

УДК 519.862.6

UDC 519.862.6

5.2.2. Математические, статистические и инструментальные методы в экономике (физико-математические науки, экономические науки)

5.2.2. Mathematical, statistical and instrumental methods in economics (physical and mathematical sciences, economic sciences)

МОДЕЛИРОВАНИЕ ЦЕНООБРАЗОВАНИЯ НА РЫНКЕ ЖИЛОЙ НЕДВИЖИМОСТИ В РОССИИ

PRICING MODELING IN THE RUSSIAN RESIDENTIAL REAL ESTATE MARKET

Ханенков Павел Вячеславович
Аспирант 1 курс кафедры прикладной математики

Hanenkov Pavel Vyacheslavovich
1st year postgraduate student of the Department of Applied Mathematics

SPIN-код: 4445-9320

RSCI SPIN-code: 4445-9320

phanenkov@mail.ru

phanenkov@mail.ru

Университет «Синергия», Москва, Россия

Synergy University, Moscow, Russia

Рынок жилой недвижимости является одним из основополагающих аспектов экономики страны, поэтому важной задачей является определение динамики и тренда развития рынка в будущем. В работе рассматриваются ключевые факторы ценообразования жилой вторичной недвижимости, а также взяты актуальные данные с две тысячи двадцать первого года по две тысячи двадцать четвёртый год. На основании них в работе строятся регрессионные модели и модель SARIMA. Разработанная модель SARIMA позволяет учитывать сезонность и взаимосвязь предыдущих значений в рамках временного ряда, что делает её более подходящим инструментом для анализа временных рядов, чем регрессионные модели. Работа также включает построение модели регрессии на переменных временного лага для повышения точности прогноза. Проведенная оценка точности построенных моделей на реальных данных демонстрирует, что модель SARIMA имеет приемлемую точность прогнозирования (MAPE составил 2.83%), однако модель регрессии на переменных временного лага показала более высокий показатель MAPE равный 2.67%, что говорит о том, что даже более простые модели способны эффективно решать задачу прогнозирования временных рядов

The residential real estate market is one of the fundamental aspects of the country's economy, so an important task is to determine the dynamics and trend of market development in the future. The paper examines the key factors in pricing secondary residential real estate, and also takes current data from two thousand twenty-one to two thousand twenty-four. Based on them, regression models and the SARIMA model are constructed in the work. The developed SARIMA model allows taking into account seasonality and the relationship of previous values within a time series, which makes it a more suitable tool for analyzing time series than regression models. The work also includes building a regression model on time lag variables to improve forecast accuracy. The assessment of the accuracy of the constructed models on real data demonstrates that the SARIMA model has acceptable forecasting accuracy (MAPE was 2.83%), but the regression model on time lag variables showed a higher MAPE of 2.67%, which indicates that even simpler models are capable of effectively solving the problem of forecasting time series

Ключевые слова: НЕДВИЖИМОСТЬ, ВТОРИЧНЫЙ РЫНОК, ЦЕНА ЖИЛЬЯ, ПАРАМЕТРЫ СТОИМОСТИ, АНАЛИЗ ВРЕМЕННЫХ РЯДОВ, МОДЕЛЬ SARIMA

Keywords: REAL ESTATE, SECONDARY MARKET, HOUSING PRICE, COST PARAMETERS, TIME SERIES ANALYSIS, SARIMA MODEL

<http://dx.doi.org/10.21515/1990-4665-209-043>

Введение

Рынок жилой недвижимости в России является одним из ключевых секторов экономики, оказывающим существенное влияние на социально-экономическое развитие страны. Его динамика определяется комплексом факторов, таких как макроэкономическая ситуация, демографические тренды, государственная политика и инвестиционный климат.

Актуальность исследования обусловлена следующими факторами:

1) Нестабильность рынка. В последние годы рынок жилой недвижимости в России столкнулся с рядом вызовов, включая экономические кризисы, пандемию COVID-19 и геополитическую напряженность. Данные факторы оказывают существенное влияние на динамику цен, объемы строительства и спрос на жилье.

2) Необходимость принятия взвешенных решений. Информация о текущем состоянии и прогнозах развития рынка жилой недвижимости является критически важной для широкого круга участников рынка: частных инвесторов, финансовых организаций и государственных органов.

3) Продвижение новых инструментов анализа. Анализ больших данных и машинное обучение дают новые возможности для исследования рынка жилой недвижимости, что позволяет прогнозировать состояние рынка. Таким образом модель SARIMA поможет более точно определять цену недвижимости, чем это делают регрессионные модели. Преимущество модели SARIMA перед регрессионными моделям заключается в ряде факторов: она учитывает сезонность; моделирует динамику временных рядов, то есть учитывает зависимость текущих значений от предыдущих; строится на стационарных данных; в случае наличия тренда модель будет более эффективной.

Методы и материалы

Целью исследования является построение моделей для определения тренда развития рынка, построение модели по оценке жилой недвижимости и изучение рынка жилой недвижимости.

Исходя из целей исследования в работе поставлены следующие, связанные друг с другом, задачи:

1. изучение работ в области анализа рынка жилой недвижимости;
2. изучение ключевых параметров рыночной стоимости жилой недвижимости;
3. основываясь на ценообразующих факторах, построение пространственной модели для оценки стоимости квартир;
4. проведение анализа временных рядов для построения моделей тренда и прогнозирование будущего состояния рынка с помощью модели SARIMA.

Объектом исследования в работе выбран рынок жилого недвижимого имущества.

Предметом исследования являются тенденции, характеристики и показатели рынка жилой недвижимости.

Для исследования будут использованы методы эмпирического исследования, методы анализа и синтеза материала, инструменты корреляционно-регрессионного анализа и анализа временных рядов.

Научная новизна работы заключается в разработке и апробации модели прогнозирования цен на жилую недвижимость в России с использованием модели SARIMA, которую другие авторы в ходе своих исследований не использовали. Данная модель значима для исследования цен недвижимости, потому что она учитывает сезонность, а регрессионные модели – нет. Предложенная модель позволяет оценить тенденцию развития рынка жилой недвижимости в России за 2021-2024 года и прогнозировать будущее состояние рынка, что является актуальной информацией.

Результаты исследования

Рынку жилой недвижимости уделяется большое внимание, потому что эта отрасль является одной из ключевых для каждой страны.

Поскольку рынок недвижимости является активно развивающимся сегментом российской экономики, то законодательная база, регламентирующая действия на нем, постоянно совершенствуется. Основным нормативным документом является жилой кодекс Российской Федерации от 29.12.2004, согласно которому жилым помещением признается изолированное помещение, которое является недвижимым имуществом и пригодно для постоянного проживания граждан (отвечает установленным санитарным и техническим правилам и нормам, иным требованиям законодательства) [1].

Предварительно изучив работы других учёных, можно понять, от чего стоит отталкиваться, и какая работа уже была проделана.

Так, в своей диссертации «Статистический анализ и прогнозирование развития рынка жилой недвижимости города Самара» Мамаева О. А. выделила ключевые параметры, характеризующие рынок жилой недвижимости: месторасположение, район, количество комнат, этаж, материал стен дома, тип квартиры, общая площадь, жилая площадь, квартал продажи.

В её работе произведено сравнение линейной и степенной регрессии, построенные для квартир первичного рынка [2]. Автор предпочёл линейную регрессию по критерию Вильямса и Клута, основанному на оценке с помощью линейной регрессии углового коэффициента λ линии.

Полученное ею уравнение говорит о том, что стоимость квартир по результатам исследования определяется ключевым образом по таким параметрам, как: месторасположение недвижимости, общая площадь и квартал, в который осуществляется продажа.

В статье «Статистический анализ рынка жилой недвижимости России» Перова А. Н. построила полиномиальный тренд, в котором коэффициент детерминации равен 0,9916 [3]. Продолжая ряд по

подобранному уравнению тренда до конца 2022 года, автор сделал вывод, что рост продолжится. По прогнозу автора цена составит 114, 121, 129 тысяч рублей за 1 кв.м. на 2, 3, 4 квартал 2022 года соответственно, что говорит о непрекращающемся росте стоимости недвижимости.

В статье «Статистический анализ динамики средней цены жилой недвижимости на первичном рынке в Российской Федерации» Трифоновко А. И. была построена матрица парных коэффициентов корреляции для выявления факторов, оказывающие наибольшее влияние на динамику цен на квартиры на первичном рынке России, провела регрессионный анализ, где пришла к выводу, что с ростом уровня инвестиций в РФ на 1% увеличится рост цен на жилье на первичном рынке на 0,78% [4].

На основании проанализированных работ можно выделить, что авторы часто обращаются к регрессионным моделям и трендам при анализе ценообразования жилой недвижимости. Поэтому стоит также построить регрессионную модель и сравнить её с моделью SARIMA.

Итак, ключевым параметром для анализа является цена жилой недвижимости. На основе имеющихся данных была построена модель множественной линейной регрессии, чтобы от её показателей можно было отталкиваться для оценки других моделей.

Исходный набор данных содержит 13 переменных. Для сокращения количества переменных использовался метод обратного исключения, который сократил количество переменных до 10, а затем для оставшихся переменных рассчитывался фактор инфляции дисперсии, который определяет наличие мультиколлинеарности переменных и итерационно были исключены переменные, у которых фактор был более 10, что говорит о наличии мультиколлинеарности.

Набор исходных и конечных переменных представлен на рисунке 1.

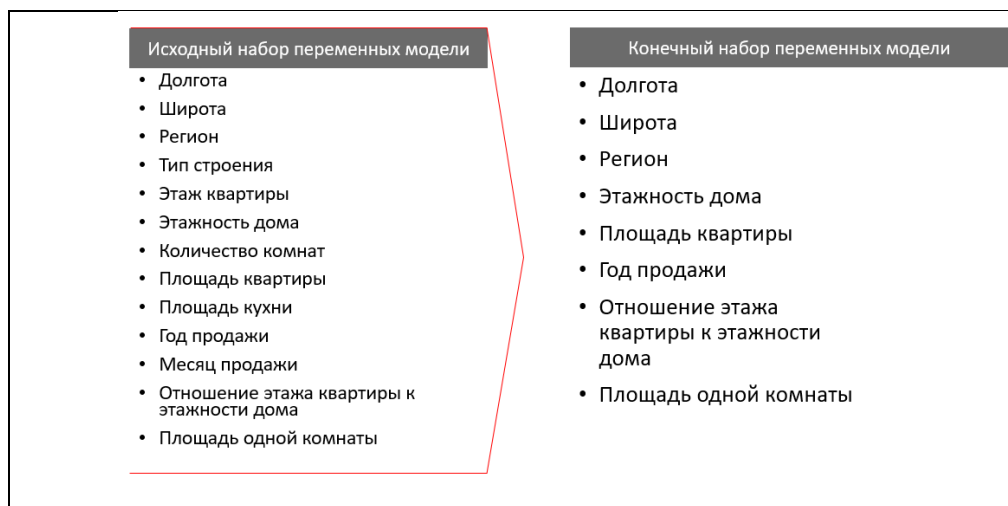


Рисунок 1– Исходный и конечный наборы переменных модели множественной линейной регрессии

Результаты построенной множественной линейной регрессии, которая была разработана для определения стоимости недвижимости, отображены на рисунке 2. Модель была построена с помощью библиотеки statsmodels языка программирования Python.

OLS Regression Results						
=====						
Dep. Variable:	price	R-squared:	0.401			
Model:	OLS	Adj. R-squared:	0.401			
Method:	Least Squares	F-statistic:	1.804e+05			
Date:	Sat, 18 May 2024	Prob (F-statistic):	0.00			
Time:	11:26:12	Log-likelihood:	8.2801e+05			
No. Observations:	2158471	AIC:	-1.656e+06			
Df Residuals:	2158462	BIC:	-1.656e+06			
Df Model:	8					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-0.0872	0.001	-134.152	0.000	-0.088	-0.086
geo_lat	0.3095	0.001	431.012	0.000	0.308	0.311
geo_lon	-0.2057	0.001	-177.844	0.000	-0.208	-0.203
region	-0.7906	0.003	-236.239	0.000	-0.797	-0.784
levels	0.3908	0.001	470.663	0.000	0.389	0.392
area	0.4422	0.001	767.362	0.000	0.441	0.443
year	0.0889	0.000	229.401	0.000	0.088	0.090
level_to_levels	-0.0697	0.005	-13.778	0.000	-0.080	-0.060
area_to_rooms	0.0151	0.001	11.344	0.000	0.013	0.018
=====						
Omnibus:	389336.659	Durbin-Watson:	2.000			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	764934.371			
Skew:	1.102	Prob(JB):	0.00			
Kurtosis:	4.909	Cond. No.	59.7			

Рисунок 2 – Результаты модели множественной линейной регрессии

Коэффициент детерминации у модели составил 0,4, что говорит о том, что модель не полностью описывается набором данных. Тем временем F-статистика указывает на статистическую значимость результата.

Для оценки точности модели были взяты 2 реальных примера, представленных в таблице 1.

Таблица 1 – Оценка точности модели на реальных данных

Тест	Факт	Предсказание	Дельта
Квартира в Москве ближе к МКАД-у	11 000 000	5 350 000	5 650 000
Квартира в Волгограде в центре города	6 500 000	5 000 000	1 500 000

Представленная модель была использована для оценки реальных квартир на рынке, не входивших в исходный набор данных. Как видим, модель оценивает квартиры в Москве с высокой погрешностью, но точнее в городах с более однородной стоимостью в рамках страны.

В ходе анализа временных рядов была построена линия тренда стоимости недвижимости по скользящей средней с помощью библиотеки Pandas языка программирования Python. Из рисунка 3 следует, что каждый январь цены на недвижимость падают. Также наблюдается стабильный небольшой спад в мае и августе-сентябре.

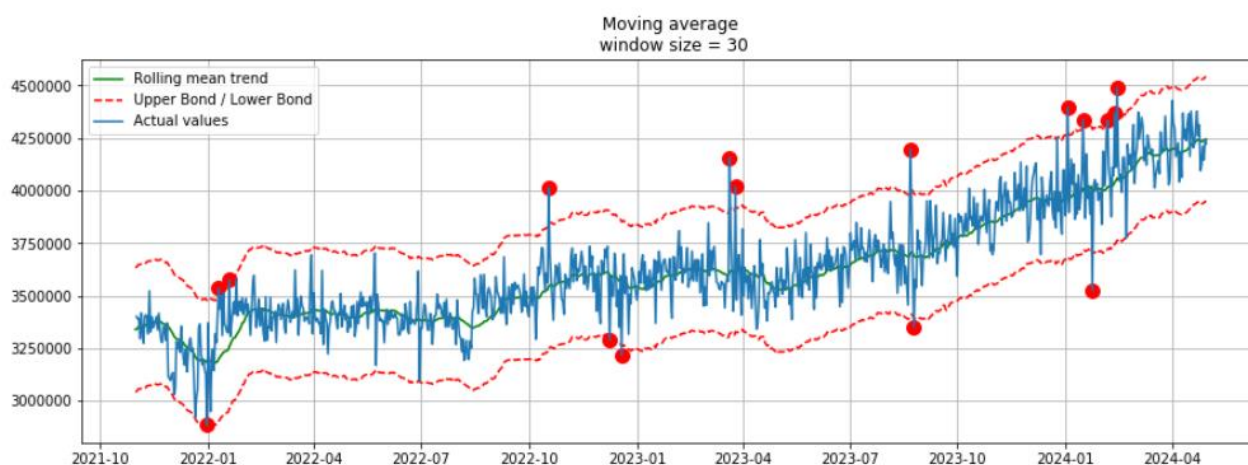


Рисунок 3 – Тренд стоимости недвижимости по скользящей средней (руб)

Также была построена модель Холта-Уинтера с помощью базового функционала языка Python. Судя по рисунку 4, модель корректно обучилась на тренировочных данных, учитывая сезонность и тренд. Резкие скачки модели говорят о том, что модель учитывает изменение структуры ряда, однако общая тенденция совпадает с набором данных. Параметр MAPE модели составляет 2,45%, что говорит о том, что модель качественно обучилась входящим данным и не переобучилась, уловив сезонные тренды.

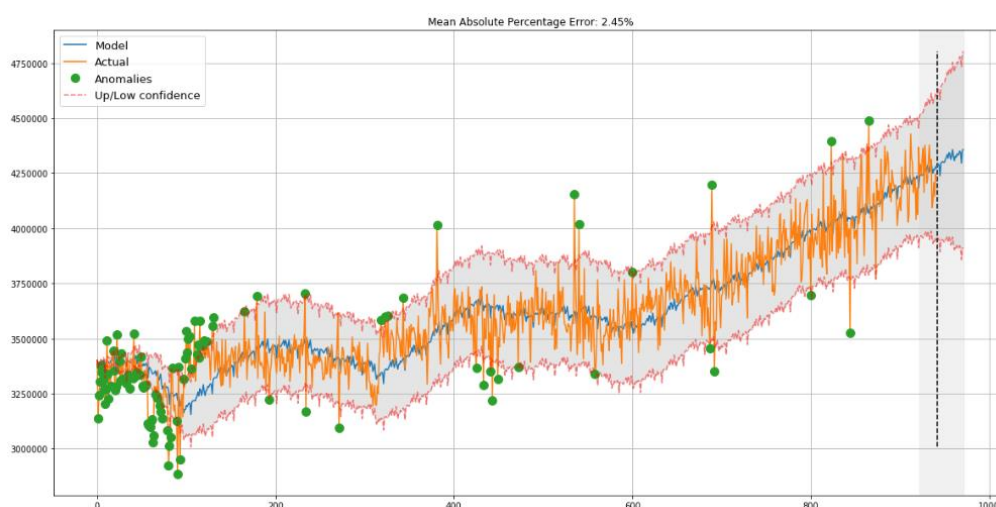


Рисунок 4 – Модель Холта-Уинтера для цен на недвижимость (руб)

Построенная с помощью пакета `statspace.SARIMAX` библиотеки `statsmodels` модель SARIMA выделена красной линией на рисунке 5. MAPE модели составило 2.83%, что говорит об отличном результате и о том, что модель не переобучилась на обучающей выборке. Судя по прогнозу на следующие 50 дней, стоимость недвижимости вырастет на примерно 100000 рублей.

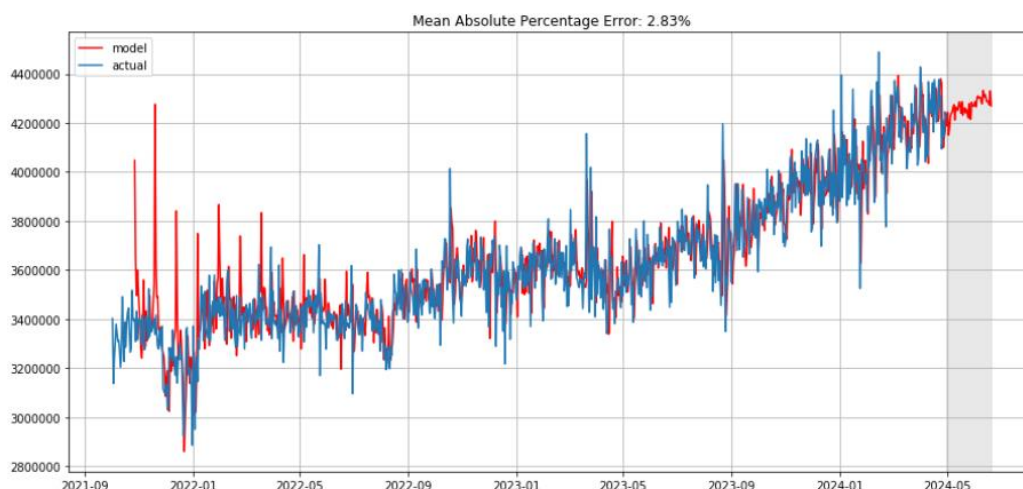


Рисунок 5 – Прогнозирование стоимости недвижимости с помощью модели SARIMA (руб)

Исходный набор данных был расширен с помощью метода временного лага. На полученных данных появилась возможность построить регрессию, изображенную на рисунке 6, с помощью библиотеки sklearn языка Python. Точность модели оказалась выше, чем у SARIMA, поэтому модель можно усовершенствовать, попробовав понизить MAPE.

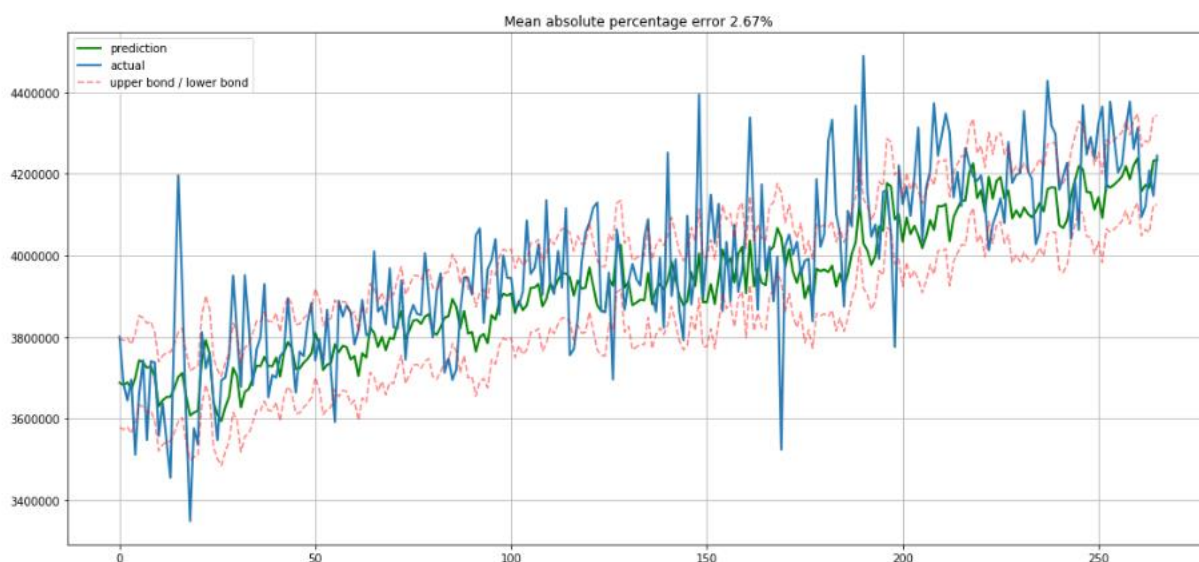


Рисунок 6 – Регрессия стоимости недвижимости на переменных временного лага (руб)

Заключение

Получились точные модели прогнозирования на временных рядах. По прогнозам, цена недвижимости в ближайшие 2 месяца вырастет на 100000 рублей. На это указывает и модель Холта-Уинтера и SARIMA. Было выявлено, что более простые модели вроде регрессий тоже способны успешно справляться с задачами прогнозирования на временных рядах.

Для работы с временными рядами можно использовать и регрессионные модели и модель SARIMA, они равнозначно справляются с задачей прогнозирования цены недвижимости. Ошибка прогнозирования по модели SARIMA составила 2,83%, а регрессионной модели на переменных временного лага 2,67%.

Литература

1. Жилищный кодекс Российской Федерации от 29.12.2004 (ред. от 28.12.2016) (с изм. и доп., вступ. в силу с 01.01.2017) N 188-ФЗ [Электронный ресурс] // Справочно-информационная система «Консультант Плюс». – Режим доступа: <http://www.consultant.ru>
2. Мамаева О. А. Статистический анализ и прогнозирование развития рынка жилой недвижимости города Самара [Электронный ресурс]. – URL: <https://core.ac.uk/download/pdf/197388774.pdf>
3. Перова А. Н. Статистический анализ рынка жилой недвижимости России [Электронный ресурс]. – URL: <https://www.elibrary.ru/item.asp?id=48623256>
4. Трифоненко А. И. Статистический анализ динамики средней цены жилой недвижимости на первичном рынке в Российской федерации [Электронный ресурс]. – URL: <https://www.elibrary.ru/item.asp?id=41804292>

References

1. Zhilishnyj kodeks Rossijskoj Federacii ot 29.12.2004 (red. ot 28.12.2016) (s izm. i dop., vstup. v silu s 01.01.2017) N 188-FZ [Elektronnyj resurs] // Spravochno-informacionnaya sistema «Konsultant Plyus». – Rezhim dostupa: <http://www.consultant.ru>
2. Mamaeva O. A. Statisticheskij analiz i prognozirovanie razvitiya rynka zhiloy nedvizhimosti goroda Samara [Elektronnyj resurs]. – URL: <https://core.ac.uk/download/pdf/197388774.pdf>
3. Perova A. N. Statisticheskij analiz rynka zhiloy nedvizhimosti Rossii [Elektronnyj resurs]. – URL: <https://www.elibrary.ru/item.asp?id=48623256>
4. Trifonenko A. I. Statisticheskij analiz dinamiki srednej ceny zhiloy nedvizhimosti na pervichnom rynke v Rossijskoj federacii [Elektronnyj resurs]. – URL: <https://www.elibrary.ru/item.asp?id=41804292>