

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
“ВЫСШАЯ ШКОЛА ЭКОНОМИКИ”

Домашняя работа
**“Оценка пространственной зависимости на первичном
рынке недвижимости в регионах России”**

В рамках курса
Пространственная эконометрика

**Выполнена студентами группы МЭКЭП231
ОП Экономика и Экономическая политика:**

Грачев Илья Сергеевич
Журба Анна Александровна
Козловская Полина Александровна
Петрова Екатерина Никитична

Москва, 2024

Оглавление

Введение.....	3
Описательная статистика.....	5
Пространственный анализ.....	10
Регрессионный анализ панельных данных.....	16
Пространственные модели панельных данных.....	21
Заключение.....	38
Список литературы.....	39
Приложения.....	40
Приложения для динамических моделей.....	42

Введение

Первичный рынок недвижимости является одним из ключевых секторов экономики государства. Это связано с тем, что постоянное обновление жилищного фонда и доступность личного жилья во многом определяют успех государственных проектов в других сферах, например, демографических. Однако, первичный рынок недвижимости России и его развитие вызывает активную дискуссию среди исследователей начиная с 2020 года: за последние 4 года было проведено более 5,000 исследований первичного рынка недвижимости в РФ¹.

Пространственная эконометрика и ее основные методы основываются на пространственной зависимости объектов. Согласно основным теоретическим предпосылкам пространственных моделей, значения зависимой переменной рассматриваемых объектов, помимо прочих объясняющих переменных, определяются значениями других рассматриваемых объектов с помощью добавления в вычисляемую линейную комбинацию матрицы пространственных весов. Таким образом, применение данных моделей требует использования панельных данных, а зависимая переменная может зависеть в большей степени как от данных соседних регионов, так и от данных удаленных регионов. Подробно основные методы пространственной эконометрики были описаны в множественных научных работах, начиная с 1980 года. [2]. Рассмотренные научные работы зачастую используют геометрически-взвешенную регрессию для расчета коэффициентов.

Исследования рынка жилья с применением методов пространственной эконометрики проводились и до 2020 года. Так, еще в 2016 году с помощью модели пространственного лага была отмечена зависимость средней цены квадратного метра жилья в регионе от доходов населения (положительная), плотности населения (положительная) и уровня преступности в регионе (отрицательная) [3]. Влияние на стоимость жилья на первичном рынке уровня занятости населения, доходов населения, реальной процентной ставки и доступности ипотечного кредитования также выделяют зарубежные эконометрические исследования, основанные на данных польского [6] и американского [7] рынков жилья. Последнее также отражает статистически значимую обратную зависимость стоимости жилья от факторов, связанных с пандемией 2020 года. В контексте российской экономики схожих исследований после 2020 года не

¹ Cyberleninka : сайт. – URL: <https://cyberleninka.ru>

проводилось, однако исследования с использованием линейных и нелинейных моделей (с использованием полиномиального расширения регрессоров) подтверждают выбор объясняющих переменных в данном исследовании [1].

Таким образом, цель данного исследования - оценить возможную пространственную зависимость цены квадратного метра на рынке первичного жилья в регионах РФ от социально-экономических показателей этих регионов. Основными задачами исследования выступают сбор и предварительная обработка данных, оценка наличия пространственной зависимости в зависимой переменной или ошибках с последующим применением соответствующей модели пространственной эконометрики, интерпретация результатов в контексте цели исследования.

Актуальность исследования обусловлена значительными изменениями в экономике России в связи с последствиями пандемии 2020 года. Данная статья позволит учесть эти изменения при определении эконометрической зависимости цены на жилищном рынке РФ от прочих факторов. При этом, ограничения в исследуемых данных не позволяют учесть влияние санкций 2022 года.

Описательная статистика

Используемые при моделировании переменные перечислены в Приложении 1. Источниками данных являются сайт Федеральной службы государственной статистики [4], а также сайт Банка России [5]. Таким образом, датасет состоит из 996 наблюдений по 83 регионам России в период с 2011 года. Выбор периода обусловлен наличием данных по всем используемым переменным за исключением некоторых регионов. В модели используются непрерывные или дискретные переменные, при этом среди переменных нет инвариантных ни по времени, ни по объекту.

Пропуски данных возникают в нескольких регионах России (Чукотский автономный округ, Мурманская область) по множеству переменных, включая целевую переменную: primary, secondary, а также переменных, связанных со статистикой ипотечного кредитования. Поскольку нет возможности произвести замену пропущенных значений ввиду отсутствия наблюдений в течении длительного периода, было принято решение не рассматривать регионы в данном исследовании. Всего в данных 81 регион и 12 лет. Пропуски в некоторых переменных составляли меньше 1%, они были заменены на среднее значение региона за прошлый и будущий периоды, в ином случае проставлялись средние значения региона за весь период.

Таблица 1. Описательная статистика

vars	n	mean	sd	median	min	max
pop	972	1775,55	1832,95	1189,85	41,40	13104,20
wforce_gr	972	-0,89	1,47	-1,27	-3,90	4,38
unempl	972	6,52	4,41	5,50	0,80	48,10
income_pc	972	28323,89	13287,46	25411,50	8829,00	116639,00
primary	972	54857,32	28268,78	46716,00	18206,00	374658,00
secondary	972	53939,11	24766,22	46902,50	24029,00	297454,00
constructprice	972	106,42	6,38	106,00	74,30	144,30
avg_mortgage_rate	972	10,66	2,15	10,86	5,19	14,63
migration_rate	972	-11,20	70,97	-14,00	-307,00	274,00
crime_gr	972	-1,59	8,28	-2,23	-22,76	40,27
mquant_gr	972	15,83	32,74	22,08	-64,16	405,23

При рассмотрении описательной статистики стоит отметить необходимость логарифмического масштабирования переменных с целью избежать большой разницы в масштабах рассматриваемых переменных. Рисунок 1 демонстрирует примеры переменных, к которым применяется логарифмическое масштабирование.

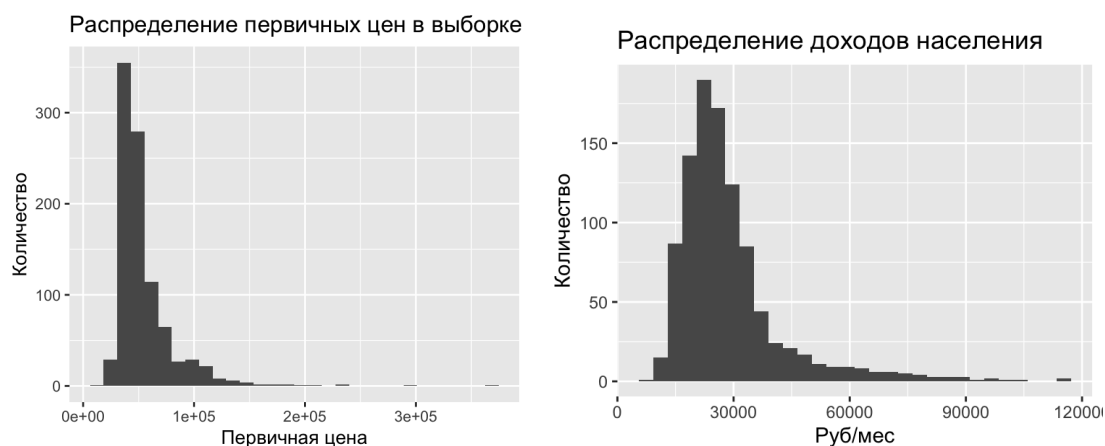


Рисунок 1. Распределение переменных `primary` и `income_pc`

Другими переменными, к которым применяется масштабирование, являются `secondary`, `total_mortgage_volume`, `total_mortgage_quantity`, `crime`. Таким образом, распределение переменных принимает вид, аналогичный представленному на Рисунке 2.

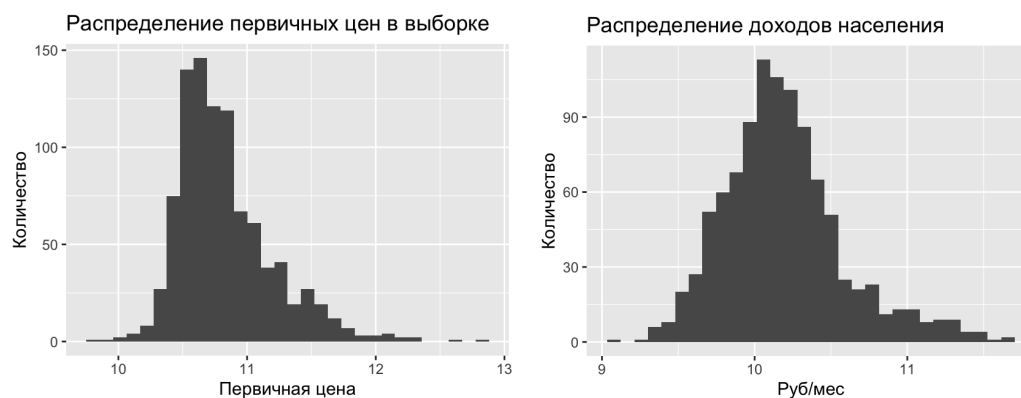


Рисунок 2. Распределение масштабированных переменных `primary` и `income_pc`

Динамика цен на первичное жилье за каждый год продемонстрирована на Рисунке 3. В динамике целевой переменной можно наблюдать растущий тренд, а также рост межквартильного размаха, что указывает на растущую волатильность цены квадратного метра на первичном рынке жилья.

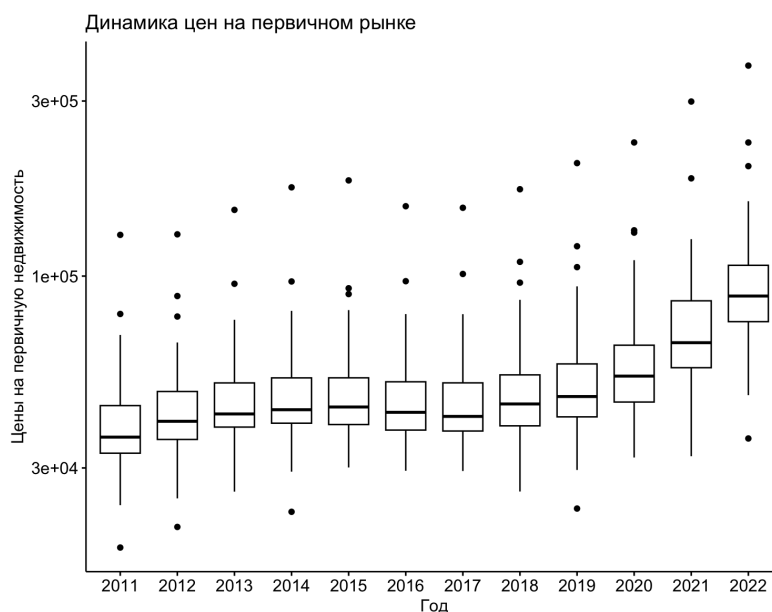


Рисунок 3. Динамика цен на первичном рынке жилья

На карте ниже можно увидеть как распределены средние значения цен на первичную недвижимость по регионам. Можно отметить, что в центральной России цены ниже, чем в Сибири, а в Сибири выше, чем на Дальнем Востоке, чуть поднимаются цены на жилье у приграничных южных регионов и у городов федерального значения.

Средние цены на первичное жилье по регионам

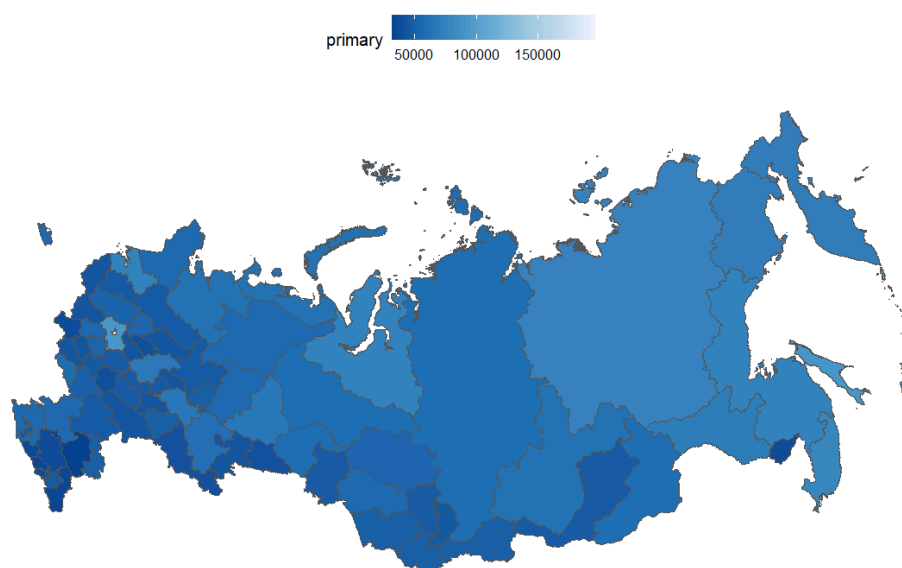


Рисунок 4. Карта средних за 2011-2022г. цен на первичную недвижимость

В Приложении 2 представлена матрица парных графиков исследуемых данных. Все переменные распределены нормально, кроме ипотечной ставки, которая имеет несколько пиков, но можно увидеть, что распределения по годам уже имеют привычную форму.

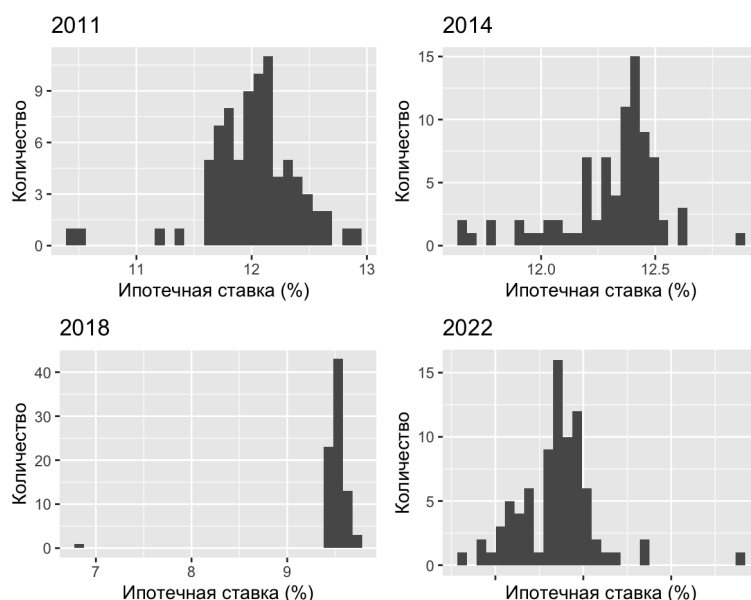


Рисунок 5. Распределение ставки ипотечного кредитования

При исследовании матрицы корреляций можно выделить несколько гипотез относительно данных. Переменная, отражающая объем выданных ипотечных кредитов, может быть эндогенна, поскольку объем зависит от зависимой переменной. С целью улучшения качества модели стоит рассмотреть замену объема выданных ипотечных кредитов на прирост количества выданных ипотечных кредитов. Для учета эндогенности цен на первичном и вторичном рынке, в качестве переменных-инструментов могут быть рассмотрены лаговые значения цен на вторичном рынке жилья.

Также стоит отметить отрицательную корреляцию между статистикой населения и количеством выданных ипотечных кредитов в регионе. Это экономически интуитивно, населения определяет его экономическую активность в реальных значениях. В теории, можно рассмотреть повышение качества модели за счет рассмотрения прироста количества выданных ипотек в переменной `mquant_gr`. Аналогичная логика может

объяснить положительную корреляцию между среднемесячными доходами населения и количеством выданных ипотек.

В итоговой модели исходя из статистического и корреляционного анализа и литературного обзора было принято решение использовать $\log(\text{primary})$, $\log(\text{secondary})$, $wforce_gr$, $unempl$, $\log(\text{income_pc})$, $constructprice$, $crime_gr$, $\log(\text{pop})$, $mquant_gr$, $avg_mortgage_rate$, $migration_rate$. Их точные описания можно найти в Приложении. Тем самым, мы считаем, что на среднюю цену первичного жилья в регионе влияют цены на вторичное жилье прошлого периода, приросты рабочей силы, преступлений, количества выданных ипотек субъекта, а также уровень безработицы, доходы населения, цены на строительные материалы, население, ипотечная ставка и уровень региональной миграции.

Пространственный анализ

Для проведения пространственной автокорреляции необходимо выбрать пространственную матрицу весов, которая позволит оценить корреляцию между значениями атрибутов объектов, основываясь на их пространственном расположении.

В пространственной матрице весов каждому объекту присваивается вес, который зависит от смежности с объектами. Это позволяет оценить степень пространственной зависимости между объектами и выявить пространственные закономерности в данных. В нашем исследовании используется матрица весов на основе соседства, стандартизированная по строкам. Мы выбрали данный вид матрицы, поскольку при исследовании рынка недвижимости нас интересует непосредственная близость и соседство объектов, поскольку мотивируем это тем, что покупатель из одного региона будет брать во внимание соседние регионы, поэтому веса у каждого соседа стандартизированы.

Глобальная пространственная автокорреляция учитывает общую тенденцию в наборе данных, и ее анализ позволяет сделать выводы о степени кластеризации в наборе. Пространственная автокорреляция – инструмент, который оценивает, есть ли в объектах кластеры или они распределены разбросанно, или случайно.

Одним из самых популярных показателей для выявления пространственных кластеров является индекс Морана. Индекс Морана – статистический показатель, который оценивает общую структуру данных и используется при определении пространственной автокорреляции. Данный индекс рассчитывается по формуле:

$$I(Y) = \frac{\sum_{i,j=1}^n w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_i^n (Y_i - \bar{Y})^2}$$

При значении индекса Морана больше нуля наблюдается положительная пространственная автокорреляция — то есть значения наблюдений в соседних территориях похожи. При отрицательных значениях индекса Морана пространственная корреляция отрицательная, наблюдения на соседних территориях отличаются. В случае, когда индекс Морана равен нулю, наблюдения на соседних территориях расположены случайным образом.

Глобальный индекс Гири, так же, как и индекс Морана, показывает наличие положительной или отрицательной автокорреляции. Данный индекс рассчитывается по формуле:

$$C(Y) = \frac{(n-1) \sum_{i,j=1}^n w_{ij}(Y_i - Y_j)^2}{\sum_i^n (Y_i - \bar{Y})^2}$$

С помощью индекса Гири проверяется, насколько близки значения выбранного показателя в соседних регионах.

Кроме этого, был рассчитан индекс Гетиса – Орда, который в отличие от индексов Морана и Гири, позволяет измерить не положительную или отрицательную пространственную автокорреляцию, а кластеризацию регионов с высокими или низкими значения показателя. Индекс Гитеса – Орда рассчитывается по формуле:

$$G(Y) = \frac{\sum_{i,j=1}^n w_{ij} (Y_i - Y_j)^2}{\sum_i^n (Y_i - \bar{Y})^2}$$

Рассчитанные индексы Морана, Гири и Гитеса-Орда представлены в таблице ниже:

	Глобальные индексы		
Год	Индекс Морана	Индекс Гири	Индекс Гитеса-Орда
2011	0.33***	0.43***	0.01*
2012	0.36***	0.46***	0.01**
2013	0.32***	0.45***	0.01**
2014	0.31***	0.44***	0.01**
2015	0.29***	0.43***	0.01
2016	0.33***	0.43***	0.01*
2017	0.29***	0.46***	0.01
2018	0.30***	0.45***	0.01*
2019	0.27***	0.45***	0.01*
2020	0.32***	0.43***	0.01**
2021	0.32***	0.44***	0.01**
2022	0.30***	0.48***	0.01**

Таблица 2. Индексы пространственных корреляций

Исходя из полученных результатов можно сделать вывод что в исследуемых данных наблюдается значимая положительная пространственная зависимость на протяжении всего периода 2011-2022 гг.

Важным этапом анализа полученных данных является построение пространственной диаграммы рассеяния Морана. По оси абсцисс откладываются стандартизированные z-значения исследуемого показателя, а по оси ординат – значения пространственного фактора Wz. На диаграмме отображается линия регрессии Wz на z, наклон который равен коэффициенту общей пространственной автокорреляции I при стандартизированной матрице весов.

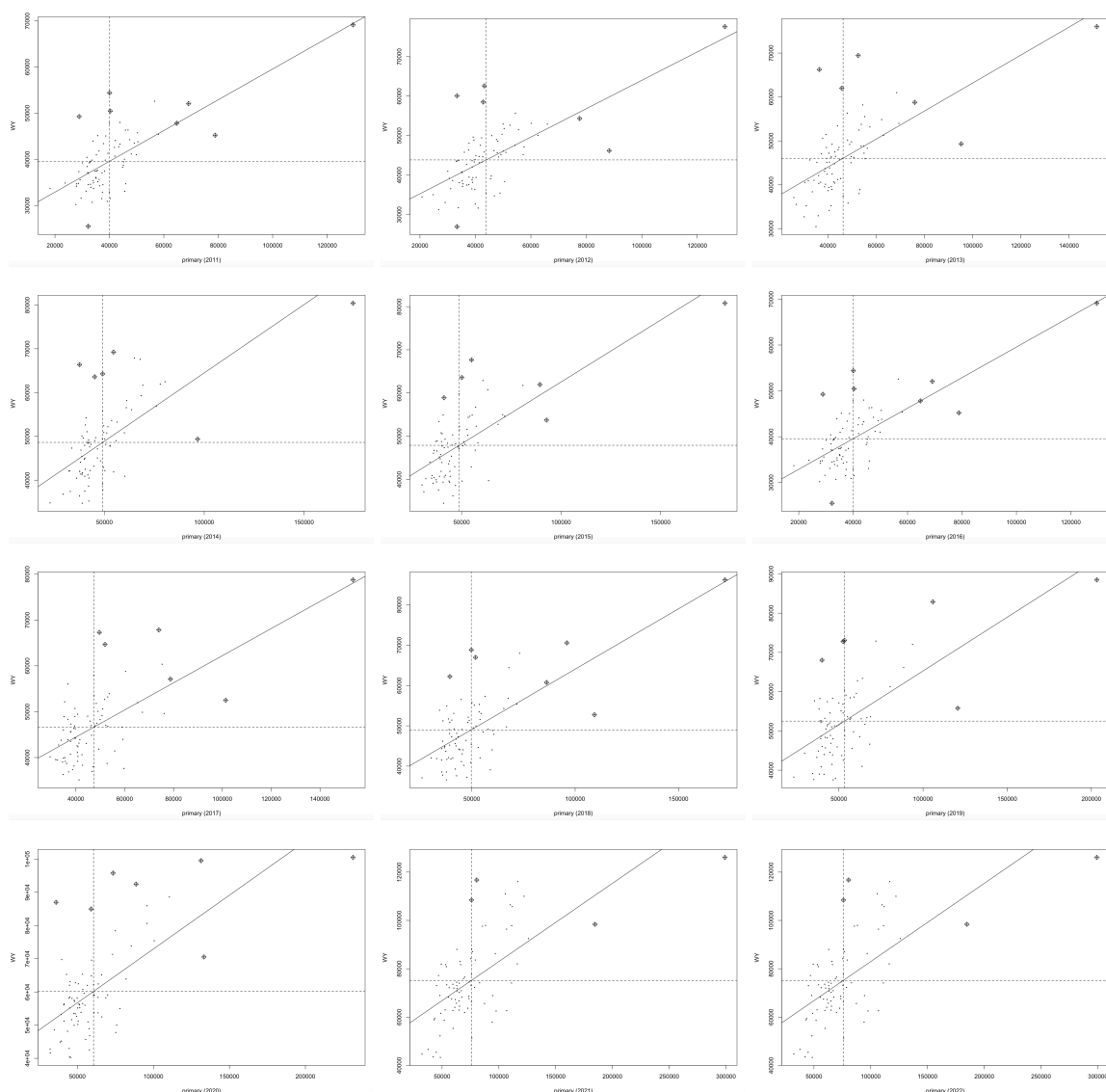


Рисунок 6. Пространственные диаграммы рассеяния Морана

На диаграммах можно заметить положительную корреляцию между переменными, что указывает на наличие положительной пространственной автокорреляции. Другими словами, средняя цена на первичном рынке жилья среди всего населения регионов демонстрирует положительную пространственную автокорреляцию — регионы с высокой средней ценой на первичном рынке жилья расположены близко друг к другу, как и регионы с низкой средней ценой.

Глобальная пространственная корреляция не показывает, где значения сконцентрированы в кластеры или имеют рассеянную структуру. Чтобы детектировать такие значения необходимо использовать локальные индексы корреляции. Они фокусируются на взаимосвязях между каждым наблюдением и его окружением. Одной из метрик для измерения локальной пространственной корреляции является LISA (Local Indicators of Spatial Association).

LISA – это модификация глобального индекса Морана. Этот показатель рассчитывается аналогично глобальному индексу Морана, но для каждого значения в отдельности. С помощью него можно определить является ли каждый элемент данных статистически значимым с точки зрения пространственной связи с соседними элементами. Это помогает выявить кластеры и разреженные области в данных.

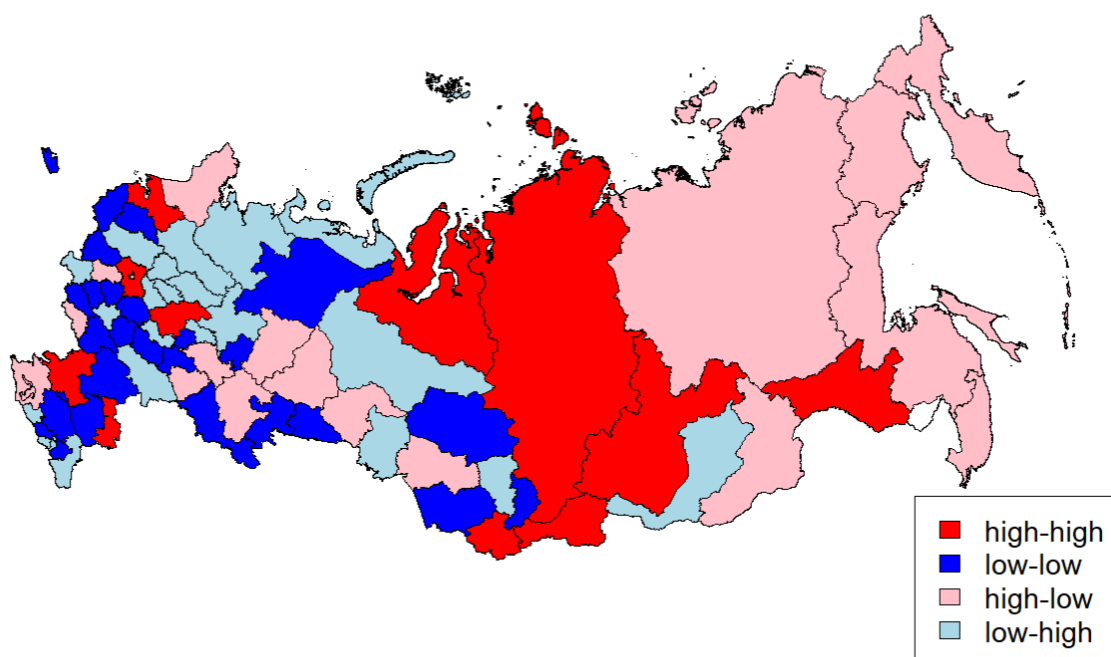
Квадрант “high-high” – территории в нем имеют относительно высокие собственные значения анализируемого показателя, окружены территориями также с относительно высокими значениями анализируемого показателя. Автокорреляция положительная.

Квадрант “low-low” – территории в нем имеют относительно низкие собственные значения анализируемого показателя, окружены территориями также с относительно низкими значениями анализируемого показателя. Автокорреляция положительная.

Квадрант “high-low” – территории в нем имеют относительно высокие собственные значения анализируемого показателя, окружены территориями с относительно низкими значениями анализируемого показателя. Автокорреляция отрицательная.

Квадрант “low-high” – территории в нем имеют относительно низкие собственные значения анализируемого показателя, окружены территориями с относительно высокими значениями анализируемого показателя. Автокорреляция отрицательная.

Карта локальных индексов Морана для первичных цен на недвижимость, 2022 г.:



К кластеру “high-high” (красный) относятся следующие регионы: Красноярский край, Ямало-Ненецкий АО, Иркутская область, Амурская область, Республика Тыва, Республика Алтай, Ленинградская область, Московская область, Нижегородская, Астраханская область, Ростовская область. Данные регионы демонстрируют высокие значения на рынке недвижимости, и соседние регионы также характеризуются высокими показателями. Это свидетельствует о пространственной зависимости: данные регионы формируют ядро активного роста на рынке недвижимости, что, вероятно, связано с их экономическим развитием, природными ресурсами и демографической активностью.

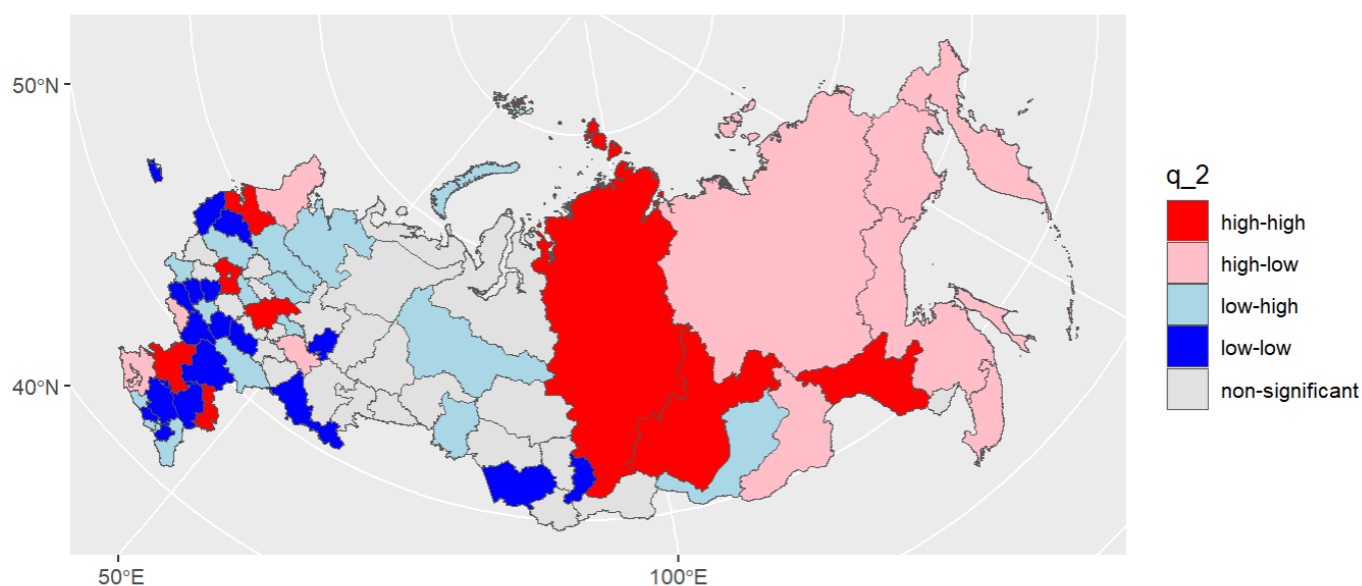
К кластеру “high-low” (розовый) относятся следующие регионы: Москва, Республика Саха (Якутия), Магаданская область, Камчатский край, Сахалинская область, Хабаровский край, Приморский край, Забайкальский край, Новосибирская область, Тюменская область, Свердловская область, Пермский край, Республика Башкортостан, Республика Татарстан, Самарская область, Республика Карелия, Белгородская область, Калужская область, Краснодарский край. Эти регионы показывают высокие показатели на рынке недвижимости, несмотря на то, что соседние регионы (например, те, что относятся к low-low) демонстрируют низкие показатели. Это отражает контрасты в экономическом развитии между столичным регионом и окружающими областями, где центр притягивает инвестиции и миграцию, а периферийные регионы могут не успевать за этим ростом.

К кластеру “low-low” (синий) относятся следующие регионы: Республика Коми, Томская область, Алтайский край, Республика Хакасия, Курганская область, Челябинская область, Оренбургская область, Республика Удмуртия, Псковская область, Новгородская область, Калининградская область, Смоленская область, Курская область, Орловская область, Тульская область, Рязанская область, Воронежская область, Тамбовская область, Пензенская область, Ульяновская область, Республика Чувашия, Волгоградская область, Ставропольский край, Республика Калмыкия, Чеченская Республика, Республика Кабардино-Балкария. Регионы имеют низкие значения на рынке недвижимости. Их соседние регионы также демонстрируют аналогичные низкие показатели, что свидетельствует о пространственной зависимости среди регионов с менее развитым рынком недвижимости. Это может быть связано с экономическими трудностями, демографическим спадом или низким уровнем инвестиционной привлекательности.

К кластеру “low-high” (голубой) относятся следующие регионы: Республика Бурятия, Кемеровская область, Ханты-Мансийский АО, Омская область, Ненецкий АО, Брянская область, Архангельская область, Вологодская область, Владимирская область, Тверская область, Ивановская область, Костромская область, Кировская область, Ярославская область, Республика Марий Эл, Республика Мордовия, Липецкая область, Саратовская область, Республика Дагестан, Республика Карачаево-Черкессия, Республика Северная Осетия, Республика Ингушетия. Такие регионы демонстрируют обратную зависимость, где низкие значения соседствуют с высокими. Это может указывать на дисбаланс в

экономическом развитии: регионы с высокой активностью на рынке недвижимости могут оказывать давление на соседние менее активные регионы, но эти соседи пока не реагируют аналогичным экономическим подъёмом.

Карта локальных индексов Морана с симуляцией (сэмплингом) наблюдений для первичных цен на недвижимость, 2022:



Большая часть регионов остались в своих кластерах. Что касается регионов с незначимой пространственной корреляцией (non-significant, серый), к которым относятся: Ямало-Ненецкий АО, Ненецкий АО, Республика Коми, Томская область, Новосибирская область, Кемеровская область, Республика Алтай, Тюменская область, Республика Тыва, Курганская область, Челябинская область, Республика Башкортостан, Пермский край, Кировская область и др. В этих регионах либо отсутствует автокорреляция, либо не является статистически значимой. Это означает, что рынок недвижимости может развиваться независимо от соседних территорий.

Регрессионный анализ панельных данных

В начале исследования мы решили оценить модели панельных данных со случайными, фиксированными эффектами и фиксированными по годам эффектами.

1. Модель со случайными эффектами (Random Effects Model, re)

В модели со случайными эффектами предполагается, что индивидуальные эффекты (не наблюдаемые) могут быть случайными и независимыми от других объясняющих переменных.

Форма модели:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \dots + \beta_k x_{kit} + u_i + \epsilon_{it}$$

где:

u_i - случайный индивидуальный эффект (не наблюдается);

ϵ_{it} - ошибка, предполагающая, что u_i независима от ϵ_{it} .

2. Модель с фиксированными эффектами (Fixed Effects Model, fe)

В модели с фиксированными эффектами предполагается, что индивидуальные эффекты могут коррелировать с объясняющими переменными, и они фиксированы для каждого объекта.

Форма модели:

$$y_{it} = \alpha_i + \beta_1 x_{it} + \dots + \beta_k x_{kit} + \epsilon_{it}$$

где:

$y_{it}, x_{it}, \beta_1, \dots, \beta_k, \epsilon_{it}$ — такие же, как и в предыдущей модели;

α_i — фиксированный индивидуальный эффект для объекта i , который уникален и не меняется во времени.

3. Модель с фиксированными эффектами по годам, (Time Fixed Effects Model, fe_year)

Эта модель учитывает фиксированные эффекты как для объектов, так и для временных периодов. Это может быть важно, если время также оказывает значительное влияние на зависимую переменную.

Форма модели:

$$y_{it} = \alpha_i + \gamma_t + \beta_1 x_{it} + \dots + \beta_k x_{kit} + \epsilon_{it}$$

где:

– γ_t – фиксированные эффекты (или "тренды") для времени t. Это позволяет контролировать специфические эффекты во времени, которые могут влиять на y_{it}

Такая модель позволяет более точно оценивать влияние независимых переменных, учитывая как индивидуальные, так и временные эффекты.

Результаты оценки моделей re, fe, fe_year представлены в таблице ниже.

	re	fe	fe_year
VARIABLES	logprimary	logprimary	logprimary
L.loglagsecondary	0.723***	0.687***	0.416***
	(0.0357)	(0.0441)	(0.0415)
wforce_gr	-0.00994**	-0.0124***	0.0267*
	(0.00456)	(0.00472)	(0.0139)
unempl	0.00416*	0.00258	0.00660**
	(0.00227)	(0.00322)	(0.00284)
logincome_pc	0.125***	-0.0317	0.337***
	(0.0357)	(0.0501)	(0.0823)
constructprice	0.00214***	0.00204***	0.000285
	(0.000667)	(0.000656)	(0.000574)
crime_gr	-0.000885*	-0.000722	6.02e-05
	(0.000518)	(0.000511)	(0.000484)

logpop	0.0366***	-0.122	0.212
	(0.0131)	(0.162)	(0.137)
mquant_gr	-0.00124***	-0.00163***	-0.000545***
	(0.000150)	(0.000167)	(0.000178)
avg_mortgage_rate	-0.0488***	-0.0603***	-0.0286***
	(0.00284)	(0.00368)	(0.00948)
migration_rate	0.000135	8.32e-05	-0.000542***
	(0.000136)	(0.000163)	(0.000177)
Constant	1.789***	5.044***	1.734
	(0.376)	(1.251)	(1.300)
Observations	891	891	891
R-squared		0.776	0.849
Number of reg_id	81	81	81

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Таблица - “ Модели панельных данных со случайными, фиксированными и фиксированными на году эффектами”

При оценке панельных моделей коэффициенты при переменных $L.loglagsecondary$, $mquant_gr$ и $avg_mortgage_rate$ имеют постоянную значимость. При этом, прирост трудоспособного населения в моделях без учета дамми-переменных $year$ имеет отрицательный коэффициент, что противоречит экономическому смыслу роста спроса на жилье на первичном рынке. Аналогичный вывод можно сделать о коэффициенте при уровне доходов населения $logincome_pc$, который оказался значимым только в последней модели, однако его значимость имеет явную экономическую интерпретацию.

Таким образом, можно предположить, что учет в модели особенностей данных в разрезе по рассматриваемым периодам, то есть введение дамми переменных (индикаторов) на каждый рассматриваемый год, может повысить качество оценок модели. Из введенных в модели fe_year подобных переменных значимыми оказываются 2015-2019 годы. Среди прочих причин, это может быть объяснено

отложенным эффектом санкций на российскую экономику. Однако, стоит отметить, что эта модель имеет наибольшие оценки стандартного отклонения коэффициентов.

year	fe_year
2013.year	-0.0220
	(0.0180)
2014.year	-0.0157
	(0.0222)
2015.year	-0.0784**
	(0.0352)
2016.year	-0.0994***
	(0.0307)
2017.year	-0.152***
	(0.0316)
2018.year	-0.141***
	(0.0383)
2019.year	-0.180***
	(0.0540)
2020.year	-0.0765
	(0.0517)
2021.year	-0.0148
	(0.0677)
2022.year	0.158**
	(0.0668)

Таблица - “Временные фиксированные эффекты в моделях панельных данных ”

Коэффициент при приросте количества выданных ипотечных кредитов *mquant_gr*, несмотря на индикацию роста спроса на жилье, имеет отрицательный коэффициент, который значим на любом уровне значимости во всех рассмотренных моделях. Это позволяет сформировать гипотезу о необходимости учета лагов рассматриваемых переменных в модели.

Также мы проводили тест Хаусмана, для того чтобы определить какая модель предпочтительнее re или fe:

Test: Ho: difference in coefficients not systematic

$$\chi^2(10) = (b-B)'[(V_b - V_B)^{-1}](b-B) = 29.91$$

$$\text{Prob} > \chi^2 = 0.0009$$

На основании теста Хаусмана мы сделали вывод, что для исследуемых данных больше подходит модель с фиксированными эффектами (fe).

	RE	FE	FE_YEAR
AIC	-	-1111.298	-1616.933
BIC	-	-1058.582	-1516.294

Таблица - “Информационные критерии Акаике и Байеса для моделей панельных данных”

Опираясь на информационные критерии AIC, BIC более подходящей оказалась модель fe с фиксированными эффектами.

Пространственные модели панельных данных

Когда речь идет о пространственной зависимости в моделях панельных данных, можно использовать модели, учитывающие как случайные эффекты, так и пространственные зависимости. Начнем наш анализ с построения моделей со случайными эффектами.

Пространственные модели панельных данных со случайными эффектами.

1. Модель со случайными эффектами SEM (Spatial Error Model)

Пространственную зависимость шоков для соседних географических единиц можно учитывать с помощью модели с пространственной зависимостью в ошибках SEM (spatial error model):

$$Y_{it} = X_{it}\beta + u_i + \epsilon_{it}$$

u_i — случайный эффект, который учитывает пространственную зависимость (например, через простую пространственную автокорреляцию),

ϵ_{it} — ошибка, которая нормально распределена и независима
, $\epsilon_{it} \sim N(0, \sigma^2)$

$$\epsilon = \lambda W\epsilon + u$$

2. Модель со случайными эффектами SAR (Spatial Autoregressive Model)

Модель SAR включает пространственную авторегрессию в зависимости самой модели. Формула данной модели может быть записана следующим образом:

$$Y_{it} = \rho WY_{it} + X_{it}\beta + u_i + \epsilon_{it}$$

, где:

ρWY - пространственный лаг зависимой переменной (среднее значение показателя в соседних регионах, не считая текущий регион).

Далее мы решили оценить на наших данных модели SEM и SAR.

Для модели SEM коэффициент пространственной зависимости лямбда оказался значим на 1% уровне значимости, что говорит о наличии пространственной корреляции ошибок. Аналогично, оценивая модель SAR коэффициент при пространственной лаге первичной цены на недвижимость оказался значим. Таким образом, мы можем сделать вывод, что действительно в наших данных присутствует пространственная связь, что подтверждалось индексами пространственной корреляции Морана, Гир и Геттиса-Орда.

Также видно, что в обеих моделях одинаковые значимые переменные, а их знаки сонаправлены, что говорит о верной и постоянной спецификации модели

	sem_re			sar_re					
VARIABLES	Main	Spatial	Variance	Main	Spatial	Variance	LR_Direct	LR_Indirect	LR_Total
loglagsecondary	0.644***			0.560***			0.576***	0.257***	0.833***
	(0.0321)			(0.0290)			(0.0287)	(0.0297)	(0.0368)
wforce_gr	-0.0116**			-0.00471			-0.00446	-0.00199	-0.00644
	(0.00493)			(0.00360)			(0.00352)	(0.00160)	(0.00510)
unempl	0.00685***			0.00814***			0.00832***	0.00374***	0.0121***
	(0.00189)			(0.00177)			(0.00175)	(0.000967)	(0.00263)
logincome_pc	0.227***			0.126***			0.129***	0.0574***	0.186***
	(0.0326)			(0.0253)			(0.0247)	(0.0125)	(0.0353)
constructprice	0.000443			0.000406			0.000437	0.000196	0.000633
	(0.000531)			(0.000507)			(0.000508)	(0.000234)	(0.000739)
crime_gr	-0.000212			-0.000355			-0.000365	-0.000161	-0.000526
	(0.000440)			(0.000398)			(0.000421)	(0.000188)	(0.000607)
logpop	0.0451***			0.0630***			0.0641***	0.0289***	0.0930***
	(0.0133)			(0.0131)			(0.0132)	(0.00789)	(0.0205)
mquant_gr	-0.000711***			-0.000446***			-0.000447***	-0.000198***	-0.000645***
	(0.000150)			(0.000129)			(0.000130)	(5.75e-05)	(0.000182)
avg_mortgage_rate	-0.0299***			-0.0145***			-0.0149***	-0.00660***	-0.0215***
	(0.00317)			(0.00287)			(0.00272)	(0.00117)	(0.00362)
migration_rate	0.000125			5.41e-05			5.46e-05	2.47e-05	7.92e-05
	(0.000118)			(0.000109)			(0.000112)	(5.18e-05)	(0.000164)
lambda		0.308***							
		(0.0494)							
ln_phi			-0.0858						
			(0.189)						
sigma2_e			0.00998***			0.00924***			

			(0.000481)			(0.000445)			
rho					0.326***				
					(0.0310)				
lgt_theta						-0.951***			
						(0.123)			
Constant	1.462***			-0.431					
	(0.362)			(0.358)					
Observations	972	972	972	972	972	972	972	972	972
R-squared	0.856	0.856	0.856	0.855	0.855	0.855	0.855	0.855	0.855
Number of reg_id	81	81	81	81	81	81	81	81	81

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Таблица - “Сравнение моделей SEM и SAR со случайными эффектами”

Положительный коэффициент при вторичных ценах на недвижимость (loglagsecondary) экономически верен, так как свидетельствует что повышение цен на вторичном рынке недвижимости приводит к повышению цен на первичном. А с ростом среднедушевых доходов населения (logincome_pc) увеличиваются цены на первичном рынке. Экономической интуиции также соответствует обратная зависимость ставки по ипотеке с первичными ценами, так как действительно, чем ниже ставка по ипотеке, тем дешевле становится брать ипотеки, спрос на них растет и это приводит к росту цен на недвижимость (и наоборот). Странно выглядят коэффициент в SEM модели и предельные эффекты в SAR- модели перед приростом рабочей силы (wforce_gr), так как отрицательный знак говорит, чем меньше прирост рабочего населения, тем выше цены на первичную недвижимость, что противоречит экономической интуиции.

3. Модель Дарбина со случайными эффектами (Spatial Darbin Model, SDM)

Формально, модель можно записать следующим образом:

$$y_{it} = \rho * W * y_{it} + \beta * X_{it} + D * Z_{it} * \theta + \alpha_i + \gamma_t + u_{it}$$

где:

$D*Z$ - пространственный лаг регрессоров X

Модель Дарбина позволяет использовать разные весовые матрицу для пространственной зависимой переменной (W) и пространственных регрессоров (D) вместе с разными наборами объясняющих переменных $D* (X_{it}) = (Z_{it})$. По умолчанию мы использовали $W=D$ и $X_{it}=Z_{it}$.

После мы решили добавить в модель пространственный лаги объясняющих переменных. Для этого мы построили модель Дарбина с пространственными лагами всех регрессоров, посмотрели на их значимость и потом построили модель Дарбина со значимыми пространственными лагами объясняющих переменных. Среди них оказались следующие объясняющие переменные:

loglagsecondary - Средние цены на вторичном рынке жилья, на конец года; рублей за квадратный метр общей площади

logpop - Среднегодовая численность населения, тыс. чел.

wforce_gr - Прирост трудоспособного населения, %

unemp - Уровень безработицы, %

logincom_pc - Среднедушевые денежные доходы населения, руб. в мес

crime_gr - Прирост количества зарегистрированных преступлений, %

mquant_gr - Прирост количества выданных за год ипотечных жилищных кредитов, %

В модели Дарбина сохранились значимыми те же регрессоры, что и в моделях SEM и SAR. Знаки остались сонаправленными.

Среди пространственных лагов WX значимыми остались все те же регрессоры, что мы подавали в модель, кроме $unemp_l$ - уровня безработицы.

$Wlogincome_{pc}$: если уменьшается среднедушевой доход в соседних регионах, то растет первичная цена. Возможно есть миграция из соседних регионов, например, студентов, что приводит к снижению дохода в регионах из которых мигрируют.

$Wlogpop$: с уменьшением населения в соседних регионах, в выбранном регионе растет цена на первичную недвижимость. Это может быть объяснено тем, что возможно если население убывает в соседних регионах, значит есть миграция из этих регионов в исходный.

	sdm_re						
VARIABLES	Main	Wx	Spatial	Variance	LR_Direct	LR_Indirect	LR_Total
loglagsecondary	0.512***	0.120**			0.530***	0.333***	0.863***
	(0.0308)	(0.0490)			(0.0304)	(0.0453)	(0.0440)
logpop	0.0728***	-0.0685***			0.0692***	-0.0612**	0.00802
	(0.0138)	(0.0233)			(0.0131)	(0.0302)	(0.0317)
wforce_gr	-0.0212**	0.0174*			-0.0195**	0.0148*	-0.00470
	(0.00940)	(0.00903)			(0.00864)	(0.00853)	(0.00503)
unempl	0.0112***	-0.00558			0.0110***	-0.00349	0.00748*
	(0.00203)	(0.00354)			(0.00189)	(0.00417)	(0.00408)
logincome_pc	0.243***	-0.177***			0.236***	-0.147***	0.0889**
	(0.0451)	(0.0518)			(0.0407)	(0.0532)	(0.0374)
constructprice	0.000107				0.000134	4.33e-05	0.000177
	(0.000504)				(0.000505)	(0.000175)	(0.000677)
crime_gr	-4.90e-05	-0.00187***			-0.000171	-0.00248***	-0.00266***
	(0.000426)	(0.000683)			(0.000436)	(0.000842)	(0.000932)
avg_mortgage_rate	-0.0178***				-0.0182***	-0.00616***	-0.0244***
	(0.00308)				(0.00295)	(0.00129)	(0.00363)
mquant_gr	-0.000132	-0.000756***			-0.000169	-0.00104***	-0.00121***
	(0.000187)	(0.000254)			(0.000175)	(0.000286)	(0.000246)
migration_rate	0.000172				0.000172	5.84e-05	0.000231
	(0.000125)				(0.000129)	(4.73e-05)	(0.000174)
rho			0.267***				
			(0.0405)				
lgt_theta				-0.905***			
				(0.125)			
sigma2_e				0.00901***			

				(0.000432)			
Constant	0.548						
	(0.459)						
Observations	972	972	972	972	972	972	972
R-squared	0.870	0.870	0.870	0.870	0.870	0.870	0.870
Number of reg_id	81	81	81	81	81	81	81

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Таблица - “Модель Дарбина со случайными эффектами”

4. Модель GSPRE (Generalized Spatial Panel Regression Equation)

Также мы решили оценить модель с пространственной зависимостью в ошибках и в случайных эффектах.

GSPRE используется в пространственной эконометрике для анализа данных, которые имеют пространственную и временную структуру. Эта модель объединяет элементы как пространственной регрессии, так и панельных данных, что позволяет учитывать как временные, так и пространственные зависимости в данных.

$$y_{it} = \beta * X_{it} + a_i + v_{it}$$

$$a_i = \phi * W * a_i + \mu_i$$

$$v_{it} = \lambda * E * v_{it} + u_{it}$$

Форма модели:

Одним из основных преимуществ GSPRE является гибкость, так как модель может включать различные типы пространственных эффектов, например, пространственную автокорреляцию), а также взаимодействия между переменными, что делает модель достаточно гибкой для различных исследований.

GSPRE			
VARIABLES	Main	Spatial	Variance
loglagsecondary	0.627***		
	(0.0327)		

logpop	0.0528***		
	(0.0134)		
wforce_gr	-0.0110**		
	(0.00499)		
unempl	0.00743***		
	(0.00197)		
logincome_pc	0.216***		
	(0.0337)		
constructprice	0.000438		
	(0.000530)		
crime_gr	-0.000184		
	(0.000440)		
avg_mortgage_rate	-0.0315***		
	(0.00328)		
mquant_gr	-0.000745***		
	(0.000152)		
migration_rate	0.000111		
	(0.000120)		
phi		0.445***	
		(0.134)	
lambda		0.313***	
		(0.0492)	
sigma_mu			0.0908***
			(0.00809)
sigma_e			0.0997***
			(0.00240)
Constant	1.715***		
	(0.380)		

Observations	972	972	972
R-squared	0.854	0.854	0.854
Number of reg_id	81	81	81

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Таблица - “Модель GSPRE (Generalized Spatial Panel Regression Equation)”

Радует, что у нас сохранились знаки перед значимыми регрессорами

Пространственные модели панельных данных с фиксированными эффектами

1. SEM (Spatial Error Model) с фиксированными эффектами.

Модель SEM с фиксированными эффектами можно записать следующим образом:

$$y_{it} = X_{it}\beta + \lambda W\epsilon + u_{it}$$

λ — параметр пространственной зависимости.

ϵ — ошибка модели, которая также может иметь пространственную зависимость.

u_{it} — специфическая ошибка для единицы i в период t , которая включает фиксированный эффект.

2. SAR (Spatial Autoregressive Model) с фиксированными эффектами

$$y_{it} = \rho W y + X_{it}\beta + u_{it}$$

ρ — параметр пространственной зависимости, который измеряет эффект соседей.

u_{it} — специфическая ошибка для единицы i в период t , которая включает фиксированный эффект.

В моделях со случайными эффектами пространственной зависимости предполагается, что ошибки распределены нормально и независимы, тогда как фиксированные эффекты более чувствительны к выбросам и нарушению нормальности.

Сохранилась значимость таких переменных как логарифм цен на вторичном рынке недвижимости, прирост рабочей силы, уровень безработицы, логарифм среднедушевых доходов населения, средняя ставка по ипотеке и количество выданных ипотек.

Значимость потерял логарифм населения как и в коэффициентах модели SEM, так и в предельных эффектах модели SAR.

	sem_fe			sar_fe					
VARIABLES	Main	Spatial	Variance	Main	Spatial	Variance	LR_Direct	LR_Indirect	LR_Total
loglagsecondary	0.608***			0.523***			0.541***	0.272***	0.813***
	(0.0363)			(0.0307)			(0.0306)	(0.0299)	(0.0406)
logpop	0.0670			0.0482			0.0446	0.0232	0.0678
	(0.125)			(0.109)			(0.109)	(0.0570)	(0.166)
wforce_gr	-0.0116**			-0.00445			-0.00419	-0.00209	-0.00628
	(0.00502)			(0.00354)			(0.00349)	(0.00179)	(0.00525)
unempl	0.00675***			0.00891***			0.00920***	0.00465***	0.0139***
	(0.00229)			(0.00208)			(0.00209)	(0.00124)	(0.00323)
logincome_pc	0.184***			0.0684**			0.0704***	0.0352***	0.106***
	(0.0388)			(0.0272)			(0.0270)	(0.0134)	(0.0399)
constructprice	0.000382			0.000296			0.000325	0.000162	0.000486
	(0.000507)			(0.000484)			(0.000486)	(0.000247)	(0.000731)
crime_gr	-9.96e-05			-0.000251			-0.000259	-0.000128	-0.000387
	(0.000424)			(0.000380)			(0.000404)	(0.000205)	(0.000607)
avg_mortgage_rate	-0.0352***			-0.0186***			-0.0191***	-0.00957***	-0.0286***
	(0.00363)			(0.00296)			(0.00280)	(0.00148)	(0.00389)
mquant_gr	-0.000865***			-0.000596***			-0.000604***	-0.000303***	-0.000907***
	(0.000154)			(0.000129)			(0.000130)	(6.84e-05)	(0.000190)
migration_rate	4.21e-05			-4.56e-05			-5.02e-05	-2.57e-05	-7.59e-05
	(0.000128)			(0.000117)			(0.000123)	(6.48e-05)	(0.000188)
lambda		0.321***							

		(0.0493)							
sigma2_e			0.00907***			0.00835***			
			(0.000418)			(0.000382)			
rho					0.356***				
					(0.0314)				
Observations	972	972	972	972	972	972	972	972	972
R-squared	0.846	0.846	0.846	0.837	0.837	0.837	0.837	0.837	0.837
Number of reg_id	81	81	81	81	81	81	81	81	81

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Таблица - “Сравнение моделей SEM и SAR с фиксированными эффектами”

3. Модель SAC (Spatial Autoregressive Combined). Это модель, которая объединяет пространственную авторегрессию и пространственные эффекты в ошибках ($\theta = \tau = \psi = 0$).

$$y_{it} = \rho * W * y_{it} + \beta * X_{it} + a_i + \gamma_t + v_{it}$$

$$v_{it} = \lambda * E * v_{it} + u_{it}$$

Далее мы построили модель SAC с фиксированными эффектами. Интерпретация коэффициентов не будет проводиться, так как знаки остались сопоставлены.

Отметим только, что при оценке предельных эффектов (прямых, косвенных и общих) потеряли значимость такие переменные как логарифм населения, прирост рабочей силы, логарифм среднедушевых доходов.

	sac_fe					
VARIABLES	Main	Spatial	Variance	LR_Direct	LR_Indirect	LR_Total
loglagsecondary	0.440***			0.475***	0.437***	0.912***
	(0.0345)			(0.0327)	(0.0482)	(0.0464)
logpop	0.0521			0.0516	0.0501	0.102
	(0.0943)			(0.0991)	(0.0963)	(0.195)
wforce_gr	-0.00124			-0.00102	-0.000866	-0.00189
	(0.00270)			(0.00278)	(0.00262)	(0.00538)

unempl	0.00902***			0.00972***	0.00901***	0.0187***
	(0.00184)			(0.00193)	(0.00229)	(0.00400)
logincome_pc	0.0340			0.0364*	0.0329*	0.0693*
	(0.0209)			(0.0216)	(0.0192)	(0.0404)
constructprice	0.000197			0.000229	0.000205	0.000434
	(0.000445)			(0.000465)	(0.000435)	(0.000896)
crime_gr	-0.000456			-0.000492	-0.000458	-0.000950
	(0.000336)			(0.000375)	(0.000361)	(0.000730)
avg_mortgage_rate	-0.0126***			-0.0135***	-0.0124***	-0.0259***
	(0.00247)			(0.00241)	(0.00205)	(0.00398)
mquant_gr	-0.000488***			-0.000516***	-0.000474***	-0.000990***
	(0.000109)			(0.000114)	(0.000111)	(0.000211)
migration_rate	-0.000107			-0.000120	-0.000112	-0.000232
	(0.000104)			(0.000115)	(0.000114)	(0.000227)
rho		0.516***				
		(0.0410)				
lambda		-0.352***				
		(0.0724)				
sigma2_e			0.00836***			
			(0.000383)			
Observations	972	972	972	972	972	972
R-squared	0.802	0.802	0.802	0.802	0.802	0.802
Number of reg_id	81	81	81	81	81	81

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Таблица - “Модель SAC (Spatial Autoregressive Combined) с фиксированными эффектами”

4. Модель Дарбина с фиксированными эффектами (Spatial Darbin Model, SDM)

Формально, модель можно записать следующим образом:

$$y_{it} = \rho * W * y_{it} + \beta * X_{it} + D * Z_{it} * \theta + a_i + \gamma_t + u_{it}$$

где:

$D * Z$ - пространственный лаг регрессоров X

Модель Дарбина позволяет использовать разные весовые матрицы для пространственной зависимой переменной (W) и пространственных регрессоров (D) вместе с разными наборами объясняющих переменных $D * (X_{it}) = (Z_{it})$. По умолчанию мы использовали $W=D$ и $X_{it}=Z_{it}$.

С моделями фиксированных эффектов, мы поступили также как и с моделями случайных эффектов.

Сначала построили модель Дарбина с фиксированными эффектами со всеми пространственными лагами независимых переменных X , отобрали значимые пространственные лаги X и затем построили модель Дарбина со значимыми регрессорами.

Значимыми оказались:

loglagsecondary - Средние цены на вторичном рынке жилья, на конец года; рублей за квадратный метр общей площади

unemp - Уровень безработицы, %

crime_gr - Прирост количества зарегистрированных преступлений, %

avg_mortgage_rate - Средняя ставка по ипотечному кредитованию за год, %

mquant_gr - Прирост количества выданных за год ипотечных жилищных кредитов, %

В модели Дарбина сохранились значимыми те же регрессоры, что и в моделях SEM и SAR. Знаки остались сонаправленными. Значимость потеряли WX - avg_mortgage_rate и unemp. Остались значимы пространственные лаги WX loglagsecondary, crime_gr и mquant_gr.

loglagsecondary: рост цен на вторичном рынке недвижимости в период $t-1$ в соседних регионах положительно влияет на рост цен первичной недвижимости в нашем регионе в период t .

crime_gr: рост преступности в соседних регионах приводит к падению цен на первичном рынке недвижимости в нашем регионе.

mquant_gr: чем больше выдают ипотек в соседних регионах, тем меньше цена на первичную недвижимость в нашем регионе. Однако коэффициент перед mquant_gr значим и близок к нулю, то есть эффект незначителен.

sdm_fe							
VARIABLES	Main	Wx	Spatial	Variance	LR_Direct	LR_Indirect	LR_Total
loglagsecondary	0.475***	0.141***			0.491***	0.334***	0.825***
	(0.0328)	(0.0497)			(0.0326)	(0.0462)	(0.0465)
logpop	0.0285				0.0241	0.00880	0.0329
	(0.109)				(0.107)	(0.0348)	(0.141)
wforce_gr	-0.00517				-0.00486	-0.00154	-0.00640
	(0.00364)				(0.00353)	(0.00120)	(0.00468)
unempl	0.0115***	-0.000679			0.0116***	0.00264	0.0142***
	(0.00228)	(0.00400)			(0.00217)	(0.00484)	(0.00501)
logincome_pc	0.0452				0.0459*	0.0145	0.0605
	(0.0283)				(0.0278)	(0.00949)	(0.0369)
constructprice	0.000105				0.000126	4.01e-05	0.000166
	(0.000482)				(0.000478)	(0.000157)	(0.000632)
crime_gr	-2.60e-05	-0.00220***			-0.000158	-0.00285***	-0.00301***
	(0.000408)	(0.000655)			(0.000418)	(0.000791)	(0.000875)
avg_mortgage_rate	-0.0117	-0.0117			-0.0128	-0.0180**	-0.0308***
	(0.00899)	(0.00936)			(0.00808)	(0.00886)	(0.00417)
mquant_gr	-0.000230	-0.000947***			-0.000275	-0.00129***	-0.00157***
	(0.000184)	(0.000256)			(0.000174)	(0.000287)	(0.000256)
migration_rate	-4.55e-05				-4.83e-05	-1.51e-05	-6.33e-05
	(0.000118)				(0.000121)	(3.95e-05)	(0.000160)
rho			0.251***				
			(0.0413)				
sigma2_e				0.00821***			
				(0.000375)			
Observations	972	972	972	972	972	972	972

R-squared	0.813	0.813	0.813	0.813	0.813	0.813	0.813
Number of reg_id	81	81	81	81	81	81	81

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Таблица - “Модель Дарбина с фиксированными эффектами”

Коэффициент пространственной автокорреляции остается значим и положительно влияет на Y , то есть если цены на первичную недвижимость выросли в соседних регионах, то такой же эффект проявится и в нашем регионе.

Также нами был проведен тесты Хаусмана, которые сравнивали одинаковые модели случайных и фиксированных эффектов. В результате тестов фиксированные эффекты оказались более предпочтительными. Вот пример теста Хаусмана для моделей Дарбина с фиксированными и случайными эффектами.

Но: difference in coeffs not systematic $\chi^2(16) = 60.58$ Prob $\geq \chi^2 = 0.0000$. P-value = 0, отвергаем H_0 , fe предпочтительнее re.

Динамические модели панельных данных с пространственной авторегрессией (Spatial Autoregressive Dynamic Panel Models)

Далее мы подумали про влияние временных лагов Y ($primary_{t-1}$) и временных лагов X_{t-1} . Для этого мы построили динамические модели пространственной авторегрессионной зависимости. Динамические модели панельных данных с пространственной авторегрессией (Spatial Autoregressive Dynamic Panel Models) представляют собой расширение стандартных панельных регрессионных моделей с учетом временной динамики и пространственных зависимостей.

Общая форма моделей выглядит следующим образом:

$$y_{it} = \alpha + rho W y_{it} + k y_{it-1} + \beta_1 * X_{it} + \beta_2 * X_{it-1} + \mu_i + v_t + \epsilon_{it}$$

где Wy_t - пространственный лаг зависимой переменной в период t

y_{t-1} - временной лаг объясняющих переменной первого порядка (период $t-1$)

X_{it} - объясняющие переменные в период t

X_{it-1} - объясняющие переменные в период $t-1$

μ_i - фиксированный эффект для наблюдения i

v_t - фиксированный эффект для периода t

Мы построили следующие 4 варианта динамических моделей:

dlag_1 - динамическая модель панельных данных, где присутствует временной лаг зависимой переменной $primary$ (y_{t-1})

$$y_{it} = \alpha + rho W y_{it} + k y_{it-1} + \beta_1 * X_{it} + \mu_i + v_t + \epsilon_{it}$$

y_{t-1} - временной лаг объясняющих переменной первого порядка (период $t-1$)

Wy_t - пространственный лаг зависимой переменной в период t

В модели $dlag_1$ y_{t-1} и Wy_t значимы на 1% уровне значимости и влияют положительно на результат, что логично: чем больше цена на первичном рынке недвижимости в период $t-1$ y_{t-1} , тем больше она будет в t , так как на российском рынке недвижимости наблюдается рост цен.

Чем больше Wy_t цены на первичную недвижимость в соседних регионах, тем больше она в нашем регионе.

dlag_2 - динамическая модель панельных данных, где присутствует пространственный лаг зависимой переменной primary (Wy_t) и временной, пространственный лаг зависимой переменной primary (Wy_{t-1}).

$$y_{it} = \alpha + rho W y_{it} + rho W y_{it-1} + \beta_1 * X_{it} + \mu_i + v_t + \epsilon_{it}$$

Wy_t - пространственный лаг зависимой переменной в период t

Wy_{t-1} - пространственный лаг зависимой переменной в период t-1

Wy_t и Wy_{t-1} значимы на 1% уровне значимости и влияют положительно на результат, что логично: чем больше цена на первичном рынке недвижимости в соседних регионах в период t-1, тем больше она будет в нашем регионе в период t. Также чем больше Wy_t цены на первичную недвижимость в соседних регионах, тем больше цена на первичную недвижимость в нашем регионе.

dlag_3 - динамическая модель панельных данных, где присутствует временной лаг зависимой переменной primary (y_{t-1}), пространственный лаг зависимой переменной primary (Wy_t) и временной пространственный лаг зависимой переменной primary (Wy_{t-1}).

$$y_{it} = \alpha + rho W y_{it} + rho W y_{it-1} + k y_{it-1} + \beta_1 * X_{it} + \mu_i + v_t + \epsilon_{it}$$

Wy_t - пространственный лаг зависимой переменной в период t

Wy_{t-1} - пространственный лаг зависимой переменной в период t-1

y_{t-1} - временной лаг объясняющих переменных первого порядка (период t-1)

В данной модели Wy_{t-1} , y_{t-1} и Wy_t значимы на 1% уровне значимости и положительно влияют на зависимую переменную, что логично.

Wy_t и y_{t-1} уже интерпретировались ранее и логика осталась прежней.

А при росте цен на первичную недвижимость в соседних регионах в период времени t-1 (Wy_{t-1}) в нашем регионе цены на первичную недвижимость будут расти в периоде t, то есть наблюдается временной эффект влияния изменения цен в соседних регионах.

Также нами была построена динамическая модель Дарбина пространственной автокорреляции как dlag2, но с учетом пространственных лагов X. Оценки данной модели можно найти в приложении под названием **dlag_2 + Darbin**.

Форма модели:

$$y_{it} = \alpha + rho W y_{it} + rho W y_{it-1} + \beta_1 * X_{it} + \beta_2 * WX_{it} + \mu_i + v_t + \epsilon_{it}$$

WX_{it} - пространственные лаги переменных X в момент времени t.

Wy_t - пространственный лаг зависимой переменной в период t

Wy_{t-1} - пространственный лаг зависимой переменной в период t-1

Она показала хороший результат, так как оценка временного лага Wy_{t-1} оказалась значима и положительна, как и оценка Wy_t . Пространственные лаги следующих регрессоров оказались значимы: уровень безработицы, логарифм лага цен на вторичную недвижимость, количество выданных ипотек и прирост количества, зарегистрированных преступлений.

Подробнее оценки, краткосрочные и долгосрочные прямые и косвенные эффекты (SR_Direct, SR_Indirect, SR_Total, LR_Direct, LR_Indirect, LR_total) каждой из трех моделей можно найти в приложении.

В таблице ниже представлены информационные критерии Акаике (AIC) и Байеса (BIC) для всех моделей представленных выше. По результатам информационных критериев наилучший результат показали динамические модели пространственной авторегрессии - **DLAG_1** и **DLAG_3**.

dlag_1 - динамическая модель панельных данных, где присутствует временной лаг зависимой переменной primary (y_{t-1}).

$$y_{it} = \alpha + \rho W y_{it} + \rho y_{it-1} + \beta_1 * X_{it} + \mu_i + v_t + \epsilon_{it}$$

dlag_3 - динамическая модель панельных данных, где присутствует временной лаг зависимой переменной primary (y_{t-1}), пространственный лаг зависимой переменной primary (Wy_t) и временной пространственный лаг зависимой переменной primary (Wy_{t-1}).

$$y_{it} = \alpha + \rho W y_{it} + \rho W y_{it-1} + \beta_1 * X_{it} + \mu_i + v_t + \epsilon_{it}$$

	SEM_RE	SAR_R E	SDM_R E	GSPRE	SEM_FE	SAR_FE	SDM_FE	SAC_FE	DLAG_1	DLAG_2	DLAG_3	DLAG_2 + Darbin
AIC	-1409.02	-1414.0	-1462.53	-1413.52	-1702.79	-1728.81	-1754.18	-1742.29	-1965.26	-1597.89	-1964.47	1629.58
BIC	-1340.70	-1345.69	-1360.06	-1340.33	1644.238	-1670.26	-1671.23	-1678.86	-1902.96	-1535.59	-1897.38	-1543.31
Obs.	972	972	972	972	972	972	972	972	891	891	891	891

Таблица - "Информационные критерии Акаике (AIC) и Байеса (BIC) для построенных моделей"

Результаты логичны, так как по тестам Хаусмана мы отдаем предпочтение моделям с фиксированными эффектами, а построенные динамические модели как базируются на fe. А также для борьбы с эндогенностью мы строили динамические модели с временными лагами, что дало положительный результат.

Заключение

В итоге лучшей основываясь на тесте Хаусмана и на информационных критериях Акаике и Байеса стала модель `dlag_3` - динамическая модель панельных данных, которая учитывает влияние прошлых значений переменных, значений зависимой переменной в соседних регионах в нынешнем и прошлом периодах. Удалось выявить на краткосрочном уровне, прямые и косвенные положительные пространственные эффекты для цен на вторичную недвижимость, только прямой эффект прироста рабочей силы и отрицательные эффекты среднедушевых доходов населения, ипотечной ставки и прироста количества взятых ипотек. Эта модель лишь немного уступает динамической модели `dlag_1`, учитывающей только временной лаг зависимой переменной `primary` (Y_{t-1}). Она выявляет такие же краткосрочные эффекты как и прошлая модель кроме оказавшегося незначимым эффект прироста количества взятых ипотек и цен на вторичное жилье, вместо чего был выявлен эффект населения. Что насчет долгосрочных эффектов, значимы только общие эффекты таких переменных, как население, трудоспособное население, доходы, ипотечная ставка. Причем все они в долгосрочном выражении меняют свой знак по отношению к краткосрочному предельному эффекту.

Предельные эффекты в модели в основном заточены на краткосрочный период, который в нашем случае имеет годовые отрезки. В среднем предельный эффект от изменения населения или прироста рабочей силы в данном регионе положительно влияют на цены в наблюдаемом регионе. И этот эффект в среднем выше, чем эффект от изменения этих показателей в соседних регионах, что подтверждается экономической теорией.

Отрицательный предельный эффект подушевых доходов на цены первичной недвижимости и в прямом, и косвенном влиянии может говорить о стратегии застройщиков, в которой даже при уменьшении доходов населения у них есть стимулы повышать цены на первичное жилье. Отрицательные предельные эффекты увеличения ипотечной ставки при прочих равных, включая спрос на недвижимость у физических лиц, и снижении цен на первичную недвижимость, что может говорить о специальном снижении базовой стоимости на первичную недвижимость, чтобы снять с покупателей груз от ипотеки. В то же время, мы наблюдаем отрицательный эффект прироста количества выданных ипотек, что говорит о том, что при увеличении прироста покупок цена на квадратный метр уменьшается, но этот эффект выявляется при учете временных пространственных лагов переменных. Причем везде при изменении переменных в самом регионе значимый эффект больше, чем при изменении соседних.

Заметим, что наши модели не учитывают введение временных субсидий на ипотеки, на субсидии определенным слоям населения, которых может быть больше в регионе независимо от географического положения.

Список литературы

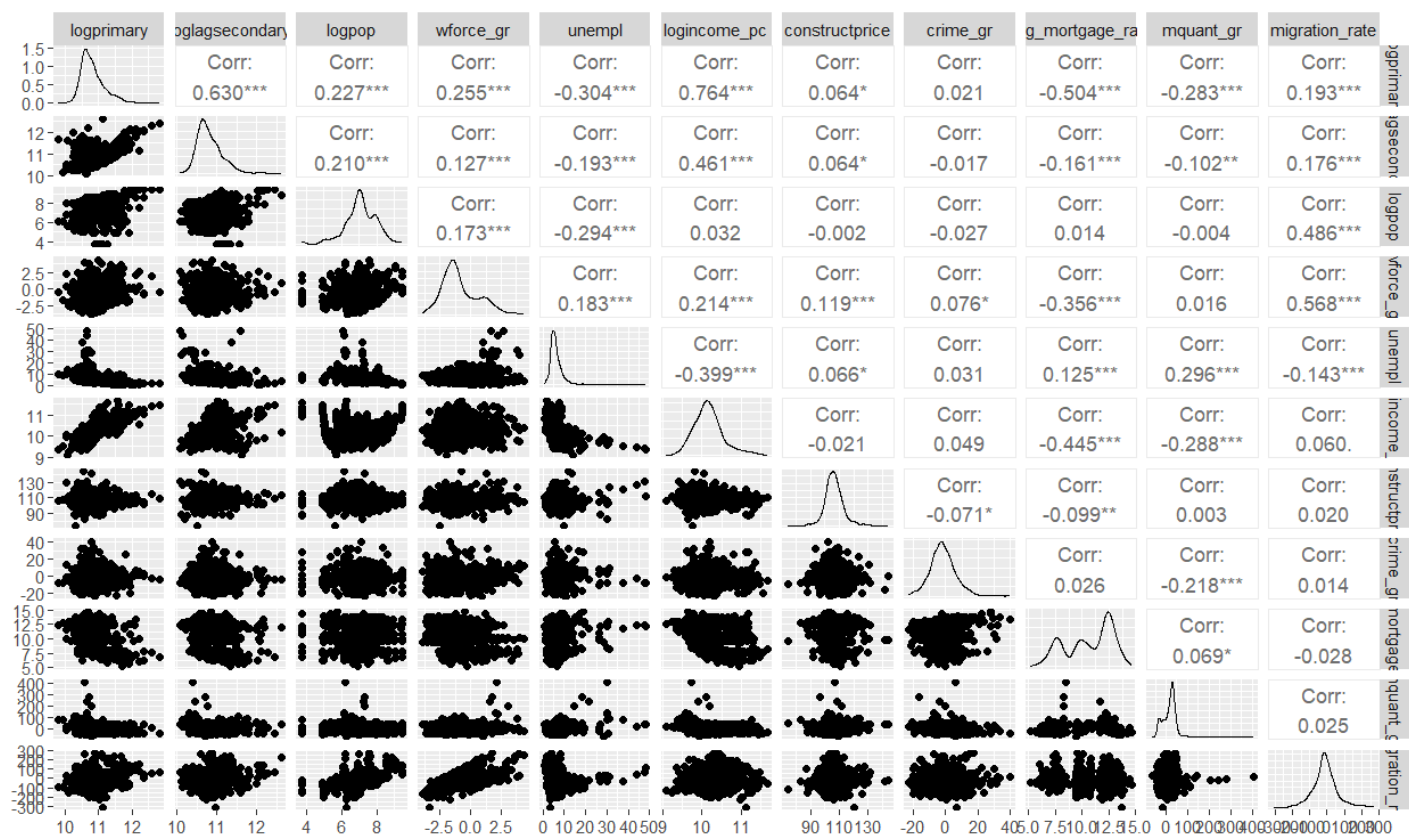
1. Боченина, М. В. Прогнозирование цен на рынке жилья в условиях изменения основной тенденции / М. В. Боченина // и практика общественного развития. – 2023. – Т. 8. – С. 137-142.
2. Демидова, О. А. Модели экономического роста с неоднородными пространственными эффектами (на примере российских регионов) / О. А. Демидова, Д. С. Иванов // Экономический журнал ВШЭ. – 2016. – Т. 20, № 1. – С. 52-75.
3. Луньков, А.Д. Пространственные методы в эконометрических исследованиях на рынке жилья / А.Д. Луньков // Развитие экономики и менеджмента в современном мире . – Воронеж, 2016. – С. 215.
4. Федеральная служба государственной статистики : сайт. – URL: <https://rosstat.gov.ru>
5. Центральный Банк Российской Федерации : сайт. – URL: <https://www.cbr.ru>
6. Donald, J.G. The Dynamics of Metropolitan Housing Prices / J.G. Donald, D. Winkler // Journal of Real Estate Research. – 2020. – V. 23, № 1-2. – pp. 29-46.
7. Li, X Spatial Analysis and Modeling of the Housing Value Changes in the U.S. during the COVID-19 Pandemic / X Li, C Kao // Risk Financial Management. – 2022. – V. 15, № 3. – pp. 139.

Приложения

Приложение 1. Переменные

Название переменной	Показатель
NAME_1	Код региона для карты
region	Регион
year	Год
pop	Численности населения
pop_yearly	Среднегодовая численность населения, тыс. чел.
wforce_gr	Прирост трудоспособного населения, %
workforce	Население в трудоспособном возрасте, тыс. чел. от числ. населения
wforce_perc	Население в трудоспособном возрасте, % от общей числ. населения
unempl	Уровень безработицы, %
income_pc	Среднедушевые денежные доходы населения, руб. в мес
primary	Средние цены на первичном рынке жилья, на конец года; рублей за квадратный метр общей площади
secondary	Средние цены на вторичном рынке жилья, на конец года; рублей за квадратный метр общей площади
constructprice	Индексы цен производителей на строительную продукцию, декабрь к декабрю предыдущего года; %
migration_rate	Коэффициенты миграционного прироста на 10,000 человек населения
crime	Количество зарегистрированных преступлений, шт. на тыс. чел. населения
crime_gr	Прирост количества зарегистрированных преступлений, %
avg_mortgage_rate	Средняя ставка по ипотечному жилищному кредитованию за год, %
total_mortgage_volume	Объем выданных за год ипотечных жилищных кредитов, млн. руб
mquant_gr	Прирост количества выданных за год ипотечных жилищных кредитов, %
avg_mortgage_term	Средний срок ипотечного жилищного кредитования, месяцев

Приложение 2. Матрица парных графиков и корреляций.



Приложения для динамических моделей

	dlag_1								
VARIABLES	Main	Spatial	Variance	SR_Direct	SR_Indirect	SR_Total	LR_Direct	LR_Indirect	LR_Total
L.logprimary	0.983***								
	(0.0290)								
loglagsecondary	-0.0145			-0.0173	-0.0163	-0.0336	-3.501	3.533	0.0317
	(0.0293)			(0.0308)	(0.0276)	(0.0583)	(61.22)	(61.23)	(0.0592)
logpop	0.204**			0.231**	0.201**	0.432**	26.18	-26.63	-0.447**
	(0.102)			(0.105)	(0.0977)	(0.201)	(396.3)	(396.3)	(0.203)
wforce_gr	0.0101***			0.0108***	0.00937***	0.0202***	1.038	-1.059	-0.0210***
	(0.00297)			(0.00309)	(0.00299)	(0.00599)	(16.76)	(16.76)	(0.00600)
unempl	0.00142			0.00155	0.00135	0.00290	0.263	-0.266	-0.00302
	(0.00204)			(0.00214)	(0.00186)	(0.00399)	(3.869)	(3.869)	(0.00416)
logincome_pc	-0.294***			-0.313***	-0.271***	-0.585***	-21.88	22.49	0.611***
	(0.0302)			(0.0317)	(0.0420)	(0.0687)	(311.5)	(311.5)	(0.0710)
constructprice	-0.000499			-0.000533	-0.000464	-0.000997	-0.0265	0.0275	0.00103
	(0.000415)			(0.000460)	(0.000406)	(0.000863)	(0.634)	(0.634)	(0.000896)
crime_gr	-0.000330			-0.000367	-0.000317	-0.000684	-0.0230	0.0237	0.000718
	(0.000321)			(0.000337)	(0.000291)	(0.000626)	(0.386)	(0.386)	(0.000664)
avg_mortgage_rate	-0.00797***			-0.00830***	-0.00706***	-0.0154***	-0.374	0.390	0.0163***
	(0.00274)			(0.00281)	(0.00224)	(0.00498)	(4.742)	(4.742)	(0.00604)
mquant_gr	5.39e-05			6.58e-05	5.98e-05	0.000126	0.00491	-0.00504	-0.000124
	(0.000114)			(0.000120)	(0.000107)	(0.000226)	(0.0984)	(0.0984)	(0.000230)
migration_rate	-0.000167			-0.000180	-0.000157	-0.000337	-0.0180	0.0183	0.000349
	(0.000103)			(0.000111)	(0.000101)	(0.000211)	(0.267)	(0.267)	(0.000213)
rho		0.495***							
		(0.0293)							

sigma2_e			0.00595***						
			(0.000266)						
Observations	891	891	891	891	891	891	891	891	891
R-squared	0.709	0.709	0.709	0.709	0.709	0.709	0.709	0.709	0.709
Number of reg_id	81	81	81	81	81	81	81	81	81

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Таблица - “dlag_1 - динамическая модель панельных данных, где присутствует временной лаг зависимой переменной primary (y_{t-1})”

dlag_2									
VARIABLES	Main	Spatial	Variance	SR_Direct	SR_Indirect	SR_Total	LR_Direct	LR_Indirect	LR_Total
L.Wlogprimary	0.212***								
	(0.0637)								
loglagsecondary	0.480***			0.489***	0.191***	0.680***	0.515***	0.461***	0.975***
	(0.0335)			(0.0324)	(0.0365)	(0.0519)	(0.0335)	(0.0757)	(0.0916)
logpop	0.0457			0.0608	0.0247	0.0855	0.0640	0.0594	0.123
	(0.124)			(0.122)	(0.0495)	(0.170)	(0.128)	(0.118)	(0.246)
wforce_gr	-0.00198			-0.00205	-0.000801	-0.00285	-0.00216	-0.00194	-0.00409
	(0.00364)			(0.00359)	(0.00143)	(0.00499)	(0.00377)	(0.00344)	(0.00718)
unempl	0.00796***			0.00817***	0.00319***	0.0114***	0.00860***	0.00771***	0.0163***
	(0.00247)			(0.00248)	(0.00114)	(0.00349)	(0.00261)	(0.00264)	(0.00509)
logincome_pc	0.00496			0.00645	0.00346	0.00991	0.00689	0.00802	0.0149
	(0.0418)			(0.0416)	(0.0171)	(0.0584)	(0.0438)	(0.0411)	(0.0845)
constructprice	0.000522			0.000537	0.000205	0.000742	0.000564	0.000497	0.00106
	(0.000505)			(0.000537)	(0.000212)	(0.000744)	(0.000565)	(0.000509)	(0.00107)
crime_gr	-0.000145			-0.000163	-6.46e-05	-0.000228	-0.000172	-0.000156	-0.000328
	(0.000392)			(0.000391)	(0.000154)	(0.000542)	(0.000411)	(0.000370)	(0.000778)

avg_mortgage_rate	-0.0188***			-0.0190***	-0.00732***	-0.0263***	-0.0200***	-0.0177***	-0.0377***
	(0.00337)			(0.00326)	(0.00147)	(0.00416)	(0.00337)	(0.00323)	(0.00594)
mquant_gr	-0.000532***			-0.000535***	-0.000207***	-0.000743***	-0.000563***	-0.000502***	-0.00106***
	(0.000138)			(0.000138)	(6.28e-05)	(0.000189)	(0.000144)	(0.000144)	(0.000275)
migration_rate	-8.00e-05			-8.32e-05	-3.38e-05	-0.000117	-8.77e-05	-8.10e-05	-0.000169
	(0.000125)			(0.000128)	(5.31e-05)	(0.000181)	(0.000135)	(0.000127)	(0.000261)
rho		0.292***							
		(0.0420)							
sigma2_e			0.00907***						
			(0.000397)						
Observations	891	891	891	891	891	891	891	891	891
R-squared	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830
Number of reg_id	81	81	81	81	81	81	81	81	81

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Таблица - “dlag_2 - динамическая модель панельных данных, где присутствует пространственный лаг зависимой переменной $primary(Wy_t)$ и временной, пространственный лаг зависимой переменной $primary(Wy_{t-1})$ ”

dlag_3									
VARIABLES	Main	Spatial	Variance	SR_Direct	SR_Indirect	SR_Total	LR_Direct	LR_Indirect	LR_Total
L.logprimary	0.822***								
	(0.0294)								
L.Wlogprimary	-0.0698***								
	(0.0545)								
loglagsecondary	0.169***			0.177***	0.0856***	0.262***	1.781	-3.834	-2.054
	(0.0292)			(0.0288)	(0.0158)	(0.0406)	(11.61)	(11.75)	(1.842)
logpop	0.0279			0.0256	0.0130	0.0386	0.133	-0.400	-0.267

	(0.102)			(0.100)	(0.0491)	(0.149)	(6.064)	(6.450)	(1.778)
wforce_gr	0.00494*			0.00514*	0.00250	0.00764*	0.0711	-0.130	-0.0586
	(0.00298)			(0.00303)	(0.00153)	(0.00452)	(0.666)	(0.672)	(0.0654)
unempl	0.00150			0.00167	0.000829	0.00250	0.0132	-0.0315	-0.0183
	(0.00203)			(0.00206)	(0.00104)	(0.00309)	(0.101)	(0.110)	(0.0399)
logincome_pc	-0.193***			-0.198***	-0.0964***	-0.295***	-2.145	4.412	2.267
	(0.0349)			(0.0346)	(0.0205)	(0.0514)	(14.51)	(14.65)	(1.949)
constructprice	-0.000266			-0.000292	-0.000145	-0.000437	-0.00349	0.00670	0.00320
	(0.000413)			(0.000420)	(0.000212)	(0.000630)	(0.0383)	(0.0395)	(0.00588)
crime_gr	-0.000466			-0.000456	-0.000222	-0.000678	-0.00514	0.0103	0.00519
	(0.000321)			(0.000322)	(0.000165)	(0.000484)	(0.0300)	(0.0309)	(0.00608)
avg_mortgage_rate	-0.0151***			-0.0154***	-0.00747***	-0.0229***	-0.159	0.339	0.179
	(0.00279)			(0.00279)	(0.00142)	(0.00389)	(0.994)	(1.008)	(0.166)
mquant_gr	-0.000238**			-0.000243**	-0.000118**	-0.000361**	-0.00267	0.00543	0.00276
	(0.000113)			(0.000114)	(5.81e-05)	(0.000170)	(0.0179)	(0.0183)	(0.00292)
migration_rate	-0.000116			-0.000114	-5.56e-05	-0.000169	-0.00166	0.00298	0.00132
	(0.000102)			(0.000110)	(5.61e-05)	(0.000165)	(0.0137)	(0.0139)	(0.00151)
rho		0.347***							
		(0.0368)							
sigma2_e			0.00603***						
			(0.000265)						
Observations	891	891	891	891	891	891	891	891	891
R-squared	0.925	0.925	0.925	0.925	0.925	0.925	0.925	0.925	0.925
Number of reg_id	81	81	81	81	81	81	81	81	81

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Таблица - “dlag_3 - динамическая модель панельных данных, где присутствует временной лаг зависимой переменной $primary(y_{t-1})$, пространственный лаг зависимой переменной $primary(Wy_t)$ и временной пространственный лаг зависимой переменной $primary(Wy_{t-1})$ ”

dlag_2 + Darbin										
VARIABLES	Main	Wx	Spatial	Variance	SR_Direct	SR_Indirect	SR_Total	LR_Direct	LR_Indirect	LR_Total
L.Wlogprimary	0.584***									
	(0.0699)									
loglagsecondary	0.449***	0.120**			0.460***	0.280***	0.740***	0.472***	0.400***	0.872***
	(0.0350)	(0.0571)			(0.0330)	(0.0660)	(0.0701)	(0.0331)	(0.0781)	(0.0849)
logpop	0.0337				0.0481	0.0156	0.0637	0.0490	0.0264	0.0754
	(0.124)				(0.120)	(0.0384)	(0.158)	(0.122)	(0.0648)	(0.186)
wforce_gr	-0.00401				-0.00410	-0.00121	-0.00531	-0.00418	-0.00208	-0.00626
	(0.00375)				(0.00367)	(0.00117)	(0.00478)	(0.00373)	(0.00197)	(0.00565)
unempl	0.0111***	-0.00448			0.0111***	-0.00204	0.00901	0.0111***	-0.000504	0.0106
	(0.00268)	(0.00479)			(0.00259)	(0.00609)	(0.00637)	(0.00260)	(0.00700)	(0.00752)
logincome_pc	0.00805				0.00944	0.00312	0.0126	0.00963	0.00525	0.0149
	(0.0443)				(0.0440)	(0.0137)	(0.0573)	(0.0448)	(0.0232)	(0.0676)
constructprice	0.000333				0.000342	9.88e-05	0.000441	0.000348	0.000171	0.000519
	(0.000504)				(0.000534)	(0.000163)	(0.000692)	(0.000543)	(0.000278)	(0.000816)
crime_gr	4.53e-05	-0.00198***			-7.58e-05	-0.00245***	-0.00253***	-0.000140	-0.00284***	-0.00298***
	(0.000425)	(0.000675)			(0.000405)	(0.000828)	(0.000882)	(0.000407)	(0.000952)	(0.00104)
avg_mortgage_rate	-0.0146	-0.00842			-0.0145	-0.0158	-0.0303***	-0.0151*	-0.0206**	-0.0357***
	(0.00951)	(0.00997)			(0.00890)	(0.00983)	(0.00467)	(0.00877)	(0.0101)	(0.00553)
mquant_gr	-0.000224	-0.000844***			-0.000266	-0.00113***	-0.00139***	-0.000298*	-0.00134***	-0.00164***
	(0.000190)	(0.000277)			(0.000181)	(0.000311)	(0.000277)	(0.000179)	(0.000347)	(0.000328)
migration_rate	-5.79e-05				-6.10e-05	-1.85e-05	-7.95e-05	-6.21e-05	-3.17e-05	-9.38e-05
	(0.000126)				(0.000129)	(3.93e-05)	(0.000167)	(0.000131)	(6.69e-05)	(0.000197)
rho			0.233***							

			(0.0466)							
sigma2_e				0.00888***						
				(0.000388)						
Observations	891	891	891	891	891	891	891	891	891	891
R-squared	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811
Number of reg_id	81	81	81	81	81	81	81	81	81	81

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Таблица - “dlag_2 + Darbin - динамическая модель панельных данных, где присутствует пространственный лаг зависимой переменной primary (Wy_t) и временной, пространственный лаг зависимой переменной primary (Wy_{t-1}) с пространственными лагами регрессоров X”