

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра Дифференциальных уравнений и математической экономики

Регрессионный анализ кривой заработной платы

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 441 группы

направления 09.03.03 «Прикладная информатика»

механико-математического факультета

Шевченко Вероники Андреевны

Научный руководитель:

доцент, к.ф.-м.н.,
доцент

Новиков В.В.

подпись, дата

Заведующий кафедрой:

зав.кафедрой, д.ф.-м.н.,
профессор

Дудов С.И.

подпись, дата

Введение. Актуальность выбранной темы определяется тем, что в настоящее время регрессионный анализ используется как в естественнонаучных исследованиях, так и в обществоведении.

Во многих практических задачах прогнозирования, изучая различного рода связи в экономических, производственных системах, необходимо на основании экспериментальных данных выразить зависимую переменную в виде некоторой математической функции от независимых переменных – регрессоров, то есть построить регрессионную модель.

Практически, речь идет о том, чтобы, анализируя множество точек на графике (т.е. множество статистических данных), можно найти линию, по возможности точно отражающую заключенную в этом множестве закономерность (тренд, тенденцию), линию регрессии.

В бакалаврской работе я буду применять регрессионный анализ для рассмотрения концепции кривой заработных плат, описывающую негативную взаимосвязь между уровнем безработицы и уровнем заработных плат.

Предметом исследования явились математические основы регрессионного анализа.

Объект исследования бакалаврской работы – практическая задача по применению регрессионного анализа в эконометрике.

Информационную базу составили труды ученых-экономистов в области эконометрических исследований, публикации и интернет источники.

Основная часть. Основная часть бакалаврской работы содержит в себе: изучение основ эконометрического моделирования, в частности методов оценивания моделей панельных данных; изучение экономических и математических аспектов анализа кривых заработной платы; построение регрессионных моделей соответствующих кривых на основе реальных данных по регионам РФ.

В рамках работы был сделан обзор методов анализа и прогнозирования регрессионных зависимостей на основе учебной литературы; обзор методов построения и анализа кривых заработной платы с использованием учебной литературы и научной периодики; а также на основе реальных данных по регионам РФ были построены и проанализированы в среде Gretl модели зависимости заработной платы от уровня безработицы.

Глава 1. Математические основы регрессионного анализа. Дадим определение основным понятиям, которые являются важными для пол-

ноценного понимания, изучения и исследования заявленной темы. Начнем с понятия эконометрики.

Эконометрика как наука расположена где-то между экономикой, статистикой и математикой. Один из ответов на вопрос, что такое эконометрика, может звучать так: это наука, связанная с эмпирическим выводом экономических законов. То есть мы используем данные или «наблюдения» для того, чтобы получить количественные зависимости для экономических соотношений. Данные, как правило, не являются экспериментальными, так как в экономике мы не можем проводить (многократные) эксперименты.

Но это — только малая часть работы эконометриста. Он также формулирует экономические модели, основываясь на экономической теории или на эмпирических данных, оценивает неизвестные величины (параметры) в этих моделях, делает прогнозы (и оценивает их точность) и дает рекомендации по экономической политике.

Математические модели широко применяются в бизнесе, экономике, общественных науках, исследовании экономической активности и даже в исследовании политических процессов. Математические модели полезны для более полного понимания сущности происходящих процессов, их анализа. Модель, построенная и верифицированная на основе (уже имеющихся) наблюдаемых значений объясняющих переменных, может быть использована для прогноза значений зависимой переменной в будущем или для других наборов значений объясняющих переменных. Можно выделить три основных класса моделей, которые применяются для анализа и/или прогноза:

- Модели временных рядов
- Регрессионные модели
- Системы одновременных уравнений

В своей же работе я буду рассматривать линейные регрессионные модели с двумя переменными, то есть многомерные регрессионные модели (multiple regression model), или модели множественной регрессии.

Часть 1.1 Оценивание регрессий по методу наименьших квадратов Основные гипотезы.

В начале этой части обозначены гипотезы, лежащие в основе модели множественной регрессии и являющиеся естественным обобщением модели парной регрессии. Их удобно записать в матричной форме, которая главным образом и будет использоваться в бакалаврской работе.

Далее идет знакомство и разбор **Метода наименьших квадратов** и **Теоремы Гаусса-Маркова**.

Как и в случае регрессионного уравнения с одной переменной, целью метода является выбор вектора оценок $\hat{\beta}$, минимизирующего сумму квадратов остатков e_t (т. е. квадрат длины вектора остатков e):

$$e = y - \hat{y} = y - X\hat{\beta}, ESS = \sum e_t^2 = e'e \rightarrow \min.$$

Теорема Гаусса-Маркова. Предположим, что

1. $y = X\beta + \varepsilon$; 2. X — детерминированная $n \times k$ матрица, имеющая максимальный ранг k ; 3. $E(\varepsilon) = 0$; $V(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I_n$

Тогда оценка метода наименьших квадратов $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$ является наиболее эффективной (в смысле наименьшей дисперсии) оценкой в классе линейных (по y) несмещанных оценок (Best Linear Unbiased Estimator, BLUE).

Часть 1.2. Фиктивные переменные.

Регрессионные модели являются достаточно гибким инструментом, позволяющим, в частности, оценивать влияние качественных признаков (пол, профессия, наличие детей и т. п.) на изучаемую переменную. Это достигается введением в число регрессоров так называемых фиктивных переменных, принимающих, как правило, значения 1 или 0 в зависимости от наличия или отсутствия соответствующего признака в очередном наблюдении. С формальной точки зрения фиктивные переменные ничем не отличаются от других регрессоров. Однако следует обратить особое внимание на правильное их использование и интерпретацию оценок.

Как правило, независимые переменные в регрессионных моделях имеют «непрерывные» области изменения (национальный доход, уровень безработицы, размер зарплаты и т.п.). Однако теория не накладывает никаких ограничений на характер регрессоров, в частности, некоторые переменные могут принимать всего два значения или, в более общей ситуации, дискретное множество значений. Необходимость рассматривать такие переменные возникает довольно часто в тех случаях, когда требуется принимать во внимание какой-либо качественный признак. Например, при исследовании зависимости зарплаты от различных факторов может возникнуть вопрос, влияет ли на ее размер, и если да, то в какой степени, наличие у работника высшего образо-

вания. Также можно задать вопрос, существует ли дискриминация в оплате труда между мужчинами и женщинами. В принципе можно оценивать соответствующие уравнения внутри каждой категории, а затем изучать различия между ними, но введение дискретных переменных позволяет оценивать одно уравнение сразу по всем категориям.

Часть 1.3. Обобщенный метод наименьших квадратов.

В этой части рассматривается, как классическая регрессионная схема обобщается в метод наименьших квадратов. Это направление изучает линейную модель, в которой ковариационная матрица Ω вектора ошибок ε не обязательно является скалярной (т.е. имеет вид $\sigma^2 I_n$), а может быть произвольной симметричной положительно определенной матрицей.

С помощью линейного преобразования исходную систему можно свести к обычному регрессионному уравнению и построить для него МНК-оценку вектора коэффициентов. Эта оценка зависит от матрицы ковариаций ошибки ε , а способ оценивания носит название обобщенный метод наименьших квадратов (ОМНК). (Соответствующий термин на английском языке — Generalized Least Squares, GLS.)

Для ОМНК-оценки устанавливается аналог теоремы Гаусса-Маркова, а именно, доказываем, что в классе всех несмещенных линейных оценок она обладает наименьшей матрицей ковариаций. Обобщенный метод наименьших квадратов позволяет с единых позиций изучать некоторые важные классы регрессионных моделей: так называемые модели с гетероскедастичностью, когда матрица Ω является диагональной, но имеет разные элементы на главной диагонали, и модели, в которых наблюдения имеют смысл временных рядов, а ошибки коррелированы по времени. Следует подчеркнуть, что практическое использование обобщенного метода наименьших квадратов усложнено тем обстоятельством, что для построения ОМНК-оценки требуется знать матрицу Ω , которая реально почти всегда неизвестна.

Одно из предположений классической регрессионной модели состоит в том, что случайные ошибки некоррелированы между собой и имеют постоянную дисперсию. В тех случаях, когда наблюдаемые объекты достаточно однородны, не сильно отличаются друг от друга, такое допущение оправдано. Однако во многих ситуациях такое предположение нереалистично. Например, если исследуется зависимость расходов на питание в семье от ее общего дохода, то естественно ожидать, что разброс в данных будет выше для семей

с более высоким доходом. Это означает, что дисперсии зависимых величин (а следовательно, и случайных ошибок) не постоянны. Это явление в эконометрике называется гетероскедастичностью (в отличие от гомоскедастичности — равенства дисперсий). Кроме того, при анализе временных рядов в довольно редких случаях можно считать, что наблюдения некоррелированы во времени. Как правило, значение исследуемой величины в текущий момент времени статистически зависит от ее значений в прошлом, что означает наличие корреляции между ошибками. Поэтому естественно изучать модели регрессии без предположения, что $V(\varepsilon) = \sigma^2 I$

В настоящей работе рассматривается так называемая обобщенная регрессионная модель

$$y = X\beta + \varepsilon,$$

где y — $n \times 1$ вектор зависимой переменной, X — $n \times k$ матрица независимых переменных, β — $k \times 1$ вектор неизвестных параметров, ε — $n \times 1$ вектор случайных ошибок, причем:

- 1) матрица X неслучайна и имеет полный ранг;
- 2) $E\varepsilon = 0$;
- 3) $V(\varepsilon) = \Omega$, и матрица Ω положительно определена.

Иными словами, обобщенная модель отличается от классической только условием 3.

Часть 1.4. Регрессионная модель с автокорреляцией ошибок.

В этой части рассматривается один из классов обобщенных регрессионных моделей, используемых при анализе данных и имеющих характер временных рядов. В этих случаях часто приходится принимать во внимание то обстоятельство, что наблюдения в разные моменты времени статистически зависимы (типичный пример — ежедневный обменный курс доллара по отношению к рублю). Следовательно, ошибки, относящиеся к разным наблюдениям (разным моментам времени), могут быть коррелированы, и ковариационная матрица вектора ошибок не является диагональной. Формально проблему оценивания неизвестных параметров решает обобщенный метод наименьших квадратов, рассмотренный выше.

Авторегрессионный процесс первого порядка.

При анализе временных рядов часто приходится учитывать статистическую зависимость наблюдений в разные моменты времени. Иными словами, для многих временных рядов предположение о некоррелированности ошибок не выполняется. В своей работе я рассматриваю наиболее простую модель, в которой ошибки образуют так называемый авторегрессионный процесс первого порядка.

Глава 2. Обобщенный метод моментов. Вторая глава посвящена обобщенному методу моментов, который в 80-ые годы получил большое распространение не только в теоретической, но и в прикладной эконометрике. ОММ основан на асимптотических свойствах оценок, поэтому может быть использован эффективно лишь для больших выборок. Преимуществом ОММ является отсутствие привязки к априори заданному (обычно нормальному) распределению ошибок. Также надо отметить большое методологическое значение метода. Практически все известные оценки могут быть получены при его применении.

Глава 3. Анализ панельных данных.

Третья глава посвящена такой важной для бакалаврской работы теме "Анализ панельных данных".

Термин "панельные данные" (panel data) пришел из обследований индивидов, и в этом контексте "панель" представляла собой группу индивидов, за которыми регулярно осуществляли наблюдения в течение определенного периода времени. В настоящее время методы анализа панельных данных получили большое распространение, и понимание панельных данных стало намного шире. Наряду с термином "панельные данные" иногда также используется термин "лонгитюдные данные" (longitudinal data).

Панельные данные состоят из повторных наблюдений одних и тех же выборочных единиц, которые осуществляются в последовательные периоды времени. В качестве объектов наблюдения могут выступать индивиды, домашние хозяйства, фирмы, страны и т.д. Примером панельных данных могут быть ежегодные обследования одних и тех же домашних хозяйств или индивидов (например, для определения изменения их благосостояния), ежеквартальные данные об экономической деятельности отдельных компаний, ежегодные социально-экономические показатели для регионов одной страны или для группы стран и т.д.

Панельные данные совмещают в себе как пространственные данные, так и временные ряды и сочетают достоинства каждого из этих видов данных. Это позволяет строить более адекватные и содержательные модели для изучения истинной причинно-следственной связи между различными переменными, что представляется невозможным в рамках только временных или только пространственных данных.

В настоящее время панельные обследования в разных формах проводятся во многих странах. Впервые панельные данные начали формироваться в США в 1960-х гг. Среди наиболее известных баз панельных данных США можно выделить PSID и NLS.

В России панельные обследования стали проводиться в 1990-х г. Наиболее известной базой панельных данных является РМЭЗ – Российский мониторинг экономического положения и здоровья населения (Russia Longitudinal Monitoring Survey (RLMS), [rsc.unc.edu /projects / rims](http://rsc.unc.edu/projects/rims)). РМЭЗ представляет собой серию общенациональных репрезентативных опросов домохозяйств и индивидов, проводившихся в России с 1992 г. Данные обследований содержат ответы на более чем 3000 вопросов, касающихся доходов и расходов, материального благосостояния, занятости, уровня образования, состояния здоровья и т.д.

Глава 4. Обзор теоретических концепций кривых заработной платы.

Изучив необходимый для исследования материал, в 4 главе можно перейти к его рассмотрению в рамках темы бакалаврской работы.

В трудах ученых-экономистов в области эконометрических исследований можно найти объяснение кривой заработных плат при помощи ряда регрессионных моделей рынка труда. В классическом представлении рынок труда является конкурентным, как и другие рынки. Как следствие этого, безработица в "конкурентных" моделях считается результатом избыточного предложения рабочей силы, связанного с высокой заработной платой. В реальности рынок труда несовершенен, и классический механизм взаимосвязи между безработицей и заработной платой часто не находит эмпирического подтверждения.

Возможным эмпирическим отражением "неконкурентных" моделей является кривая заработных плат (wage curve), концепция которой была предложена в статье Blanchflower and Oswald (1990). Формирование концепции происходило на основе эмпирических исследований, проведенных в работе

Blanchflower and Oswald (1990) для локальных рынков труда Великобритании. Были исследованы различные виды уравнения Минцера (Mincer wage equation) для локальных рынков труда Великобритании (отраслевых или территориальных). В результате была получена зависимость, из которой следует, что при удвоении безработицы заработная плата на данном рынке уменьшается на 10 процентов. Эта величина, $-0,1$, является эластичностью заработной платы.

Неоклассическая модель рынка труда предполагает равновесие между спросом и предложением на конкурентном рынке. Рост заработной платы ведет к росту предложения и одновременно к снижению спроса на труд. Как результат избыточного предложения возникает безработица. Данные, полученные при определении кривой заработных плат, напротив, связывают низкие заработные платы и высокую безработицу, что полностью противоречит неоклассической модели избыточного предложения. Таким образом, возникновение зависимости между высокой безработицей и низкими заработными платами необъяснимо при использовании конкурентной модели.

Глава 4. Оценивание кривой заработной платы на основе панельных данных по регионам РФ.

В бакалаврской работе было проведено оценивание регрессионной модели кривой заработных плат на основе реальных данных. Были использованы сведения о номинальной средней заработной плате в регионах России в рублевом выражении, а также количество безработных в этих регионах в период с 2000 по 2017г.

Данные о количестве безработных были взяты не из числа зарегистрированных безработных, а из публикуемого Росстатом (<https://rosstat.gov.ru/>) количества безработных (оценочная величина по методике Международной организации труда), что более соответствует реальному предложению на рынке труда. Несмотря на оценочный характер переменной количество безработных, она точнее описывает ситуацию на рынке труда. Этот феномен связан с тем, что в европейских странах безработные стремятся зарегистрироваться для получения пособия в отличие от РФ, где размер пособий по безработице настолько незначителен, что не стимулирует безработных вставать на учет. На основе описанных данных была сконструирована сбалансированная модель, и проведен эконометрический анализ при помощи программы Gretl.

Заключение.

В настоящее время регрессионный анализ используется как в естественнонаучных исследованиях, так и в обществоведении. Уравнение регрессии позволяет найти значение зависимой переменной, если величина независимой или независимых переменных известна.

Практически, речь идет о том, чтобы, анализируя множество точек на графике (т.е. множество статистических данных), найти линию, по возможности точно отражающую заключенную в этом множестве закономерность (тренд, тенденцию), линию регрессии.

В бакалаврской работе рассмотрены теоретические основы некоторых методов построения регрессионных моделей, в частности моделей с фиктивными переменными и моделей панельных данных. Кроме того, на основе имеющейся научной периодики рассмотрены подходы к построению и анализу кривых заработной платы. Практическая часть посвящена построению кривой для регионов России, полученной на основе анализа региональных данных.