

МИНОБРНАУКИ РОССИИ

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

Кафедра теории функций и стохастического анализа

**РЕГРЕССИОННОЕ МОДЕЛИРОВАНИЕ БИЗНЕС-ПРОЦЕССОВ НА
РЫНКЕ НЕДВИЖИМОСТИ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 452 группы
направления 38.03.05 — Бизнес-информатика

механико-математического факультета
Хасаханова Идриса Руслановича

Научный руководитель
старший преподаватель

А. Д. Лульнов

Заведующий кафедрой
д. ф.-м. н., доцент

С. П. Сидоров

Саратов 2019

ВВЕДЕНИЕ

Актуальность темы. В научных трудах многократно отмечалось, что рынок жилой недвижимости занимает особое место среди элементов рыночной экономики как системы и имеет важное значение в формировании и функционировании экономических механизмов обеспечения жителей России жильем, что имеет непосредственное отношение к повышению качества жизни народонаселения страны. Рынок недвижимости оказывает влияние на рыночное отношение, участвует в экономических процессах, выполняет ценообразующую функцию. В стоимости квартир концентрируется информация о насыщенности данного рынка, покупательной способности населения страны.

Цель бакалаврской работы Данная работа посвящена разработке и анализу эконометрических моделей бизнес-процессов на рынке недвижимости субъектов Российской Федерации.

Объект исследования

Для построения эконометрических моделей использовались данные с сайта росстата.

- Доля городского населения - отношение общего числа населения к числу проживающее в городских населенных пунктах.
- Индекс цен на вторичном рынке жилья Индекс цен на вторичном рынке жилья рассчитывается на основе зарегистрированных цен на квартиры функционирующего жилого фонда, находящиеся в собственности, если они являются объектами совершения рыночных сделок. Наблюдение ведется по выборочному кругу организаций, осуществляющих операции с недвижимостью в территориальных центрах и отдельных городах субъектов Российской Федерации. При регистрации цены за 1 м² квартиры учитываются ее количественные и качественные характеристики.
- Общий прирост - подсчитывался, как сумма естественного и внутрироссийского миграционного приростов, и зависит от числа рабочих мест, величины доходов. Данные о миграции населения получены в результате разработки документов статистического учета прибытия и выбытия, при регистрации и снятии с регистрационного учета населения по месту жительства. Внутрироссийская миграция делится на миграцию внутри региона и из региона в регион. Прирост для каждой миграции получен

разницей между прибывшими и убывшими. Прирост внутри региона и прирост из региона в регион образуют внутривосприимчивый прирост. Естественный прирост - абсолютная разница между рождаемостью и смертностью в конкретном регионе. Лидером по абсолютному приросту является г. Москва (84847,08571) и г. Санкт-Петербург (50195,47143).

- Плотность населения - это количество человек на 1 км². Она зависит от размера доходов, климатических условий, а также географического положения конкретного региона. Ясно, что самая большая плотность населения в г.Москва (4623,626374) и в г.Санкт-Петербург (3644,081633), т.к. это самые густонаселенные города РФ с небольшой площадью, относительно других регионов, а самая маленькая - Магаданская область (0,325559), Республика Саха (0,310639) и Чукотский АО (0,07029) [у них большая площадь, и небольшая численность населения]. Общая плотность населения России на 1 января 2017 года составляет 8,57 чел/км². Население распределено крайне неравномерно: 68,3 % россиян проживают в европейской части России, составляющей 20,82 % территории. Плотность населения европейской России — 27 чел/км², а азиатской — 3 чел/км².

- Среднедушевые денежные доходы населения. Денежные доходы населения включают доходы лиц, занятых предпринимательской деятельностью, выплаченную заработную плату наемных работников (начисленную заработную плату, скорректированную на изменение просроченной задолженности), социальные выплаты (пенсии, пособия, стипендии, страховые возмещения и прочие выплаты), доходы от собственности в виде процентов по вкладам, ценным бумагам, дивидендов и другие доходы («скрытые» доходы, доходы от продажи иностранной валюты, денежные переводы, а также доходы, не имеющие широкого распространения).

Предмет исследования

В самостоятельной части бакалаврской работы рассмотрены доля городского населения, индекс цен на вторичное жилье, индекс цен на первичное жилье, общий прирост населения, плотность населения, средний доход населения в 80 регионов РФ. По данным полученным с сайта росстата проведен эконометрический анализ, а именно была построена регрессионная модель

рынка недвижимости.

Структурное содержание бакалаврской работы

Работа состоит из введения, пяти разделов, заключения, списка использованных источников и приложений. Общий объем работы составляет более 40 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введение** было сформулировано определение эконометрики, основная цель работы, объекты исследования, предметы исследования и вывод по результатам полученных эконометрическим анализом.

В **первом разделе** приводятся основные понятия, используемые в анализе данных.

Меры отклонения функции $f(X, \beta)$ от набора наблюдений:

1) сумму квадратов отклонений $F = \sum_{t=1}^n (Y_t - f(X_t, \beta))^2$,

2) сумму модулей отклонений $F = \sum_{t=1}^n |Y_t - f(X_t, \beta)|$,

3) $F = \sum_{t=1}^n g(Y_t - f(X_t, \beta))$,

где g любое преобразование отклонения $Y_t - f(X_t, \beta)$ входящего в функционал F

Метод наименьших квадратов

Рассматривается задача наилучшей аппроксимации набора наблюдений $X_t, Y_t, t = 1, \dots, n$, линейной функцией $f(X) = a + bX$ в смысле минимизации функционала

$$F = \sum_{t=1}^n g(Y_t - f(X_t, \beta))^2. \quad (1)$$

Укажем необходимые условия экстремума (First Order Condition, FOC):

$$\frac{\partial F}{\partial a} = -2 \sum_{t=1}^n (Y_t - a - bX_t) = 0, \quad \frac{\partial F}{\partial b} = -2 \sum_{t=1}^n X_t(Y_t - a - bX_t) = 0. \quad (2)$$

Раскрыв скобки, получим *стандартную форму нормальных уравнений* (для краткости опустим индексы суммирования у знака суммы $\sum$):

$$\hat{a}n + \hat{b} \sum X_t = \sum Y_t, \quad \hat{a} \sum X_t + \hat{b} \sum X_t^2 = \sum X_t Y_t. \quad (3)$$

Решения $\hat{a}$, $\hat{b}$ системы (3) найдем:

$$\hat{b} = \frac{n \sum X_t Y_t - (\sum X_t)(\sum Y_t)}{n \sum X_t^2 - (\sum X_t)^2} = \frac{Cov(X, Y)}{Var(X)}, \quad (4)$$

$$\hat{a} = \frac{1}{n} \sum Y_t - \frac{1}{n} \sum X_t \hat{b} = \bar{Y} - \bar{X} \hat{b}. \quad (5)$$

Обобщенный метод наименьших квадратов

В данном разделе рассматривается, так называемая обобщенная регрессионная модель

$$y = X\beta + \epsilon, \quad (6)$$

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) $E\epsilon = 0$;
- 3) $V(\epsilon) = \Omega$, и матрица Ω положительно определена.

Обобщенная модель отличается от классической только условием 3).

Обычный метод наименьших квадратов. К системе (6) можно применить обычный метод наименьших квадратов. Пусть $\beta_{OLS} = \hat{\beta} = (X'X)^{-1} X'y$ — МНК-оценка вектора β , $\epsilon = My = I - X(X'X)^{-1}X'y$ — вектор остатков. Тогда нетрудно проверить, что

$$E(\hat{\beta}) = \beta + (X'X)^{-1}X'E(\epsilon) = \beta, \text{ т. е. } \beta \text{ является несмещенной оценкой,}$$

$$V(\hat{\beta}) = (X'X)^{-1}X'V(y)X(X'X)^{-1} = (X'X)^{-1}X'\omega X(X'X)^{-1},$$

$$E(\epsilon) = ME(y) = 0, V(\epsilon) = M\Omega M' = M\Omega M,$$

$M^2 = M$, $M' = M$ и $MX = 0$. Поэтому для математического ожидания суммы квадратов остатков получаем следующее выражение:

$$E(e'e) = \text{tr}(V(e)) = \text{tr}(M\Omega M) = \text{tr}(M^2\Omega) = \text{tr}(M\Omega).$$

Следовательно,

$$E(\hat{\sigma}^2) = E\frac{e'e}{n-k} = \frac{\text{tr}(M\Omega)}{n-k}.$$

Таким образом, если в качестве оценки матрицы ковариаций $V(\hat{\beta})$ взять стандартную оценку $\hat{V}\hat{\beta} = \hat{\sigma}^2$, то $E(\hat{V}(\hat{\beta})) = (1/(n-k))\text{tr}(M\Omega)(X'X)^{-1}$, что, в общем случае, не совпадает с $V(\hat{\beta})$. Значит, оценка матрицы ковариации

ций вектора β , получаемая при использовании обычного метода наименьших квадратов, является смещенной.

Заметим, что с очевидными изменениями, можно получить аналогичные результаты для стохастических регрессоров X . В частности,

$$V(\hat{\beta}) = E((X'X)^{-1}X'\Omega X(X'X)^{-1}) = \frac{1}{n}E((\frac{1}{n}X'X)^{-1}(\frac{1}{n}X'\Omega X)(\frac{1}{n}X'X)^{-1}),$$
 откуда следует, что если при $n \rightarrow \infty$ матрицы $(\frac{1}{n})X'X(1/n)'\Omega X$ стремятся к положительно определенным матрицам, то $V(\hat{\beta}) \rightarrow 0$ и, значит, оценка $\hat{\beta}$ будет состоятельной. Однако в отличие от классической модели, она не будет оптимальной в смысле теоремы Гаусса-Маркова. Для получения эффективной оценки надо воспользоваться так называемым обобщенным, методом наименьших квадратов (ОМНК).

Статистика R^2 — коэффициент детерминации.

Определение. Коэффициентом детерминации, или долей объясненной дисперсии, называется

$$R^2 = 1 - \frac{ESS}{TSS} = \frac{RSS}{TSS}. \quad (7)$$

F-статистика.

Ранее было показано, что s^2 и $\hat{b}$ — независимые случайные величины, поэтому по определению распределений Фишера и χ -квадрат, получаем

$$F = \frac{\left(\frac{\hat{b}-b}{\sigma_b}\right)^2 \frac{1}{1}}{\frac{\sum e_t^2}{\sigma^2} \frac{1}{n-2}} = \frac{(\hat{b}-b)^2 \sum x_t^2}{\sum e_t^2 / (n-2)} \sim \frac{\frac{1}{1}\chi^2(1)}{\frac{1}{n-2}\chi^2(n-2)} = F(1, n-2). \quad (8)$$

Полученную F-статистику можно использовать для проверки нулевой гипотезы $H_0 : b - b_0 = 0$. При этой гипотезе статистика выглядит следующим образом:

$$F = \frac{(\hat{b} - b_0)^2 \sum x_t^2}{\sum e_t^2 / (n-2)}. \quad (9)$$

Во втором разделе рассмотрены Теорема Гаусса-Маркова и основные гипотезы и тесты на гетероскедастичность используемые в исследовании и сделаны небольшие выводы по разделам.

Теорема 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):

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

Находим оценку метода наименьших квадратов:

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y. \quad (10)$$

Справедлива формула для суммы квадратов остатков

$$e'e = y'y - \hat{\beta}'X'y.$$

Предположим, что:

1. $y = X\beta + \varepsilon$;
2. X — детерминированная $p \times k$ 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) несмещенных оценок.

Проверка гипотез. Доверительные интервалы и доверительные области

$H_0 : H\beta = r$ линейное ограничение общего вида

Статистика для проверки гипотезы имеет вид:

$$F = \frac{(H\hat{\beta} - r)'(H(X'X)^{H^{-1}}(H\hat{\beta} - r)/q}{e'e/(n - k)} \sim F(q, n - k).$$

F -статистику можно выразить через коэффициенты детерминации R^2 для «короткой» и «длинной» регрессий:

$$F = \frac{(R_{UR}^2 - R_R^2)/q}{(1 - R_{UR}^2)/(n - k)} \sim F(q, n - k).$$

В этом случае «длинной» регрессией является регрессия без ограничений на параметры β , а «короткой» — регрессия с ограничениями $H\beta = r$. МНК

при этом состоит в минимизации функции ESS при условии $H\beta = r$.

Тесты на гетероскедастичность

Тест Уайта (White). Содержательный смысл этого теста состоит в следующем. Если в модели присутствует гетероскедастичность, то очень часто это связано с тем, что дисперсии ошибок некоторым образом зависят от регрессоров, а гетероскедастичность должна как-то отражаться в остатках обычной регрессии исходной модели.

- Сначала к исходной модели применяется обычный метод наименьших квадратов и находятся остатки регрессии e_t , $t = 1, \dots, n$.
- Затем осуществляется регрессия квадратов этих остатков e_t^2 на все регрессоры X , их квадраты, попарные произведения и константу (если ее не было в составе исходных регрессоров).
- При гипотезе H_0 величина nR^2 асимптотически имеет распределение $\chi^2(N-1)$, где R^2 — коэффициент детерминации, а N — число регрессоров второй регрессии.

Однако если гипотеза H_0 отвергается, этот тест не дает указания на функциональную форму гетероскедастичности, и единственным способом коррекции на гетероскедастичность является применение стандартных ошибок в форме Вайта.

Выводы 1) Применение обобщенного метода наименьших квадратов при наличии гетероскедастичности сводится к минимизации суммы взвешенных квадратов отклонений; 2) использование доступного обобщенного метода наименьших квадратов в общем случае требует оценивания n параметров по наблюдениям, что не позволяет получать состоятельные оценки; 3) в некоторых ситуациях (ошибка пропорциональна одной из независимых переменных, дисперсии ошибок принимают два значения) можно применять доступный обобщенный метод наименьших квадратов и получать состоятельные оценки коэффициентов регрессии; 4) если в модели с гетероскедастичностью использовать обычный метод наименьших квадратов, то для получения состоятельной оценки соответствующей матрицы ковариаций можно применять оценки ошибок в форме Вайта.

В третьем разделе работы рассмотрен программный пакет Gretl.

Основные возможности

- Оценивание параметров с помощью метода наименьших квадратов (OLS), метода максимального правдоподобия (ML), обобщённого метода моментов (GMM) и др.
- Выделение сезонности при помощи встраиваемых пакетов.
- Модели временных рядов: авторегрессия скользящего среднего (ARMA), авторегрессия интегрированного скользящего среднего (ARIMA), обобщенная авторегрессия условной гетероскедастичности (GARCH), векторная авторегрессия (VAR), векторная модель коррекции ошибок (VECM) и др.;
- Модели с ограниченными зависимыми переменными: логит (logit), пробит (probit), тобит (tobit), интервальная регрессия и др.;
- Вывод моделей в формате LaTeX;
- Скриптовый язык сценариев с поддержкой циклов для реализации метода Монте-Карло и итерационных процедур оценки;
- Создание графиков с помощью Gnuplot;
- Интеграция с R, GNU Octave и Oct для дальнейшего анализа данных;

Реализует основные операции с файлами (открытие, сохранение, настройки).

- Tools (Инструменты). Реализует функции статистических вычислений (вероятности, обратные величины - квантили и др.), калькулятор статистических тестов, вызов командной строки, вызов консоли;
- Data (Данные). Основные операции над данными: вывод; редактирование; сортировка; добавление; преобразование;
- View (Вид; Представление). Графическое представление; расчет основных статистик; корреляционная матрица; главные компоненты;
- Add (Дополнительно).
- Sample (Выборка). Ограничение значений; удаление наблюдений с отсутствующими значениями и т.д.;
- Variable (Переменная). Описательные статистики; распределения; гистограмма; коррелограмма; спектральный анализ; анализ сезонных колебаний;
- Model (Модель). Простой метод наименьших квадратов; другие линейные модели; анализ временных рядов; нелинейные модели; оценка устой-

чивости;

Помимо этого, основное окно имеет ряд кнопок, расположенных на нижней панели:

- 1) вызывает системный калькулятор;
- 2) открывает новое окно для скриптов GRETЛ;
- 3) открывает новое окно для инструкций GRETЛ;
- 4) открывает окно иконок для быстрого вызова окон текущей сессии;

В программе GRETЛ имеется простой редактор для ввода данных, однако он не очень удобный и требует большого времени для освоения. Наиболее простой путь - импорт данных через текстовый файл в формате *.txt или через таблицы MS Excel в формате *.xls. Мы рассмотрим именно этот способ.

Формат исходных данных в таблице Excel следующий: переменные вводятся по столбцам; в первой строке над каждым столбцом следует указать имя переменной с использованием только латиницы и цифр (не более 8 символов). Эти данные должны находиться в первом листе книги.

Процесс импорта данных. Последовательность пунктов меню следующая: [File] > [Opendata] > [Import] — * — [Excel...].

Далее в стандартном окне ввода выберут нужный файл.

После импорта данных целесообразно сохранить их в формате *.gdt (формат файлов данных программы GRETЛ). В этом случае снимаются ограничения на использование кириллицы как в имени файла, так и в именах папок. Последовательность пунктов меню следующая: [File] > [Savedataas] > [Standardformat...].

После этого появляется окно выбора переменных следует по очереди отметить в левом окне имя нужной переменной и нажать кнопку [Select—>].

В четвертом разделе описана эмпирическая часть бакалаврской работы.

В условиях современной российской экономики применение эконометрических методов становится все более актуальным способом проведения экономических исследований, анализа взаимосвязей экономических показателей, явлений и процессов.

Для эконометрического анализа будем использовать статистические данные по рынку первичного жилья представленные на рисунке 1.

Все представленные результаты получены с использованием программ-

ных продуктов Excel, Gretl.

	a	b	c	d	e	f
	2017	proportion of urban population(%)	house price index (primary)	Total population growth	Population density	Average income
1	2017	67.4	102.3	-0.002324	87.3097508	307797
2	2018	70.2	105.0	-0.001714	24.46997153	300824
3	2019	70.1	107.7	-0.000697	47.74652302	274236
4	2020	67.8	100.1	-0.002123	44.71180077	354828
5	2021	67.4	102.3	-0.001148	27.4077882	281148
6	2022	70.1	99.7	-0.002590	34.02846564	341024
7	2023	72.2	99.4	-0.002380	40.7461195	287460
8	2024	67.9	99	-0.000802	37.0231033	300780
9	2025	66.3	101.1	-0.002346	44.0888697	341460
10	2026	64.7	104.3	-0.001021	149.7402097	448108
11	2027	66.8	97.8	-0.003444	34.06805162	278844
12	2028	71.8	97.7	-0.004477	24.45978566	264344
13	2029	72.1	103.4	-0.007795	18.1004411	290128
14	2030	65.6	100.8	-0.000628	20.04407754	314028
15	2031	72.6	93.2	-0.004990	13.02102424	280896
16	2032	74.9	93.2	-0.000604	18.32484123	230004

Рисунок 1 – Статистические данные по рынку первичного жилья

На рисунке 1 рассмотрено современное состояние рынка недвижимости в России, а также факторы, которые оказывают на него наибольшее влияние. В качестве причин, которые сформировали такие тенденции, выделены, например:

- повышение себестоимости возводимых новостроек;
- снижение платежеспособности потенциальных покупателей и отсутствие стабильного роста их доходов, трудности в оплате задолженности по ипотечным кредитам.

В настоящее время рынок жилья в России испытывает серьезные затруднения. Квартиры продаются тяжело, цены на жилье сравнительно низкие в связи с отсутствием у среднестатистического жителя России необходимых наличных денежных средств на его приобретение. В то же время продавцы жилья не всегда ведут себя адекватно сложившимся на рынке условиям: одни предлагают необоснованно завышенные цены, другие – их слишком занижают.

В пятом разделе проведены эконометрические исследования применяя разделы 1-3. Сделаны выводы и прогнозы.

- проверено наличие автокорреляции остатков;
- проверено наличие гетероскедастичности с помощью теста Вайта;
- проверено на нормальность остатков;
- сделано соответствующие выводы.

С помощью приложения Gretl построена модель и оценено ее качество.

В качестве исходных данных были выбраны:

- Доля городского населения (proportion of urban population);
- Индекс цен на первичное жилье (house price index (primary));
- Общий прирост населения (Total population growth);
- Плотность населения (Population density);

- Средний доход населения (Average income).

В качестве объясняемой переменной будет индекс цен на первичное жилье.

В качестве объясняющих переменных – доля городского населения, общий прирост населения, плотность населения, средний доход населения.

Для построения регрессионной модели рынка недвижимости импортируем файл Excel в Gretl.

ЗАКЛЮЧЕНИЕ

В ходе данной работы была рассмотрена регрессионная модель рынка недвижимости. В эмпирической части курсовой работы была построена регрессионная модель рынка недвижимости. В программе Gretl оценено качество построенной модели. По корреляционной матрицы были выбраны значимые переменные по вероятности принятия нулевой гипотезы, это позволило вывести значимое уравнение для построенной модели. Определено отсутствие гетероскедастичности с помощью теста Вайта. Далее проверили модель на тест нормальности остатков. **ОСНОВНЫЕ РЕЗУЛЬТАТЫ**

- Определены основные понятия для эконометрического исследования на рынке недвижимости.
- Приведен материал по данной теме, а именно модель парной регрессии, метод наименьших квадратов, теорема Гаусса-Маркова, обобщенный метод наименьших квадратов, оценка максимального правдоподобия, модель множественной регрессии.
- Изучен программный пакет gretl, основные его функции и применение.
- Сделаны выводы из приведенных эконометрических исследований.