**Правительство Российской Федерации**

**федеральное государственное автономное образовательное учреждение высшего профессионального образования**

**"Национальный исследовательский университет   
"Высшая школа экономики"**

**Факультет экономики**

**Кафедра прикладной математики и моделирования в социальных системах**

Допускаю к защите

Заведующий кафедры ПМиМСС, к. э. н., доцент кафедры ПМиМСС,

Потапов Д. Б.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

«\_\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2013 г.

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

На тему: **ЭМПИРИЧЕСКАЯ ОЦЕНКА СПРОСА НА ИПОТЕЧНЫЕ КРЕДИТНЫЕ ПРОДУКТЫ**

Студент группы Э-09-1

Кылосова Мария Евгеньевна

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

Научный руководитель

Преподаватель кафедры Прикладной математики и моделирования в социальных системах

Ожегов Евгений Максимович

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

Пермь, 2013 г.

Оглавление

[Введение 3](#_Toc357470661)

[Глава 1. Теоретические подходы к моделированию функции спроса 6](#_Toc357470662)

[1.1. Обзор литературы 6](#_Toc357470663)

[1.2. Модель Хекмана и инструментальные переменные 17](#_Toc357470664)

[Глава 2. Эмпирическая оценка спроса по данным ПАИЖК 24](#_Toc357470665)

[2.1. Описание переменных 24](#_Toc357470666)

[2.2. Описание модели 31](#_Toc357470667)

[2.3. Построение модели 33](#_Toc357470668)

[Заключение 53](#_Toc357470669)

[Список использованной литературы 55](#_Toc357470670)

[Приложения 58](#_Toc357470671)

# Введение

За последние несколько лет количество ипотечных организаций в России значительно возросло. По сравнению с концом 2009 года, количество кредитных организаций, предоставляющих ипотечные жилищные кредиты, увеличилось на 15% [29]. Таким образом, вопрос о возможных конкурентных преимуществах и разработке оптимальных продуктов встает особенно остро.

Именно спрос на продукт в большей степени определяет поведение организации на рынке, поэтому любой производитель товаров или услуг должен стремиться уменьшить информационную асимметрию, которая может быть связана с потребительскими рисками и предпочтениями. Наличие качественной информации о функции спроса и особенностях потребителей является значимым конкурентным преимуществом для банка.

Однако в настоящее время в большинстве случаев условия кредита разрабатываются банками без учета потребительских характеристик: их предпочтений и ограничений. Таким образом, основная проблема заключается в отсутствии модели и, как следствие, индивидуальных оптимальных условий кредита, устанавливаемых по единой общепринятой схеме. В связи с этим существующие кредитные продукты могут быть не оптимальными для клиентов с разными социально-демографическими чертами.

Следовательно, оценка спроса на продукты ипотечного кредитования является актуальной и достаточно новой темой для России.

Целью данной выпускной квалификационной работы является нахождение функции спроса на ипотечные кредитные продукты, а так же нахождение и интерпретация коэффициентов эластичности спроса по процентной ставке и сроку жизни кредита (т. е. по основным условиям контракта).

Основой для настоящего исследования является база данных о заемщиках, предоставленная ОАО «Пермское агентство по ипотечному жилищному кредитованию», где собраны социально-демографические характеристики заемщиков и характеристики выданных им ипотечных кредитов.

В соответствии с поставленной целью можно сформулировать следующие задачи:

а) Анализ общих моделей на основе имеющихся данных и проведенных ранее исследований в аналогичной области;

б) Введение предпосылок для построения моделей, максимально приближенных к реальной ситуации, но позволяющих работать с данными на доступном исследователю уровне;

в) Разработка кода для моделирования данного процесса в программном пакете Stata;

г) Выбор оптимальной модели и получение адекватных коэффициентов;

д) Экономическая интерпретация результатов.

В соответствии с поставленной целью и задачами работа состоит из следующих частей: во-первых, анализ существующих наиболее значимых исследований по данной теме, общей теоретической базы и ограничений, в рамках которых были сделаны предшествующие работы; во-вторых, практическая часть исследования. В этом разделе произведен анализ имеющихся данных, построена модель, с помощью которой в конечном итоге моделируется уравнение спроса на ипотечные кредитные продукты.

Предполагается, что уровень дохода заемщика значим на любой стадии процесса получения ипотечного кредита. Также выдвигается гипотеза о том, что эластичность спроса по сроку кредита выше, чем по ставке, а так же гипотеза о значимости смоделированного показателя смещения в итоговом уравнении спроса.

Объектом настоящего исследования является спрос на продукты ипотечного кредитования в городе Пермь.

Предмет исследования — социально-демографические характеристики потребителей, а так же характеристики кредитных продуктов, оказывающие влияние на спрос.

Результаты работы могут быть использованы при разработке новых ипотечных продуктов и для прогнозирования реакции потребителей на новинки или изменения в текущих кредитных продуктах.

# Глава 1. Теоретические подходы к моделированию функции спроса

## 1.1. Обзор литературы

Эмпирическая оценка спроса на кредитные продукты представляет собой процесс, состоящий из нескольких шагов. В целом, можно выделить три основных этапа.

Во-первых, потенциальный заемщик осуществляет выбор: подавать ему заявки на получение кредита или не подавать. Данное уравнение участия должно быть смоделировано, чтобы итоговые оценки параметров в уравнении спроса не были смещенными.

Второй шаг – анализ процесса принятия решений банком. Иными словами, на втором этапе оценивается наличие дискриминации по какому-либо признаку при одобрении заявок на кредит и дальнейшей ценовой политики.

Последний шаг – моделирование итогового уравнения спроса на кредитные продукты в случае прохождения двух предыдущих шагов и согласия клиента с предлагаемыми условиями контракта.

В большинстве работ анализ всех трех этапов не проводится: внимание фокусируется либо на комбинации первого и третьего этапов, либо отдельно на втором.

Первая группа авторов исследует значения эластичности спроса на кредитные продукты по сроку жизни кредита и ставке процента в зависимости от уровня дохода потребителей. Вторая группа анализирует критерии, которые использует банк во время принятия решений о выдаче кредита и во время выбора условий контракта.

Впервые исследование по теме влияния ограничений ликвидности потребителей (то есть уровня их достатка) на эластичность спроса на кредиты по процентной ставке и сроку жизни было проведено в работе «Consumer Sensitivity to Finance Rates: An Empirical and Analytical Investigation» (Juster et al., 1964).

Исследователи фокусировалось на двух основных вопросах. Во-первых, это изучение взаимосвязей между указанными (в анкетах) планами покупки и другими переменными: ожидания касательно персонального финансового состояния и общих условий для ведения бизнеса, бытовой статус домохозяйств в отношении прошлых трат, долги, сбережения и имеющиеся активы. Во-вторых, это причины, по которым домохозяйства отклоняются от запланированных покупок, а также факторы, ведущие к незапланированным тратам.

На основе выборки, содержащей 25 000 наблюдений, были выявлены основные характеристики домохозяйств, влияющие на планы и непосредственно на фактические покупки: это доход домохозяйства и возраст главы семьи. В частности, для домохозяйств с одинаковым доходом и возрастом главы семьи изменение дохода на 20% или меньше за последнее время имело небольшое влияние на фактические или планируемые покупки товаров длительного пользования.

Также достаточно старой, но известной работой является статья  **«**Consumption and liquidity constraints: An empirical investigation» (Zeldes, 1985). В исследовании автор проверял гипотезу о значимости ограничений ликвидности домохозяйств в процессе максимизации их полезности.

С помощью системы уравнений Эйлера автор рассматривает максимизацию функции полезности без учета ограничений ликвидности и с их учетом. Для оценки был использован двухшаговый метод наименьших квадратов.

В большинстве случаев результаты проведенных тестов подтверждали выдвинутую автором гипотезу. В частности, найденная чувствительность потребления к текущему нерегулярному доходу была выше, чем в рамках гипотезы перманентного дохода.

В статье «The Credit-Constrained Consumer: An Empirical Study of Demand and Supply in the Loan Market» (Perraudin et al., 1992) произведен эмпирический анализ домохозяйств с ограниченной ликвидностью в моделях спроса и предложения на рынке ипотечного кредитования.

Основным выводом работы является выявление основных факторов, влияющих вероятность подачи заявки на выдачу ипотечного кредита в банк. Этими факторами являются пол, раса, материальный статус, размер домохозяйства. Темнокожий болезненный глава семьи, проживающий в большом городе, наименее склонен обращаться за ипотечным кредитом в банк.

Авторы моделируют проблему межвременной оптимизации полезности, которая является квадратичной функцией от потребления домохозяйства в условиях бесконечной жизни индивида и совершенного рынка. Затем моделируется порядковая нелинейная логит-модель, которая включает в себя фиктивные переменные, учитывает деление потребителей на три типа, и где социально-демографические характеристики представлены в виде линейных функций. Также вводилась предпосылка об экзогенности всех переменных, включенных в уравнение.

Более прогрессивный анализ значимости кредитных ограничений потребителей и анализ эластичности спроса по ставке процента был проведен в работе «Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data» (Gross et al., 2002).

В работе было доказано, что ослабление кредитных ограничений ведет к быстрому и значительному увеличению объема долга, что противоречит гипотезе перманентного дохода. Эластичность спроса к ставке процента составила в среднем -1.3 в долгосрочном периоде и -0,8 в краткосрочном периоде, однако значение эластичности варьируется для групп потребителей, которые имеют разные ограничения ликвидности.

В исследованиях в данной области чаще всего встречается наличие эндогенных параметров. В данном случае, например, если эмитент увеличивает количество выпущенных кредитных карт в преддверье увеличения спроса.

Для решения проблемы эндогенности были введены инструментальные переменные. Более детально метод инструментальных переменных будет рассмотрен после обзора литературы. Через метод наименьших квадратов были оценены параметры из двух уравнений зависимости изменения объема кредитов от изменения (по сравнению с прошлым месяцем) кредитного лимита конкретного заемщика (в модели учтены 12 лагов, контролирующие и фиктивные переменные) и от изменения процентной ставки (аналогичное уравнение, но включающее только 9 лагов).

В работе «A structural econometric model of price discrimination in the French mortgage lending industry» (Gary-Bobo et al., 2003) исследуется схожая область, но с точки зрения ценовой дискриминации. То есть в статье рассматриваются две основные модели пересечения спроса и предложения на ипотечном рынке: первая модель построена в рамках конкурентного рынка, вторая – в рамках дискриминационной монополии. Конкурентная модель основывается на предельных условиях с добавленным условием нулевой прибыли, дискриминационная модель основывается на максимизации прибыли.

Для модели дискриминационной монополии, которая была выбрана как наиболее удачная, существует «социальная дискриминация», то есть условия кредита зависят и от заработной платы заемщика, и от величины задатка, а так же от профессионального статуса заемщика. Также было выявлено, что ценовая эластичность спроса разнится для групп людей с различным профессиональным статусом и обратно пропорциональна ставке процента.

В моделях процентная ставка и величина кредита являются функцией от видимых характеристик заемщиков. Зависимые переменные выводятся из двух уравнений (предложения и спроса). Также авторы ввели предпосылку о том, что эндогенной переменной является величина выплат по погашению кредита, а не непосредственно процентная ставка по кредиту.

Сначала моделировались отдельно функции спроса, куда включаются характеристики домохозяйства, функция полезности и условие того, что кредит был оформлен, и предложения в двух случаях. Для конкурентного рынка функция предложения была выведена через условия нулевой прибыли, для дискриминационного – с помощью меню контрактов.

Каждая версия представляла собой бивариативную нелинейную модель, оцениваемую с помощью метода максимального правдоподобия.

Еще одной работой, в которой проводилось исследование эластичности спроса на рынке ипотечных кредитов по ставке процента, является статья «The impact of mortgage interest-rate subsidies on household borrowing» (Martins et al., 2005). Исследователями было проведена оценка эффекта реформы по субсидированию процентных ставок на ипотечные кредиты для домохозяйств с низким и средним достатком.

По результатам исследования эластичность спроса на ипотечные займы по процентной ставке варьируется от 1,33% до 2,8%. Также было найдено, что молодые домохозяйства с низким уровнем дохода (соответственно) менее чувствительны к изменениям процентных ставок чему другие группы потребителей.

При условии того, что индивид оформил ипотечный кредит, было сформировано уравнение, включающие: размер долга, величина получаемой субсидии (3 уровня), время выдачи кредита относительно прохождения реформы, дороговизну жилья, социально-демографические характеристики заемщика. Переменные времени и стоимости приобретаемой недвижимости были заменены фиктивными переменными. Затем коэффициенты полученного уравнения были оценены с помощью тройных разностей и пробит-моделей.

В работе «The Home Purchase Mortgage Preferences of Low- and Moderate-Income Households» (LaCour-Little, 2007) автор фокусируется на исследовании предпочтений в выборе ипотечной программы домохозяйствами с низким и среднем уровнем дохода.

В целом, результаты показали, что потребители избегают более дорогих альтернатив, кредитные оценки заемщиков и другие факторы риска определяют выбор контракта. Было выявлено, что все дополнительные возможности, описанные в статье и отличные от традиционных черт ипотечного кредита (длительное время на ожидание ответа банка по заявке, большой первоначальный взнос и др.) делают кредит более доступным.

На основе выборки, содержащей 226,933 кредитных договоров, автор оценил пять уравнений, зависимыми переменными в которых были: LTV (отношение размера ипотечного займа к стоимости недвижимости), вероятность того, что заемщик выберет ипотечную программу, требующую минимальное количество документов при подаче заявления, цена кредита, время от подачи заявки до решения банка, вероятность выбора заемщиком определенной программы.

Для оценивания параметров уравнений использовались логистическая регрессия и полиномиальная логит-регрессия.

Для контроля над эндогенностью были использованы расчетные значения эндогенных величин, посчитанные по методу инструментальных переменных.

Статья «Observing Unobservables: Identifying Information Assymetries with a Consumer Credit Field Experiment» (Karlan et al., 2009) посвящена исследованию информационной асимметрии и морального риска. В работе рассматриваются реальные последствия кредитных ограничений через наличие и величину двух провалов рынка кредитов: неблагоприятного отбора и морального риска.

Было найдено доказательство существования морального риска и неблагоприятного отборы. Ориентировочно от 7% до 16% невыплат заемщиками долга связано с асимметрией информации при заключении сделок.

Авторами по особой методике была смоделирована несмещенная выборка, затем на основании этих данных были построены регрессии, оцененные с помощью МНК.

Один из наиболее интересных трудов, посвященных эмпирической оценке спроса на товары длительного пользования – это работа «Credit Constraints in the market for consumer durables: evidence form micro data on car loans» (Attansio et al., 2008).

Для авторов наиболее интересной представлялась оценка эластичности спроса на кредиты по процентной ставке и сроку жизни относительно различных потребительских групп, сформированных по их вероятности наличия ограничения ликвидности, то есть по уровню дохода.

Было выявлено, что группы потребителей с более низким доходом более чувствительны к изменениям в сроках жизни кредита и менее – к изменениям процентной ставки. В частности, увеличение срока жизни кредита на 1 года ведет к увеличению спроса в среднем на 88,5%. Группы потребителей с высоким уровнем дохода наоборот более чувствительны к изменению процентной ставки, а не к сроку жизни кредита. Различий в чувствительности разных возрастных групп потребителей выявлено не было.

Другое значимое отличие это работы заключается в том, что исследование проводилось не на основе данных по потреблению, а на основе данных по выданным кредитам на покупку машины. База данных для исследования была получена через «Опрос Потребительских Расходов» (CES), проведенным Бюро статистики труда. Данные являются ротационной панелью, где каждое домохозяйство было опрошено 4 раза за один год. Каждый квартал 25% выборки замещалось новыми домохозяйствами.

Согласно статье, условия кредита являются эндогенными и варьируются для различных групп потребителей.

В ходе работы с данными исследователи столкнулись с некоторыми трудностями, которые будут свойственны всем аналогичным работам. Во-первых, крайне вероятно наличие смещения выборки, поскольку наблюдения доступны лишь по тем потребителям, которые выбрали автокредит среди всех альтернатив. Во-вторых, важным свойством кредитного контракта является то, что финансирование будет варьироваться от 0 до полной стоимости машины. В-третьих, также важна одновременность вопросов: наблюдаемая процентная ставка и срок жизни кредита, вероятнее всего, являются эндогенными переменными (с экономической и эконометрической точки зрения). Наконец, предположение о нормальности распределения переменных в рамках данной работы является неуместным.

Для решения проблемы эндогенности переменных были введены следующие инструменты: налоговая реформа (1986) для посленалоговой процентной ставки и срок службы машины (темп амортизации) для срока жизни кредита.

Оцениваемая зависимость имеет вид (1):

(1)

Зависимая переменная (L\* – желаемая сумма кредита) выражена через логарифм, поскольку сумма кредита не может быть отрицательной. Вектор x – вектор переменных, отражающих социально-демографические черты и макроэкономические эффекты, воздействующие на процесс. Переменные r и m – это процентная ставка и срок жизни кредита, являющиеся эндогенными параметрами, ε – ненаблюдаемая ошибка.

Для получения корректных результатов при наличии вышеуказанных проблем выборки, авторы далее работали с порядковой моделью бинарного выбора. В данной модели каждому домохозяйству в выборке соответствует переменная d, которая принимает 3 значения: 0, 1 и 2. Каждое значение означает одно решение, соответственно «не покупать/не оплачивать», «оплатить меньше, чем 100% стоимости машины», «оплатить 100%стоимости машины». Промежутки (−∞, a1), (a1, a2), и (a2, ∞) ― отрезки прямой, где а1 и а2 являются неизвестными величинами. Следовательно, было получено следующее уравнение участия домохозяйства в кредитной сделке (2):

d = (2)

Эндогенность двух параметров была учтена следующим образом (3), (4):

(3)

(4)

Таким образом, по методу максимального правдоподобия будет оцениваться полином первого порядка относительно переменных r и m (включенных в вектор X) следующего вида (5):

(5)

В дальнейшем в статье были приведены два трехшаговых полупараметрических метода.

Первым этапом для всех методов является оценка параметров уравнения участия в сделке.

В первом методе использовалась модель Хекмана (тобит-модель второго типа). На втором шаге параметры, представляющие значения процентной ставки и срока жизни кредита, были оценены через взвешенный метод наименьших квадратов с использованием ядерного оценивания функции плотности вероятности.

В основе второго метода лежит предположение, что условное среднее значение ошибки в уравнениях отбора и экзогенных регрессоров зависит только от вероятности p. Далее параметр p в полученных с учетом предположенной зависимости уравнениях был оценен через ряд аппроксимаций, а оценка параметров при процентной ставке и сроке жизни кредита была произведена по методу наименьших квадратов степенного полинома и с условием участия индивида в сделке.

Не менее важно исследование другой стороны процесса ипотечного кредитования: процесс принятия решения кредиторами.

В работе «Unfair Lending: The Effect of Race and Ethnicity on the Price of Subprime Mortgages» (Bocian et al., 2008) произведен анализ различий в ипотечном ценообразовании. Основной вопрос исследования ― вопрос о существовании различий в процессе ипотечного ценообразования (после проверки основных факторов риска) для потребителей с разной расовой и этнической принадлежностью.

Результаты, полученные по 177000 наблюдений, предоставленных Home Mortgage Disclosure Act (HMDA), показали, что афроамериканские и латиноамериканские заемщики, скорее всего, получат ипотеку по более высокой ставке процента, даже с учетом основных факторов риска (отношение LTV, кредитная оценка заемщика).

Исследование было проведено с помощью множественного регрессионного анализа: были построены логистические модели с использованием показателя соотношения шансов.

Особое внимание в данной работе было уделено проблеме эндогенности. Для переменных LTV и размером кредита, которые являются эндогенными, была построена система из двух одновременных уравнений и использована оценка с помощью трехшагового МНК (3 SLS). Для учета эндогенности переменных итоговой суммы выплат, первоначального взноса и вида ставки по кредиту (фиксированная или плавающая) исследователи разделили всю выборку на подвыборки, затем проведя анализ по полученным группам.

Логика регрессионного анализа в данной работе схожа с анализом, проведенным в статье «The Effect of Conforming Loan Status on Mortgage Yield Spreads: A Loan Level Analysis» (Ambrose et al., 2004), где особое внимание также было уделено проблеме эндогенности и смещениям выборки.

Исследователи сравнивали ипотечный спрэд доходности кредитов, выдаваемых компаниями, которые финансируются правительством, и обычными компаниями на первичном ипотечном рынке с учетом различий в кредитных рисках.

Авторы использовали систему из двух одновременных уравнений для инструментирования суммы кредита и LTV (Loan-To-Value Ratio, отношение величины долга к стоимости квартиры).

Затем предсказанный через систему показатель LTV был использован в итоговом уравнении с зависимой переменной в виде ставки спрэда, смоделированном при помощи метода OLS, то есть весь процесс оценки является усложненным двухшаговым методом наименьших квадратов (2SLS).

В работе «Fair Lending Analysis of Mortgage Pricing: Does Underwriting Matter?» (Zhang, 2010) исследовалось наличие возможной взаимосвязи между ипотечным андеррайтингом и ценовыми решениями для анализа ценообразования в данной сфере.

В статье доказывается наличие потенциальных различий в ценообразовании на ипотечные кредиты при принятии во внимание решений по другим кредитам. В частности, обосновывается наличие систематической ошибки выборки, связанной с андеррайтингом и ценообразованием.

Основой исследования послужила выборка, составленная по информации от двух крупных банков. Результаты были получены с помощью метода Хекмана и метода инструментальных переменных.

В подтверждении свойств полученных оценок были произведены 2 симуляции по методу Монте-Карло.

Работой, также посвященной вопросу различного ценообразования для людей разных рас и этнических принадлежностей, является исследование «Race, Redlining, and Subprime Loan Pricing» (Ghent et al., 2012).

Было доказано применения практики "красной черты" (отказ в выдаче ссуды по закладной на дома в старых или трущобных районах; часто по расовым соображениям) и неблагоприятном ценообразовании на ипотечные кредиты для чернокожих людей и латиноамериканцев. Более того, географическое окружение также влияет на установление цены для конкретного заемщика.

База данных была скомпилирована из двух баз, предоставленных корпорацией «CoreLogic Information Solutions» и Home Mortgage Disclosure Act.

Результаты были получены с помощью сэмплирования по Гиббсу, в основе которого лежит алгоритм Метрополис-Гиббса. Процесс состоял из трех шагов: оценка параметров с помощью пробит-модели, вычисление вероятностей с поправками на сходимость на основе предыдущего шага, проведение симуляций для оценки каждого параметра.

В заключении стоит отметить, что большинство исследований подтвердили то, что эластичность спроса на кредитные продукты по ставке процента и сроку жизни кредита изменяется в зависимости от уровня дохода потребителей. Также подавляющее большинство авторов представили убедительные доказательства наличия ценовой дискриминации в различных формах.

## 1.2. Модель Хекмана и инструментальные переменные

В исследованиях по эмпирической оценке спроса достаточно часто присутствуют две проблемы: проблема эндогенных переменных и изначальное смещение выборки.

Теперь следует остановиться на каждой проблеме и путях ее решения более детально.

Стоит начать с того, что достаточно часто выборка имеет не случайный характер. Другими словами, определенные элементы общей совокупности имеют разные вероятности попадания в формируемую выборку, то есть возникает проблема, связанная с ограничениями на процесс формирования выборки.

В рамках данного исследования наблюдаемыми являются условия контракте только для тех потребителей, которые его подпишут. Однако гипотетически в экономических целях интерес представляют условия контрактов, предлагаемых любому человеку без ограничения этим условием.

На практике такого рода проблемы возникают из-за двух причин. Во-первых, данное смещение выборки может образоваться из-за самоотбора индивидов или объектов изучения. Вторая причина возникновения не случайного характера выборки – некоторые решение исследователей или процесс автоматической обработки данных могут носить такие же последствия, как и самоотбор.

Хекман (Heckman, 1979) предложил в своем исследовании весьма простую, но эффективную модель, помогающую учесть данное смещение.

Стандартная модель представляет собой систему из двух уравнений, где первое – уравнение участия с бинарной зависимой переменной (2), второе – результирующее уравнение с непрерывной зависимой переменной, скорректированное на смещение, полученное из первого уравнения.

Данная модель имеет вид, описанный формулами (6 – 9):

(6)

(7)

(8)

(9)

где: – экзкогенные регрессоры, то есть не коррелированны с ошибками ;

– бинарная зависимая переменная;

– латентная переменная.

Также предполагается, что совместное распределение ошибок является нормальным со средним, равным нулю, стандартными отклонениями, равными 1 и σ, и ненулевой корреляцией, равной ρ (то есть ) (10):

(10)

То есть подвыборка не носит случайный характер.

Следует отметить, что оценивание может происходить двумя методами: методом максимального правдоподобия и упрощенным двухшаговым способом, где из второго уравнения рассчитывается лямбда Хекмана (Inverse Mill’s Ratio) по формуле (11) (Heckman, 1979):

, (11)

где: – плотность стандартного нормального распределения;

– функция стандартного нормального распределения.

Затем к первому уравнению (1) добавляется еще один регрессор (12):

(12)

где остаток равен (13):

(13)

Однако в этом случае оценки получаются неэффективными, но состоятельными (Вербик, 2008), при условии, что распределение не зависит от xi (но не от hi), остаток не коррелирован с и по построению.

В настоящее время существует множество расширенных версий модели Хекмана с более слабыми изначальными предпосылками о нормальности распределения, включением эндогенных переменных и т. д.

Например, в статье «Nonparametric Estimation of Sample Selection Model» (Das et al., 2003) авторами рассмотрено несколько расширенных моделей, основанных на модели Хекмана, при определенных жестких предпосылках.

Первая модель – непараметрическая версия стандартной модели Хекмана, позволяющая функциональной форме уравнения быть неопределенной, точно так же как и функции совместного распределения.

Вторая модель включает в себя несколько правил отбора, где поправка на смещение зависит от вектора переменных, отражающих вероятность участия, одна переменная для одного правила отбора.

Третья модель позволяет ввести инструментальные переменные для контроля над эндогенностью, когда в коррекцию на смещение включается не только уравнение участия, но и остатки из уравнения с инструментальными переменными.

Поскольку проблема, связанная с эндогенностью переменных и дальнейшим их инструментированием, является актуальной для данной работы, следует более подробно остановиться на этом пункте.

Условие о некоррелированности регрессоров и случайно ошибки, которое является одним из основных для построения несмещенных и состоятельных оценок стандартными методами (МНК, ОМНК), на практике достаточно часто не выполняется.

Другими словами, в случае коррелированности случайной ошибки и объясняющих переменных данные переменные задаются внутри системы, то есть являются эндогенными, а не экзогенными.

Данная зависимость может возникать по ряду причин. Основной проблемой, влекущей за собой эндогенность, принято считать пропуск значимых переменных, а так же ошибки измерения, самоотбор, одновременность, серийная корреляция ошибок вместе с наличием лагированных значений зависимой переменной среди регрессоров (Эббес, 2007).

Для решения проблемы эндогенности были разработаны методы инструментальных переменных.

Считается, что метод инструментальных переменных был впервые сформулирован Райтом (Wright, 1928) как метод оценки кривых спроса и предложения. Сам термин "инструментальные переменные" был впервые использован Риерсолом (Riersol, 1941) при обсуждении ошибок в переменных. Далее метод получил развитие в работах Дарбина (Durbin, 1954) и др. В контексте систем одновременных уравнений метод развивался параллельно под названием "двухшаговый МНК".

Инструментальные переменные – это экзогенные переменные, которые коррелируют с эндогенным регрессором (или регрессорами), но не коррелируют с ошибкой модели.

Стандартная модель с инструментальными переменные получается добавлением к обычной регрессии уравнения, которое связывает эндогенные регрессоры и инструментальные переменные (14), (15).

(14)

(15)

где: определяются как в обычной линейной регрессии;

– матрица инструментальных переменных;

– матрица ошибок для уравнения с инструментами.

Стандартные способы оценивания коэффициентиов в подобной модели – двухшаговый метод наименьших квадратов и метод максимального правдоподобия с ограниченной информацией или более обобщенным методом: ОММ (Обобщенный Метод Моментов) (Hansen, 1982).

Рассмотри наиболее распространенную причину появления эндогенности в модели.

Модель с пропущенными переменными выглядит следующим образом (16) (Judge et al., 1985):

(16)

где – ненаблюдаемые (латентные) переменные.

В случае, если ожидание берется только от наблюдаемой переменной, уравнение выглядит следующим образом (17):

(17)

Данное выражение не равно при условии, что и , то есть пропущенные и включенные независимые переменные не ортогональны и пропущенные не влияют на зависимую переменную.

Следовательно, в данном случае смещение МНК-оценки для параметра будет равно (18):

(18)

Таким образом, пропуск значимых независимых переменных влияет на все оцениваемые коэффициенты.

Метод инструментальных переменных на практике часто затруднен поиском адекватных инструментов.

Выбранные инструменты могут быть «слабыми» в случае, если их корреляция с эндогенными регрессорами достаточно низкая. При работе с такими инструментами снижается точность оценок (Эббес, 2007), а так же не работают стандартные асимптотические приближения, как следствие, построение стандартных тестирующих гипотез и доверительных интервалов не работает корректно (Hahn et al., 2003).

Также выбранные инструменты потенциально могут коррелировать с ошибкой модели. Такие оценки обладают относительной несостоятельностью по сравнению с МНК-оценкой (Bound et al., 1995).

Резюмируя вышесказанное, следует сказать, что исследователи, занимающиеся эмпирической оценкой функции спроса как на рынке кредитования, так и на смежных рынках, сталкиваются с двумя основными проблемами. Во-первых, это выборки не случайного характера, а во-вторых, это эндогенность регрессоров.

Для получения корректных оценок параметров регрессии исследователи используют различные методы: непараметрическое оценивание, многошаговые алгоритмы и их комбинации. Однако в основе практически любого процесса моделирования лежат две модели: модель Хекмана и модель инструментальных переменных, чей обзор был также сделан в данной главе.

Подведя итог, следует сказать, что и модель Хекмана, и метод инструментальных переменных активно используются на практике, быстро развиваясь. Таким образом, оба метода можно считать эффективными и подходящими для использования в данной работе.

# Глава 2. Эмпирическая оценка спроса по данным ПАИЖК

## 

## 2.1. Описание переменных

Для настоящего исследования была использована база данных, предоставленная Пермским агентством по ипотечному жилищному кредитованию. База была дополнена переменными, характеризующими общую экономическую ситуацию в Перми. Данные являются кросс-секцией и содержат 4897 наблюдений по заемщикам в период с 01.08.2008 до 31.08.2012.

Имеющиеся параметры можно условно разделить на 4 группы: социально-демографические данные заемщика, параметры квартиры, параметры контракта и макро-показатели.

Первая группа: социально-демографические характеристики заемщика.

1. borrower\_age – возраст заемщика (в годах);
2. sex – пол заемщика: 0 – женский, 1– мужской;
3. familystatus – семейное положение: 0 – нет информации; 1 – не замужем/холостяк; 2 – замужем/женат; 3 – вдовец/вдова; 4 – в разводе;
4. activity\_category– вид занятости: 0 – нет информации; 1 – не работает; 2 – пенсионер; 3 – военнослужащий; 4 – наемный работник; 5 – индивидуальный предприниматель; 6 – госслужащий;
5. work\_is\_budget – является ли заемщик работником бюджетной сферы: 0 – нет, 1 – да;
6. education\_level – уровень образования: 0 – нет информации; 1 – начальное; 2 – среднее; 3 – незаконченное высшее; 4 – высшее;
7. income – указанный доход заемщика в месяц (в рублях);
8. income\_cat – доход заемщика по категориям: 1 – не указан; 2 – до 10000 рублей; 3 – от 10000 до 20000 рублей; 4 – 20000-40000 рублей; 5 – свыше 40 тысяч рублей;
9. Number\_of\_CoBorrowers – числосозаемщиков;
10. coborrower\_income – доход созаемщика/созаемщиков (в рублях);
11. coborrower\_income\_cat – доход созаемщиков по категориям: 0 – созаемщиков нет; 1 – доход не указан; 2 – до 10 000 рублей, 3 – от 10000 до 20000 рублей;4 –20000-40000 рублей; 5 – свыше 40 тысяч рублей;
12. outlier – является ли наблюдение выбросом, который необходимо исключить: 1 – да.

После краткого описания переменных следует перейти к краткому анализу их описательных статистик.



Рис. 1. Плотность распределения заемщиков согласно возрасту

На значение переменной «borrower\_age» был наложен фильтр (см. Рис. 1), поскольку возраст менее 21 года встречался крайне редко, то такие переменные были рассмотрены как выбросы (неточности заполнения или обработки анкет).

Таблица 1.

Описательные статистики переменной "borrower\_age"

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Obs | Mean | Std. Dev. | Min | Max | Pr(Skewness) | Pr(Kurtosis) |
| borrower\_age | 4300 | 33.99558 | 7.599215 | 21 | 61 | 0.0000 | 0.0068 |

Коэффициент вариации в 22% свидетельствует о том, что выборка однородна. Однако при коэффициентах асимметрии и эксцесса (probabilityкоторых равна 0% и 0,7% соответственно) гипотеза о нормальном распределении отвергается (см. Таблица 1).



Рис. 2. Распределение заемщиков по половому признаку

Как видно их гистограммы (см. Рис. 2), мужчин среди тех, кто подал заявку на ипотечный кредит, больше.



Рис. 3. Распределение заемщиков согласно их семейному статусу

Чаще всего заявку на получение ипотеки подают люди, находящиеся в браке, чем те, кто свободен или уже разведен (см. Рис. 3).

Также гистограммы показывают (Рис. 4, 5), что заявки подаются чаще всего людьми, работающими по найму в не бюджетной сфере.



Рис. 4. Распределение заемщиков согласно виду деятельности



Рис. 5. Распределение заемщиков согласно сфере работы

Теперь стоить рассмотреть наиболее важную переменную – ранговую переменную уровня дохода заемщика. Как видно из гистограммы (Рис. 6), большинство заемщиков не указывают в документах свой доход (либо допускаются серьезные ошибки в обработке информации). Однако по оставшимся данным можно сделать вывод, что наиболее часто заемщики имеют доход в размере от 20 тысяч до 40 тысяч рублей. Реже всех заявку на получение ипотечного кредита подают люди с минимальным уровнем дохода: до 10 000 рублей.

Исключать первую категорию не имеет смысла из-за большого числа входящих туда наблюдений, поэтому глядя на гистограмму можно утверждать, что распределение не является нормальным.



Рис. 6. Распределение заемщиков по уровню дохода

Вторая группа показателей включает в себя параметры заключенного контракта.

1. dti – соотношение ежемесячного платежа к ежемесячному доходу;
2. dti\_cat – соотношение ежемесячного платежа к ежемесячному доходу по категориям: 1 – нет данных; 2 – 0-20% включительно; 3 – 20-40%; 4 – 40-60%; 5 – 60-80%; 6 – 80-100%;
3. ltv – отношение суммы кредита к стоимости объекта недвижимости;
4. flag\_of\_endorsement – факт одобрения заявки на кредит агентством: 0 – заявка не одобрена; 1 – одобрена;
5. flag\_of\_contract\_agreement – факт подписания контракта заемщиком: 0 – контракт не подписан; 1 – контракт подписан;
6. max\_loan – максимальная сумма кредита в рублях;
7. loan\_amount – выданная сумма в рублях;
8. monthly\_payment – ежемесячный платеж (в рублях);
9. creditperiod – количество месяцев, в течение которых погашается кредит;
10. downpayment – итоговая сумма, которую необходимо вернуть заемщику;
11. rate – ставка по кредиту (в процентах);
12. type\_of\_rate – тип ставки: 1 – фиксированная; 2 – плавающая;
13. aim\_of\_credit – цель кредита: 1 – индивидуальное жилищное строительство; 2 – кредит под залог имеющегося жилья; 3 – оплата пая в потребительском кооперативе; 4 – погашение ранее предоставленного кредита; 5 – приобретение жилья на первичном рынке; 6-приобретение жилья на вторичном рынке; 7 – участие в долевом строительстве.

Поскольку часть переменных является бинарными, а также в этом блоке содержаться в основном переменные, которые являются эндогенными регрессорами в дальнейших моделях, детальные описательные статистики проведены не будут.

Третья группа показателей: характеристики приобретаемого объекта недвижимости.

1. flat\_value – стоимость объекта недвижимости;
2. totalsquare – общая площадь (м2);
3. livingsquare – жилая площадь (м2);
4. kitchensquare – площадь кухни (м2);
5. floorsquantity – количество этажей в доме;
6. roomsquantity – количество комнат;
7. dispositionfloor – этаж, на котором находится квартира;
8. buildyear – год постройки дома;
9. lavatory – расположениесанузла: 0 – объединенный; 1 – раздельный;
10. city – расположение объекта недвижимости: 1 – в Перми, 2 – в области; 3 – в городе Пермского края; 4 – в поселке Верещагино; 5 – в деревни, селе.
11. Typeofbuilding – тип дома:1 – панельный: панельный, железобетонные плиты, стеновые панели,твиноблоки с утепленными минераловатными плитами, трехслойные ж/б панели; 2 – кирпичный; 3 – смешанный: крупнопанельный, монолитный, смешанный, иное.

Данные параметры будут использоваться лишь в качестве инструментов в уравнении, оценивающем стоимость квартиры.

Характеристики макро среды и ипотечного рынка в целом по региону [28]:

1. unemployment\_rate – уровень безработицы в определенный месяц;
2. mortgage\_volume – объем выданных ипотек, млн. руб.;
3. mortgage\_amount – объем выданных ипотек в регионе;
4. mean\_loan – средняя сумма ипотечного кредита, в тыс. руб.;
5. median\_maturity – средний срок погашения ипотечного кредита, мес.;
6. median\_rate – средняя величина процентной ставки;
7. mean\_ltv – среднее отношение суммы кредита к стоимости объекта недвижимости;
8. mean\_dti – среднее отношение суммы ежемесячного платежа к доходу заемщика;
9. mean\_m2\_value – средняя стоимость квадратного метра, руб.;
10. lodging\_volume – жилищный фонд, млн. кв. м.;
11. lodging\_coef – обеспеченность населения жильем, кв. м. / чел.;
12. lodging\_coef\_in\_years – коэффициент доступности жилья, лет;
13. Population – население в регионе, тыс. человек;
14. nominal\_earnings – средний доход на человека по региону, руб;
15. ahml\_amount – количество рефинансированных в АИЖК кредитов из региона;
16. exact\_ahml\_amount – количество заявок на получение ипотечного кредита в ПАИЖК, поданное в определенном месяце.

Данные характеристики «привязаны» к месяцам: то есть по каждому наблюдению проставлен месяц (то есть месяц, когда человек пришел в ПАИЖК), для каждого из которых показатели разняться.

Таким образом, четыре группы данных были детально описаны выше, позволяя перейти к описанию и построению модели.

## 2.2. Описание модели

Теперь следует перейти непосредственно к моделированию спроса на ипотечные кредитные продукты ПАИЖК. Процесс получения ипотеки состоит из следующих этапов.

Во-первых, необходимо вычислить вероятность того, что человек придет в ПАИЖК с заявкой на получение кредита. Поскольку эта вероятность будет достаточно обобщенной для всего населения и не основана на каких-либо социальных характеристиках потенциальных клиентов, то независимыми параметрами являются показатели макро-среды. Для построения данного смещения будет использоваться стандартная пробит-регрессия.

Второй шаг состоит в оценивании смещения, возникающего в процессе одобрения или не ободрения поданной заявки отделом безопасности ПАИЖК. На данном этапе также строится пробит-регрессия, в независимых параметрах которой в частности находится лямбда Хекмана, полученная на первом шаге.

После одобрения заявки заемщику выставляется максимальная сумма кредита, которую он может получить. Данный показатель («max\_loan») является эндогенным, поэтому его необходимо инструментировать. Необходимо построить робастную регрессию [27, с. 69], поскольку предполагается, что условия Гаусса-Маркова выполняются не всегда, поэтому стандартный метод наименьших квадратов корректно работать не будет.

Следующее уравнение описывает этап, на котором заемщик выбирает, подписывать ему договор с ПАИЖК. В данное уравнение в качестве регрессоров включаются социально-демографические показатели заемщика, инструментированная переменная максимальной суммы, а так же смещение из второго шага.

Примерно в это же время заемщик должен подобрать подходящую квартиру. Считая, что в этом случае стоимость квартиры является эндогенным параметром (выбор квартиры зависит от максимально одобренной суммы), необходимо вводить инструментальные переменные, в данном случае это характеристики приобретаемого жилья.

Затем необходимо инструментировать такие характеристики контракта, как срок погашения, ставку процента, ежемесячный платеж, как эндогенные величины, которые являются независимыми параметрами в уравнении спроса.

Наконец, моделируется итоговое уравнение спроса, где зависимая переменная – желаемая сумма кредита (Attanasio et al, 2008), т. е. та, которая указана в контракте. Поскольку все переменные строго положительны, необходимо построить лог-линейную модель. Более того, при построении такой модели коэффициенты перед регрессорами равны показателю эластичности, что упрощает интерпретацию результатов.

Таким образом, итоговое уравнение модели будет иметь следующий вид (19):

(19)

Где – размер ипотечного кредита;

– экзогенные показатели (социально-демографические характеристики заемщика, макроэкономические характеристики и др.);

– эндогенные показатели (цена квартиры, условия контракта);

– итоговый показатель смещения, который не наблюдается в реальности и не имеет интерпретации.

Поскольку данный процесс моделируется с помощью нескольких уравнений участи и нескольких уравнений, содержащих инструментальные переменные, то в стандартных процедурах любого программного пакета данного алгоритма нет. Поэтому для реализации данного процесса был написан код для программного пакета Stata (см. Приложение 1).

## 2.3. Построение модели

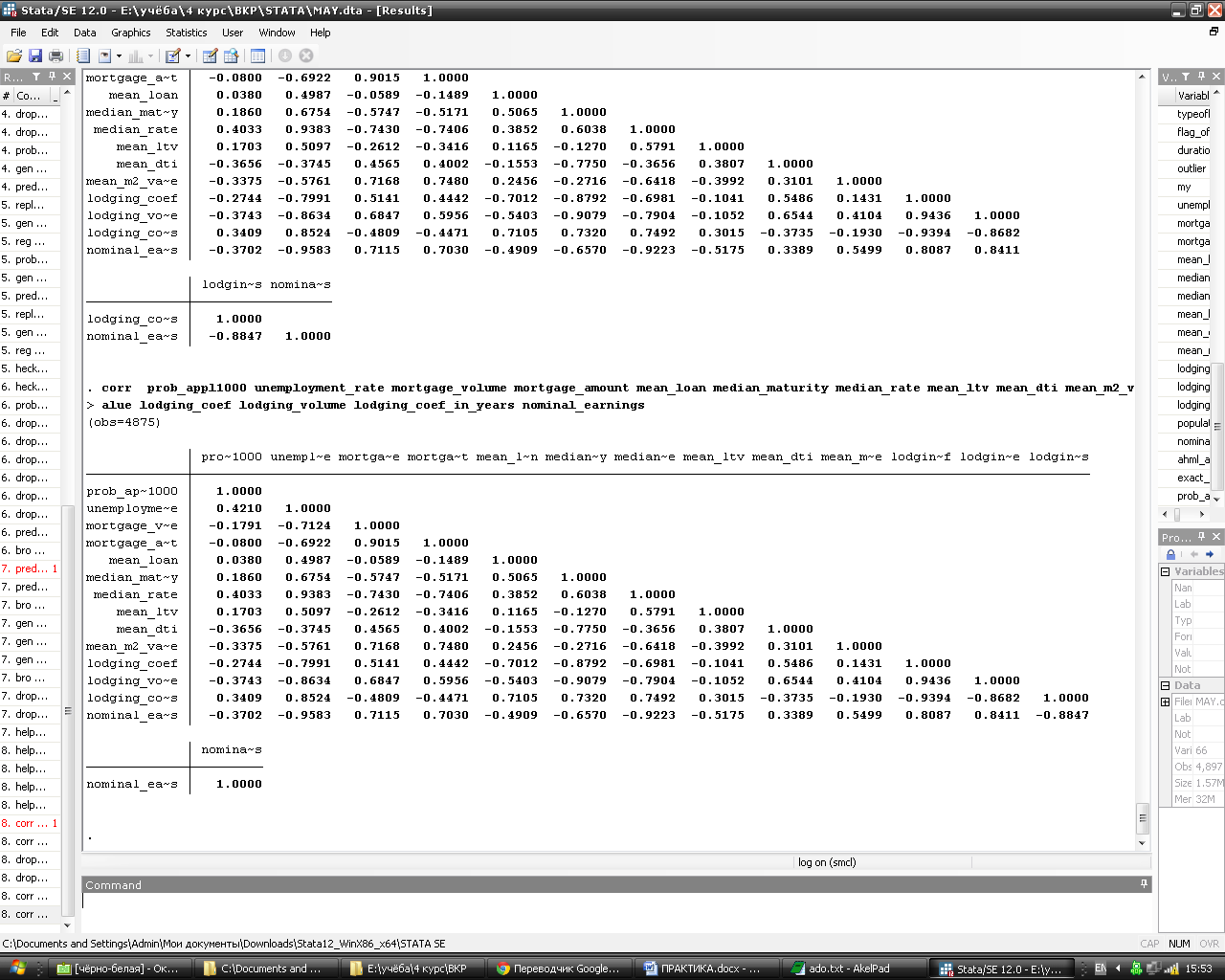
**Уравнение 1.** Необходимо смоделировать вероятность того, что человек придет в ПАИЖК в какой-либо месяц. В независимые переменные включены макро-факторы, то есть показатели экономического состояния заемщиков и ситуация на ипотечном рынке в целом, которую можно рассмотреть как поведение остальных банков-конкурентов. Зависимая переменная – собственно, и есть эта вероятность, рассчитанная как количество людей, подавших заявки на ипотеку в агентство в определенный месяц, деленная на общее число жителей (домноженная на 1000 вследствие крайне малых значений, с учетом того, что переменная «population» измеряется в тыс. человек). Сгенерируем данную переменную с помощью следующей команды:

«genprob\_appl1000= exact\_ahml\_amount\*1000/ population»

Построим корреляционную матрицу для отбора переменных в модель (Таблица 2).

Таблица 2

Корреляционная матрица для переменных четвертой группы



После анализа корреляционной матрицы построим линейную модель Eq.1 (см. Таблицу 2) для зависимой переменной «prob\_appl1000» (Таблица 3). Уравнение имеет следующий вид (20):

(20)

Таблица 3

Eq.1

|  |  |  |
| --- | --- | --- |
|  |  | Eq.1 (robust regression) |
| Oбъем выданных ипотек в регионе | mortgage\_amount | .0224348\*\*  (.0000477) |
| Cредний срок погашения ипотечного кредита | median\_maturity | -1.253809\*\*  (.0028467) |
| Cреднее отношение суммы кредита к стоимости объекта недвижимости | mean\_ltv | 1.590069\*\*  (.0081673) |
| Cреднее отношение суммы ежемесячного платежа к доходу заемщика | mean\_dti | -25.03713\*\*  (.0389848) |
| Средняя стоимость квадратного метра | mean\_m2\_value | -.0015538\*\*  (2.65e-06) |
| Коэффициент доступности жилья | lodging\_coef\_in\_years | 20.99284\*\*  (.0485395) |
| Константа | \_cons | 1043.922\*\*  (1.644245) |
|  | Adj R-squared | 0.6317 |
|  | Prob>F | 0.0000 |

(\* - уровень значимости составляет 10%, \*\* - уровень значимости в 5%)

Таким образом, вероятность того, что заемщик подаст заявку на получение ипотечного кредита в ПАИЖК, напрямую зависит от изменения объемов ипотечного рынка и среднего отношения суммы к стоимости квартиры в регионе, обратно зависит от среднего срока погашения кредита в регионе и стоимости квадратного метра жилья. Также прямая зависимость наблюдается с коэффициентом доступности жилья. Если обобщить результаты данного этапа, то чем лучше состояние экономики и жизни в регионе, тем выше вероятность, что потенциальный потребитель захочет взять ипотечный кредит.

Также в построении регрессии была использована команда fweigth (frequency weights), «взвешивающая» согласно частоте прихода людей в зависимости от месяца.

Стоит сказать, что в данном случае моделируется смещение в виде обычной линейной вероятности, то есть пробит-модель для данного случая не подходит, так как значение вероятности не принимает строго значения 0 или 1: следует использовать классический МНК. Однако вероятнее всего, что не все условия Гаусса-Маркова выполняются, поэтому, как и говорилось выше, в данном случае будет использована робастная регрессия.

Поскольку коэффициенты корреляции между некоторыми регрессорами высоки, необходимо провести VIF(Variance Inflation Factor) тест на проверку проблем мультиколлинеарности, рассчитывающийся по формуле (21).

Этот тест применим для МНК-регрессий и показывает, в какой мере дисперсия оцененных коэффициентов регрессии увеличивается из-за мультиколлинеарности.

(21)

При работе в программном пакете Stata пороговое значение VIFсоставляет 10 (для tolerance, соответственно, 0,1) [Baun, c. 79]: то есть при превышении порогового значения в модели возможна проблема мультиколлинеарности и необходимо дальнейшее исследование.

По результатам теста (см. Таблица 4) можно утверждать, что в модели нет проблем, связанных с мультиколлинеарностью регрессоров, поэтому можно переходить к интерпретации результатов.

Таблица 4

Variance Inflation Factor (Eq. 1)



Интерпретировать полученные результаты можно следующим образом. Если другие банки выдают кредиты на небольшие сроки (средние показатели по рынку), то больше человек будут приходить в ПАИЖК. При росте общего рынка кредитования количество людей, пришедших в агентство, будет также расти.

Рассмотрим статистику предсказанных и реальных величин вероятности подачи заявки на получение ипотечного кредита в ПАИЖК без учета выбросов (Таблица 5).

Таблица 5

Статистика показателей pxb1 и prob\_appl1000



Статистики двух переменных схожи.

Из первого уравнения получены оцененные значения вероятности подачи заявки на кредит человеком в определенный месяц в ПАИЖК, которые затем становятся одним из регрессоров во втором уравнении, описывающим смещение при подтверждении или отвержении заявки клиента агентством.

**Уравнение 2.** Важным допущением является то, что у кредитора нет никаких ограничений по количеству выданных ипотечных кредитов. Поэтому при такой предпосылке отдел безопасности банка при одобрении или отказе в подавляющем большинстве случаев смотрит лишь на социально-демографические характеристики конкретного заемщика.

Так как на основании полученных оцененных значений зависимой переменной в последствие будет выводиться лямбда Хекмана, то на этом этапе будет использоваться пробит-модель.

Можно предположить, что одобрение заявки от некоторых переменных будет зависеть нелинейно. Такой переменной является переменная возраста заемщика: молодым людям одобряют заявки реже, чем более взрослым, но с определенного момента при увеличении возраста вероятность одобрения уменьшается. Таким образом, зависимость описывается параболой с ветвями вниз. Поэтому в качестве аппроксимирующей функции следует использовать полином второй степени.

В связи с этим была введена новая переменная «bor\_age2», являющаяся квадратом переменной «borrower\_age» (возраст заемщика).

Поскольку большинство характеристик заемщика – ранговые переменные, их возможно включить в регрессию присваивая определенные «базовые» значения. Например, для переменной «familystatus» таким значением будет 2, поскольку большинство заемщиков состоят в законном браке, данное утверждение подтверждается гистограммой переменной. Присвоение такого базового ранга, то есть familystatus=2, помогает проанализировать, как повлияет на зависимую переменную отклонения этого показателя от данного значения.

По такой же логике будут расставлены базовые значения у других ранговых переменных, включаемых в регрессию.

Рассмотрим корреляционные матрицы первой группы показателей (см. Таблицы 6-8).

Таблица 6

Корреляционная матрица по первой группе показателей

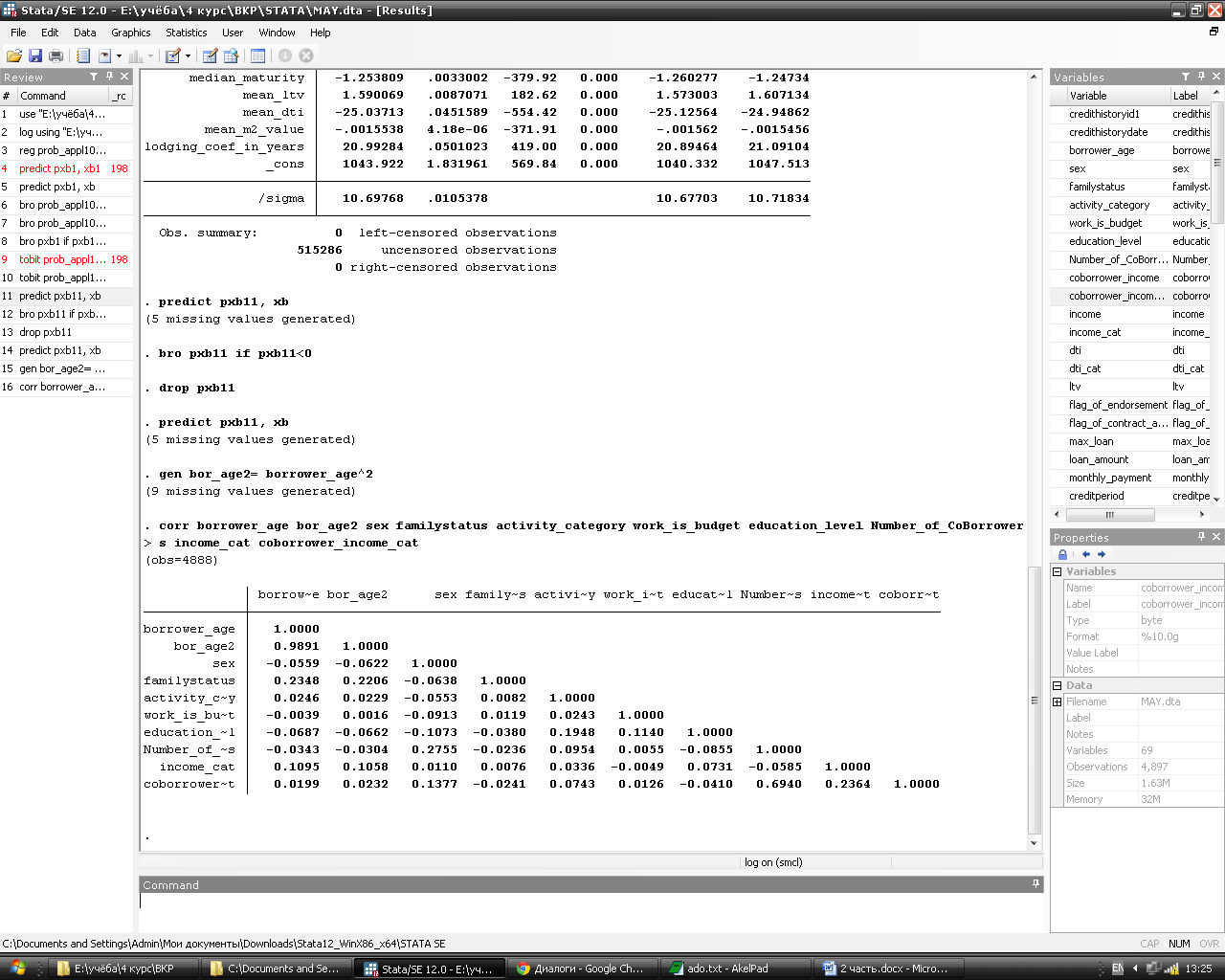


Таблица 7

Корреляция Спирмена

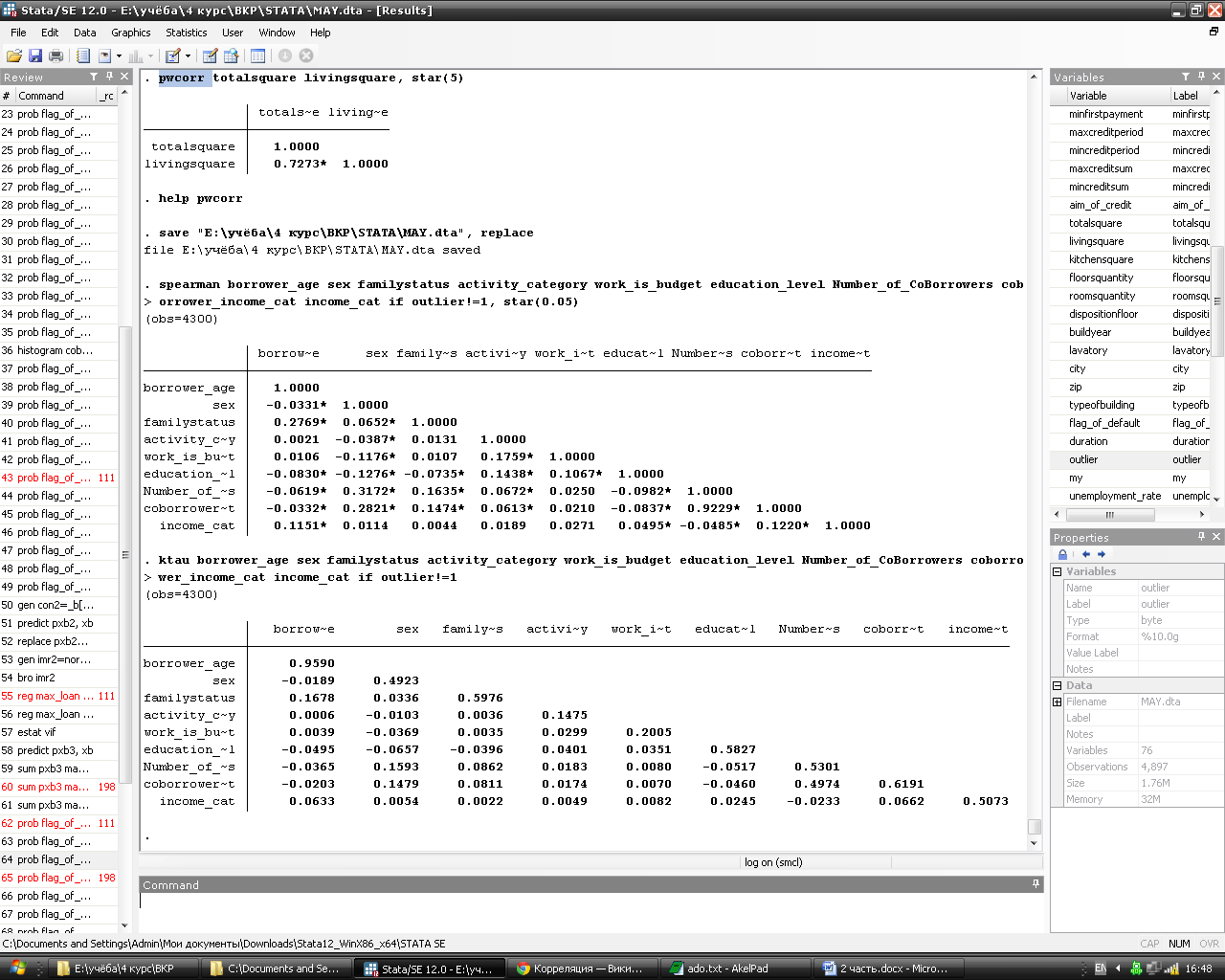
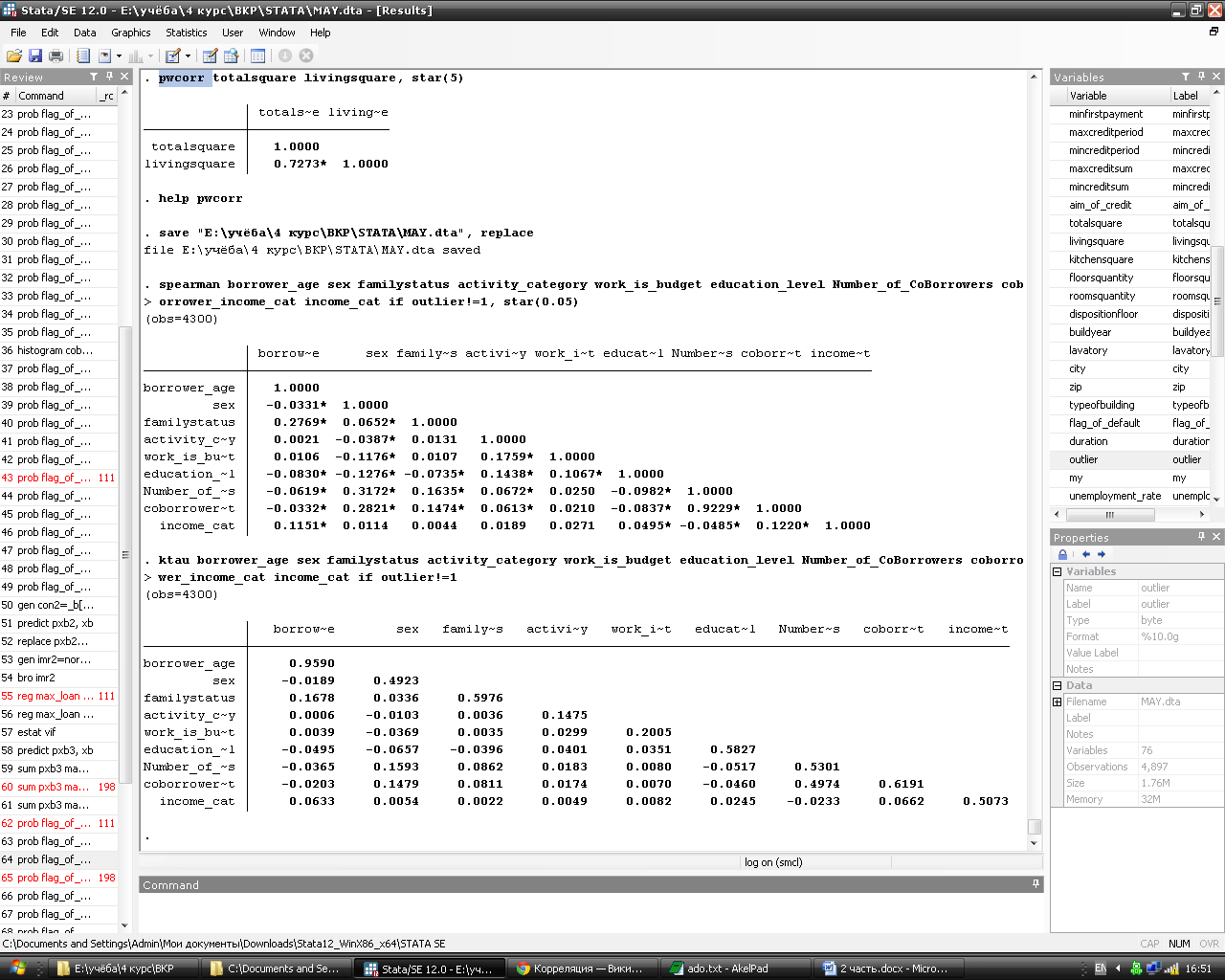


Таблица 8

Корреляция Кендалла



Построим несколько моделей для нахождения оптимальной спецификации: начиная с уравнения с максимальным числом независимых переменных (Eq.2.1), затем постепенно исключая незначимые (Eq.2.2, Eq.2.3) (Приложение 2).

Оптимальной была выбрана третья модель. Явный отрицательный эффект заметен в случае некачественного заполнения заявки: не указания места работы (значение activity\_category, education\_level, income\_cat равные 0, 0 и 1 соответственно). Также в большинстве случаев играет большую роль уровень образования при одобрении заемщика.

Уравнение Eq.2.3 имеет следующий вид (22):

,

(22)

где: xi – ранговая независимая переменная с i-тым рангом;

– коэффициент для i-го ранга ранговой переменной x.

После построения регрессии была посчитана лямбда Хекмана (imr2), отражающая смещение.

Интерпретация данного уравнения может звучать следующим образом. На вероятность одобрения банком поданной заявки влияет образование потенциального заемщика (чем лучше уровень образования, тем выше вероятность одобрения), информация о доходе (как видно из регрессии, значимым коэффициентом является только отрицательный коэффициент при income\_cat=1, то есть при неуказанном доходе вероятность одобрения ипотеки меньше, во всех остальных случаях уровень дохода не влияет на решение банка) и вид занятости. Оцененная вероятность подачи заявления с прошлого шага имеет коэффициент с отрицательным знаком: иными словами, чем больше подано заявок в определенный промежуток времени, тем меньше вероятность, что конкретному заемщику банк примет решение об одобрении.

**Уравнение 3**. На данном шаге моделируется процесс выставления максимальной суммы кредита (кредитного лимита) (Приложение 3). Поскольку данная величина задается банком (внутри системы), она является эндогенной, следовательно, необходимо вводить инструментальные переменные (Эббес, 2007). В качестве таких переменных будут использованы социально-демографические показатели заемщиков и общие показатели ипотечного рынка, являющиеся экзогенными параметрами.

Таким образом, модель будет иметь следующий вид (23):

(23)

Другими словами, наибольший кредитный лимит выставляется людям, которые находятся в официальных отношениях и с наиболее качественным уровнем образовании (с высшим). Прямую зависимость величины суммы также можно наблюдать от уровня дохода заемщика и созаемщиков, обратную – от средней величины ипотечного кредита по региону (прокси-переменная для поведения других банков-конкурентов). Также при выставлении кредитного лимита важен вид занятости и возраст заемщика, зависимость от которого описывается полиномом второй степени. Иначе говоря, сначала с увеличением возраста кредитный лимит увеличивается, а с определенного момента при увеличении возраста заемщика величина максимальной суммы кредита уменьшается. Данная зависимость подтверждается в реальности.

При инструментировании кредитного лимита показатель смещения (лямбда Хекмана) оказался незначимым и не был включен в модель.

Проверим регрессию (Таблица 10) на наличие возможных проблем, связанных с мультиколлинеарностью факторов (Таблица 9).

Таблица 9

VIF тест (Eq. 3)



Показателей VIF, больших 10, не обнаружено (кроме значений, характеризующих полином), следовательно, мультиколлинеарности в модели нет.

Рассмотрим статистику предсказанных и реальных величин максимальной суммы ипотеки по одобренным заявкам без учета выбросов (Таблица 10).

Таблица 10

Статистика реальных значений и предсказанных зависимой переменной 3 уравнения



Как видно из таблицы 10, значимых различий между средними значениями нет, дисперсия предсказанных значений меньше, чем реальных.

Предсказанные моделью значения максимальной суммы контракта затем входят в следующее уравнение как одна из объясняющих переменных.

**Уравнение 4.** На следующем шаге моделируется последнее смещение, образующееся в результате того, что не все одобренные заемщики подписывают непосредственно контракт на получение ипотечного кредита.

В качестве регрессоров в эту probit-регрессию (см. Таблицу 13) включаются индивидуальные характеристики заемщика, предсказанная на предыдущем шаге максимальная сумма кредита, некоторые показатели макро-среды и лямбда Хекмана из уравнения 2. Зависимая переменная – «flag\_of\_contract\_agreement». Уравнение имеет следующий вид (24):

(24)

Значимость коэффициентов уравнения видна из таблицы 11.

Как и на предыдущих шагах, уровень дохода играет важную роль при подписании контракта. С увеличением дохода увеличивается вероятность того, что заемщик подпишет предлагаемый контракт (не считая первой категории, которая означает то, что доход заемщика не указан). Обратная зависимость с кредитный лимитом, вероятнее всего, можно интерпретировать следующим образом: чем выше установили кредитный лимит заемщику, тем более надежным его считают, тем больше выбора у самого заемщика среди других банков, и, следовательно, более низкая вероятность придти именно в ПАИЖК. Также на вероятность подписания контракта влияет семейное положение заемщика и коэффициент доступности жилья (влияние, аналогичное влиянию на первом шаге).

Таблица 11

Eq. 4

|  |  |  |
| --- | --- | --- |
|  |  | Eq.4 |
| Оцененная максимальная сумма кредита | pxb3 | -8.28e-7\*\*  (2.25e-7) |
| Лямбда Хекмана | imr2 | 1.17722\*\*  (.2661366) |
| Семейное положение | familystatus |  |
|  | 0 | .0998423  (.295161) |
|  | 1 | -.3495763\*\*  (.0587021) |
|  | 3 | -.4609425\*\*  (.2151922) |
|  | 4 | -.2973502\*\*  (.0726922) |
| Категория дохода | income\_cat |  |
|  | 1 | .2673595  (.3551253) |
|  | 2 | -1.398315\*\*  (.2041894) |
|  | 3 | -.4562223\*\*  (.1517506) |
|  | 5 | .5016053\*\*  (.2049936) |
| Коэффициент доступности жилья | lodging\_coef\_in\_years | .5082988\*\*  (.0422014) |
| Константа | \_cons | -.470169\*  (.314756) |

Из данного уравнения итоговая лямбда Хекмана перейдет в регрессоры уравнений параметров контракта для корректировки результатов на смещение.

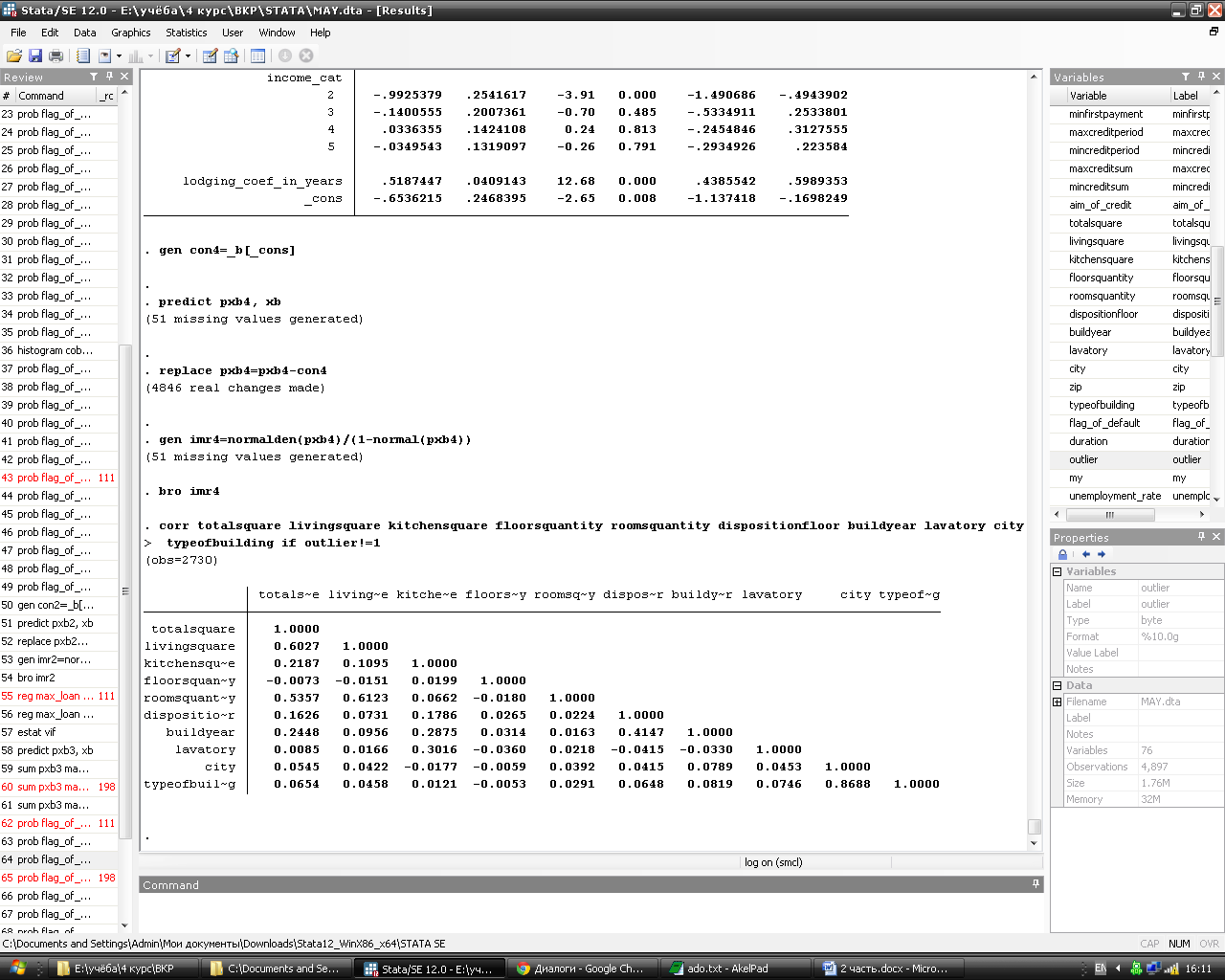
**Уравнение 5.** Поскольку стоимость квартиры является эндогенным параметром, ее необходимо инструментировать. В качестве инструментальных переменных были выбраны характеристики объекта недвижимости.

Необходимо рассмотреть корреляционную матрицу для выявления переменных, коэффициент корреляции между которыми высок (Таблица 12).

Ожидались высокие коэффициенты корреляции между тремя показателями площадей квартиры, однако сильной линейной зависимости между этими показателями нет (коэффициент корреляции между общей площадью и жилой площадью составляет 0,6027 и является наибольшим для трех рассматриваемых характеристик). Высокая корреляция существует между типом дома и типом населенного пункта, где находится здание (city) и составляет 0,8688.

Таблица 12

Корреляционная матрица для третьей группы переменных



Можно предположить, что такой фактор, как этаж, на стоимость напрямую не влияет. Однако в случаях, когда квартиры находятся на первом или последнем этажах, их стоимость чаще всего ниже аналогичных объектов, находящихся не на крайних этажах. Поэтому будет введена новая переменная «floordum», принимающая значение 0, когда квартира находится на первом или последнем этаже (или данных об этаже нет), и 1 в ином случае.

Для улучшения качества модели можно ввести новую ранговую переменную, которая делила бы временные промежутки на периоды для группировки домов по времени строительства. Введем переменную «buildperiod» и присвоим ей следующие значения в зависимости от года постройки: до 1960 года – 0, 1961-1985 года – 1 (в это время строились «хрущевки»), 1986-2005 – 2, 2006-2012 – 3. Создадим данную новую переменную (см. Рисунок 7).

Построим модель Eq.5, заменив год постройки дома на сгенерированную ранговую переменную с присвоением базового ранга, равного 1, и с введением переменной «lavatory» в регрессоры (Таблица 13).



Рис. 7. Гистограмма переменной "buildperiod"

Также следует отметить, что данная регрессия является робастной, поскольку в этом случае предполагается ослабление условий Гаусса-Маркова.

Таким образом, модель имеет вид (25):

(25)

Модель показывает положительную зависимость от показателей площадей (общей, жилой, кухни), количества комнат, этажа (как в количественном значении, так и для исключения «крайних» этажей). Также важен период застройки: квартиры в домах новее хрущевского типа стоят дороже; а также характеристики сан. узла и местоположении объекта (дома в Перми стоят не дешевле аналогов в других населенных пунктах края.

При инструментировании цены квартиры показатель смещения (лямбда Хекмана) оказалась незначима и не была включена в модель.

Следует убедиться, что в модели отсутствует мультиколлинеарность.

Значений VIF, больших 10, не наблюдается, следовательно, мультиколлинеарности в модели нет (Таблица 14).

Значимых различий между средними значениями предсказанной и реальной стоимостями объектов недвижимости не обнаружено (Таблица 15).

Таблица 13

Eq. 5

|  |  |  |
| --- | --- | --- |
|  |  | Eq. 5 (robust regression) |
| Общая площадь | totalsquare | 11353.33\*  (5053.697) |
| Жилая площадь | livingsquare | 10804.72\*  (55109.1) |
| Площадь кухни | kitchensquare | 22231.09\*\*  (7226.275) |
| Количество комнат | roomsquantity | 104341.6\*  (52275.7) |
| Этаж | dispositionfloor | 39223.38\*\*  (4997.95) |
| Дамми-переменная этажа | floordum | 160594.9\*\*  (27145.48) |
| Период постройки | buildperiod |  |
|  | 0 | 143186  (116071.5) |
|  | 2 | 182847\*\*  (61363.42) |
|  | 3 | 367098.1\*\*  (103167.9) |
| Сан. узел | lavatory | -75552.78\*  (38517.62) |
| Населенный пункт | city |  |
|  | 0 | -364318.3\*\*  (45720.56) |
|  | 2 | -361782.6\*\*  (89510.45) |
|  | 3 | 44725.8  (78544.04) |
|  | 4 | -519704.5  (412827.5) |
| Константа | \_cons | 598518.6\*\*  (145248.8) |

Проведем тест Two-sample mean-comparison test (paired) на сравнение средних значений по двум переменным с неравными дисперсиями (Таблица 16).

Так как значение Pr больше пороговых 5% для трех случаев, принимается нулевая гипотеза о равенстве средних значений по двум переменным.

Таблица 14

VIF тест для уравнения 5



Таблица 15

Статистики для переменных flat\_value и pxb5



Таблица 16

Two-sample t test with unequal variances для flat\_value и pxb5



**Уравнения для инструментирования условий контракта**. Конечный этап представляет собой построение регрессии (Eq. 6.1 – 6.2) для моделирования основных условий контракта: сумму ипотеки, величину процента и срок.

Регрессия со ставкой по кредиту в качестве зависимой переменной (см. Таблица 17) предсказывает значения крайне схожие с реальными (см. Таблица 18).

Таблица 17

Eq. 6.1

|  |  |  |
| --- | --- | --- |
|  |  | Eq. 6.1 (robust regression) |
| Лямбда Хекмана | imr4 | 2.133253\*\*  (.1107651) |
| Средняя величина процентной ставки | median\_rate | -.8989232\*\*  (.0439045) |
| Категория дохода | income\_cat |  |
|  | 1 | -.0053209  (.0996643) |
|  | 2 | 1.689371\*\*  (1.689371) |
|  | 3 | -.0124432  (.1092043) |
|  | 5 | .2483492\*  (.1396905) |
| Возраст заемщика | borrower\_age | .153335\*\*  (.0274425) |
| Возраст заемщика во второй степени | bor\_age2 | -.0019067\*\*  (.056085) |
| Пол | sex | -.2077843\*\*  (.056085) |
| Константа | \_cons | 16.72941\*\*  (.7360522) |

Таблица 18

Описательные статистики для переменных rate и predrate



Таким образом, уравнение ставки процента можно записать следующим образом (25):

(25)

Ставка по ипотечному кредиту инструментируется с помощью средней ставки по региону (прокси-переменная для поведения банков конкурентов, обратная зависимость), уровня дохода (если потребитель имеет очень маленький доход или, наоборот, высокий, ставки для него выше), возраста и пола (женщинам предоставляют более низкие ставки).

Регрессия со сроком в качестве зависимой переменной описана в таблице 19.

Таблица 19

Eq. 6.2

|  |  |  |
| --- | --- | --- |
|  |  | Eq. 6.2 (robust regression) |
| Категория дохода | income\_cat |  |
|  | 1 | 3.74421  (3.034305) |
|  | 2 | -.9913474  (5.506842) |
|  | 3 | .6411427  (3.848724) |
|  | 5 | -9.284528\*\*  (4.009183) |
| Коэффициент доступности жилья | lodging\_coef\_in\_years | -12.64319\*\*  (2.439214) |
| Средний срок погашения ипотечного кредита | median\_maturity | .3569055\*\*  (.1498797) |
| Возраст заемщика | borrower\_age | 9.23236\*\*  (1.062467) |
| Возраст заемщика во второй степени | bor\_age2 | -.0019067\*\*  (.0136234) |
| Константа | \_cons | 33.08344  (32.54153) |

На срок выплаты кредита влияют такие показатели, как доход, коэффициент доступности жилья (обратная зависимость), средний срок погашения ипотечного кредита (прямая зависимость), возраст заемщика (параболическая зависимость).

Стоит отметить, что лишь при переходе к наиболее высокому уровню дохода (income\_cat=5) срок погашения кредита снижается. То есть для людей низкого и среднего достатка изменения срока погашения кредита не является значимым.

Уравнение срока ипотечного кредита представляет собой (26):

(26)

Предсказанное с помощью инструментов значение срока кредита также схоже с реальным значением в среднем, и меньше в дисперсии (см. Таблица 20).

Таблица 20

Статистики переменных creditperiod и predperiod



**Уравнение спроса**. Моделирование данного уравнения является последним шагом (27):

(27)

где: – размер ипотечного кредита (желаемый и выданный);

– экзогенные показатели (социально-демографические характеристики заемщика, макроэкономические характеристики и др.) и оценки эндогенных параметров, смоделированные через инструментальные переменные;

– итоговый показатель смещения, который не наблюдается в реальности и не имеет интерпретации.

В качестве независимых переменных в уравнение входят лямбда Хекмана из четвертого уравнение, категориальная переменная дохода и инструментированные значения стоимости квартиры и двух параметров договора (ставка и срок кредита) (Таблица 21). Поскольку зависимая величина не может быть отрицательна, построим логистическую форму.

Таблица 21

Уравнение спроса

|  |  |  |
| --- | --- | --- |
|  |  | Eq. 6.3 (log. robust regression) |
| Логарифм лямбды Хекмана | ln\_imr4 | -.3705301\*\*  (.0451649) |
| Логарифм смоделированной цены квартиры | ln\_pxb5 | .5358237\*\*  (.0268954) |
| Логарифм оцененной ставки | ln\_predrate | .2022722  (.1460863) |
| Логарифм оцененного срока ипотеки | ln\_predperiod | .1491938\*\*  (.0575658) |
| Категория дохода | income\_cat |  |
|  | 1 | -.0125867  (.0265775) |
|  | 2 | -.91939\*\*  (.0547058) |
|  | 3 | -.3878425\*\*  (.0303713) |
|  | 5 | .3931735\*\*  (.0322966) |
| Константа | \_cons | 4.992986\*\*  (.5634276) |

В данном уравнении imr4 описывает лишь смещение, не имея реального выражения, pxb5 – оцененная стоимость квартиры, income\_cat – ранговая переменная, отражающая уровень дохода заемщика, predrate, predperiod – смоделированный ставка и срок жизни кредита.

Таким образом, уравнение спроса на ипотечные кредитные продукты выглядит следующим образом (28):

(28)

В логистических моделях коэффициент при независимой переменной показывает, на сколько процентов изменится Y при изменении X на 1%.

Таким образом, коэффициент перед показателем, характеризующим смещение, значим: смоделированное смещение важно для данной модели. При изменении стоимости объекта недвижимости на 1% спрос на ипотеки увеличится на 0,54%. При увеличении срока жизни кредита на 1% спрос, соответственно, возрастет на 0,15%. К изменению ставки кредита спрос не чувствителен, поскольку коэффициент не значим. Наконец, возрастание индивидуального спроса при переходе заемщика от 4 группы по доходу к пятой составит 0,39%, при переходе от 4 группы к третьей индивидуальный спрос упадет на 0,39%, а при переходе ко второй – на 0,92%.

Итоговое уравнение спроса было смоделировано, а коэффициенты регрессии проинтерпретированы в рамках коэффициентов эластичности спроса по процентной ставке и сроку жизни кредита. Гипотезы, выдвигаемые в начале работы, подтверждены: коэффициент перед уровнем дохода заемщиков значим во всех уравнениях, описывающих процесс формирования спроса; коэффициент при лямбде Хекмана значим в итоговом уравнении спроса, а так же значимость коэффициентов эластичности спроса по ставке процента и сроку кредита подтверждает изначально выдвигаемую гипотезу.

# Заключение

На сегодняшний день банковская сфера в России находятся на стадии интенсивного роста и развития, поэтому исследования в этой области крайне важны для всех участников рынка. Поэтому моделирование уравнения спроса на ипотечные кредитные продукты, безусловно, является актуальной работой.

В ходе исследования были проанализированы многие труды, сделанные в данной области ранее, где подтверждалось наличие дискриминации по различным признакам во время принятия решений банком, а так же различия в эластичности спроса по ставке процента и сроку кредита в зависимости от уровня дохода заемщиков. Далее был проведен анализ ограничений и проблем, с которыми сталкивались исследователи: в данном случае это выборки не случайного характера и эндогенные переменные. Также были описаны методы решения этих проблем и применение их на практике: смещение оценок параметров регрессии при неслучайной выборке контролируется с помощью модели Хекмана, а проблема эндогенности регрессоров решается методом инструментальных переменных.

После изучения теоретической базы были введены необходимые предпосылки и была описана модель, после чего данная модель была применена на практике. Также были выполнены остальные задачи исследования: был разработан программный код для эконометрического пакета Stata, были выбраны оптимальные спецификации моделей на каждом шаге и получены результаты, не противоречащие экономической логике и подходящие для интерпретации.

В конечном итоге был смоделирован процесс, состоящий из трех уравнений участия, с помощью которых была получена итоговая лямбда Хекмана, и четырех уравнений для ввода инструментальных переменных для оценивания эндогенных показателей: цены квартиры и условий контракта. Восьмым уравнением являлось итоговое уравнение спроса.

Все гипотезы, выдвигаемые в начале работы, были подтверждены, что доказывается значимостью коэффициентов в результирующем уравнении спроса. Была подтверждена значимость такой социально-демографической характеристики заемщика, как уровень дохода, на всех этапах моделирования смещения. Также коэффициент перед логарифмом срока кредита значим, в то время как коэффициент перед логарифмом ставки нет, что говорит о значимом показателе эластичности спроса по сроку жизни кредита. Более того, было подтверждено предположение о необходимости моделирования смещения, то есть о значимости коэффициента перед итоговой лямбдой Хекмана в конечном уравнении спроса.

Безусловно, данная модель может быть расширена. В качестве продолжения и усложнения данного исследования могут быть использованы полупараметрические и непараметрические методы оценивания для ослабления предпосылок, необходимых для построения пробит-моделей и получения качественных показателей смещения.

# Список использованной литературы

**Специальная литература**

1. Вербик М. Путеводитель по современной эконометрике / пер. с англ. В. Банникова В. Научн. ред. и предисл. С. Айвазяна // М: Научная книга, 2008. 616 с.
2. Магнус Я. и др. Эконометрика. Начальный курс: Учеб. – 4-е изд / Я. Магнус, , П. Катышев, А. Пересецкий // М.: Дело, 2000. 400 с.
3. Эббес П. (2007) Инструментальные переменные и эндогенность: нетехнический обзор // Квантиль. №2. С. 3-20.
4. Ambrose B. et al. 2004. The Effect of Conforming Loan Status on Mortgage Yield Spreads: A Loan Level Analysis / B. Ambrose, M. LaCour-Little, A. Sanders // Real Estate Economics / Volume 32. Number 4. P. 541–569.
5. Attanasio O. P. et al. Credit Constraints in the market for consumer durables: evidence form micro data on car loans / O. P. Attanasio, P. K. Goldberg, E. Kuriazidou // International Economic Review / Volume 49. Number 2. May 2008. P. 401–436.
6. Baun C. An introduction to modern econometrics using Stata: Stata Press, 2006. 341 p.
7. Bocian D. G. et al. Unfair Lending: The Effect of Race and Ethnicity on the Price of Subprime Mortgages / D. G. Bocian, K. S. Ernst, W. Li // Journal of Economics and Business. 60 (2008). P. 110–124.
8. Bound J. et al. Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak /J. Bound, D. A. Jeager, R. M. Baker // Journal of American Statistical Association / Volume 90. Number 430. June 1995. P. 443-450
9. Das M. et al. Nonparametric Estimation of Sample Selection Model / M. Das, W. K. Newey, F. Vella// Review of Economic Studies (2003) 70, 33-58
10. Durbin J. Errors in Variables // Review of the International Statistical Institute. 22 (1954). pp. 23–32
11. Gary-Bobo R. J. et al. A structural econometric model of price discrimination in the French mortgage lending industry / R. J. Gary-Bobo, S. Larribeau // International Journal of Industrial Organization. 22 (2004). P. 101 – 134.
12. Ghent A. C. et al. Race, Redlining, and Subprime Loan Pricing / A. C. Ghent, R. Hernández-Murillo, M. T. Owyang // Federal Reserve Bank of St. Louis Working Paper № 2011-033A. 12 July 2012. 49 p.
13. Gross D. B. et al. Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data / D. B. Gross, N. S. Souleles // The Quarterly Journal of Economics / Volume 117. Number 1. February 2002. P. 149–185.
14. Hahn J. et al. Weak Instruments: Diagnosis and Cures in Empirical Econometrics / J. Hahn, J. Hausman // Recent Advances in Econometric Methodology. 93 (2003). P. 118-125.
15. Hansen L. P. Large Sample Properties of Generalized Method of Moments Estimators // Econometrica. 50 (1982). P. 1029-1054.
16. Heckman J. J. Sample Selection Bias as a Speciﬁcation Error // Econometrica. 47 (1979). P. 153–161.
17. Judge G. G. et al. The Theory and Practice of Econometrics / G. G. Judge, W. E. Griffiths, R. C. Hill, H. Lütkepohl, T.C. Lee. New York: John Wiley & Sons, 1985. 1050 pages.
18. Juster F. T. et al. Consumer Sensitivity to Finance Rates: An Empirical and Analytical Investigation / F. T. Juster, R. P. Shay // NBER Occasional Paper No. 88. 1964. 105 p.
19. Karlan D. et al. Observing Unobservables: Identifying Information Assymetries with a Consumer Credit Field Experiment / D. Karlan, J. Zinman // Econometrica. Volume 77. Number 6. P.1993–2008.
20. LaCour-Little M. The Home Purchase Mortgage Preferences of Low- and Moderate-Income Households // Real Estate Economics / Volume 35. Issue 3. Fall 2007. P. 265–290.
21. Martins N. C. et al. The impact of mortgage interest-rate subsidies on household borrowing / N. C. Martins, E. Villanueva // Journal of Public Economics. 90 (2006). P. 1601 – 1623.
22. Perraudin W. R. M. et al. The Credit-Constrained Consumer: An Empirical Study of Demand and Supply in the Loan Market / W. R. M. Perraudin, B. E. Sørensen // Journal of Business & Economic Statistics / Volume 10. Number 2. April 1992. P. 179-192.
23. Reiersøl O. Conﬂuence Analysis by Means of Lag Moments and Other Methods of Conﬂuence Analysis // Econometrica. 9 January 1941. P. 1–24.

Wright Ph. G. The Tariff on Animal and Vegetable Oils. New York: MacMillan (1928). 56 p.

1. Zeldes S. Consumption and liquidity constraints: An empirical investigation // Journal of Political Economy / Volume 97. Number 2. April 1989. P. 305-346.
2. Zhang, Y. Fair Lending Analysis of Mortgage Pricing: Does Underwriting Matter? // Office of the Comptroller of the Currency / OCC Economics Working Paper 2010-1. 2010. 31 p.

**Электронные ресурсы**

1. Шведов А. Робастная регрессия с применением t-распределения и ЕМ-алгоритма [Эл. ресурс]. Режим доступа: http://uisrussia.msu.ru/docs/nov/hse\_ejournal/2011/1/15\_01\_05.pdf
2. ahml.ru / Агентство по ипотечному жилищному кредитования [Эл. ресурс]. Режим доступа: http://www.ahml.ru/ru/agency/analytics/statsis/
3. cbr.ru / Сайт Центрального Банка Российской Федерации [Эл. ресурс]. Режим доступа: http://www.cbr.ru/statistics/UDStat.aspx?TblID=3-1&pid=ipoteka&sid=ITM\_8591.

# Приложения

Приложение 1

Код для программного пакета Stata

gen prob\_appl1000= exact\_ahml\_amount\*1000/ population

reg prob\_appl1000 mortgage\_amount median\_maturity mean\_ltv mean\_dti mean\_m2\_value lodging\_coef\_in\_years [fweight= exact\_ahml\_amount ] if outlier!=1, vce(hc3)

predict pxb1, xb

gen bor\_age2=borrower\_age^2

prob flag\_of\_endorsement pxb1 b4.activity\_category b4.education\_level b4.income\_cat if outlier!=1

gen con2=\_b[\_cons]

predict pxb2, xb

replace pxb2=pxb2-con2

gen imr2=normalden(pxb2)/(1-normal(pxb2))

reg max\_loan b4.income\_cat b2.familystatus b4.education\_level i.coborrower\_income\_cat borrower\_age bor\_age2 mean\_loan if outlier!=1 & flag\_of\_endorsement==1, vce(hc3)

predict pxb3, xb

prob flag\_of\_contract\_agreement pxb3 imr2 b2.familystatus b4.income\_cat lodging\_coef\_in\_years if outlier!=1 & flag\_of\_endorsement==1

gen con4=\_b[\_cons]

predict pxb4, xb

replace pxb4=pxb4-con4

gen imr4=normalden(pxb4)/(1-normal(pxb4))

gen floordum=1

replace floordum=0 if dispositionfloor==1

replace floordum=0 if floorsquantity==dispositionfloor

replace floordum=0 if dispositionfloor==.

gen buildperiod=.

replace buildperiod=0 if buildyear<=1960

replace buildperiod=1 if buildyear>1960 & buildyear<=1985

replace buildperiod=2 if buildyear>1985 & buildyear<=2005

replace buildperiod=3 if buildyear>2005 & buildyear<=2013

reg flat\_value totalsquare livingsquare kitchensquare roomsquantity dispositionfloor floordum b1.buildperiod lavatory b1.city if outlier!=1 & flag\_of\_contract\_agreement==1, vce(hc3)

predict pxb5, xb

reg rate imr4 median\_rate b4.income\_cat borrower\_age bor\_age2 sex if outlier!=1 & flag\_of\_contract\_agreement==1, vce(hc3)

predict predrate, xb

reg creditperiod b4. income\_cat lodging\_coef\_in\_years median\_maturity borrower\_age bor\_age2 if outlier!=1 & flag\_of\_contract\_agreement==1, vce(hc3)

predict predperiod, xb

gen ln\_loan=ln(loan\_amount)

gen ln\_imr4=ln(imr4)

gen ln\_pxb5=ln(pxb5)

gen ln\_predrate=ln(predrate)

gen ln\_predperiod=ln(predperiod)

reg ln\_loan ln\_imr4 ln\_pxb5 ln\_predrate ln\_predperiod b4.income\_cat if outlier!=1 & flag\_of\_contract\_agreement==1, vce(hc3)

Приложение 2

Eq.2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Eq.2.1 | Eq.2.2 | Eq.2.3 |
| Вероятность подачи заявления | pxb1 | -.0032683\*  (.0017302) | -.0031479\*  (.0017283) | -.0031417\*  (.0017188) |
| Возраст заемщика | borrower\_age | -.0371138  (.0293189) | -.0363699  (.0290749) | - |
| Возраст заемщика во второй степени | bor\_age2 | .0006137  (.0004012) | .0006038  (.0003979) | - |
| Пол | sex | .0901104  ( .0567855) | .1047297\*\*  (.0528755) | - |
| Вид занятости | activity\_category |  |  |  |
|  | 0 | -.8206339\*\*  (.1247885) | -1.419857\*\*  (.2253564) | -.8291221\*\*  ( .1391574) |
|  | 1 | 0 (empty) | 0 (empty) | 0 (empty) |
|  | 3 | 0 (empty) | 0 (empty) | 0 (empty) |
|  | 5 | .5674622  (.4562643) | .5829592  (.4603843) | .6286374  (.4652923) |
|  | 6 | .5085807\*\*  (.1864853) | .5864818\*\*  ( .1805733) | .5780568\*\*  (.1800182) |
| Уровень образования | education\_level |  |  |  |
|  | 0 | -.8206339\*\*  (.1247885) | -.8191336\*\*  (.1238422) | -.8037639\*\*  (.1233847) |
|  | 1 | -.2638155  (.2091639) | -.2623572  (.2090354) | -.2380637  (.2085748) |
|  | 2 | -.3751748\*\*  ( .0570491) | -.3731107\*\*  (.0564111) | -.3508977\*\*  ( .0555548) |
|  | 3 | -.3161796\*\*  (.1416648) | -.3189168\*\*  (.1415353) | -.3121832\*\*  (.1406767) |
| Уровень дохода | income\_cat |  |  |  |
|  | 1 | -1.069563\*\*  (.1254186) | -1.104891\*\*  (.1171298) | -1.114828\*\*  (.116802) |
|  | 2 | .5013503  (.4081827) | .5793376  (.4008189) | .5554083  (.3956694) |
|  | 3 | .2551492  (.2123289) | .3260043  (.2106774) | .3343336  (.2115143) |
|  | 5 | -.1943274  ( .1853174) | -.2396389  (.1834851) | -.2240662  (.1829078) |
| Бюджетная ли сфера работы | work\_is\_budget | .1470346  (.0984434) | - | - |
| Количество созаемщиков | Number\_of\_CoBorrowers | -.0206797  (.1423579) | - | - |
| Уровень дохода созаемщиков | coborrower\_income\_cat |  |  |  |
|  | 0 | -.0727748  (.1603401) | - | - |
|  | 2 | .4756281  (.3148098) | - | - |
|  | 3 | .1821615  (.2068386) | - | - |
|  | 4 | -.1122698  (.210231) | - | - |
|  | 5 | -.1260732  (.4255882) | - | - |
| Константа | \_cons | 3.346814\*\*  (.6088965) | 3.401588\*\*  (.5658112) | 2.36101\*\*  (.1469437) |
|  | Pseudo R2 | 0.1501 | 0.1480 | 0.1451 |
|  | LR chi2 | 521.18  chi2(22) | 513.61  chi2(15) | 503.63  chi2(12) |
|  | Prob > chi2 | 0.0000 | 0.0000 | 0.0000 |

Приложение 3

Eq. 3

|  |  |  |
| --- | --- | --- |
|  |  | Eq.3 (robust regression) |
| Семейное положение | familystatus |  |
|  | 0 | -86410.02  (88797.97) |
|  | 1 | -60265.2\*\*  (24424.79) |
|  | 3 | -208702.3\*\*  (58628.01) |
|  | 4 | -78360.68\*\*  (30861.12) |
| Вид занятости | activity\_category |  |
|  | 0 | 86101.29  (120228.8) |
|  | 5 | 534597.7\*\*  (161168.4) |
|  | 6 | -62229.29 \*  (37277.03) |
| Уровень образования | education\_level |  |
|  | 0 | -86246.23\*  (48545.72) |
|  | 1 | -199933.4\*\*  (47646.51) |
|  | 2 | -157030\*\*  (20632.45) |
|  | 3 | 134330.1  (147522.87) |
| Уровень дохода | income\_cat |  |
|  | 1 | 136229.3  (163712.13) |
|  | 2 | -337829.2\*\*  (33676.84) |
|  | 3 | -272772.5\*\*  (22333.15) |
|  | 5 | 646404.7\*\*  (208318.9) |
| Возраст заемщика | borrower\_age | 36830.34\*\*  (9141.738) |
| Возраст заемщика во второй степени | bor\_age2 | -452.4067\*\*  (120.4377) |
| Уровень дохода созаемщиков | coborrower\_income\_cat |  |
|  | 1 | 58565.19  (77398.18) |
|  | 2 | -81998.81  (136086.9) |
|  | 3 | 9486.173  (32707.55) |
|  | 4 | 188374.4\*\*  (42466.2) |
|  | 5 | 553684.3\*\*  (208318.9) |
| Средняя сумма ипотечного кредита | mean\_loan | -266.637\*\*  (42.39893) |
| Константа | \_cons | 659554.6\*\*  (176906.7) |