

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»**

Кафедра теории функций и стохастического анализа

Регрессионные модели с переключением режимов

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 451 группы
направления 38.03.05 Бизнес-информатика

механико-математического факультета

Гоба Марьи Алексеевны

Научный руководитель:

ст. преп. _____

А. Д. Луньков

Заведующий кафедрой:

д.ф.-м.н., доцент _____

С. П. Сидоров

Саратов 2020

СОДЕРЖАНИЕ

Стр.

ВВЕДЕНИЕ	3
1 Регрессионные модели.....	5
1.1 Линейная вероятностная модель	5
1.2 Уравнение парной линейной регрессии. Оценка параметров	6
2 Модель множественной регрессии	8
2.1 Оценка качества уравнения множественной линейной регрессии.....	8
3 Модели двоичного выбора и с ограничениями для зависи- мой переменной.....	12
3.1 Логит-модель	12
3.2 Пробит-модель.....	12
3.3 Тобит-модель	14
4 Описание программного продукта GRETL.....	16
5 Построение моделей	17
5.1 Описание переменных	17
5.2 Построение логит-моделей и пробит-моделей	18
ЗАКЛЮЧЕНИЕ	22
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	23

ВВЕДЕНИЕ

Эконометрика - это наука, изучающая конкретные количественные и качественные взаимосвязи экономических объектов и процессов с помощью математических методов и моделей. Эконометрические методы - это методы статистического анализа конкретных экономических данных. С помощью эконометрики выявляют новые, ранее неизвестные связи, уточняют или отвергают гипотезы о существовании определенных связей между экономическими показателями. Основной предмет исследования эконометрики - это массовые экономические явления и процессы.

Рынок недвижимости находится под значительным влиянием: факторов государственного регулирования, общеэкономической ситуации, микроэкономической ситуации, социального положения и природных условий в регионе. На уровень цен первичного и вторичного рынка влияют следующие переменные: фактор уровня жизни, политика городских властей в сфере недвижимости и нового строительства, уровень доходов населения, доступность кредитов и жилищных субсидий, ввод жилья, количество преступлений против личности, макроэкономические факторы (курс доллара, акции, цены на нефть и т.д.) [1].

Калибровка эконометрических моделей в данной работе проведена на примере российского рынка недвижимости по регионам за 2017-2019 г.г. Используются следующие данные: цена жилья (первичного, вторичного рынка); доходы населения; количество преступлений, направленных против личности; ввод жилья. В работе рассмотрены логит-, пробит- и тобит-модели. С помощью указанных моделей находим факторы, влияющие на ценообразование жилья в регионах Российской Федерации, и устанавливаем, как конкретно они влияют.

Структура работы имеет несколько разделов. В работе описан математический аппарат вышеперечисленных моделей, даны общие сведения о переключении режимов. В качестве источника для самостоятельных исследований выбран портал Федеральной службы государственной статистики [2]. С помощью программного обеспечения GRETЛ построены регрессионные модели. Результатам оценивания дана содержательная интерпретация. Проанали-

зированы отличия в результатах по годам, проведен сравнительный анализ по результатам, получаемым при анализе моделей с переключением режимов.

Рынок недвижимости невероятно подвижен. Изменения происходят на разных уровнях, а влияние на них оказывают и сами участники рынка. Если говорить о географии цен на жилье, то в разных регионах цены существенно отличаются друг от друга. В настоящее время на рынке недвижимости заметно выросло количество объектов, введенных в эксплуатацию, увеличился спрос на ипотечное кредитование. Если бы не ситуация с возникшей эпидемией, повлиявшей на экономику в стране и в мире, 2020 г. мог стать самым успешным за всю историю рынка жилой недвижимости.

1 Регрессионные модели

1.1 Линейная вероятностная модель

Рассмотрим линейную вероятностную модель. Она является простейшей моделью двоичного выбора [3]

$$Y_i = \begin{cases} 1, & \text{если событие произошло,} \\ 0, & \text{если не произошло.} \end{cases}$$

$$Y_i = \beta_1 + \beta_2 X_i + u_i, E(u_i) = 0 \quad (1.1)$$

$$Y_i = E(Y_i|X_i) + u_i \quad (1.2)$$

Величина зависимой переменной Y_i в i -м наблюдении содержит нестохастическую и случайную составляющие. Так, нестохастическая составляющая зависит от X_i и от параметров модели. Она равна математическому ожиданию Y_i при заданном X_i , $E(Y_i|X_i)$. Итак, найдем математическое ожидание Y_i , учитывая, что оно принимает два значения. Рассматриваем его как сумму детерминированной части и случайного возмущения

$$E(Y_i) = 0 \cdot P(Y_i = 0) + 1 \cdot P(Y_i = 1) = 0 \cdot (1 - p_i) + 1 \cdot p_i = p_i \quad (1.3)$$

Вероятность события p является линейной функцией объясняющей переменной

$$p_i = p(Y_i = 1) = \beta_1 + \beta_2 X_i \quad (1.4)$$

Таким образом, вероятностная функция представляет нестохастическую составляющую зависимость между X и Y . Если результирующая переменная

$Y_i = 1$, то случайный член $(1 - \beta_1 - \beta_2 X_i)$, если результат равен нулю, то случайный член $(-\beta_1 - \beta_2 X_i)$ [4]. Следовательно, распределение случайного члена включает только два его специфических значения. Распределение не является нормальным, оно не непрерывно. А возможное значение случайного члена меняется при изменении X , поэтому распределение гетероскедастично, то есть теоретическая дисперсия u_i равна $(\beta_1 + \beta_2 X_i) \cdot (1 - \beta_1 - \beta_2 X_i)$ и она меняется вместе с X_i .

Существуют две наиболее популярные формы зависимости - это логистическая функция, которая используется в логит-оценивании, и кумулятивное нормальное распределение, используемое в пробит-оценивании [5].

1.2 Уравнение парной линейной регрессии. Оценка параметров

Наиболее разработанной в эконометрике и наиболее простой, с точки зрения понимания, интерпретации и техники расчетов, является линейная форма регрессии [7].

Уравнение парной линейной регрессии

$$y = b_0 + b_1 x, \text{ где } b_0, b_1 - \text{параметры уравнения.} \quad (1.5)$$

Свободный член регрессионного уравнения

b_0 не имеет экономического смысла и показывает значение результативного признака y , если факторный признак $x = 0$.

Коэффициент регрессии

b_1 показывает на какую величину в среднем изменится результативный признак y , если переменную x увеличить на единицу измерения. Знак при коэффициенте регрессии показывает направление связи: при $b_1 > 0$ - связь прямая; при $b_1 < 0$ - обратная связь [8].

Нахождение значений параметров b_0, b_1 производится на основе совокупности наблюдений (выборки, матрицы наблюдений). Для разных выборок (даже одного объема) из генеральной совокупности будут найдены разные значения b_0, b_1 . Поэтому их рассматривают как приближенные значения (оценки) истинных параметров регрессионного уравнения. Сама процедура

нахождения приближенных значений также называется оценкой параметров. Для получения оценок параметров линейного регрессионного уравнения можно использовать метод наибольшего правдоподобия, метод наименьших модулей, метод минимакса и другие. Однако, согласно теореме Гаусса-Маркова, наилучшие результаты дает метод наименьших квадратов (МНК).

2 Модель множественной регрессии

Естественным обобщением линейной регрессионной модели с двумя переменными является многомерная регрессионная модель или модель множественной регрессии

$$y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, \quad t = 1, \dots, n,$$

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, \quad t = 1, \dots, n, \quad (2.1)$$

где x_{tk} — значения регрессора x_k в наблюдении t , а $x_{t1} = 1$, где $t = 1, \dots, n$. С учетом этого замечания далее не различаются модели вида (2.1) со свободным членом или без свободного члена.

2.1 Оценка качества уравнения множественной линейной регрессии

Необходимо оценить тесноту множественной линейной регрессионно-корреляционной связи и провести оценку качества модели, которая заключается в проверке уравнения регрессии и анализе остатков. Построение уравнения парной линейной регрессии дополняется оценкой тесноты связи между зависимой и независимыми переменными.

Регрессионная модель представляет сумму уравнения регрессии и остатков. Проверяется качество обоих слагаемых. Оценка качества уравнения линейной регрессии состоит из следующих этапов [14]:

1. Оценка математической точности уравнения. Для оценки математической точности уравнения рассчитывается средняя относительная ошибка аппроксимации

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \hat{y}_i}{y_i} \cdot 100, \quad (2.2)$$

где y_i и \hat{y}_i — соответственно фактические и теоретические значения переменной y .

2. Проверка статистической значимости уравнения регрессии в целом с помощью F-критерия Фишера. Выдвигается гипотеза H_0 : уравнение регрессии статистически незначимо, при конкурирующей гипотезе H_1 : уравнение регрессии статистически значимо. Находится расчетное значение (статистика) критерия

$$F_{\text{расч}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (\hat{y}_i - y_i^2)} \cdot \frac{n - m - 1}{m}, \quad (2.3)$$

$$F_{\text{расч}} = \frac{S_{\text{факт}}^2}{S_{\text{ост}}^2} \cdot \frac{n - m - 1}{m}, \quad (2.4)$$

$$F_{\text{расч}} = \frac{R_2}{1 - R_2} \cdot \frac{n - m - 1}{m}, \quad (2.5)$$

где y_i, \hat{y}_i, \bar{y} - соответственно фактическое (наблюдаемое), теоретическое и среднее значение y ; n - объем выборки, m - число параметров уравнения регрессии при независимых переменных, R^2 - коэффициент детерминации.

Табличное (критическое) значение $F_{\text{табл}}$ находится по таблице критических значений распределения Фишера-Снедекора (F-распределения) по уровню значимости α и двум числам степеней свободы $df_1 = m$ и $df_2 = n - m - 1$. Если $F_{\text{расч}} > F_{\text{табл}}$, то гипотеза H_0 отвергается с вероятностью ошибки α , то есть уравнение регрессии признается, в целом, статистически значимым. В противном случае уравнение считается незначимым.

3. Проверка статистической значимости оценок параметров $b_0, b_1, b_2, \dots, b_m$ с помощью t-критерия Стьюдента. Критерий Стьюдента проверяется только для линейного уравнения. Выдвигается гипотеза H_0 : параметр $b_j = 0$ ($j = 0, 1, 2, \dots, m$) (статистически незначим), при

конкурирующей гипотезе H_1 : параметр $b_j \neq 0$ (статистически значим).
Находятся расчетные значения критерия

$$t_{bj} = \left| \frac{b_j}{S_\varepsilon \sqrt{a_j}} \right|, \quad (2.6)$$

где S_ε - стандартное (среднее квадратическое) отклонение (ошибка) уравнения регрессии, которое определяется по формуле:

$$S_\varepsilon = \sqrt{\frac{\sum_{i=1}^n (y_{zi} - \hat{y}_i^2)}{n - m - 1}} \quad (2.7)$$

a_j - диагональные элементы матрицы $(X'X)^{-1}$. X - матрица значений независимых переменных x_1, x_2, \dots, x_m , размерность матрицы равна $n \times (m + 1)$. Первый столбец матрицы является единичным

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nm} \end{pmatrix} \quad (2.8)$$

X' - транспонированная матрица X , а $(X'X)^{-1}$ - обратная матрица

$$X' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1m} & x_{2m} & \dots & x_{nm} \end{pmatrix} \quad (2.9)$$

Теоретическое значение критерия $t_{\text{табл}}$ находится по таблице критических значений распределения Стьюдента по уровню значимости α и числу степеней свободы $df = n - m - 1$. Если $t_{bj} > t_{\text{табл}}$, то гипотеза H_0 отвергается с вероятностью ошибки α , то есть оценка коэффициента регрессии b_j признается статистически значимой и не является результатом действия внешних случайных факторов, в противном случае ($t_{bj} < t_{\text{табл}}$) - незначимой. Если статистическая значимость коэффици-

ента регрессии b_j не подтверждается, то следует вывод о несущественности в модели факторного признака x_j и необходимости его устранения из модели или замены на другой факторный признак.

4. Построение интервальных оценок (доверительных интервалов) параметров регрессии. Интервальные оценки параметров регрессии

$$b_j - m_{bj} \cdot t_{\text{табл}} < \beta_j < b_j + m_{bj} \cdot t_{\text{табл}} \quad (j = 0, 1, 2, \dots, m), \quad (2.10)$$

которые с надежностью (вероятностью) $\gamma = 1 - \alpha$ покрывают истинные параметры β_j . Если границы некоторого доверительного интервала имеют разные знаки, соответствующий параметр уравнения регрессии статистически незначим. Здесь $t_{\text{табл}}$ — значение, найденное по таблице критических точек распределения Стьюдента по уровню значимости $1 - \alpha$ и числу степеней свободы $df = n - m - 1$, где m_{bj} — стандартные ошибки коэффициентов регрессии. В случае уравнения регрессии с двумя независимыми переменными ($m = 2$) $\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2$. m_{bj} , они находятся по формулам

$$m_{b0} = \sqrt{S_\varepsilon^2 \cdot \frac{\frac{1}{n} + \bar{x}_1^2 \sum_{i=1}^n (x_{i2} - \bar{x}_2^2) + \bar{x}_2^2 \sum_{i=1}^n (x_{i1} - \bar{x}_1^2) - 2\bar{x}_1\bar{x}_2 \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\sum_{i=1}^n (x_{i2} - \bar{x}_2^2) \cdot \sum_{i=1}^n (x_{i1} - \bar{x}_1^2) - (\sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2))^2}}, \quad (2.11)$$

$$m_{b1} = \sqrt{S_\varepsilon^2 \frac{1}{\sum_{i=1}^n (x_{1i} - \bar{x}_1^2) \cdot (-r_{x_1x_2}^2)^2}}, \quad (2.12)$$

$$m_{b2} = \sqrt{S_\varepsilon^2 \frac{1}{\sum_{i=1}^n (x_{2i} - \bar{x}_2^2) \cdot (-r_{x_1x_2}^2)^2}} \quad (2.13)$$

Если уравнение регрессии имеет хорошую математическую точность, статистически значимо в целом и по отдельным параметрам, то оно признается качественным.

3 Модели двоичного выбора и с ограничениями для зависимой переменной

3.1 Логит-модель

Для логит-модели нужно рассмотреть логистическое распределение. Логистическая модель имеет зависимую переменную с двумя возможными значениями. В лог-коэффициентах значение «1» представляет собой линейную комбинацию из одного или нескольких независимых переменных. Соответствующая вероятность значения «1» изменяется в диапазоне от 0 до 1. Функция, которая преобразует лог-коэффициенты вероятности, является логистической функцией, отсюда и название логит-модели. Логистическая регрессия разработана в 1958 г. статистиком Дэвидом Коксом. В логит-оценивании предполагается, что вероятность события определяется функцией [16]

$$p = F(Z) = \frac{1}{1 + e^{-Z}}, Z = \beta_1 + \beta_2 X \quad (3.1)$$

Если Z стремится к бесконечности, то e^{-Z} стремится к нулю, и p будет ограничена сверху единицей. Если Z стремится к минус бесконечности, то e^{-Z} стремится к бесконечности, и p ограничена снизу нулем. Таким образом, нельзя предсказать, что искомая вероятность будет меньше нуля и больше единицы. Функция $f(Z)$ — это предельное воздействие величины Z на вероятность, представленное производной этой функции по Z . Если Z по модулю достаточно большое, то воздействие на вероятность достаточно мало, оно возрастает для небольших значений. F нелинейно зависит от β_1 и β_2

$$f(Z) = F'(Z) = \frac{e^{-Z}}{(1 + e^{-Z})^2} \quad (3.2)$$

3.2 Пробит-модель

Цель модели пробит в том, чтобы оценить вероятность попадания наблюдения с конкретными характеристиками в конкретную категорию [17]. Модель была введена Честером Блиссом в 1934 г., а быстрый метод для вычисле-

ния оценки максимального правдоподобия пробит был предложен Рональдом Фишером в качестве приложения к работе Честера Блиссона в 1935 г.

Пробит-анализ — это альтернативный подход к модели двоичного выбора. В качестве F выступает функция стандартного нормального распределения с нулевым математическим ожиданием и нулевой дисперсией. Определим переменную Z — линейную функцию переменных, определяющих искомую вероятность

$$Z = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (3.3)$$

Функция $F(Z)$ выражает вероятность данного события для каждого значения Z

$$p_i = F(Z_i) \quad (3.4)$$

Чтобы найти оценки параметров используется метод максимального правдоподобия [18]. Предельный эффект переменной X_i

$$\frac{\partial p}{\partial X_i} = \frac{dp}{dZ} \cdot \frac{\partial Z}{\partial X_i} = f(Z)\beta_i \quad (3.5)$$

Далее, так как $F(Z)$ — функция стандартного нормального распределения, тогда $f(Z)$ — производная $F(Z)$ и функция плотности вероятности стандартного нормального распределения

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} \quad (3.6)$$

В целом, логит-анализ и пробит-анализ демонстрируют схожие предельные эффекты, однако результат оценивания может сильно отличаться, если выборка не сбалансирована или содержит большое количество сходных наблюдений.

3.3 Тобит-модель

Тобит-модель была предложена Джейсом Тобином в 1958 г. для описания зависимости, которая появляется между неотрицательной зависимой переменной и независимой переменной. Тобит используется, когда зависимая переменная модели подвергается цензурированию, то есть за переменной ведется наблюдение до наступления некоторого события. Существует некоторая латентная переменная. Нам доступно Y^* , с предположением о наличии зависимости вида

$$Y^* = \beta_1 + \beta_2 X + u, \quad (3.7)$$

где зависимая переменная лежит выше нижней границы Y_L или ниже верхней границы Y_U [19]

$$Y = \begin{cases} Y^* & \text{для } Y^* > Y_L, \\ Y_L & \text{для } Y^* \leq Y_L. \end{cases} \quad (3.8)$$

Чтобы определить в какую категорию, регрессионную или ограниченную, входит наблюдение, нужно ответить на вопрос: полностью ли оно зависит от регрессоров и случайного члена или принадлежность наблюдения к регрессионной категории зависит от факторов, которые отличны от непосредственно включенных в модель. В этом случае может потребоваться общая спецификация модели с явным двухшаговым процессом. Модель выглядит следующим образом: на первом шаге принадлежность к одной из двух категорий зависит от чистого выигрыша латентной переменной B^* , которая зависит от $(m - 1)$ переменных Q_j и некоторого случайного члена ε

$$B_i^* = \sigma_1 + \sum_{j=2}^m \sigma_j Q_{ji} + \varepsilon_i \quad (3.9)$$

На втором шаге

$$Y_i^* = \beta_1 \sum_{j=2}^k \beta_j X_{ji} + u_1, \quad (3.10)$$

где $Y_i = Y_i^*$ для $B_i^* > 0$. Для наблюдения из выборки верно следующее

$$E(u_i | B_i^* > 0) = E\left(u_i | e_i > -\sigma_1 - \sum_{j=2}^m \sigma_j Q_{ji}\right) \quad (3.11)$$

Если ε_i и u_i распределены независимо, то вторая часть равенства превращается в безусловное $E(u_i)$, а сам процесс отбора не отражается на регрессионной модели.

Если ε и u коррелированы, проблема смещения при отборе возникает из-за того, что математическое ожидание $u \neq 0$ для наблюдений из выбранной категории.

4 Описание программного продукта GRETL

Пакет программ GRETL (GNU Regression Econometrics and Time Series Library) представляет собой инструментарий для практической реализации сложных вычислительных процедур эконометрического моделирования [21]. С 2002 г. GRETL является общедоступным, бесплатным продуктом с возможностью дальнейшей доработки открытых кодов.

Возможности программы:

1. Основные описательные статистики (среднее арифметическое, медиана, минимальное и максимальное значения, среднеквадратическое отклонение, коэффициент изменчивости (вариации), коэффициент асимметрии, коэффициент эксцесса).
2. Проверка нормальности распределения, распределение частот случайной величины, распределение плотности вероятностей, определение коэффициентов корреляции и т.д.
3. Предусматривает непосредственный доступ к статистическим таблицам. Существует возможность вычисления критических значений, *p*-value.
4. Регрессионный анализ, взвешенный МНК, двухшаговый МНК — оценка систем одновременных уравнений, методы оценивания логитовых, пробитовых и тобитовых моделей, нелинейных моделей и т.д.
5. Экспорт и импорт GRETL-Microsoft Excel и текстовые редакторы (Notepad и т.д.).

5 Построение моделей

5.1 Описание переменных

В работе рассматриваются данные по 84 регионам России 2017-2019 г.г., за исключением Чукотского АО, официальные данные по которому на сайте Федеральной службы государственной статистики для рассматриваемых показателей отсутствуют. Также, не учитывались автономные округа, которые являются частью других регионов и имеют общую с ними статистику.

В работе были рассмотрены следующие факторы:

1. Индекс цен первичного рынка жилья. Он измеряет изменение средней цены на жилые здания и помещения, которые еще не находятся в собственности.
2. Индекс цен вторичного рынка жилья. Он рассчитывается на основе зарегистрированных цен на квартиры функционирующего жилого фонда, которые находятся в собственности и являются объектами совершения рыночных сделок. Наблюдение ведется по выборочному кругу организаций, осуществляющих операции с недвижимостью в территориальных центрах и отдельных городах субъектов Российской Федерации.
3. Среднедушевые денежные доходы населения. Они включают доходы лиц, занятых предпринимательской деятельностью, выплаченную заработную плату наемных работников, социальные выплаты (пенсии, пособия, стипендии, страховые возмещения и прочие выплаты), доходы от собственности в виде процентов по вкладам, ценным бумагам, дивидендам и другие доходы («скрытые» доходы, доходы от продажи иностранной валюты, денежные переводы).
4. Преступления против личности. Это преступные деяния, посягающие на безопасность жизни, здоровья, свободу, честь и достоинство, половую неприкосновенность, конституционные права и свободы человека и гражданина, интересы семьи и несовершеннолетних.
5. Ввод жилья. Ввод общей площади жилых домов — это общая площадь жилых помещений, во введенных в эксплуатацию жилых и нежилых зданиях, жилых домах, которая определяется как сумма площадей всех частей жилых помещений, включая площадь помещений вспомога-

ного использования, предназначенных для удовлетворения гражданами бытовых и иных нужд, связанных с их проживанием в жилом помещении, площадей лоджий, балконов, веранд, террас, подсчитываемых с соответствующими понижающими коэффициентами, а также жилых и подсобных помещений в построенных населением индивидуальных жилых домах.

Уровень ввода жилья является интегральным показателем. Он характеризует состояние отдельных секторов экономики региона, уровень платежеспособного спроса в регионе, эффективность усилий региональных властей по развитию жилищного рынка и строительного сектора. Более высокие объемы жилищного строительства соответствует более высокому уровню экономического здоровья региона.

Эконометрические расчеты были проведены с помощью GRETL. Были проанализированы модели со следующими объясняющими значениями: показатели преступности, среднедушевые доходы населения и ввод жилья. С их помощью можно объяснить, что влияет на зависимые переменные, связанные с рынком цен жилья.

5.2 Построение логит-моделей и пробит-моделей

В случае логит-модели и пробит-модели, режим это выбор между нулем и единицей. Переключение осуществляется детерминированно, по тому, больше или меньше некоторая линейная комбинация нуля.

В качестве зависимых переменных для моделей логит и пробит были выбраны следующие:

1.

$$\{y1 = index_{primary_1} > index_{secondary_1}\}$$

— бинарный индикатор того, что индекс первичной цены жилья больше индекса вторичной за 2017 г.

2.

$$\{y2 = index_{primary_2} > index_{secondary_2}\}$$

— бинарный индикатор того, что индекс первичной цены жилья больше индекса вторичной за 2018 г.

3.

$$\{y3 = index_{primary3} > index_{secondary3}\}$$

— бинарный индикатор того, что индекс первичной цены жилья больше индекса вторичной за 2019 г.

4.

$$\{y4 = index_{primary1} > 98.16\},$$

где 98.16 — среднее значение индекса первичной цены за 2017 г.

5.

$$\{y5 = index_{primary2} > 99.9\},$$

где 99.9 — среднее значение индекса первичной цены за 2018 г.

6.

$$\{y6 = index_{primary3} > 104.96\},$$

где 104.06 — среднее значение индекса первичной цены за 2019 г.

7.

$$\{y7 = index_{secondary1} > 96.92\},$$

где 96.92 — среднее значение индекса вторичной цены за 2017 г.

8.

$$\{y8 = index_{secondary2} > 98.6\},$$

где 98.6 — среднее значение индекса вторичной цены за 2018 г.

9.

$$\{y9 = index_{secondary3} > 103.47\},$$

где 103.47 — среднее значение индекса вторичной цены за 2019 г.

Были построены логит-модели по данным за 2017-2019 г.г. В работу включены модели только со значимыми переменными.

В 2017 г. на $y1$ влияет фактор преступности с уровнем значимости (*) 10 % (таблица 1).

Зависимая переменная	Фактор	Коэффициент	р-значение	Корректно предсказанных случаев	Критерий отношения правдоподобия
y1	Преступность	0,00487269	0,0988 (* -10%)	69%	0,2980

Таблица 1 - Логит-модель по данным за 2017 г.

Также с (*) 10 % уровнем на y_4 влияет фактор дохода населения (таблица 2). Так как коэффициенты при обоих регрессорах положительны, при их увеличении растут зависимые переменные. Количество корректно предсказанных случаев и критерий отношения правдоподобия с фактором преступности составляют 69 % и 0,2980, что превышает аналогичные показатели с фактором дохода населения.

Зависимая переменная	Фактор	Коэффициент	р-значение	Корректно предсказанных случаев	Критерий отношения правдоподобия
У4	индексы по доходам населения	0,252494	0,0758 (* -10%)	61,9%	0,1212

Таблица 2 - Логит-модель по данным за 2017 г.

На y_5 в 2018 г. влияет фактор преступности с уровнем значимости (*) 10 % (таблица 3). При увеличении регрессора растет зависимая переменная. Количество корректно предсказанных случаев 56 %. Критерий отношения правдоподобия составляет 0,2758.

Зависимая переменная	Фактор	Коэффициент	р-значение	Корректно предсказанных случаев	Критерий отношения правдоподобия
y5	Преступность	0,00458651	0,0733 (* -10%)	56%	0,2758

Таблица 3 - Логит-модель по данным за 2018 г.

В 2019 г. на y_3 влияет фактор ввода жилья с (* *) 5 % уровнем значимости (таблица 4).

Зависимая переменная	Фактор	Коэффициент	р-значение	Корректно предсказанных случаев	Критерий отношения правдоподобия
у3	Индекс ввода жилья на душу населения	0,0027521	0,0355 (**-5%)	66,7%	0,1309

Таблица 4 - Логит-модель по данным за 2019 г.

На y_9 влияет показатель преступности также с (* *) 5 % уровнем значимости (таблица 5), тогда как на y_6 влияет показатель преступности с (* *) 1 % уровнем значимости (таблица 6).

Зависимая переменная	Фактор	Коэффициент	р-значение	Корректно предсказанных случаев	Критерий отношения правдоподобия
у9	Преступность	0,00636592	0,0295 (**-5%)	66,7%	0,1205

Таблица 5 - Логит-модель по данным за 2019 г.

Зависимая переменная	Фактор	Коэффициент	р-значение	Корректно предсказанных случаев	Критерий отношения правдоподобия
у6	Преступность	0,0100868	0,0035 (***-1%)	66,7%	0,0025

Таблица 6 - Логит-модель по данным за 2019 г.

Так как в моделях 2019 г. коэффициенты при всех трех регрессорах положительны, при их увеличении растут зависимые переменные. Количество корректно предсказанных случаев во всех трех моделях одинаково и составляет 66,7 %.

За весь период с 2017 по 2019 г.г. коэффициенты при регрессоре в логит-модели положительны. Это свидетельствует о том, что при увеличении регрессора, растет зависимая переменная, а именно вероятность, связанная с индексами цен жилья, от которых зависит сама цена недвижимости. Таким образом, наибольшее влияние на вероятность, связанную с ценой жилья, за весь период оказывал фактор преступности, а наименьшее фактор ввода жилья и фактор дохода населения.

ЗАКЛЮЧЕНИЕ

В дипломной работе были рассмотрены теоретические аспекты эконометрики, регрессионные модели, критерии их оценивания, а также модели с переключением режимов. В качестве моделей с переключением режимов изучены логит-модель, пробит-модель и тобит-модель. В рассмотренных моделях выбор (переключение) происходит по вероятности события, в котором линейная комбинация больше или меньше нуля. Для построения моделей был изучен и применен программный продукт Gretl. В процессе построения моделей на основе данных Федеральной службы государственной статистики были сделаны выводы о влиянии регрессоров на зависимые переменные.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Грибовский, С.В.* Оценка стоимости недвижимости: Учебное пособие / С. В. Грибовский — М.: Маросейка, 2009. - 432 с.
2. *Федеральная служба государственной статистики* [Электронный ресурс]: - URL: <http://www.gks.ru/> (дата обращения 01.03.2020) - Яз. рус.
3. *Буравлёв, А.И.* Эконометрика : учебное пособие / А. И. Буравлёв. — М.: БИНОМ. Лаборатория знаний, 2012. — 164 с
4. *Бородин, С.А.* Эконометрика : учеб, пособие / С. А. Бородин. — Минск : Новое знание, 2001.
5. *Кремер, Н.Ш.* Эконометрика : учебник и практикум для академического бакалавриата / Н. Ш. Кремер, Б. А. Путко ; под ред. Н. Ш. Кремера. — 4-е изд., испр. и доп. — М. : Издательство Юрайт, 2019. — 308 с.
6. *Бабешко, Л.О.* Основы эконометрического моделирования / Л. О. Бабешко. — М. : КомКнига, 2006.
7. *Домбровский, В.В.* Эконометрика: учебник / В. В. Домбровский. — М.: Новый учебник, 2004. — 342 с.
8. *Соколов, Г.А.* Введение в регрессионный анализ и планирование регрессионных экспериментов в экономике: Учебное пособие / Г.А. Соколов, Р.В. Сагитов. - М.: Инфра-М, 2016. - 352 с.
9. *Горидько, Н.П.* Современный экономический рост: теория и регрессионный анализ: Монография / Н.П. Горидько, Р.М. Нижегородцев. - М.: Инфра-М, 2017. - 444 с.
10. *Александрова, Н.В.* История математических терминов, понятий, обозначений: словарь-справочник. / Н.В. Александрова — 3-е изд.. — М.: ЛКИ, 2008. — 248 с.
11. *Айвазян, С.А.* Прикладная статистика и основы эконометрики. / С.А. Айвазян, Мхитарян, В.С. — М.: ЮНИТИ, 1998.

12. *Козинова, А.Т.* Эконометрика. Курс лекций. – Учебно-методическое пособие. / А. Т. Козинова, А.А. Отделкина. – Н. Новгород, 2004 – 95 с.
13. *Новиков, А.И.* Эконометрика: Учебное пособие. / А.И. Новиков. – М.: ИНФРАМ, 2003. – 106 с.
14. *Доугерти, К.* Введение в эконометрику: Пер. с англ. / К. Доугерти. — М: ИНФРА-М, 2009— XIV, 465 с.
15. *Елисеева, И.И.* Эконометрика : учебник / И.И. Елисеева. — М. : Финансы и статистика, 2007.
16. *Носко, В.П.* Эконометрика. Кн. 2. Ч. 3, 4: учебник / В.П. Носко. – М. : Издательский дом “Дело”, 2011. – 576 с.
17. *Вербик, М.* Путеводитель по современной эконометрике. Пер. с англ. В. А. Банникова. Научн. ред. и предисл. С. А. Айвазяна. / М. Вербик. – М. : Научная книга, 2008. – 616 с.
18. *Чучуева, И.А.* Модель прогнозирования по выборке максимального подобия / И.А. Чучуева.- М.: Московский государственный технический университет им. Н.Э. Баумана. Москва, 2012.
19. *Магнус Я.* Эконометрика. Начальный курс. / Я. Магнус, П. Катышев, А. Пересецкий. М.: «Дело», 2004, с.337-349
20. *Гладилин, А.В.* Эконометрика: учебное пособие для вузов / А.В. Гладилин, А.Н. Герасимов, Е.И. Громов.- 2-е изд., стереотип.- М.:КНОРУС,2008.-226.
21. *Руководство пользователя Gretl* [Электронный ресурс]: - URL: <http://gretl.sourceforge.net/ru.html> (дата обращения 01.03.2020). - Загл. с экрана. - Последнее изменение страницы: 11 апреля 2020 года. - Яз. рус.