2005

Ратникова Т.А. Введение в эконометрический анализ панельных данных. ЭЖ ВШЭ, т.10, №2, 2006

Примерные варианты проекта:

Предлагаемые данные представляют собой результаты 5-ой, 7-ой, 8-ой и 9-ой волн обследования РМЭЗ (Российского мониторинга экономического положения и здоровья населения), охватывающие период с 1994 г. по 2000 г. В опросе приняло участие свыше 10000 человек.

Информация, собранная в РМЭЗ, касается размеров, источников и структуры доходов и расходов домохозяйств и индивидуумов, занятости, распределения времени, уровня образования, состояния здоровья и других характеристик (свыше 500 переменных).

Используя данные предъявленных выборок, необходимо выполнить приведенные ниже упражнения.

Выборки построены по региональному принципу:

Описание переменных.

Тип переменной обозначен C для непрерывных, D для фиктивных (dummy) переменных и никак не обозначен для переменных промежуточного типа.

idind – индивидуальный номер респондента, сохраняемый за ним на протяжении всего периода наблюдения;
aid_i - индивидуальный номер респондента в 1994 году;
cid_i - индивидуальный номер респондента в 1996 году;
did_i - индивидуальный номер респондента в 1998 году;
eid_i - индивидуальный номер респондента в 2000 году;
year –год;
sex (D)– пол, 1 если респондент – женщина;
age (C)– возраст;
age2 (C)– возраст в квадрате;
education – образование, 1 – начальное, …, 7 – аспирантура.
stagna – стаж работы на данном предприятии: 1 – менее года; 2 – от 1 года до 2-х лет, 3 – до 3-х лет, 4 – до 5-ти лет, 5 – до 10-ти лет, 6 – свыше 10-ти лет;
settlement (D) – место жительства: 1 – город, 0 – село;
psu – код региона проживания;
isco_1 (D) – управляющие;
isco_2 (D) – специалисты с высшим образованием, творческие и научные работники;
isco_3 (D) – технический персонал;
isco_4 (D) – конторские служащие;
isco_5 (D) – работники сферы услуг и торговли;
isco_6 (D) – ориентированные на рынок работники сельского хозяйства;
isco_7 (D) - квалифицированные рабочие;
isco_8 (D) – водители и операторы;
isco_9 (D) – неквалифицированные рабочие;
logrealwage (C)– логарифм реальной заработной платы, полученной за последние 30 дней до момента обследования;
debt (D) – наличие долга по заработной плате;
d94 (D) – фиктивная переменная для 1994 года;
d96 (D) – фиктивная переменная для 1996 года;
d98 (D) – фиктивная переменная для 1998 года;
d00 (D) – фиктивная переменная для 2000 года;
mic_unempp (C)– вероятность потери работы для респондента с данными полом, возрастом и образованием в данном регионе;
subord (D)- наличие подчиненных
secondjob (D)- наличие второй работы;
workhour (C)- количество часов, проработанных на основном рабочем месте за последние 30 дней;
satisfied – степень удовлетворенности жизнью: 1 – полная удовлетворенность, …, 5 – полная неудовлетворенность.

Проект №1. Оценивание моделей с качественными зависимыми переменными.
Исследование детерминант задолженности по заработной плате.

В 90-х годах практика задержек выплат заработной платы была одним из способов адаптации российских предприятий к рыночным условиям, позволившим избежать роста уровня безработицы с одной стороны, но отрицательно сказавшимся на уровне благосостояния основной части населения с другой.

Вам предлагается выяснить, какие группы населения выделенного вам региона наиболее пострадали от практики невыплат. Для этого вы должны подобрать и оценить известными вам способами регрессионную зависимость, где в качестве регрессанта выступает **бинарная** переменная **debt** (наличие долга по заработной плате). Выбор независимых переменных и обоснование этого выбора – ваша самостоятельная задача.

Для содержательной интерпретации оцененной модели вам необходимо вычислить предельные эффекты независимых переменных, влияние которых окажется значимым.

Проект №2. Оценивание моделей по панельным данным.

Оценивание отдачи от человеческого капитала в условиях переходного периода.

Вам предлагается оценить уравнение Минцера, взяв в качестве зависимой переменной **logrealwage** – логарифм реальной заработной платы респондентов, полученной за месяц, предшествующий опросу, на основном месте работы. Выбор независимых переменных и обоснование этого выбора – ваша самостоятельная задача.

Вам необходимо оценить три модели: сквозную регрессионную модель, модель со случайными индивидуальными эффектами и модель с детерминированными индивидуальными эффектами. Затем с помощью соответствующих тестов необходимо будет выбрать наиболее адекватную модель и дать ее экономическую интерпретацию.

Программное обеспечение

Выполнение поставленных задач предполагается в пакетах STATA (также возможно использование EViews и R).

Справочная литература.

1. С. Колеников. Прикладной эконометрический анализ в статистическом пакете STATA. М.: РЭШ, 2000.
2. Т. Ратникова. Анализ панельных данных в пакете «STATA». Методические указания к компьютерному практикуму по курсу "Эконометрический анализ панельных данных". М., ГУ-ВШЭ, 2004.

Студенты могут сами предложить объект исследования для выполнения домашнего задания, согласовав его с преподавателем.

Примерные образцы задач и вопросов зачета, домашних заданий и опросов:

Теоретическая часть

(тут не нужны доказательства, только краткие пояснения)

1. (2 балла) Уравнение $y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t$ оценивают методом наименьших квадратов и получают значение статистики Дарбина-Уотсона DW=3.53. Что можно сказать об автокорреляции ошибок?

2. (2 балла) Перечислите свойства МНК-оценок в модели множественной регрессии при гетероскедастичности ошибок.

3. (2 балла) Пусть x_t - инвестиции, y_t - выпуск фирмы в год *t*. В году t_0 сменилась администрация фирмы. Предложите способ проверки гипотезы о наличии структурных изменений в момент t_0.

4. (2 балла) Уравнение $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ оценивают методом наименьших квадратов и получают значение статистики Дарбина-Уотсона DW=1.03. Что можно сказать об автокорреляции ошибок?

5. (2 балла) Может ли быть в парной регрессии $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ МНК-оценка $\widehat{\beta}_1$ положительной, а оценка коэффициента при *y* в регрессии *x* на *y* и константу отрицательной?

6. (2 балла) Оценивание зарплаты в зависимости от возраста (*age*), пола (*sex*) и уровня образования (*edu*) дало следующий результат:

$$z = \underset{(4.5)}{20.1} + \underset{(2.1)}{5.1}age - \underset{(8.1)}{0.7}sex + \underset{(5.2)}{3.2}edu$$

(в скобках даны значения *t*-статистик). Можно ли на основании этой регрессии говорить о дискриминации женщин по зарплате (*sex*=1 для женщин и =0 для мужчин)?

7. (2 балла) Уравнение $y_i = \alpha + \beta x_i + \varepsilon_i$ оценивается МНК. Может ли коэффициент детерминации быть малым (<0.05), а статистика $t_{\hat{\beta}} = \hat{\beta} / \sigma_{\hat{\beta}}$ большой (>10)?

8. (2 балла) Верно ли, что $R^2_{adj} = 1 - (1 - R^2)\dfrac{(n-1)}{(n-k)}$ распределен по F(n-k,n-1)? Если да, то объясните, почему, если нет, то тоже объясните, почему.

9. (2 балла) Пусть $Y = X\beta + \varepsilon, \quad \varepsilon : \ N(0, \sigma_\varepsilon^2 I_{n\times n})$, $\beta \in \mathbb{R}^{k\times 1}$, $h \in \mathbb{R}^{k\times 1}$.

Как распределена случайная величина $\dfrac{h'\hat{\beta} - h'\beta}{\sqrt{\hat{\sigma}_\varepsilon^2 h'(X'X)^{-1}h}}$?

(Требуется четкое обоснование).

Задачи

10. (6 баллов) Пусть $Y = X\beta + \varepsilon$, $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2 I$. Пусть к-мерная квадратная матрица A является невырожденным линейным преобразованием регрессоров: $Z = XA$. В преобразованных регрессорах уравнение выглядит так: $Y = Z\gamma + \varepsilon$, $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2 I$.

а) Как связаны между собой МНК-оценки $\hat{\beta}$ и $\hat{\gamma}$?
б) Как связаны между собой векторы остатков регрессий?
в) Как связаны между собой прогнозные значения, полученные по двум регрессиям?

11. (6 баллов) Рассмотрим оценку вида $\tilde{\beta} = \left((X'X)^{-1} + \gamma I\right)X'y$ для вектора коэффициентов регрессионного уравнения $y = X\beta + \varepsilon$, удовлетворяющего условиям классической регрессионной модели.
Найдите $E(\tilde{\beta})$ и $V(\tilde{\beta})$.
Можно ли найти такое γ, что оценка $\tilde{\beta}$ более эффективна, чем оценка МНК $\hat{\beta}$?

12. (6 баллов) Пусть есть набор данных (x_i, y_i), $i = 1,\ldots,n$ $(x_i > 0, y_i > 0)$, порожденных уравнением $y_i = \alpha + \beta x_i + \varepsilon_i$, удовлетворяющим условиям стандартной модели парной регрессии. Рассматриваются следующие оценки параметра β:

$$\tilde{\beta}_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i}, \qquad \tilde{\beta}_2 = \frac{\bar{y}}{\bar{x}}.$$

1) Найти дисперсию и смещение каждой из оценок.
2) Сравните смещения и дисперсии оценок. Какая из оценок более эффективна?

13. (6 баллов) По данным для 15 фирм (n=15) была оценена производственная функция Кобба-Дугласа: $\ln Q_i = \gamma + \alpha \ln L_i + \beta \ln K_i + \varepsilon_i$.

$$\underset{s.e.}{\widehat{\ln Q}} = \underset{(4.48)}{0.5} + \underset{(0.7)}{0.76} \ln L + \underset{(0.138)}{0.19} \ln K,$$

где Q- выпуск, L- трудозатраты, K- капиталовложения.

Матрица обратная к матрице регрессоров имеет вид:

$$\left(X^T X\right)^{-1} = \begin{pmatrix} 121573 & -19186 & 3718 \\ & 3030 & -589 \\ & & 116 \end{pmatrix}.$$

Требуется:

1) написать формулу для несмещенной оценки ковариации $\widehat{\mathrm{cov}}(\hat{\alpha}, \hat{\beta})$ и вычислить её по имеющимся данным (если это возможно);
2) проверить $H_0 : \alpha + \beta = 1$ при помощи t-статистики (обязательно требуется указать формулу для статистики, а также указать число степеней свободы);
3) построить 95% доверительный интервал для величины $\alpha + \beta$.

Примерный образец экзаменационной работы

Задача 1. (10 баллов)

Пусть есть регрессия $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$. По выборке оценены коэффициенты и прогнозные значения $\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 x_t$. Рассмотрим регрессию $y_t = \gamma_1 + \gamma_2 \hat{y}_t + u_t$. Можете ли вы найти оценки коэффициентов в этой регрессии?

Задача 2. (15 баллов)

Рассмотрим модель $y_t = \alpha + \beta x_t + \varepsilon_t$, где $Cov(x_t, \varepsilon_t) = \gamma \neq 0$. Пусть z_t — экзогенная инструментальная переменная, принимающая только 2 значения: 0 и 1. Найдите формулы для IV-оценки параметров α, β через выборочные средние $\bar{x}_1$, $\bar{y}_1$ и $\bar{x}_0$, $\bar{y}_0$ двух подвыборок ($z_t = 1$ и $z_t = 0$).

Задача 3 (10 баллов)

Пусть есть logit модель $y_i^* = x_i'\beta + \varepsilon_i$, $P(\varepsilon_i \le x) = \Lambda(x) = \dfrac{e^x}{1+e^x}$, $y_i = \begin{cases} 1, & y_i^* > 0, \\ 0, & y_i^* \le 0, \end{cases}$

Пусть $\hat{\beta}$ — оценки полученные по методу максимума правдоподобия (*ML, maximum likelihood*); $\hat{p}_i = \Lambda(x_i'\hat{\beta})$ — прогнозные значения вероятностей того, что $y_i = 1$. Набор регрессоров x_i включает константу.

Покажите, что сумма прогнозных значений вероятностей совпадает с числом $y_i = 1$ в выборке, т.е.

$$\sum_{i=1}^{n} \hat{p}_i = \sum_{i=1}^{n} y_i .$$

Задача 4. (15 баллов) (см. Описание примера MROZ)

INLF — индикатор того, что замужняя женщина работает; EDUC — уровень образования; EXPER — опыт работы; AGE — возраст; KIDSLT6, KIDSGE6 — количество детей возраста до 6 и после 6 лет;

MRT — ставка налога; HUSWAGE — зарплата мужа. Оценка logit модели участия на рынке труда приведена в таблице.

Dependent Variable: INLF
Method: ML - Binary Logit (Quadratic hill climbing)
Included observations: 753

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	8.782271	1.786486	4.915948	0.0000
EDUC	0.141681	0.045215	3.133479	0.0017
EXPER	0.215674	0.032783	6.578852	0.0000
EXPER^2	-0.003423	0.001041	-3.287138	0.0010
AGE	-0.091060	0.014784	-6.159473	0.0000
KIDSLT6	-1.288610	0.203670	-6.326959	0.0000
KIDSGE6	0.101550	0.075930	1.337424	0.1811
MTR	-9.569563	1.822471	-5.250874	0.0000
HUSWAGE	-0.173875	0.034617	-5.022885	0.0000
Mean dependent var	0.568393	S.D. dependent var		0.495630
S.E. of regression	0.416741	Akaike info criterion		1.053538
Sum squared resid	129.2130	Schwarz criterion		1.108806
Log likelihood	-387.6570	McFadden R-squared		0.247083
Obs with Dep=0	325	Total obs		753
Obs with Dep=1	428			

1) Рассчитайте вероятность того, что женщина с данными (см. табл. ниже) будет работать.

EDUC	EXPER	AGE	KIDSLT6	KIDSGE6	MTR	HUSWAGE
12	10	45	0	1	0.6	5

2) Как изменится эта вероятность, если зарплата мужа HUSWAGE станет равна 10?

3) Найдите маржинальный эффект $\frac{\partial P(INLF=1)}{\partial HUSWAGE}$.

Задача 5. (15 баллов)

Оценивание производственной функции по методу наименьших квадратов дало следующие результаты:

$$\ln Q = 1.37 + \underset{(0.257)}{0.632}\ln K + \underset{(0.219)}{0.452}\ln L, \qquad R^2 = 0.98,\ \mathrm{Cov}(\hat{\beta}_K, \hat{\beta}_L) = 0.055\ .$$

(в скобках даны стандартные ошибки). Проверьте на 5%-ном уровне значимости гипотезы:

(а) Эластичности по труду и капиталу совпадают;

(б) Выполнено свойство постоянства отдачи на масштаб.

Замечание. В задаче не указано число наблюдений. Будут ли ваши выводы зависеть от этого числа?

Задача 6. (20 баллов)

(*теоретическая часть, тут не нужны доказательства, только краткие пояснения*)

5.1. (4 балла). Оценивается модель *INV* = α + β *PROF*+ ε, где *INV* − инвестиций, а *PROF* – доход, измеренные в тыс.руб.

а) Как изменятся значения $\hat{\beta}$, $t_{\hat{\beta}}$, R^2, если *INV* измерить в млн.руб.?

б) Как изменятся значения $\hat{\beta}$, $t_{\hat{\beta}}$, R^2, если *INV* и *PROF* измерить в млн.руб.?

5.2. (4 балла). Почему в *Probit* модели рассматривается функция распределения стандартного нормального распределения $N(0,1)$, а не нормального распределения $N(0,\sigma^2)$ с неизвестной дисперсией, которая также бы оценивалась по данным, что, возможно, позволило бы сделать модель более гибкой?

5.3. (4 балла). Уравнение $y_t = \alpha + \beta x_t + \gamma y_{t-1} + \varepsilon_t$ оценивают методом наименьших квадратов и получают значение статистики Дарбина-Уотсона DW = 0.53. Что можно сказать об автокорреляции ошибок?

5.4. (4 балла). Перечислите свойства МНК-оценок в модели множественной регрессии при гетероскедастичности ошибок.

5.5. (4 балла).

a) Сформулируйте tobit-модель. В каких случаях она применяется?

b) Чем принципиально отличается модель Heckman от Tobit?

Задача 7. (15 баллов)

Данные по арестам в течение 1986 года по 2,725 мужчинам, родившимся в Калифорнии в 1960 или 1961 гг. Каждый из мужчин в выборке был арестован по крайней мере однажды до 1986 г. Вопрос: что объясняет, что мужчина снова был арестован в течение 1986 г. (сколько раз, и т.д.) Имеются следующие данные.

narr86 Количество арестов в 1986 г.
pcnv доля предыдущих арестов, приведших к осуждению. (прокси для неотвратимости наказания) avgsen среднее время срока заключения по предыдущим случаям осуждения (в месяцах) (прокси для суровости наказания)
tottime общее время проведенное в тюрьме после достижения возраста 18 лет (в месяцах)
ptime86 число месяцев в тюрьме во время 1986 (не может быть арестован, пока в тюрьме)
qemp86 количество кварталов, в которых имел работу, в течение, 1986 г., (возможности на рынке труда)
inc86 легальный доход, 1986, $100s
durat длительность последнего периода безработицы (в месяцах)
black =1 если афроамериканец
hispan =1 если латиноамериканец
born60 =1 если родился в 1960 г.
crime86 =1, если был арестован хотя бы однажды в 1986 (genr crime86 = narr86>0)
narr012 =0, если narr86 = 0, = 1, если narr86 = 1, = 2, если narr86 > 1.

1) Используйте данные таблицы для того, чтобы оценить вероятности 0, 1 и 2 и более арестов в 1986 для мужчины афроамериканца, который провел в тюрьме 20 месяцев до возраста 18 лет, родившегося в 1961 г., если доля предыдущих арестов, приведших к осуждению равна 0.3.

2) Как изменятся эти вероятности, если pcnv возрастет до 0.6?

Dependent Variable: NARR012
Method: ML - Ordered Logit (Quadratic hill climbing)
Included observations: 2725
Number of ordered indicator values: 3

Variable	Coefficient	Std. Error	z-Statistic	Prob.
BLACK	0.925821	0.112139	8.256003	0.0000
HISPAN	0.500905	0.105729	4.737645	0.0000
PCNV	–0.782682	0.118043	–6.630493	0.0000
TOTTIME	0.001969	0.009306	0.211617	0.8324
BORN60	–0.063839	0.090490	–0.705483	0.4805
Limit Points				
LIMIT_1:C(6)	0.968052	0.073949	13.09072	0.0000
LIMIT_2:C(7)	2.614744	0.096349	27.13829	0.0000

Pseudo R-squared	0.031162	Akaike info criterion	1.456103
Schwarz criterion	1.471285	Log likelihood	–1976.941
Hannan-Quinn criter.	1.461591	Restr. log likelihood	–2040.528
LR statistic	127.1741	Avg. log likelihood	–0.725483
Prob(LR statistic)	0.000000		