

5. Тиндова М. Г., Носов В. В. Алгоритм нечёткого логического вывода для определения цены земельных участков // Никоновские чтения. 2012. № 17. С. 320-322.
6. Савицкая Г. В. Анализ производственно-финансовой деятельности сельскохозяйственных предприятий : учебник - 5 изд. М. : Инфра-М, 2011. 368 с.
7. Крылов С. И. Рейтинговая оценка финансового состояния коммерческой организации экспертно-балльным методом // Финансовый вестник: финансы, налоги, страхование, бухгалтерский учет. 2010. № 6. С. 10-15.

## **МЕТОДЫ СБОРА И АНАЛИЗА ДАННЫХ ПРИ ПРОГНОЗИРОВАНИИ СТОИМОСТИ КВАДРАТНОГО МЕТРА НА РЫНКЕ НЕДВИЖИМОСТИ**

**А. А. Терников**

*Национальный исследовательский университет  
«Высшая школа экономики», Санкт-Петербург, Россия  
E-mail: ternikov.spb@mail.ru*

В настоящей работе представлены методики сбора и обработки открытых данных для описания как ценовых, так и неценовых характеристик квартир на рынке первичной недвижимости. Анализируются факторы, влияющие на стоимость квадратного метра в Санкт-Петербурге и Ленинградской области.

## **METHODS OF DATA COLLECTION AND ANALYSIS IN FORECASTING THE PRICE PER SQUARE METER ON REAL ESTATE MARKET**

**A. A. Ternikov**

This paper presents the methodology of collecting and processing open data for the description of both price and non-price characteristics of the apartments in the primary real estate market. The factors that affect the cost per square meter in St. Petersburg and Leningrad region are also analyzed.

Аналізу російського ринку нерухомості уделено достаточное большое внимание в статьях различных авторов. Например, на рынке московской недвижимости была выявлена значимая связь между вредными выбросами в атмосферу и ценой на квартиру [1]. Значимое влияние на цену квадратного метра на рынке вторичной недвижимости Санкт-Петербурга оказывают плотность населения в микрорайоне постройки, расстояние до метро и класс жилья [2]. Схожие результаты были получены для вторичного рынка московской недвижимости [3]. Однако никто из вышеупомянутых авторов не анализировал влияние различных факторов на цены на рынке первичной недвижимости. Отчасти это связано с закрытым характером получения необходимой информации. С другой стороны это связано с трудностью сбора и анализа данных из-за большого количества пропущенных значений для различного рода ценовых и неценовых факторов. Автором настоящей работы предложено несколько методик

сбора необходимой для анализа информации в сбалансированную панельную структуру данных и проведен соответствующий эконометрический анализ.

Обратимся к сайту <http://spbгuru.ru/> (раздел «новостройки»), который предлагает удобную для обработки информацию о рынке первичного жилья в Санкт-Петербурге единого формата. На начальном этапе сбора данных найдем ссылки на все существующие сейчас жилые комплексы первичного рынка. На начало 2016 года таких комплексов насчитывалось 604. Однако прежде чем приступить к обработке данных с полученных web-страниц соберем информацию о тех же самых объектах строительства во времени. Для этого воспользуемся сайтом [http://web.archive.org/web/\\*/](http://web.archive.org/web/*/), который предлагает для просмотра ряд кэшированных копий необходимых пользователю интернет-страниц. Отобрав случайным образом несколько из полученных ранее ссылок, выяснилось, что практически ни у одного жилого комплекса нет информации на сайте [spbгuru.ru](http://spbгuru.ru) ранее 2013 года. Скорее всего это вызвано характером и сроками строительства жилья. На следующем этапе сбора данных запустим алгоритм, позволяющий найти все копии заданных интернет-страниц и сформировать единую таблицу с данными (характеристики жилого комплекса и информация по существующему в данном комплексе планировкам). Объектом в исследовании является планировка квартиры в жилом комплексе и по окончании работы алгоритма было собрано 50549 наблюдений. Теперь определим временной промежуток для создания сбалансированной панельной структуры данных. Оказывается, что наиболее полным будет отображение данных по годам. Приведем данные к формату панели, убирая идентичные планировки для одних и тех же комплексов и наблюдения с пропущенными значениями стоимости квадратного метра.

Получаем после проведенной обработки 963 наблюдения (сбалансированная панель из 321 уникальной планировки по 82 жилым комплексам). Обрабатываем полученные данные, приведем их к единому формату и создадим бинарные переменные. Теперь перейдем к сбору дополнительной информации для анализа. Дополним пропущенную информацию о расстоянии жилого комплекса до метро и создадим переменную расстояния до центра города (координаты Петропавловского собора). Далее дополним базу данных переменными, отвечающими за инфраструктуру вблизи постройки. Воспользуемся аналитическим инструментом «Google Places API Web Service» из Google Developer Console, выгружая количество объектов инфраструктуры в радиусе 500 метров (согласно примечаниям Google радиус может быть расширен автоматически до 2-х километров). Далее добавляем среднюю цену за квадратный метр на конец года в зависимости от района города с сайта <http://bn.ru>. Используем схожую с вышеупомянутой методику поиска кэшированных копий страниц, производя расчет цены, исходя из доступной величины процентного прироста цен на конец года.

Перейдем к поиску информации о плотности населения. Сложность сбора данных заключается с одной стороны в отсутствии доступа информации для России, а с другой стороны в невозможности сбора данных в автоматическом режиме. Однако вышеупомянутые проблемы удастся избежать при помощи сайта <https://www.freemaptools.com/findpopulation.htm>. Единственной техниче-

ской проблемой представлялся автоматический сбор данных. Однако для этого были прописаны географические координаты полигона заданного радиуса (1.5 км – минимальный в данном случае радиус) с центром в точке с координатами жилого комплекса, в пределах которого высчитывалось количество жителей.

Последней добавленной переменной являются данные о загрязненности атмосферного воздуха в Санкт-Петербурге. Подходящие данные предоставлены только Экологическим порталом Санкт-Петербурга (<http://www.infoeco.ru>). Данные представляют собой ежедневные замеры загрязненности воздуха на каждой из 21 станции автоматического забора воздуха Санкт-Петербурга. Единственным недостатком является то, что данные представлены в виде картинок-графиков формата \*.gif. Однако заметим, что все картинки имеют одинаковый размер 338x273 и пропорциональны по высоте, что позволяет привести их к единому измерению. Воспользуемся пакетом «saTools» для R, и последовательно выгружая картинки, нормируем полученную матрицу пикселей относительно значений по оси ординат, удаляем лишние цвета с картинки и рассчитываем среднемесячное значение загрязнений между максимальным и минимальным значениями, а затем и среднегодовое. В конце концов присваиваем полученные значения загрязненности объектам наблюдения по критерию ближайшей атмосферной станции к жилому комплексу.

После очистки данных от выбросов получим 786 наблюдений по 69 жилым комплексам. Теперь проверим на выбросы зависимую переменную Price (цена квадратного метра). Для того чтобы не потерять панельной привязки данных к объектам, усредним цены для каждой планировки по годам и построим блочные диаграммы средней цены за три года от класса жилья (рис. 1). Наконец удобно представить данные после чистки на карте (рис. 2).

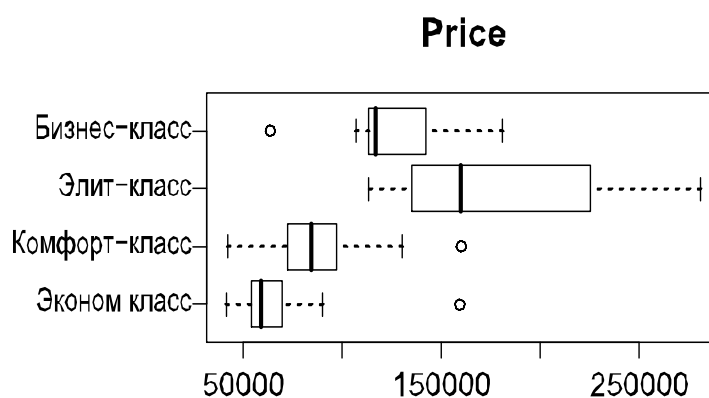


Рис. 1. Блочная диаграмма средней цены за квадратный метр

Воспользуемся в дальнейшем анализе единым набором независимых переменных: S\_live (жилая площадь), Flat (количество комнат), in\_SPB (квартира в городе), Complex (квартира в жилом комплексе), Polyclinic (количество поликлиник неподалеку), School (количество школ неподалеку), People\_density (плотность населения в округе), Ecology\_pollution (уровень загрязненности воздуха), Passed (очередь сдана в срок) и построим три регрессионные модели (1 –

OLS, 2 – Fixed-effects, 3 – Random-effects). Результаты представлены в таблице.



Рис. 2. Жилые комплексы на карте

С содержательной точки зрения на цену квадратного метра значимо влияют жилая площадь квартиры, репутация комплекса с точки зрения транспортной доступности, наличие медицинских и образовательных объектов инфраструктуры, плотность населения и репутация застройщика по сдаче в срок, увеличение комнат в квартире, репутация комплекса по экологичности и другие.

Увеличение комнат в квартире ведет к снижению стоимости квадратного метра – является наиболее понятным содержательно, то есть при большей площади (комнат) происходит относительно небольшое снижение цены за квад-

ратный метр. Однако интересным фактором представляется репутация комплекса по экологичности. Оказывается, что большая экологичность комплекса (не по материалам, а по общей ситуации в районе строительства) ведет к снижению стоимости квадратного метра. Возможным объяснением этому факту может служить нахождение более дорогих квартир в более экологически-неблагоприятных районах. То есть, чем воздух более загрязнен, тем дороже квадратный метр жилья.

### Результаты оценивания регрессионных моделей

Dependent variable:

Price

	(1)	(2)	(3)
S_live	897.977*** (317.131)	886.194*** (317.399)	888.793*** (317.546)
Flat2	-18, 588.040*** (5, 567.454)	-18, 008.260*** (5, 585.997)	-18,137.030*** (5, 590.854)
Flat3	-29, 753.230*** (8, 944.581)	-28, 921.680*** (8, 939.435)	-29,106.820*** (8, 954.793)
Flat	7, 602.229* (4, 469.148)	7, 951.565* (4, 543.170)	7, 872.366* (4, 525.623)
in_SPB	13, 989.000*** (3, 360.143)	13, 508.500*** (3, 318.193)	13, 615.740*** (3, 324.816)
Complex	-3, 916.854 (3, 295.560)	-3, 953.376 (3, 287.035)	-3, 948.848 (3, 292.711)
Polyclinic	2, 515.231*** (818.456)	2, 802.641*** (829.175)	2, 738.300*** (828.554)
School	3, 755.961*** (699.091)	3, 557.535*** (692.668)	3, 601.355*** (700.098)
People_density	0.453*** (0.162)	0.519*** (0.164)	0.505*** (0.162)
Ecology_pollution	10, 209.670*** (3,468.293)	14,163.260*** (3, 543.548)	13, 295.530*** (3, 540.620)
Passed	17, 049.690*** (3, 598.795)	12, 061.510*** (3,487.906)	13,179.910*** (3, 704.758)
Constant	22, 314.810*** (8, 264.053)		17, 825.070** (8, 373.239)
Observations	786	786	786
R <sup>2</sup>	0.495	0.497	0.496
Adjusted R <sup>2</sup>	0.487	0.488	0.489
F Statistic	68.865*** (df = 11; 774)	69.392*** (df = 11; 772)	69.287*** (df = 11; 774)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Таким образом, проанализированы различные ценовые и неценовые факторы, влияющие на цену квадратного метра на рынке первичного строительства в Санкт-Петербурге и Ленинградской области. В дальнейшем при расширении рассмотренной тематики предполагается дополнить базу данных новыми переменными и использовать приведенные в настоящей работе подходы для анализа рынка московской недвижимости с последующим сравнением результатов для двух столиц.

### СПИСОК ЛИТЕРАТУРЫ

1. Катыйшев П. К., Хакимова Ю. А. Экологические факторы и ценообразование на рынке недвижимости (на примере г. Москвы) // Прикладная эконометрика. 2012. № 4 (28). С. 113–123.
2. Красильников А. А., Щербакова А. А. Детерминанты цены на вторичном рынке недвижимости Санкт-Петербурга // Экономические науки. 2011. № 11. С. 93–99.
3. Магнус Я. Р. Цены квартир в Москве // Прикладная эконометрика. 2010. № 1 (17). С. 89–105.