

Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.
ЧЕРНЫШЕВСКОГО»

Кафедра теории функций
и стохастического анализа

ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ НА РЫНКЕ ЖИЛЬЯ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 412 группы
направления 01.03.02 Прикладная математика и информатика
Фроловой Кристины Игорьевны

Научный руководитель

к.э.н., доцент _____ А.В.Харламов

Зав. кафедрой

д.ф-м.н, доцент _____ С.П. Сидоров

Саратов, 2018

ВВЕДЕНИЕ

Актуальность темы. Ценообразование является одним из ключевых элементов экономики. Воздействие процессов ценообразования на социально-экономическую ситуацию неоспоримо. С другой стороны, цены на товары и услуги формируются под воздействием множества факторов. Поэтому анализ и прогнозирование развития рынка невозможны без исследования динамики цен.

Методы экономико-математического моделирования позволяют не только оценить и спрогнозировать экономическую ситуацию, но так же выявить факторы, влияющие на динамику цен, и подобрать инструменты для ее регулирования. Все многообразие факторов, которые воздействуют на изучаемый процесс, можно разделить на две группы: главные (определяющие уровень изучаемого процесса) и второстепенные. Последние часто имеют случайный характер, определяя специфические и индивидуальные особенности каждого объекта исследования.

Взаимодействие главных и второстепенных факторов и определяет колеблемость исследуемого процесса. В этом взаимодействии синтезируется как необходимое, типическое, определяющее закономерность изучаемого явления, так и случайное, характеризующее отклонение от этой закономерности. Случайные отклонения неизбежно сопутствуют любому закономерному явлению.

Для достоверного отображения объективно существующих в экономике процессов необходимо выявить существенные взаимосвязи и не только выявить, но и дать им количественную оценку. Этот подход требует вскрытия причинных зависимостей. Под причинной зависимостью понимается такая связь между процессами, когда изменение одного из них является следствием другого.

Не все факторы, влияющие на экономические процессы, являются случайными величинами. Поэтому при анализе экономических явлений обычно рассматриваются связи между случайными и неслучайными величинами. Такие связи называются регрессионными, а метод математической статистики называется регрессионным анализом.

Наличие множества исходных признаков, характеризующих процесс функционирования объектов, заставляет отбирать из них наиболее существенные и

изучать меньший набор показателей. Чаще исходные признаки подвергаются некоторому преобразованию, которое обеспечивает минимальную потерю информации. Сжатие информации получается за счет того, что число главных компонент - новых единиц измерения - используется значительно меньше, чем было исходных признаков.

Цель бакалаврской работы заключается в исследование рынка жилья.

Объект исследования — экономические модели на рынке недвижимости.

Предмет исследования — координаты квартир, их стоимость, этажность.

Для достижения поставленных целей в работе необходимо решить следующие **задачи**:

- Изучить методы построения и спецификации эконометрических моделей и проанализировать проблему спецификаций моделей по реальным данным: мультиколлинеарность, гетероскедастичность.
- Создать базу данных по продажам на рынке однокомнатных квартир г.Саратова за 2018гг.
- Построить и проанализировать эконометрические модели ценообразования.

Структура и содержание бакалаврской работы. Работа состоит из введения, двух разделов с подпунктами, заключения, списка использованных источников, содержащего 18 наименований. Общий объем работы составляет 49 страниц, включая 2 таблицы.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обосновывается актуальность темы работы, формулируется цель работы и решаемые задачи, отмечается практическая значимость полученных результатов.

В **первом** разделе приводятся основные понятия множественной линейной регрессии вида:

$$y_i = a_0 + a_1x_{i1} + a_2x_{i2} + \dots + a_kx_{ik} + \varepsilon_i, \quad (1)$$

где k – количество включенных в модель факторов, ε_i – взаимно некоррелированные случайные величины с ненулевым математическим ожиданием и дисперсией σ^2 , то есть $M\varepsilon_i = 0, D\varepsilon_i = \sigma^2$.

Так же рассматривается теорема Гаусса-Маркова и несмещенность коэффициентов регрессии. Рассмотрена регрессия по методу наименьших квадратов с одной независимой переменной.

Так же рассматриваются особенности оценивания модели, а именно: мультиколлинеарность, гетероскедастичность и фиктивные переменные

Под **мультиколлинеарностью** понимается высокая взаимная коррелированность объясняющих переменных. Мультиколлинеарность может проявляться в функциональной (явной) и стохастической (скрытой) формах.

При функциональной форме мультиколлинеарности по крайней мере одна из парных связей между объясняющими переменными является линейной функциональной зависимостью. В этом случае матрица $X^T X$ особенная, так как содержит линейно зависимые векторы-столбцы, и ее определитель равен нулю, т.е. нарушается предпосылка регрессионного анализа, это приводит к невозможности решения соответствующей системы нормальных уравнений и получения оценок параметров регрессионной модели.

Однако в экономических исследованиях мультиколлинеарность чаще проявляется в стохастической форме, когда между хотя бы двумя объясняющими переменными существует тесная корреляционная связь. Матрица $X^T X$ в этом случае является неособенной, но её определитель очень мал. В то же время вектор оценок b и его ковариационная матрица $\sum b$ пропорциональны обратной матрице $(X^T X)^{-1}$, а значит, их элементы обратно пропорциональны величине определителя $|X^T X|$. В результате получаются значитель-

ные средние квадратические отклонения (стандартные ошибки) коэффициентов регрессии $b_0, b_1, \dots, b_p$ и оценка их значимости по t -критерию не имеет смысла, хотя в целом регрессионная модель может оказаться значимой по F -критерию.

Оценки становятся очень чувствительными к незначительному изменению результатов наблюдений и объема выборки. Уравнения регрессии в этом случае, как правило, не имеют реального смысла, так как некоторые из его коэффициентов могут иметь неправильные с точки зрения экономической теории знаки и неоправданно большие значения.

Точных количественных критериев для определения наличия или отсутствия мультиколлинеарности не существует. Тем не менее, имеются некоторые эвристические подходы по ее выявлению.

Один из таких подходов заключается в анализе корреляционной матрицы между объясняющими переменными $X_1, X_2, \dots, X_p$ и выявлении пар переменных, имеющих высокие переменные корреляции (обычно больше 0,8). Если такие переменные существуют, говорят о мультиколлинеарности между ними. Полезно также находить множественные коэффициенты детерминации между одной из объясняющих переменных и некоторой группой из них. Наличие высокого множественного коэффициента детерминации (обычно больше 0,6) свидетельствует о мультиколлинеарности.

Другой подход состоит в исследовании матрицы $X^T X$. Если определитель матрицы $X^T X$ либо её минимальное собственное значение λ_{min} близки к нулю (например одного порядка с накапливающимися ошибками вычислений), то это говорит о наличии мультиколлинеарности. О том же может свидетельствовать и значительное отклонение максимального собственного значения λ_{max} матрицы $X^T X$ от ее минимального собственного значения λ_{min} .

Для устранения или уменьшения мультиколлинеарности используется ряд методов. Самый простой из них (но далеко не всегда возможный) состоит в том, что из двух объясняющих переменных, имеющих высокий коэффициент корреляции (больше 0,8), одну переменную исключают из рассмотрения. При этом, какую переменную оставить, а какую удалить из анализа, решают в первую очередь на основании экономических соображений. Если с экономической точки зрения ни одной из переменных нельзя отдать предпочтение, то оставляют ту из двух переменных, которая имеет больший коэффициент

корреляции с зависимой переменной.

Гетероскедастичность. Во втором условии Гаусса—Маркова утверждается, что дисперсия случайного члена в каждом наблюдении должна быть постоянной. Такое утверждение может показаться странным, и здесь требуется пояснение. Случайный член в каждом наблюдении имеет только одно значение, и может возникнуть вопрос о том, что означает его «дисперсия».

Имеется в виду его возможное поведение до того, как сделана выборка. Записывая модель (1), первые два условия Гаусса — Маркова указывают, что случайные члены $u_1, u_2 \dots u_n$ в n наблюдениях появляются на основе вероятностных распределений, имеющих нулевое математическое ожидание и одну и ту же дисперсию. Их фактические значения в выборке иногда будут положительными, иногда — отрицательными, иногда — относительно далекими от нуля, иногда — относительно близкими к нулю, но у нас нет причин априори ожидать появления особенно больших отклонений в любом данном наблюдении. Другими словами, вероятность того, что величина и примет какое-то данное положительное значение, будет одинаковой для всех наблюдений. Это условие известно, как гомоскедастичность, что означает одинаковый разброс. Оно проиллюстрировано в левой части рисунка 2.

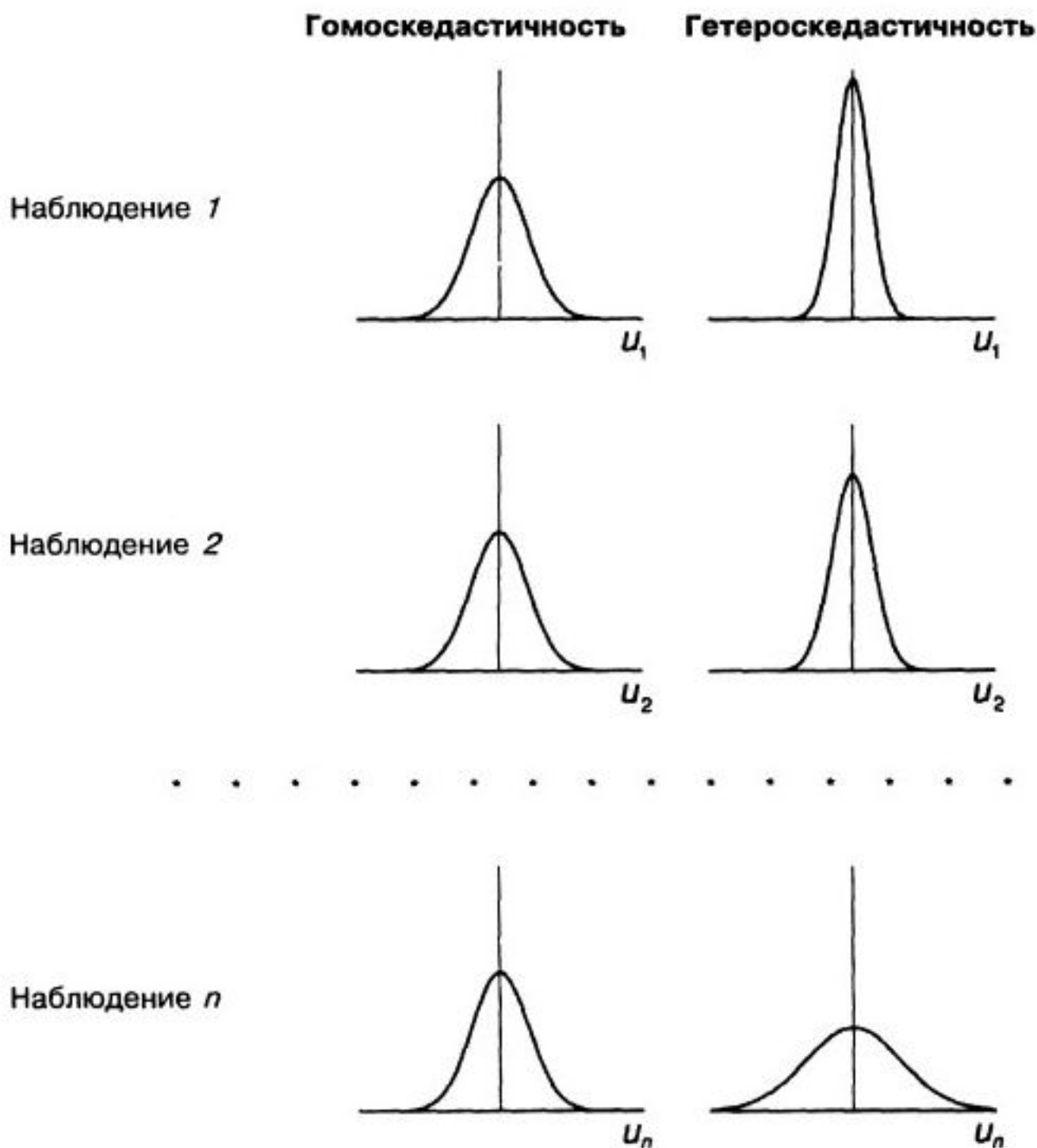


Рисунок 2 - Различия между гетероскедастичностью и гомоскедастичностью

Вместе с тем для некоторых выборок, возможно, более целесообразно предположить, что теоретическое распределение случайного члена является разным для различных наблюдений в выборке. В правой части рисунка 2 дисперсия величины u_i увеличивается по мере продолжения выборочных наблюдений. Это не означает, что случайный член обязательно будет иметь особенно большие (положительные или отрицательные) значения в конце вы-

борки, но это значит, что априорная вероятность получения сильно отклоненных величин будет относительно высока. Это пример гетероскедастичности, что означает «неодинаковый разброс». Математически гомоскедастичность и гетероскедастичность могут определяться следующим образом: *Гомоскедастичность*: $\text{pop.var}(u_i) = \sigma^2$ постоянна для всех наблюдений; *Гетероскедастичность*: $\text{pop.var}(u_i) = \sigma_i^2$ обязательно одинакова для всех i .

На рисунке 3 показано, как будет выглядеть характерная диаграмма рассеяния, если y — возрастающая функция от x и имеется гетероскедастичность типа, показанного на рисунке 2. Можно видеть, что, хотя наблюдения не обязательно все дальше отстоят от основной не стохастической составляющей линии регрессии $y = \alpha + \beta x$, по мере роста x все же имеется тенденция к увеличению их разброса.

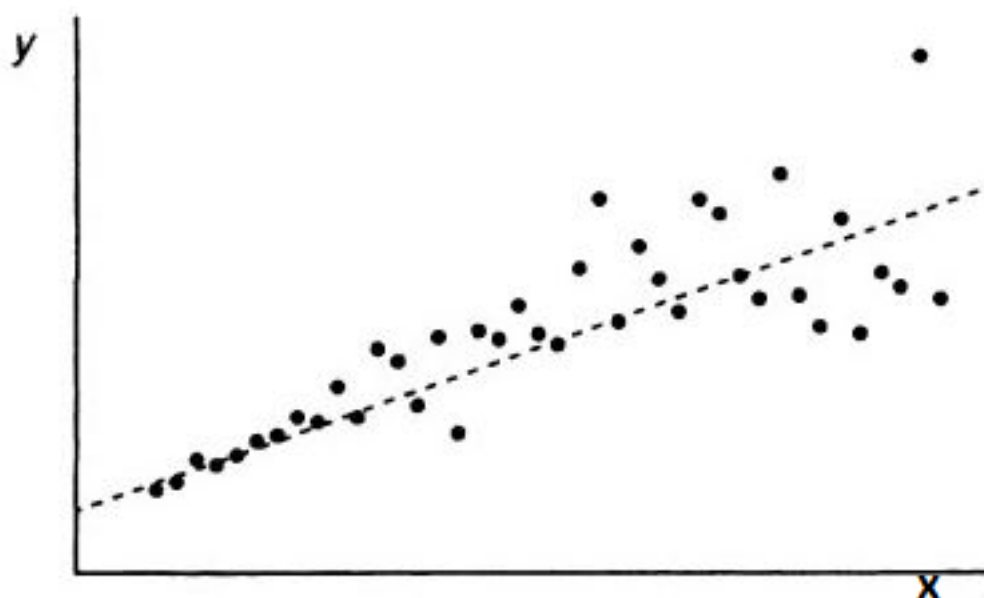


Рисунок 3 - Модель с гетероскедастичным случайным членом

Следует иметь в виду, что гетероскедастичность не обязательно относится к типу, показанному на рисунке 2. Данное понятие относится к любому случаю, в котором дисперсия вероятностного распределения случайного члена различна для разных наблюдений.

Возникает вопрос, почему гетероскедастичность имеет существенное значение. В самом деле, соответствующее условие Гаусса—Маркова пока не использовалось в проводимом анализе, и оно может показаться практически не нужным. В частности, при рассмотрении простой модели (1) и оцененного

уравнения (6) в доказательстве того, что b является несмещенной оценкой β и a — несмещенной оценкой α , это условие не использовалось.

Это объясняется двумя причинами. Первая касается дисперсии оценок, a и b . Желательно, чтобы она была как можно меньше, т.е. (в вероятностном смысле) обеспечивала максимальную точность.

При отсутствии гетероскедастичности обычные коэффициенты регрессии имеют наиболее низкую дисперсию среди всех несмещенных оценок, являющихся линейными функциями от наблюдений y . Если имеет место гетероскедастичность, то оценки МНК неэффективны. Можно, по меньшей мере в принципе, найти другие оценки, которые имеют меньшую дисперсию и, тем не менее, являются несмещенными.

Вторая, не менее важная причина заключается в том, что сделанные оценки стандартных ошибок коэффициентов регрессии будут неверны. Они вычисляются на основе предположения о том, что распределение случайного члена гомоскедастично, если это не так, то они неверны. Вполне вероятно, что стандартные ошибки будут занижены, а, следовательно, t -статистика — завышена, и будет получено неправильное представление о точности оценки уравнения регрессии. Возможно, что коэффициент значимо отличается от нуля при данном уровне значимости, тогда как в действительности это не так.

Свойство неэффективности можно легко объяснить интуитивно. Предположим, что имеется гетероскедастичность типа, показанного на рисунке 2 и на рисунке 3. Наблюдение, для которого теоретическое распределение случайного члена имеет малое стандартное отклонение будет обычно находиться близко к линии регрессии $y = \alpha + \beta x$ и, следовательно, может стать хорошим направляющим ориентиром, указывающим место этой линии. В противоположность этому наблюдение, где теоретическое распределение имеет большое стандартное отклонение, не сможет существенно помочь в определении местоположения линии регрессии. Обычный МНК не делает различия между качеством наблюдений, придавая одинаковые «веса» каждому из них независимо от того, является ли наблюдение хорошим или плохим для определения местоположения этой линии. Из этого следует, что, если можно найти способ придания большего «веса» наблюдениям высокого качества и меньшего — наблюдениям низкого качества, то, вероятно, получатся более точные оценки.

Другими словами, оценки для α и β будут более эффективными.

Фиктивные переменные являются весьма гибким инструментом при исследовании влияния качественных признаков. Рассмотрим пример. Введем новую переменную i - доход, используемый на потребления. В регрессии известно:

$$y_t = \beta_0 + \beta_1 i_t + \varepsilon_t, \quad (2)$$

коэффициент β_1 носит название "склонность к потреблению". Для этого нужно исследовать влияние сезона на склонность к потреблению.

$$y_t = \beta_0 + \beta_1 d_{t1} + \beta_2 d_{t2} + \beta_3 d_{t3} + \beta_4 d_{t1} i_t + \beta_5 d_{t2} i_t + \beta_6 d_{t3} i_t + \beta_7 i_t + \varepsilon_t, \quad (3)$$

согласно которой склонность к потреблению зимой, весной, летом и осенью есть $\beta_4 + \beta_7$, $\beta_5 + \beta_7$, $\beta_6 + \beta_7$ и β_7 соответственно. Можно тестировать гипотезы об отсутствии сезонных влияний на склонность к потреблению.

С помощью фиктивных переменных можно строить и оценивать кусочно-линейные модели.

Пусть y - зависимая переменная и есть только две независимые переменные: x и постоянный член. x и y представлены в виде временных рядов $\{(x_t, y_t), t = 1, \dots, n\}$. Из априорных соображений считается, что в момент t_0 произошла структурная перестройка и линия регрессии будет отличаться от той, что была до момента t_0 , но общая линия остается непрерывной, показанная на рисунке 4:

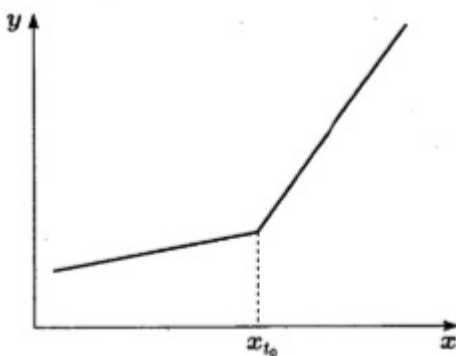


Рисунок 4 - Регрессионная линия

Чтобы оценить такую модель, введем бинарную переменную γ , полагая $\gamma_t = 0$, если $t \leq t_0$ и $\gamma_t = 1$, если $t > t_0$, и запишем регрессионное уравнение:

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 (x_t - x_{t_0}) \gamma_t + \varepsilon_t. \quad (4)$$

Регрессионная линия, показанная на рисунке 4, имеет коэффициент наклона β_2 для $t \leq t_0$ и $\beta_2 + \beta_3$ для $t > t_0$, и разрыва в точке x_{t_0} не происходит. Тестирую гипотезу $\beta_3 = 0$, мы проверяем предположение о том, что фактически структурного изменения не произошло.

С помощью фиктивных переменных можно исследовать влияние разных качественных признаков, а также их взаимное влияние.

Во **втором** разделе приведено построение и анализ модели рынка недвижимости.

Первый шаг практического задания состоит в выявление наиболее существенных факторов, влияющих на формирование стоимости объектов недвижимости, в частности рынка однокомнатных квартир г. Саратова, а также получение и анализ функциональных уравнений, описывающих эту зависимость. В данной работе попытаемся ответить на вопрос применительно к структуре цен квартир на вторичном рынке жилья Саратова, т.е. зависимости цены от характеристик квартиры: общей и жилой площади, материала постройки, этажности и т.д.

Следует отметить, что в зарубежных исследованиях при анализе рынка недвижимости используются фактические цены, по которым реализуется жилье, в данной работе используются цены предложения на рынке. Вопрос же о том, насколько правомерно распространять полученные при этом результаты на действительные процессы и явления, происходящие на рынке, остается открытым. Ведь как известно, цены предложения и реальные цены сделок купли-продажи жилья далеко не совпадают, поэтому не исключено, что картина, по которой нам приходится судить об исследуемом рынке, заметно отличается от реальной.

Для анализа рынка недвижимости была использована база объявлений о продажах однокомнатных квартир на рынке недвижимости жилья г. Саратова за 2018 гг. Данные о продажах квартир брались из разных источников, например, с сайта "Квартиры Саратова". Данные оттуда были извлечены и переведены в формат таблиц Excel.

Было рассмотрено порядка 400 однокомнатных квартир г. Саратова. Вся информация за 2018 г. содержит данные о ценах, общей и жилой площади, площади кухни и других не менее важных характеристик квартир, таких как этаж, этажность, материал постройки, наличие лоджии, балкона, санузел

совместный/раздельный и общее состояние квартиры.

Использовались следующие данные:

Таблица 1 - Используемые обозначения показателей

Обозначение	Описание
Price	цена квартиры
X1	общая площадь квартиры
X2	жилая площадь
X3	площадь кухни
X4	дополнительная площадь
X5	этаж квартиры
X6	квартира на первом этаже
X7	квартира на последнем этаже
X8	дом меньше пяти этажей
X9	пятиэтажный дом
X10	дом выше девяти этажей
X11	кирпичный дом
X12	монолитный дом
X13	квартира в отличном состоянии
X14	квартира в хорошем состоянии
X15	есть балкон
X16	есть лоджия
X17	санузел раздельный
XX, YY	координаты
X18	расстояние от центра до города
X19	район
X20	логарифм расстояния до центра города, $\ln(m)$

Построили модель МНК-оценки по всему городу и так же по отдельным районам. Каждый и районов сравнился по тесту Чоу.

В **заключении** приведены результаты бакалаврской работы.

Основные результаты

- Изучены методы построения и спецификации эконометрических моделей

и проанализирована проблема спецификаций моделей по реальным данным: мультиколлинеарность, гетероскедастичность.

- Создана база данных по продажам на рынке однокомнатных квартир г.Саратова за 2018гг.
- Построены и проанализированы эконометрические модели ценообразования.