

DOI: 10.17323/2587-814X.2023.4.7.24

# Прогнозирование распределений удельных цен на объекты недвижимого имущества на основе свойств ПСИ-процессов\*

**М.Б. Ласкин**<sup>a</sup> 

E-mail: laskinmb@yahoo.com

**О.В. Русаков**<sup>b</sup> 

E-mail: ovirusakov@yahoo.co.uk

<sup>a</sup> Санкт-Петербургский Федеральный исследовательский центр РАН  
Адрес: Россия, 199178, г. Санкт-Петербург, Васильевский остров, 14 линия, д. 39

<sup>b</sup> Санкт-Петербургский государственный университет  
Адрес: Россия, 199034, г. Санкт-Петербург, Университетская наб., д. 7/9

## Аннотация

Прогнозирование цен на рынке недвижимого имущества всегда является предметом интереса ученых-экономистов, аналитиков рынка, участников рынка (продавцов и покупателей), маркетинговых служб предприятий строительного комплекса, аналитиков банков и страховых компаний, инвесторов. В современных условиях поведение цен на объекты недвижимости приобретает особую важность под влиянием таких факторов как изменения в структуре доходов населения, изменение ставок ипотечных кредитов и их доступность, динамичные изменения макроэкономического характера и других внешних социально-экономических, политических факторов. Однако, в отличие от финансовых рынков, рынков ценных бумаг, рынок недвижимости всегда отличается замедленной реакцией на внешние возмущения, часто до полугода, что позволяет надеяться на построение адекватных прогнозов, хотя бы на время замедленной реакции. Традиционные авторегрессионные методы прогнозирования отличаются быстро нарастающей дисперсией прогноза, так как предполагают фактор стохастической волатильности. В статье предложена модель и метод построения прогноза, основанная на стохастических процессах «пуассоновского случайного индекса» с быстро выходящей на стационарный режим и стабилизирующейся дисперсией. В основе модели лежит «принцип замещения» старых цен

\* Статья опубликована при поддержке Программы НИУ ВШЭ «Университетское партнерство»

новыми. Детально рассмотрен пример применения «принципа замещения» для построения прогноза цен на вторичную жилую недвижимость в г. Санкт-Петербурге по результатам четырехлетних наблюдений за ценами предложений.

**Ключевые слова:** прогноз цен на недвижимость, логарифмически нормальное распределение цен, псевдопуассоновский процесс, процесс пуассоновского случайного индекса, процесс Орнштейна–Уленбека

**Цитирование:** Ласкин М.Б., Русаков О.В. Прогнозирование распределений удельных цен на объекты недвижимого имущества на основе свойств ПСИ-процессов // Бизнес-информатика. 2023. Т. 17. № 4. С. 7–24. DOI: 10.17323/2587-814X.2023.4.7.24

### Введение

Рынок недвижимого имущества является важным сектором экономики любой страны. Объекты капитального строительства, поступающие в оборот товарного рынка недвижимого имущества, создают цепочки хозяйственных связей в строительной отрасли, в отраслях производства строительных материалов, добывающей промышленности. Сфера оборота первичных и вторичных объектов недвижимого имущества в значительной степени является областью бизнес-интересов банковского и страхового секторов экономики. Совокупность объектов недвижимого имущества, находящихся в собственности и в коммерческом обороте является налогооблагаемой базой по налогу на имущество и, следовательно, основой для пополнения бюджета страны. По этим причинам прогнозирование цен на рынке недвижимого имущества всегда было и остается актуальной задачей для всех участников рынка. Материалы, посвященные данной тематике, регулярно появляются в периодических изданиях. Задачи, связанные с построением прогнозов, постоянно являются предметом исследований ученых и исследователей. Среди работ общего методологического характера следует выделить известную переведенную книгу [1], серии работ авторов публикаций [2–5]. В качестве примеров относительно недавних публикаций отечественных исследователей можно назвать работы [6–16], зарубежных – [17–22]. Рынки недвижимости, как справедливо отмечалось в [5], имеют характерные региональные особенности. Прогнозированию цен на региональных отечественных рынках посвящены работы отечественных авторов [3, 6, 10, 11, 13–16, 23, 24], зарубежных – [25–28]. В большинстве случаев рассматриваются традицион-

ные методы прогнозирования: фундаментального и технического анализа [11], факторные модели [14], регрессионные модели, модели авторегрессии и скользящих средних [2–5, 15, 16, 18, 27]. В тоже время, «стандартные эконометрические методы оказываются неприспособленными к прогнозированию тенденций рынка недвижимости в современных условиях», «методы, разработанные в странах с развитой рыночной экономикой непригодны к прогнозированию в странах с переходной экономикой» [5]. В последнее время появляются работы, в которых для целей прогнозирования цен на недвижимость применяются методы машинного обучения [24], в том числе методы нейросетевого моделирования [8, 29]. Основным недостатком традиционных методов, основанных на авторегрессии, является быстрое нарастание дисперсии прогноза, что делает результат прогноза малоинформативным уже через 2–3 шага. Это обусловлено введением случайного масштаба волатильности, что выражается в том, что возникает смесь распределений, а дисперсия смеси всегда больше или равна смеси дисперсий. В то же время многочисленные эмпирические наблюдения говорят о том, что разброс удельных цен на относительно сходные объекты недвижимого имущества мало изменяется во времени даже при наличии сильных восходящих или нисходящих трендов. Исключение могут составлять относительно короткие периоды воздействия сильных внешних возмущающих факторов, приводящих к заметной смене тренда. В этой связи предпочтительным выглядел бы прогноз с относительно стабильной дисперсией.

В настоящей статье мы предлагаем математическую модель процесса изменения цен на рынке недвижимого имущества при простых и, как пред-

ставляется, естественных допущениях. Такими допущениями являются следующие положения:

- 1) объявленная цена продажи на некоторый объект сохраняется неизменной в течение некоторого (случайного) промежутка времени;
- 2) в любой момент времени объект может быть снят с продажи или продан;
- 3) в любой момент времени в листинге продающихся объектов может появиться новый объект и заместить снятый с продажи, или проданный;
- 4) цены на объекты в листинге в каждый фиксированный момент времени независимы, или, по меньшей мере, условно независимы при условии некоторого внешнего фактора;
- 5) общий набор выставленных на продажу объектов в фиксированный момент времени образует наблюдаемую выборку удельных цен, которая подлежит изучению и статистической обработке.

Теоретической основой модели, предложенной ниже, являются три положения:

- ◆ принцип логарифмически нормального распределения удельных цен на сравнительно однородные объекты недвижимости;
- ◆ свойства распределения и ковариационной функции процесса пуассоновского случайного индекса (далее ПСИ-процесса);
- ◆ центральная предельная теорема для ПСИ-процессов – сходимость их нормированных сумм к стохастическому процессу Орнштейна–Уленбека (стационарному, гауссовскому, марковскому процессу).

В работах [30, 31] приводится обоснование сходимости удельных цен, образованных последовательными сравнениями, к логарифмически нормальному распределению. По-видимому, первой работой, в которой было отмечено следование удельных ставок аренды логарифмически нормальному распределению, является работа британских статистиков [32], это следование отмечено и в работе японских ученых [26]. В работе [33] исследованы свойства распределений и корреляционных функций ПСИ-процессов, там же доказывалась сходимость нормированных сумм независимых ПСИ-процессов к процессам типа Орнштейна–Уленбека. Теоретические основы, предложенной в настоящей статье модели, изложены в работах [30, 31, 33, 34].

## 1. Определение и основные свойства ПСИ-процесса

Пусть  $\xi_0, \xi_1, \xi_2, \dots$  – случайная последовательность, которую мы назовем формирующей, или ведомой;  $\Pi(t) = \Pi_\lambda(t)$ ,  $t \geq 0$  – независимый от нее пуассоновский процесс с интенсивностью  $\lambda > 0$ , который мы назовем ведущим. Зададим пуассоновский субординатор<sup>1</sup> для последовательности  $\xi_0, \xi_1, \xi_2, \dots$  следующим образом:  $\psi(t) = \psi_\lambda(t) := \xi_{\Pi(t)}$ . Полученный процесс  $\psi(t)$  с непрерывным временем  $t \geq 0$  – будем называть процессом пуассоновского случайного индекса (Poisson Stochastic Index process) или ПСИ-процессом. Заметим, что ПСИ-процессы являются естественным обобщением псевдо-пуассоновских процессов, введенных и подробно рассмотренных в главе X второго тома классического труда У. Феллера [35].

ПСИ-процесс представляет собой последовательные замещения членов формирующей последовательности, происходящие в моменты скачков пуассоновского процесса. Промежутки времени  $\{\tau_{j+1}\}$ ,  $j = 0, 1, 2, \dots$  между последовательными скачками ведущего пуассоновского процесса называются спейсингами. Известно, что спейсинги являются независимыми одинаково распределенными случайными величинами с общим показательным (или, что то же самое, экспоненциальным) распределением, имеющим интенсивность  $\lambda > 0$ . В нулевой момент времени разыгрывается  $\xi_0$ , которая «держит» свое значение в течение первого спейсинга; в момент первого скачка пуассоновского процесса она замещается на  $\xi_1$ , и так далее. На протяжении  $j$ -го спейсинга разыгранная случайная величина  $\{\xi_{j-1}\}$  не меняет своего значения до момента (включительно)  $j$ -го скачка пуассоновского процесса  $\theta_j$ . В момент смены спейсинга  $\tau_{j+1}$  на  $\tau_{j+2}$ , значение случайной величины  $\xi_j$  замещается на  $\xi_{j+1}$  (рис. 1).

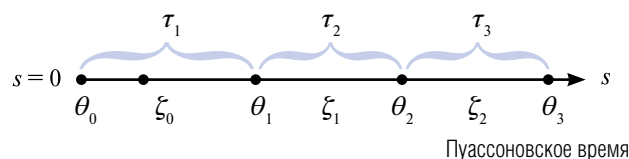


Рис. 1. Схематическое изображение случайных длин спейсингов  $\tau_0, \tau_1, \tau_2, \dots$ , моментов скачков пуассоновского процесса  $\theta_0, \theta_1, \theta_2, \dots$  и замещающихся случайных величин  $\xi_0, \xi_1, \xi_2, \dots$ .

<sup>1</sup> Субординацией в теории стохастических процессов принято называть случайную замену времени.

Заметим, что ПСИ-процесс имеет следующее представление в виде взвешенной суммы элементов случайной последовательности  $\xi_0, \xi_1, \xi_2, \dots$ , где веса — суть пуассоновские индикаторы

$$\psi_\lambda(t) = \sum_{j=0}^{\infty} \xi_j \cdot \mathbf{1}\{\Pi(t) = j\},$$

здесь и далее  $\mathbf{1}$  — индикаторная функция.

Если  $\xi_0, \xi_1, \xi_2, \dots$  представляют собой стационарную последовательность, то ПСИ-процесс стационарен. В частности, это будет выполнено, когда  $\xi_0, \xi_1, \xi_2, \dots$  — независимые одинаково распределенные величины. В последнем случае, если  $E\xi_0 = a$ , то условное математическое ожидание ПСИ-процесса

$$E(\psi_\lambda(t+s) | \psi_\lambda(s) = z) = ze^{-\lambda s} + a(1 - e^{-\lambda s})$$

для всякого произвольного, но фиксированного  $z$ , при любых неотрицательных  $t, s$ . И, в частности, при  $a = 0$

$$E(\psi_\lambda(t+s) | \psi_\lambda(s) = z) = ze^{-\lambda s}.$$

Важно отметить, что если последовательность  $\xi_0, \xi_1, \xi_2, \dots$  марковская, то соответствующий ПСИ-процесс тоже обладает марковским свойством<sup>2</sup>.

## 2. Ковариационная функция процесса $\psi(t)$

В работе [33] получен следующий результат. Пусть  $\xi_0, \xi_1, \xi_2, \dots$  — независимые одинаково распределенные случайные величины,  $E(\xi_0) = 0$ ,  $D(\xi_0) = 1$ , тогда ковариационная функция процесса  $\psi(t)$ ,  $t \geq 0$ , убывает экспоненциально и имеет вид

$$\text{cov}(\psi(t), \psi(0)) = \exp(-\lambda t).$$

Рассмотрим независимые копии одного ПСИ-процесса  $\psi_\lambda(t)$ :  $\psi_1(t), \psi_2(t), \dots$ . Из экспоненциального вида ковариационной функции следует, что в силу центральной предельной теоремы для векторов, случайный процесс, составленный из нормированных сумм ПСИ-процессов вида

$$Z_N(t) = \frac{1}{\sqrt{N}} \sum_{j=0}^N \psi_{(j)}(t),$$

сходится в смысле слабой сходимости конечномерных распределений<sup>3</sup>, при  $N \rightarrow \infty$ , к процессу Орнштейна–Уленбека — стационарному, гауссовскому, марковскому процессу, причем заданному

в «стандартизованном» виде. Последнее означает, что в произвольные моменты времени  $(t_1, \dots, t_d)$  вектор  $Z(t_1), \dots, Z(t_d)$  имеет совместное нормальное распределение с нулевым средним и с ковариациями  $\exp(-\lambda|t_i - t_j|)$ , где  $t_i, t_j$  пробегают все элементы набора  $(t_1, \dots, t_d)$ . Таким образом, распределение «стандартизованного» процесса Орнштейна–Уленбека характеризуется гауссовостью (совместной нормальностью конечномерных распределений), нулевым средним (нулевым «теоретическим» трендом) и ковариацией вида  $\exp(-\lambda t)$ . Коэффициент  $\lambda > 0$  называется «скоростью» процесса Орнштейна–Уленбека, а  $1/\lambda$  — «вязкостью». Здесь мы видим, что скорость предельного процесса Орнштейна–Уленбека в точности есть интенсивность «ведущего» пуассоновского процесса. Условное математическое ожидание процесса Орнштейна–Уленбека совпадает с условным математическим ожиданием ПСИ-процесса

$$E(U(t+s) | U(t) = z) = ze^{-\lambda s},$$

а условная дисперсия не зависит от условия  $z$  и равна

$$D(U(t+s) | U(t) = z) = 1 - e^{-2\lambda s}.$$

Более того, условное распределение процесса  $U$  при условии  $z$  нормально при любом неотрицательном  $s$ . Из стандартного процесса Орнштейна–Уленбека путем сдвига  $a$  и масштаба  $b > 0$  получаем процесс Орнштейна–Уленбека  $OU$ , у которого стационарное распределение нормально с параметрами  $a$  и  $b^2$ , ковариацией  $b^2 \exp(-\lambda t)$ . Для такого процесса Орнштейна–Уленбека, естественно, стационарность и марковское свойство сохраняются, а условное математическое ожидание и условная дисперсия соответственно равны

$$E(OU(t+s) | OU(s) = y) = a + (y - a)e^{-\lambda s},$$

$$D(OU(t+s) | OU(t) = y) = b^2(1 - e^{-2\lambda s}). \quad (1)$$

Откуда, в частности, видно, что условная дисперсия  $OU$  от условия не зависит.

Применительно к ценам на недвижимость интерпретируем рассматриваемую модель следующим образом. Листинг цен представляем как таблицу, у которой каждая строка — это временной срез теку-

<sup>2</sup> Марковское свойство — это когда будущее при фиксированном прошлом и настоящем не зависит от прошлого (то есть от прошлого зависит только через настоящее).

<sup>3</sup> Более того, имеет место сходимость  $Z_N(t)$  в функциональном пространстве Скорохода, см. [34].

щих цен предложений ( $N$  – объем среза, меняется от среза к срезу), а каждый столбец – это цена объекта в динамике, возможно, с поправкой на тренд ( $n$  – количество срезов). Мы наблюдаем срезы цен с некоторой периодичностью, определяемой выходом очередного выпуска журнала цен (обычно раз в неделю). К следующему выпуску журнала каждый объект может «уйти» (или кардинально «уйти» его цена), либо остаться с прежней ценой. Объект, или его цена, могут замещаться вновь пришедшим объектом (ценой). Причем, остающихся цен из предыдущего выпуска журнала существенно больше, чем вновь пришедших в последующем выпуске. Каждый ПСИ-процесс представляет собой столбец в таблице. Значения наблюдаются по срезам во времени, в соответствии с датами выхода очередного номера журнала. Все ПСИ-процессы предполагаются независимыми и одинаково распределенными (то есть представляют собой независимые копии одного ПСИ-процесса – в нашем контексте цены, или же логарифма цены). Распределение каждого ПСИ-процесса понимается, как распределение кусочно-постоянной функции с непрерывным временем, непрерывной справа, имеющей конечные пределы слева (такие функции называют Right Continuous Left Limits, RCLL). Пуассоновский процесс выступает в роли точечного процесса, определяющего моменты замещений, в силу его свойства «отсутствия последствий»: сколько бы времени не прошло с момента предыдущего скачка, – следующий скачок наступит через экспоненциальное время.

В нашем подходе мы производим следующую **аппроксимацию логарифмов цен** в «таблице». Пусть  $V(t)$  – цена 1 кв. м. выбранного типа недвижимости в момент времени  $t$ . В каждом срезе по времени  $t$  производим логарифмирование и наблюдаем независимые реализации процесса  $\ln(V(t))$  с добавленным трендом и коэффициентом масштаба  $\sigma > 0$ . Таким образом, в каждом временном срезе  $t$  мы имеем (свою) простую выборку размера, равного количеству цен в данном временном срезе.

Метод построения прогноза распределений цен, их числовых характеристик, предложенный в настоящей статье, основан на специфических свойствах ПСИ-процессов и пределов их нормированных сумм: процессов Орнштейна–Уленбека. Основные общие свойства здесь (как для ПСИ-процесса, так и для процесса Орнштейна–Уленбека) – марковость, вид условных математических ожиданий. Также мы используем вид условной

дисперсии и гауссовское свойство условного распределения процесса Орнштейна–Уленбека.

Постоянный мониторинг цен на рынке недвижимости, с заранее заданной периодичностью, позволяет использовать свойства процессов Орнштейна–Уленбека для прогнозирования будущих распределений цен и их числовых характеристик, таких как средние, медианные, модальные (рыночная стоимость) значения, среднеквадратичное отклонение, коридоры допустимых значений и так далее. Особенностью рынка недвижимости является то, что изменения цен на рынке могут наблюдаться, как правило, не чаще 1 раза в неделю, т.к. большинство печатных рекламных изданий и интернет ресурсы обновляются с такой же периодичностью. Вместе с тем рынок недвижимости имеет замедленную реакцию на внешние источники возмущения макроэкономического характера, так как поиск объекта, достижение договоренностей, оформление сделки, вступление в права собственности требуют значительного времени. В этой связи достаточно разумным представляется мониторинг цен с периодичностью 1 раз в месяц, т.к. именно в течение такого срока изменение цен становится заметным.

### 3. Модель

Рассмотрим случайный процесс  $V(t)$  – динамика цены 1 кв. м. недвижимости во времени и связанный с ним процесс логарифма цены  $Y(t) = \ln(V(t))$ . В качестве основной оценки  $Y(t)$  мы берем среднее по каждому срезу и обозначаем той же буквой (это касается также и  $\hat{Y}(t)$ , введенного ниже). Предположим, у процесса  $Y(t)$  наличие линейного тренда  $E(Y(t)) = \alpha t + \beta$  и постоянного во времени стандартного отклонения  $\sigma(t) = \sigma > 0$ . Центрированный и нормированный процесс

$$\hat{Y}(t) = \frac{Y(t) - \alpha \cdot t - \beta}{\sigma}$$

будем рассматривать как ПСИ-процесс. Отметим, что в сформулированных предположениях у  $\hat{Y}(t)$  математическое ожидание тождественно равно нулю, а дисперсия – единице. При выполнении условия, что распределение (как случайного процесса)  $\hat{Y}(t)$  совпадает с распределением ПСИ-процесса, или процесса Орнштейна–Уленбека, ковариационная функция имеет вид  $cov(\hat{Y}(t)(s+t), \hat{Y}(t)) = \exp(-\lambda s)$  и не зависит от  $t$ . Поскольку наблюдения цен на рынке недвижимости

доступны только в дискретном режиме, то мы говорим о значениях и наблюдениях процесса  $V(t)$  в дискретные моменты времени  $V(j)$ ,  $j = \overline{0, n-1}$ . В работах [30, 31] доказана слабая сходимость по распределению цен, образованных последовательными сравнениями, к логарифмически нормальному распределению. Таким образом, у нас имеются основания считать последовательность логарифмов цен  $Y(j) = \ln(V(j))$ ,  $j = \overline{0, n-1}$  последовательностью нормально распределенных случайных величин, откуда следует, что все члены соответствующей последовательности  $\hat{Y}(j)$  стандартно нормальны. В этом контексте физическим смыслом членов последовательности  $(\xi_j)$ ,  $j = \overline{0, \infty}$  (из определения ПСИ-процесса) могут служить центрированные, нормированные на стандартное отклонение логарифмы цены 1 кв. м. некоторого объекта, сохраняющего свое значение в течение промежутка времени  $(\tau_i)$ ,  $i = \overline{1, \infty}$ ,  $(i = j + 1)$ , т.е. последовательность вида  $\hat{Y}(j)$ ,  $j = \overline{0, n-1}$ . В предположении, что средние значения логарифмов цены следует процессу Орнштейна–Уленбека, для любой пары средних значений логарифмов случайных величин  $V(t)$ ,  $V(t + s)$  может быть рассмотрена гипотеза о многомерном совместном нормальном распределении для  $\ln(V(t))$ . Заметим, что в качестве средних значений здесь можно рассматривать квантили (в частности, медиану) и моду. При подтверждении данной гипотезы для любого, наперед заданного  $V(0) = v(0)$  по формулам условной моды, медианы и/или условного математического ожидания (см., например, [36, 37]) могут быть получены прогнозные оценки для модального, медианного и среднего значения случайной величины  $V(s)$ . Здесь момент времени 0 соответствует моменту времени последнего по времени наблюденного распределения цен,  $s$  – время, на которое дается прогноз, его отсчет начинается от момента времени последнего наблюденного распределения.

Для верификации предлагаемой интерпретации следует статистически доказать, подтвердить следующие положения.

- 1) Изучить распределение спейсингов  $(\tau_i)$ , где  $i = \overline{1, n}$  – количество периодов непрерывного нахождения цены в потоке (цель – получить статистическое подтверждение гипотезы об их экспоненциальном распределении, оценить параметр экспоненциального распределения).
- 2) Изучить распределения цен  $V(j)$ ,  $j = \overline{0, n-1}$  в каждом срезе (цель – убедиться в логарифмически нормальном виде распределения цен).
- 3) Получить подтверждение независимости копий случайной последовательности  $\hat{Y}(j)$ ,  $j = \overline{0, n-1}$ , для чего изучить поведение накопленных дисперсий в каждом срезе. Мы будем проверять некоррелированность, которая вместе с гауссовостью даст независимость.
- 4) Проверить свойство совместной нормальности (гауссовости) средних (по каждому срезу) значений  $\hat{Y}(t)$ , что вместе с условием об экспоненциальном виде ковариационной функции подтверждает марковское свойство случайной последовательности, составленной из средних (по каждому срезу) от  $\hat{Y}(j)$ ,  $j = \overline{0, n-1}$ . Это следует из того, что стационарный гауссовский случайный процесс с корреляцией, убывающей по экспоненте, есть процесс Орнштейна–Уленбека, а значит он обладает марковским свойством. Более того, гауссовость средних подтвердит марковское свойство медианы и моды в зависимости от  $j = \overline{0, n-1}$ .

Заметим, что гауссовость средних по каждому срезу равносильна гауссовости сумм по каждому срезу, нормированных на  $\sqrt{N}$  при нашем диапазоне значений  $N$ : от 254 до 729. Также заметим, что гауссовость средних может следовать из того, что процесс Орнштейна–Уленбека в дискретном времени есть авторегрессия первого порядка, и для некоррелированного шума здесь достаточно показать соответствие нормальному закону.

#### 4. Методика построения прогноза<sup>4</sup>

- 1) Построить ряд из *средних значений* всех срезов по времени, выявить тренд, вычесть его из ряда, получить оценку центрированного и стандартизованного ряда  $\hat{Y}(j)$ ,  $j = \overline{0, n-1}$ .
- 2) Для оценочного ряда  $\hat{Y}(j)$ ,  $j = \overline{0, n-1}$ , построить частную ковариационную функцию (partial autocovariance function, PACF), получить подтверждение гипотезы об условной некоррелированности  $\hat{Y}(j)$  с  $\hat{Y}(k)$ ,  $|k - j| > 0$ .
- 3) Для оценочного ряда  $\hat{Y}(j)$ ,  $j = \overline{0, n-1}$ , построить ковариационную функцию (autocovariance function, ACF), получить подтверждение гипотезы об экспоненциальном виде ковариационной

<sup>4</sup> Ниже мы продолжаем использование уже принятых обозначений, но применительно к обрабатываемым данным.

функции, оценить степенной параметр экспоненты, сравнить с параметром экспоненциального распределения наблюдаемых спейсингов.

- 4) По формулам вида (2)–(5), записанным ниже, построить прогноз.

Введем обозначения:

$$E(\ln(V(0))) = \mu(0), E(\ln(V(s))) = \mu(s),$$

$$\sigma(\ln(V(0))) = \sigma(0), \sigma(\ln(V(s))) = \sigma(s),$$

$$\rho(\ln(V(0)), \ln(V(s))) = e^{-\lambda s}.$$

В предположении, что для средних логарифмов цен имеется линейный тренд вида  $\mu(s) = \alpha \cdot s + \beta$ , и среднеквадратичное отклонение постоянно  $\sigma(s) = \sigma(0) = \sigma$ , получаем для любого заранее заданного значения  $V(0) = v(0)$  прогнозные оценки:

$$\begin{aligned} Mode(V(s)|V(0) = v(0)) &= \\ &= \exp(\alpha \cdot s + \beta + e^{-\lambda s} (\ln(v(0)) - \mu(0) - \\ &\quad - \sigma^2(1 - e^{-2\lambda s}))), \end{aligned} \quad (2)$$

$$\begin{aligned} Median(V(s)|V(0) = v(0)) &= \\ &= \exp(\alpha \cdot s + \beta + e^{-\lambda s} (\ln(v(0)) - \mu(0))), \end{aligned} \quad (3)$$

$$\begin{aligned} E(V(s)|V(0) = v(0)) &= \\ &= \exp(\alpha \cdot s + \beta + e^{-\lambda s} (\ln(v(0)) - \mu(0)) + \\ &\quad + \frac{1}{2} \sigma^2(1 - e^{-2\lambda s})). \end{aligned} \quad (4)$$

Также приведем формулу, описывающую границы  $V_{L,R}(s)$  коридора ошибок (например, в пределах одного стандартного отклонения):

$$\begin{aligned} V_{L,R}(s) &= \exp(\alpha \cdot s + \beta + \\ &+ e^{-\lambda s} (\ln(v(0)) - \mu(0)) \pm \sigma \cdot \sqrt{1 - e^{-2\lambda s}}), \end{aligned} \quad (5)$$

где  $R, L$  — индексы, обозначающие, соответственно, правую (*right*) и левую (*left*) границы коридора ошибок.

Формулы (2)–(5) являются прямым следствием того важного свойства, что условное распределение процесса ОУ нормально с параметрами, заданными (1).

### 5. Практический пример применения метода

Для примера отобраны данные о продажах квартир в г. Санкт-Петербурге, опубликованные в номерах с 1483 по 1686 Бюллетеня недвижимо-

сти Санкт-Петербурга, охватывающих период с сентября 2011 г. по октябрь 2015 г. Бюллетень недвижимости Санкт-Петербурга издавался в печатном виде до конца 2019 года еженедельно. Номера выбирались по принципу «один номер в месяц», всего 50 номеров. Очевидно, что полная подборка ежемесячного номера содержит смешанную информацию об объектах разных категорий, а для нашего примера необходимы цены на объекты приблизительно одной категории. С этой целью, из общегородской информации отобраны объекты, расположенные в Адмиралтейском районе г. Санкт-Петербурга. Адмиралтейский район отличается обширной застройкой категории «старый фонд», с относительно небольшим количеством объектов класса «премиум» в центральной части города или старого фонда с видом на большую Неву. Премиальные объекты также исключены из выборки. Нас интересуют объекты, находившиеся в журнале хотя бы один период, таким образом, поток состоит из 48 выборок (в дальнейшем будем их называть временными срезами), являющихся эмпирическими выборками реализаций случайных величин  $Y(j) = \ln(V(j)), j = 0, 47$  и охватывает период в 4 года. Стандартные отклонения всех срезов потока имеют среднее значение  $\sigma = 0,22$  с незначительными изменениями во времени (стандартное отклонение от своего среднего (0,22) имеет порядок 0,01). Учитывая незначительные колебания стандартных отклонений, будем считать стандартное отклонение в потоке постоянным, равным  $\sigma = 0,22$  и строить прогноз при таком условии.

### 6. Распределение спейсингов

Для каждого значения цены подсчитывалось количество периодов, в течение которых данная цена сохранялась от среза к срезу без перерыва. Полученное число является длиной спейсинга  $\tau$ , соответствующего данной цене, выраженной в целом количестве периодов (один период — один месяц), принимающего целые значения от 1 до 48. Затем строим эмпирическое распределение спейсингов по длине (1, 2, 3, 4, ... и т.д.), а также их относительную частоту к общему количеству всех спейсингов. Накопленные относительные частоты дают наблюдаемые значения эмпирической функции распределения для длины спейсинга. Наше предположение заключается в том, что теоретическое распределение спейсингов подчиняется экс-

по экспоненциальному закону распределения<sup>5</sup>, то есть  $F(t) = P(l \leq t) = 1 - e^{-\lambda t}$ , где  $l$  – случайная (теоретическая) длина спейсинга. Рассматривается дополнительная функция распределения вероятностей  $P(l > t) = 1 - F(t) = e^{-\lambda t}$ . Очевидно, что  $\ln(1 - F(t)) = -\lambda \cdot t$ , то есть логарифм дополнительной функции линейно зависит от  $t$ . При этом  $\ln(1 - F(0)) = 0$ . Это свойство используется для подбора параметра  $\lambda$  экспоненциального распределения спейсингов. На рисунке 2 показаны наблюдаемые значения функции  $1 - F(t)$  при значениях  $t = 0, 1, 2, 3, 4, 5, 6, 7, 8$  и их аппроксимация экспонентой вида  $e^{-\lambda t}$ .

Оценка параметра  $\lambda$  получена библиотечной функцией `lm` статистического пакета *R*, значение  $\lambda = 0,6510$ .

### 7. Распределения цен

На рисунке 3 показаны эмпирические распределения цен 1 кв. м. в первых шести срезах  $V(j), j = 0, 5$ .

На рисунке 4 показаны значения  $p$ -value тестов Колмогорова–Смирнова на соответствие эмпирических распределений в 48 срезах теоретическому логарифмически нормальному распределению, с подобранными параметрами.

Таким образом, нет оснований отвергнуть гипотезу о логарифмически нормальных распределениях цен во всех 48 срезах (или нормальных распределениях логарифмов цен).

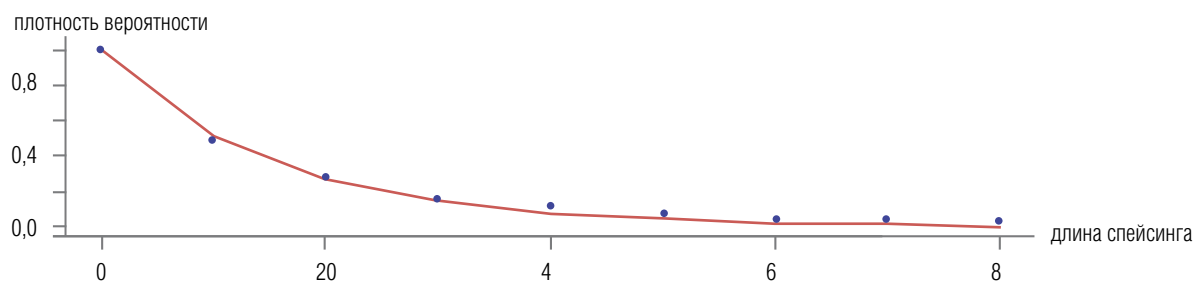


Рис. 2. Наблюдаемые значения дополнительной функции  $1 - F(t)$  при значениях  $t = 0, 1, 2, 3, 4, 5, 6, 7, 8$  (точки) и их аппроксимация экспонентой вида  $e^{-\lambda t}$ .

### 8. Независимость копий случайной последовательности

Рассмотрим независимость копий случайной последовательности  $Y(j) = \ln(V(j)), j = 0, 47$ . Сначала отметим, что количество копий в каждом срезе различно и меняется от 254 до 729. Известно, что дисперсия суммы двух случайных величин равна сумме дисперсий плюс удвоенная ковариация. Нам требуется свидетельство некоррелированности записей в срезах. С этой целью изучим поведение накопленных дисперсий в каждом из 48 срезов  $\ln(V(j)), j = 0, 47$ .

В выборке каждого среза проводятся следующие действия:

- ♦ формируется подвыборка, путем случайного выбора фиксированного количества элементов из выборки (среза), (например, 20, минимальный размер выборки в срезе 254, максимальный 729), подсчитывается дисперсия;
- ♦ из оставшихся элементов выборки выбираются следующие 20 элементов, подсчитывается дисперсия, результат прибавляется к дисперсии, полученной на предыдущем шаге;
- ♦ процедура продолжается до исчерпания выборки (на рисунке 5 по горизонтальной оси отмечается количество таких шагов).

Затем рассматривается зависимость накопленной таким образом дисперсии от количества шагов. Линейный характер накопленной дисперсии

<sup>5</sup> Т.к. мы рассматриваем срезы в дискретные моменты времени, то мы наблюдаем геометрически распределенные эмпирические спейсинги, которые являются проекцией непрерывных показательных спейсингов на дискретные моменты времени  $0, 1, 2, 3, \dots$ . Если случайная величина  $\xi \in \text{exp}(\lambda)$ , то  $[\xi] \in \text{Geom}(p)$  (где квадратные скобки обозначают целую часть числа) и в нашем случае  $p = 1 - e^{-\lambda}$ .



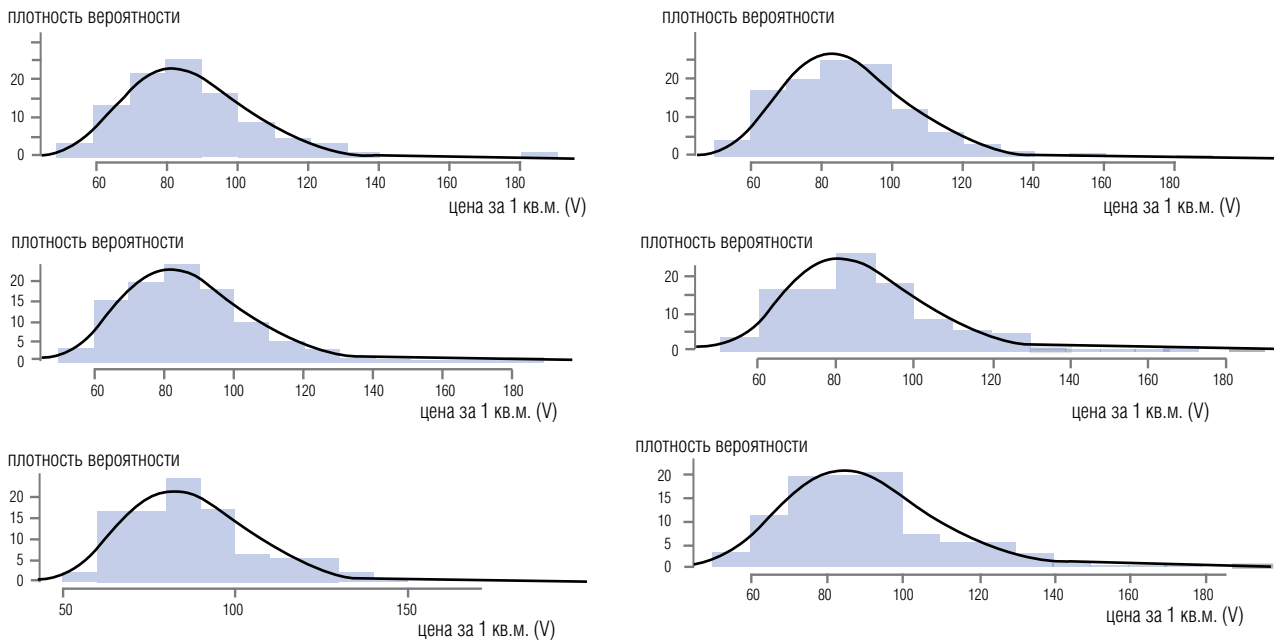


Рис. 3. Эмпирические распределения цен 1 кв. м. в первых шести срезах  $V(j), j = \overline{1, 6}$ , линия – плотность модельного логарифмически нормального распределения.

[1]	0.32078110	0.20548610	0.40557570	0.32223009	0.26319875	0.22332022
[7]	0.16753867	0.12097452	0.07531444	0.08037770	0.15068216	0.13773904
[13]	0.15556880	0.76906843	0.26749012	0.44678095	0.30453727	0.46525538
[19]	0.27115174	0.42116513	0.21523598	0.29461480	0.14404826	0.85821447
[25]	0.60624471	0.44881713	0.47551761	0.27010204	0.30031997	0.11477638
[31]	0.11761958	0.35463007	0.41267475	0.25858806	0.09053699	0.27756816
[37]	0.13531391	0.17893827	0.68449172	0.91269908	0.53167338	0.16490183
[43]	0.40555451	0.24649549	0.11753382	0.17232872	0.08156706	0.25339992

Рис. 4. Скриншот окна статистического пакета R со значениями  $p$ -value тестов Колмогорова–Смирнова на соответствие эмпирических распределений в 48 срезах  $V(j), j = \overline{0, 47}$  теоретическому логарифмически нормальному распределению с подобранными параметрами.

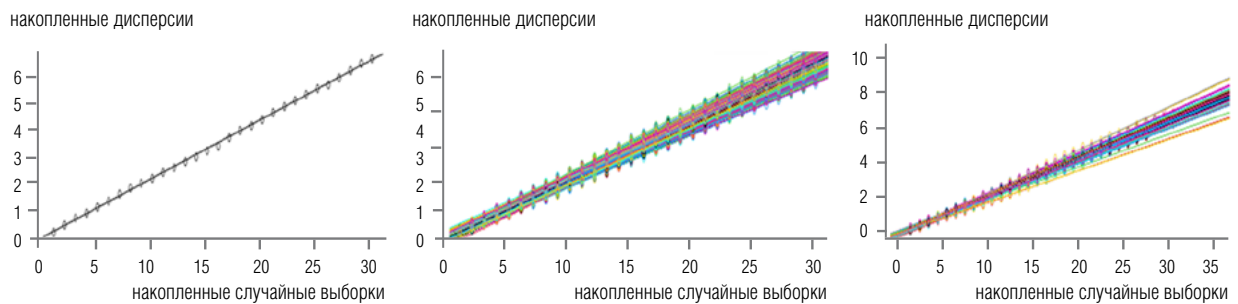


Рис. 5. Левый рисунок – накопленные дисперсии среза  $\ln(V(1))$ , рисунок в центре – накопленные дисперсии среза  $\ln(V(1))$  при 1000 повторениях процедуры случайного отбора подвыборки, правый рисунок – накопленные дисперсии всех срезов  $\ln(V(j)), j = \overline{0, 47}$ .

свидетельствует об отсутствии корреляционного слагаемого в полученных суммах. Поскольку в нашем случае суммируемые величины подчиняются нормальному закону распределения, некоррелируемость свидетельствует о независимости.

На *рисунке 5* показано поведение накопленных дисперсий в первом срезе, поведение накопленных дисперсий в первом срезе при повторениях процедуры случайного выбора 1000 раз, поведение накопленных дисперсий во всех остальных срезах.

### 9. Построение прогноза

Для построения прогноза следует выявить линейный тренд в данных, оценить параметр ПСИ-процесса  $\lambda$ , проверить совместную нормальность средних значений, которая даст подтверждение марковского свойства наблюдаемого процесса<sup>6</sup>. Стандартное отклонение процесса принято выше постоянным и равным  $\sigma = 0,22$ .

На *рисунке 6* показан ряд средних значений логарифмов  $Y(j) = \ln(V(j), j = \overline{0, 47}$ .

Линейный тренд оценен с применением библиотечной функции  $\text{lm}$  статистического пакета  $R$ . Удаляем линейный тренд, нормируем данные на стандартное отклонение  $\sigma = 0,22$ , усредняем данные в каждом срезе.

Получаем ряд (*рис. 7*) средних значений процесса  $\hat{Y}(j), j = \overline{0, 47}$ . График визуально соответствует графику траектории стационарного процесса.

На *рисунке 8* показан график автокорреляционной функции для ряда средних значений процесса  $\hat{Y}(j), j = \overline{0, 47}$ , с лагами до 7 и его аппроксимация экспонатной вида  $e^{-\lambda \cdot s}$ .

Оценка параметра  $\lambda = 0,6502$  получена применением библиотечной функции  $\text{lm}$  статистического пакета  $R$ .

На *рисунке 9* показан график частной автокорреляционной функции для ряда средних значений процесса  $\hat{Y}(j), j = \overline{0, 47}$  с лагами до 7.

### 10. Проверка совместной нормальности средних значений $\hat{Y}(j), j = \overline{0, 47}$

В случае с постоянно замещающимися объектами и ценами, различным объемом выборок в каждом срезе, проверка совместной нормальности библиотечными тестами (такими как, например, тесты пакета MVN [38, 39]) представляет значительные трудности. Для проверки на совместную нормальность нормированных сумм средних нашего ПСИ-процесса используем вид ковариационной функции и факт сходимости нормированных сумм ПСИ-процессов к процессу Орнштейна–Уленбека. Здесь мы опираемся на представление процесса Орнштейна–Уленбека с дискретным временем в виде авторегрессии первого порядка с гауссовским белым шумом.

Пусть  $\varepsilon_0, \varepsilon_1, \varepsilon_2, \dots$  – независимые одинаково распределенные стандартно нормальные случайные величины ( $\varepsilon_j \in N(0, 1), j = \overline{0, +\infty}$ ). Зададим процесс Орнштейна–Уленбека с дискретным временем и скоростью  $\lambda > 0$  в рекуррентном виде:

$$\begin{cases} u_0 = \varepsilon_0 \\ u_{n+1} = e^{-\lambda} u_n + \sqrt{1 - e^{-2\lambda}} \varepsilon_{n+1} \end{cases}$$

Очевидно, что  $u_j \in N(0, 1), j = \overline{0, +\infty}$ . Можно записать

$$\begin{cases} \varepsilon_0 = u_0 \\ \varepsilon_{n+1} = \frac{u_{n+1} - e^{-\lambda} u_n}{\sqrt{1 - e^{-2\lambda}}} \end{cases} \quad (6)$$

Количество элементов  $u$  выборок в каждый момент дискретного времени обозначим  $N(j), j = \overline{0, +\infty}$  и положим

$$u_j = \frac{1}{\sqrt{N(j)}} \sum_{i=1}^{N(j)} \hat{Y}_i(j), \quad j = \overline{0, +\infty} \quad (7)$$

Применяя формулы (6) к величинам (7), получаем оценки  $\hat{\varepsilon}_0, \hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots$  для  $\varepsilon_0, \varepsilon_1, \varepsilon_2, \dots$ . Из независимости и нормальности величин  $\varepsilon_0, \varepsilon_1, \varepsilon_2, \dots$  следует совместная нормальность величин (7).

Данные состоят из 48 выборок (каждая разной длины), соответствующих 48 дискретным моментам времени. Для каждой из них построим  $\hat{Y}(j), j = \overline{0, 47}$ , нормированные суммы средних (7), оценки  $\hat{\varepsilon}_0, \hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_{47}$  (6). Для этого надо выявить тренд (показано выше), оценить параметр  $\lambda$  (мы его

<sup>6</sup> На марковское свойство также указывает вид PACF на *рис. 9*, где ровно один значимый пик в единице.

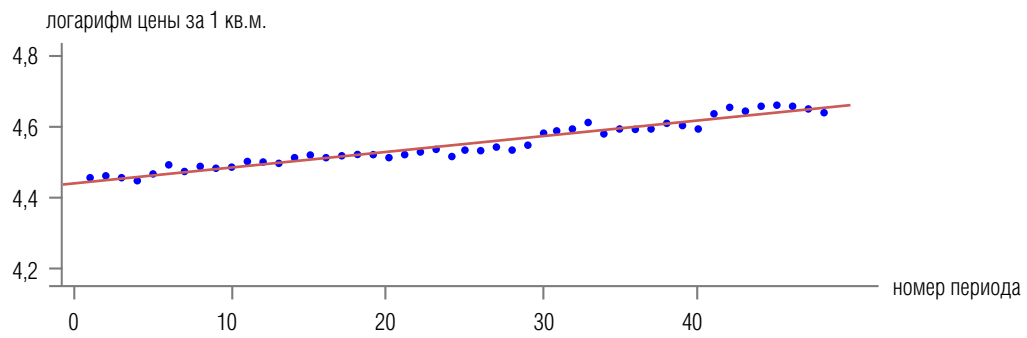


Рис. 6. Ряд из средних значений для  $Y(j) = \ln(V(j))$ ,  $j = \overline{0, 47}$ ,  
уравнение линии тренда  $4,443 + 0,00432 \cdot t$ ,  
где  $t$  – время в периодах (1 период = 1 месяц) от начала наблюдений (при  $j = 1$  значение  $t = 1/4$ ).

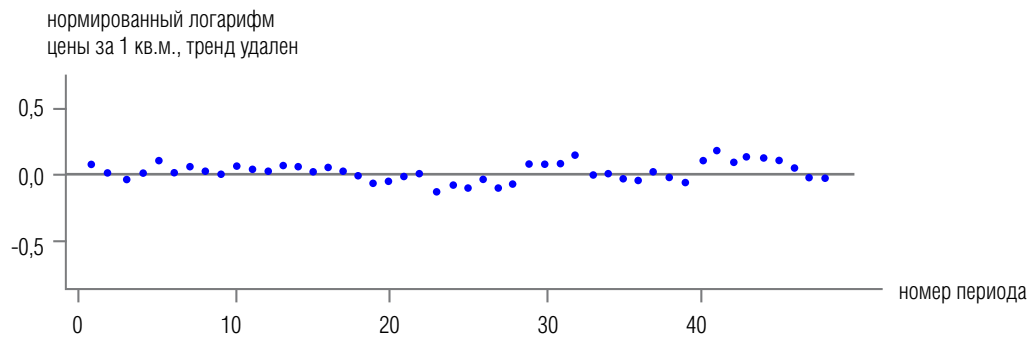


Рис. 7. Ряд средних значений процесса  $\hat{Y}(j)$ ,  $j = \overline{0, 47}$ .

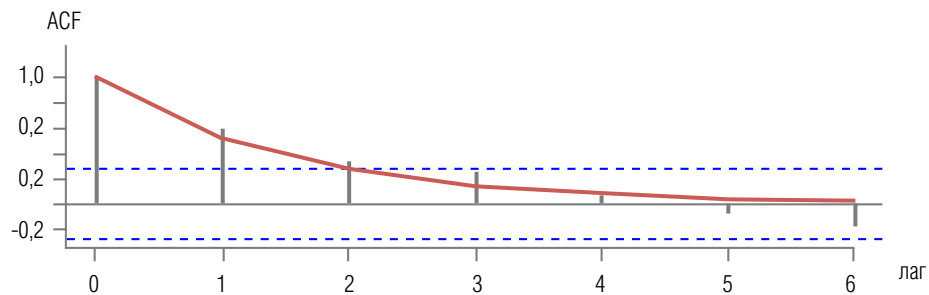


Рис. 8. График автокорреляционной функции для ряда средних значений процесса  $\hat{Y}(j)$ ,  $j = \overline{0, 47}$ , с лагами до 7 и его аппроксимация экспонатной вида  $e^{-0,6502 \cdot s}$ .

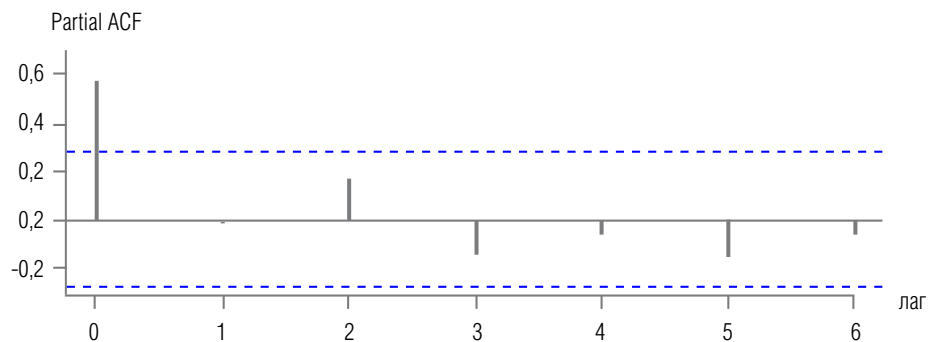


Рис. 9. График частной автокорреляционной функции для ряда средних значений процесса  $\hat{Y}(j)$ ,  $j = \overline{0, 47}$  с лагами до 7.

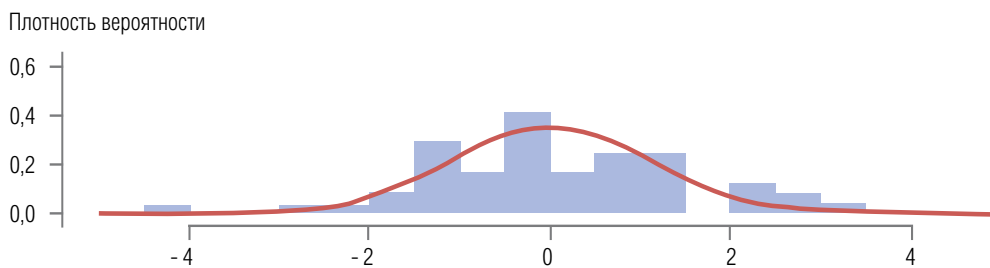


Рис. 10. Эмпирическое распределение величин  $\hat{\epsilon}_j, j = \overline{0, 47}$  и его аппроксимация плотностью нормального распределения. Параметры модельного распределения  $\mu = 0,0267, \sigma = 1,1$ .

получили дважды,  $\lambda \approx 0,65$ ), стандартное отклонение принято ранее постоянным  $\sigma = 0,22$ . На рисунке 10 показано эмпирическое распределение величин  $\hat{\epsilon}_j, j = \overline{0, 47}$  и его аппроксимация плотностью нормального распределения. На рисунке 11 показан скриншот результатов статистических тестов на нормальность величин  $\hat{\epsilon}_j, j = \overline{0, 47}$  тестами Колмогорова–Смирнова, Шапиро, Андерсена–Дарлингга.

Таким образом, проведенные проверки позволяют нам подтвердить следующие гипотезы:

- 1) спейсинги ( $\tau_i$ ),  $i = \overline{1, 48}$  подчиняются экспоненциальному закону распределения с параметром  $\lambda \approx 0,65$ ;
- 2) каждая компонента дискретного процесса  $\hat{Y}(j)$ ,  $j = \overline{0, 47}$  стандартно нормальна;
- 3) наблюдаемые копии случайного ряда  $\ln(V(j))$  являются независимыми ( $i$  – номер копии,  $j$  – номер среза);

- 4) автокорреляционная функция может быть аппроксимирована (рис. 8) функцией экспоненциального вида  $e^{-\lambda s}$ , с параметром  $\lambda \approx 0,65$ , таким образом коэффициент корреляции для прогноза на 6 шагов (то есть, на полгода вперед) может быть установлен как  $e^{-0,65 \cdot 6} = e^{-3,9}$ ;
- 5) вид частной автокорреляционной функции, представленной на рисунке 9, указывает на условную некоррелированность  $\hat{Y}(j)$  с  $\hat{Y}(k)$ ,  $|k - j| > 1$ ,  $j = \overline{0, 47}$  при условии, что фиксированы значения  $\hat{Y}$  в моменты времени строго между  $k$  и  $i$ ;
- 6) марковское свойство дискретного процесса  $\hat{Y}(j)$ ,  $j = \overline{0, 47}$  следует из совместной нормальности и вида ковариационной функции. Марковское свойство также подтверждается видом частной автоковариационной функции.

## 11. Прогноз

Приведенные выше проверки позволяют установить соответствие наблюдаемых данных модели ПСИ-процесса. Марковское свойство наблюдаемого дискретного процесса  $\hat{Y}(j)$ ,  $j = \overline{0, 47}$  позволяет строить прогноз только по последнему распределению случайной величины  $V$  (стоимость 1 кв. м. вторичной жилой недвижимости в Адмиралтейском районе, в 48 периоде, соответствующем сентябрю 2015 года). Эмпирическое распределение случайной величины  $V(48)$  удовлетворительно аппроксимируется логарифмически нормальным распределением с параметрами  $\mu(0) = 4,64$ ,  $\sigma = 0,22$  ( $p$ -value теста Колмогорова–Смирнова равно 0,2534, см. рис. 4).

Математическое ожидание модельного распределения в 48 периоде равно  $e^{4,64+0,5 \cdot 0,22^2} = 106,081$  тыс. руб. за 1 кв. м.; медиана модельного распределения в 48 периоде равна  $e^{4,64} = 103,544$  тыс. руб. за

```
> ks.test(eps, "pnorm", mean=0.0267, sd=1.1)
      Exact one-sample Kolmogorov-Smirnov test
data:  eps
D = 0.096711, p-value = 0.7239
alternative hypothesis: two-sided

> shapiro.test(eps)
      Shapiro-Wilk normality test
data:  eps
W = 0.97975, p-value = 0.5684

> ad.test(eps)
      Anderson-Darling normality test
data:  eps
A = 0.25751, p-value = 0.7054
```

Рис. 11. Скриншот результатов статистических тестов Колмогорова–Смирнова, Шапиро, Андерсена–Дарлингга. Результаты тестов подтверждают гипотезу нормальности шума (6).

1 кв. м.; мода модельного распределения в 48 периоде равна  $e^{4,64-0,22^2} = 98,652$  тыс. руб. за 1 кв. м.; отметим, что среднее значение цены превышает наиболее вероятное на 7,5%, минимум цены 1 кв. м. в эмпирической выборке в 48 периоде 63,265 тыс. руб.; максимум цены 1 кв. м. в эмпирической выборке в 48 периоде 172,556 тыс. руб.

Корреляционная функция нормированных сумм дискретного процесса  $\hat{Y}(j), j = \overline{0, 47}$  имеет экспоненциальный вид с параметром  $\lambda = 0,65$ . Причем, это значение мы получили дважды: как параметр экспоненты, аппроксимирующей корреляционную функцию, и как параметр экспоненциального распределения спейсингов.

Коэффициент корреляции для логарифма пары произвольных сечений процесса  $V(t)$ , образующих двумерный случайный вектор  $(V(0), V(s))$ , равен  $\rho(\ln(V(0)), \ln(V(s))) = e^{-0,65s}$ . Формулы (2)–(4) имеют вид:

$$\begin{aligned} \text{Mode}(V(s)|V(0) = v(0)) &= \\ &= \exp(0,00432 \cdot s + 4,443 + e^{-0,65s} (\ln(v(0)) - \mu(0) - \\ &\quad - 0,22^2(1 - e^{-2 \cdot 0,65s})), \end{aligned}$$

$$\begin{aligned} \text{Median}(V(s)|V(0) = v(0)) &= \\ &= \exp(0,00432 \cdot s + 4,443 + e^{-0,65s} (\ln(v(0)) - \mu(0))), \end{aligned}$$

$$\begin{aligned} E(V(s)|V(0) = v(0)) &= \\ &= \exp(0,00432 \cdot s + 4,443 + e^{-0,65s} (\ln(v(0)) - \mu(0)) + \\ &\quad + \frac{1}{2} 0,22^2(1 - e^{-2 \cdot 0,65s})). \end{aligned}$$

Формула (5), описывающая границы  $V_{L,R}(s)$  коридора ошибок (например, в пределах одного стандартного отклонения), принимает вид:

$$\begin{aligned} V_{L,R}(s) &= \exp\left(0,00432 \cdot t + 4,443 + \right. \\ &\quad \left. + e^{-0,65s} (\ln(v(0)) - \mu(0)) \pm 0,22 \cdot \sqrt{1 - e^{-2 \cdot 0,65s}}\right). \end{aligned}$$

Для любого объекта цена 1 кв. м. которого  $V(0) = v(0)$  была выставлена в последнем периоде (как указано выше, соответствует сентябрю 2015 года), может быть сделан прогноз на следующие 6 периодов.

Построим прогноз на 6 периодов (на полгода вперед) для объекта, у которого в последнем периоде была цена 106,081 тыс. руб. за 1 кв. м. (средняя цена по срезу с номером 47).

На *рисунке 12* показан прогноз для значения цены  $v(0) = 106,081$  (тыс. руб. за 1 кв. м.)

На *рисунке 12* точками отмечены фактические средние значения цен 1 кв. м. в этом же множестве объектов недвижимости (масс-маркет, Адмиралтейский район), которые составляют:

в октябре 2015 г.	103,100 тыс. руб. за 1 кв. м.,
в ноябре 2015 г.	103,492 тыс. руб. за 1 кв. м.,
в декабре 2015 г.	103,247 тыс. руб. за 1 кв. м.,
в январе 2016 г.	102,027 тыс. руб. за 1 кв. м.,
в феврале 2016 г.	101,044 тыс. руб. за 1 кв. м.,
в марте 2016 г.	102,046 тыс. руб. за 1 кв. м.

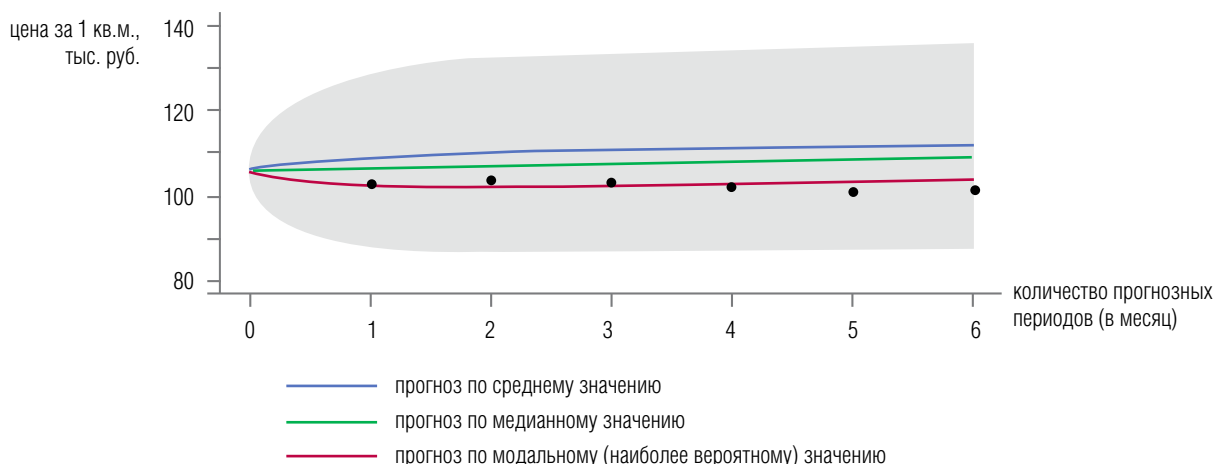


Рис. 12. Прогноз для значения цены  $v(0) = 106,081$  тыс. руб. за 1 кв. м. — средней цены в последнем срезе наблюдений. Серый фон — коридор в пределах одного стандартного отклонения.

### Заключение

1. Проведенный комплекс проверок позволил установить следование изучаемых данных модели ПСИ-процесса.
2. На *рис. 11* показан прогноз для цены, равной среднему значению в последнем срезе наблюдений. Черными точками показаны фактически наблюдавшиеся средние значения цены 1 кв. м. в последующих 6 месяцах, они расположены вблизи линии модального прогноза, что служит хорошей верификацией прогноза по наиболее вероятному значению. Линия модальных значений указывает на наиболее вероятное значение будущей цены, при условии, что цена в последнем срезе равнялась среднему значению. Это следствие свойства логарифмически нормального распределения цен и того, что усредненные значения логарифмов цен следуют процессу Орнштейна–Уленбека.
3. Параметр  $\lambda = 0,65$  указывает, в том числе, на среднее время, за которое происходит скачок цены объекта в наблюдаемом секторе рынка, он составляет  $1/\lambda = 1,54$  периода. В представленном примере периодом является 1 месяц.
4. Основные преимущества прогнозирования, основанного на свойствах ПСИ-процессов:
  - ◆ стандартное отклонение прогноза стабилизируется с течением времени на уровне постоянной дисперсии случайного процесса, в отличие от моделей скользящего среднего, накапливающих ошибку прогноза на каждом шаге;
  - ◆ возможность построить прогноз не только для средних значений, но и для любого объекта, выставленного на продажу в последнем временном периоде наблюдений по некоторой цене. ■

### Литература

1. Фридман Дж., Ордуэй Н. Анализ и оценка приносящей доход недвижимости. М.: Дело, 1997.
2. Стерник Г.М., Стерник С.Г. Методология прогнозирования Российского рынка недвижимости // Механизация строительства. 2018. № 8(830). С. 57.
3. Стерник Г.М., Стерник С.Г. Рынок жилой недвижимости Москвы и Московской области: текущее состояние и ценовой прогноз // Управление развитием территории. 2015. № 4.
4. Стерник Г.М., Печенкина А.В. Прогноз цен предложения квартир на российском рынке жилья (макроэкономический подход) // Имущественные отношения в Российской Федерации. 2007. № 10(73). С.11–18.
5. Стерник Г.М., Стерник С.Г., Свиридов А.В. Развитие и совершенствование методов прогнозирования на рынке жилой недвижимости // Урбанистика и рынок недвижимости. 2014. № 1. С. 53–93.
6. Алексеева М.О., Пайдыганова М.Ю. Анализ и прогнозирование цен на квартиры в республике Мария Эл // Аллея науки. 2019. Т. 1. № 9(36). С. 65–70.
7. Дробышевский С.М. Анализ возможности возникновения «пузыря» на российском рынке недвижимости // Научные труды ИЭПП. 2009. № 128.
8. Евстафьев А.И., Гордиенко В.А. Прогнозирование индикаторов рынка недвижимости с применением нейронных сетей // Известия вузов. Северо-Кавказский регион. Общественные науки. 2009. № 5. С. 83–89.
9. Зенчик А.С., Морозова Н.Н. Прогнозирование цен на недвижимость с учетом сезонности // Молодежный научный форум: общественные и экономические науки. 2016. № 11(40). С. 397–402.
10. Кошкин В.С., Боронина Н.Ю. Прогнозирование рыночной стоимости коммерческой недвижимости на основе показателей экономического развития территории г. Барнаул // Реализация приоритетных программ развития АПК. Сборник научных трудов по итогам X Международной научно-практической конференции, посвященной памяти заслуженного деятеля науки РФ и КБР профессора Б.Х. Жерукова, Нальчик, 2022. С. 312–315.
11. Молчанова М.Ю., Печенкина А.В. Особенности использования методов фундаментального и технического анализа при прогнозировании цен на рынке недвижимости региона // Вестник Пермского университета. Серия «Экономика». 2011. № 3(10). С. 53–54.
12. Никитина Н.С. Прогнозирование индекса цен на недвижимость с учетом сезонности // Экономическое развитие России. 2022. Т. 29. № 6. С. 23–28.
13. Нурмухаметов И.М. Анализ и прогнозирование цен на недвижимость в республике Марий-Эл // Промышленное развитие России: проблемы, перспективы. Труды XII Международной научно-практической конференции преподавателей, ученых, специалистов, аспирантов, студентов: в 3 томах. Кафедра экономики предприятия, 2014. С. 97–104.
14. Печенкина А.В. Использование многоуровневой факторной модели при прогнозировании ситуации на региональном рынке недвижимости (на примере Пермского края) // Имущественные отношения в Российской Федерации. 2010. № 11(110). С. 57–72.

15. Питулин С.С. Построение ARIMA моделей для анализа динамики цен на недвижимость в Смоленской области // Интернаука. 2019. № 20-1(102). С. 63–67.
16. Рубинштейн Е.Д., Осипенко Н.С. Анализ рынка недвижимости и его прогнозирование // Теория и практика общественного развития. 2015. № 12. С. 140–143.
17. Clapp J.M., Giaccotto C. Evaluating house price forecasts // *Journal of Real Estate Research*. 2002. Vol. 24. No. 1. P. 1–25.
18. Crawford G.W., Fratantoni M.C. Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house price // *Real Estate Economics*. 2003. Vol. 31 No. 2. P. 223–243.
19. Geltner D.M., Miller N.G., Clayton J., Eichholtz P. *Commercial real estate analysis and investments*. South-Western Educational Pub, 2006.
20. Gotham K.F. The secondary circuit of capital reconsidered: Globalization and the U.S. real estate sector // *American Journal of Sociology*. 2006. Vol. 112. No. 1. P. 231–275. <https://doi.org/10.1086/502695>
21. Green R.K. Imperfect information and the housing finance crisis: A descriptive overview // *Journal of Housing Economics*. 2008. Vol. 17. No. 4. P. 262–271. <https://doi.org/10.1016/j.jhe.2008.09.003>
22. Green R.K. Follow the leader: How changes in residential and nonresidential investment predict changes in GDP // *Real Estate Economics*. 1997. Vol. 25. No. 2. P. 253–270.
23. Ласкин М.Б., Черкесова П.А., Сравнение рыночных и кадастровых данных для прогнозирования рыночной стоимости объектов недвижимости // *Статистика и экономика*. 2020. Том 17. № 4. С. 44–54.
24. Федоров Н.И. Прогнозирование цен на жилую недвижимость на рынке Челябинска методами машинного обучения // *Студент и научно-технический прогресс. Материалы XLIV научной конференции молодых ученых, Челябинск, 2020*. С. 223–226.
25. Li Y., Xiang Z., Xiong T. The behavioral mechanism and forecasting of Beijing housing prices from multiscale perspective // *Discrete Dynamics in Nature and Society*. 2020. Vol. 2020. Article ID 5375206. <https://doi.org/10.1155/2020/5375206>
26. Ohnishi T., Mizuno T., Shimizu C., Watanabe T. On the evolution of the house price distribution // *Columbia Business School. Center of Japanese Economy and Business. Working Paper Series*. 2011. No. 296. P. 1–20.
27. Raymond Y.C.T. An application of the ARIMA model to real estate prices in Hong-Kong // *Journal of Property Finance*. 1997. Vol. 8. No. 2. P. 152–163.
28. Singh Y. *Model for forecasting price of houses in city of Stillwater*. M.S. dissertation. Oklahoma State University, 2005.
29. Sa'at N.F., Maimun N.H.A., Idris N.H. Enhancing the accuracy of Malaysian house price forecasting: A comparative analyses on the forecasting performance between the hedonic price model and artificial neural network model // *Planning Malaysia*. 2021. Vol. 19. No. 3. P. 249–259.
30. Rusakov O., Laskin M., Jaksumbaeva O. Pricing in the real estate market as a stochastic limit. Log Normal approximation // *International Journal of the Mathematical models and methods in applied sciences*. 2016. Vol. 10. P. 229–236.
31. Rusakov O., Laskin M., Jaksumbaeva O., Ivakina A. «Pricing in real estate market as a stochastic limit. Lognormal approximation» // *2015 Second International Conference on Mathematics and Computers in Sciences and in Industry (MCSI)*. Institute of Electrical and Electronics Engineers Inc., 2015. P. 235–239. <https://doi.org/10.1109/MCSI.2015.4848>
32. Aitchinson J., Brown J.A.C. *The Lognormal distribution with special references to its uses in economics*. Cambridge: University Press, 1963.
33. Русаков О.В. Псевдо-пуассоновские процессы со стохастической интенсивностью и класс процессов, обобщающих процесс Орнштейна–Уленбека // *Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия*. 2017. Т. 4. № 2. С. 247–257.
34. Yakubovich Y., Rusakov O., Gushchin A. Functional limit theorem for the sums of PSI-processes with random intensities // *Mathematics*. 2022. Vol. 10. No. 21. 3955. <https://doi.org/10.3390/math10213955>
35. Feller W. *An introduction to probability theory and its applications*. Wiley, 1971.
36. Ласкин М.Б. Корректировка рыночной стоимости по ценообразующему фактору «площадь объекта» // *Имущественные отношения в Российской Федерации*. 2017. № 8(191). С. 86–99.
37. Ласкин М.Б. Определение скидки на торг по рыночным данным и кадастровой стоимости // *Бизнес-информатика*. 2018. № 3(45). С. 53–61. <https://doi.org/10.17323/1998-0663.2018.3.53.61>
38. Korkmaz S., Goksuluk D., Zararsiz G. MVN: An R package for assessing multivariate normality // *The R Journal*. 2014. Vol. 6/2. P. 151–162.
39. Korkmaz S., Goksuluk D., Zararsiz G. Package 'MVN'. 2022. [Электронный ресурс]: <https://cran.r-project.org/web/packages/MVN/MVN.pdf> (дата обращения 30.11.2023).

### Об авторах

#### Ласкин Михаил Борисович

кандидат физико-математических наук;

доцент, с.н.с., Лаборатория интеллектуальных систем, Санкт-Петербургский Федеральный исследовательский центр РАН, Россия, 199178, г. Санкт-Петербург, Васильевский остров, 14 линия, д. 39;

E-mail: laskinmb@yahoo.com

ORCID: 0000-0002-0143-4164

#### Русаков Олег Витальевич

кандидат физико-математических наук;

доцент, Кафедра теории вероятностей и математической статистики, Санкт-Петербургский государственный университет, Россия, 199034, г. Санкт-Петербург, Университетская наб., д. 7/9;

E-mail: ovirusakov@yahoo.co.uk

ORCID: 0000-0001-7444-9896

# Prediction of distributions of unit prices for real estate properties on the basis of the characteristics of PSI-processes

#### Mikhail B. Laskin<sup>a</sup>

E-mail: laskinmb@yahoo.com

#### Oleg V. Rusakov<sup>b</sup>

E-mail: ovirusakov@yahoo.co.uk

<sup>a</sup> Saint Petersburg Federal Research Center, Russian Academy of Sciences  
Address: 39, 14th Line, Vasilyevsky Island, St. Petersburg 199178, Russia

<sup>b</sup> Saint Petersburg State University  
Address: 7/9, Universitetskaya Nab., St. Petersburg 199034, Russia

### Abstract

Real estate market price forecasting is always in the focus of interests of scientists-economists, market analysts, market participants (sellers and buyers), marketing services of building complex enterprises, analysts working for banks and insurance companies and investors. Under present day conditions, the price behavior of properties on real estate markets takes especially important meaning subject to the influence of such factors as changes in the structure of household incomes, changes in mortgage rates and their availability, dynamic changes in the macroeconomic and other external socio-economic and political type factors. However, unlike the financial and securities markets, the real estate market is always characterized by a delayed reaction to external perturbations, often up to half a year, which allows us to hope for an appropriate construction of forecasts, at least in time for the delayed reaction. Traditional autoregressive forecasting methods are characterized by rapidly increasing forecast variance, because they assume a factor of stochastic volatility.



This paper proposes a model and method of forecast construction based on stochastic processes of the “Poisson random index” having a short time for reaching a stationary stable variance. The model is based on the “principle of replacements” of current prices with new ones. We analyze in detail an example of the application of the “principle of replacements” for construction of price forecasts on secondary residential real estate in St. Petersburg which is based on data of four-year observations of offer prices.

**Keywords:** real estate price forecast, log-normal price distribution, pseudo-Poisson process, Poisson random index process, Ornstein–Uhlenbeck process

**Citation:** Laskin M.B., Rusakov O.V. (2023) Prediction of distributions of unit prices for real estate properties on the basis of the characteristics of PSI-processes. *Business Informatics*, vol. 17, no. 4, pp. 7–24. DOI: 10.17323/2587-814X.2023.4.7.24

### References

1. Friedman J., Ordway N. (1997) *Analysis and assessment of income-generating real estate*. M.: Delo (in Russian).
2. Sternik G.M., Sternik S.G. (2018) Methodology for forecasting the Russian real estate market, *Mechanization of construction*, no. 8(830), p. 57 (in Russian).
3. Sternik G.M., Sternik S.G. (2015) Residential real estate market in Moscow and the Moscow region: current state and price forecast. *Territory Development Management*, no. 4 (in Russian).
4. Sternik G.M., Pechenkina A.V. (2007) Forecast of supply prices for apartments on the Russian housing market (macroeconomic approach). *Property Relations in the Russian Federation*, no. 10(73), pp. 11–18 (in Russian).
5. Sternik G.M., Sternik S.G., Sviridov A.V. (2014) Development and improvement of forecasting methods in the residential real estate market. *Urban Studies and Real Estate Market*, no. 1, pp. 53–93 (in Russian).
6. Alekseeva M.O., Paidiganova M.Yu. (2019) Analysis and forecasting of prices for apartments in the Republic of Maria El. *Alleya Nauki*, vol. 1, no. 9(36), pp. 65–70 (in Russian).
7. Drobyshevsky S.M. (2009) Analysis of the possibility of a “bubble” in the Russian real estate market. *IEP Scientific Works*, no. 128 (in Russian).
8. Evstafiev A.I., Gordienko V.A. (2009) Forecasting indicators of the real estate market using neural networks. *University News. North-Caucasian Region. Social Sciences Series*, no. 5, pp.83–89 (in Russian).
9. Zenchik A.S., Morozova N.N. (2016) Forecasting real estate prices taking into account seasonality. *Youth Scientific Forum: Social and Economic Sciences*, no. 11(40), pp. 397–402 (in Russian).
10. Koshkin V.S., Boronina N.Yu. (2022) Forecasting the market value of commercial real estate based on indicators of economic development of the territory of Barnaul. *Implementation of priority programs for the development of the agro-industrial complex. Collection of scientific papers based on the results of the X International Scientific and Practical Conference dedicated to the memory of the Honored Scientist of the Russian Federation and Kabardino-Balkaria, Professor B.Kh. Zherukova, Nalchik*, pp. 312–315 (in Russian).
11. Molchanova M.Yu., Pechenkina A.V. (2011) Features of using methods of fundamental and technical analysis when forecasting prices on the regional real estate market. *Bulletin of Perm University. Series “Economics”*, no. 3(10), pp. 53–54 (in Russian).
12. Nikitina N.S. (2022) Forecasting the real estate price index taking into account seasonality, *Economic Development of Russia*, vol. 29, no. 6, pp. 23–28 (in Russian).
13. Nurmukhametov I.M. (2014) Analysis and forecasting of real estate prices in the Mari-El Republic. *Industrial development of Russia: problems, prospects. Proceedings of the XII International Scientific and Practical Conference of teachers, scientists, specialists, graduate students, students*, pp. 97–104 (in Russian).
14. Pechenkina A.V. (2010) Using a multi-level factor model when predicting the situation in the regional real estate market (using the example of the Perm Territory). *Property Relations in the Russian Federation*, no. 11(110), pp. 57–72 (in Russian).
15. Pitulin S.S. (2019) Construction of ARIMA models for analyzing the dynamics of real estate prices in the Smolensk region. *Internauka*, no. 20-1(102), pp. 63–67 (in Russian).
16. Rubinshtein E.D., Osipenko N.S. (2015) Analysis of the real estate market and its forecasting. *Theory and Practice of Social Development*, no. 12, pp. 140–143 (in Russian).
17. Clapp J.M., Giaccotto C. (2002) Evaluating house price forecasts. *Journal of Real Estate Research*, vol. 24, no. 1, pp. 1–25.
18. Crawford G.W., Fratantoni M.C. (2003) Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house price. *Real Estate Economics*, vol. 31, no. 2, pp. 223–243.
19. Geltner D.M., Miller N.G., Clayton J., Eichholtz P. (2006) *Commercial real estate analysis and investments*. South-Western Educational Pub.

20. Gotham K.F. (2006) The Secondary Circuit of Capital Reconsidered: Globalization and the U.S. Real Estate Sector. *American Journal of Sociology*, vol. 112, no. 1, p. 231–275. <https://doi.org/10.1086/502695>
21. Green R.K. (2008) Imperfect information and the housing finance crisis: A descriptive overview. *Journal of Housing Economics*, vol. 17, no. 4, p. 262–271. <https://doi.org/10.1016/j.jhe.2008.09.003>
22. Green R.K. (1997) Follow the leader: How changes in residential and nonresidential investment predict changes in GDP. *Real Estate Economics*, vol. 25, no. 2, p. 253–270.
23. Laskin M.B., Cherksova P.A. (2020) Comparison of market and cadastral data for predicting the market value of real estate. *Statistics and Economics*, vol. 17, no. 4, pp. 44–54 (in Russian).
24. Fedorov N.I. (2020) Forecasting prices for residential real estate on the Chelyabinsk market using machine-learning methods. *Student and scientific and technological progress. Materials of the XLIV scientific conference of young scientists, Chelyabinsk*, pp. 223–226 (in Russian).
25. Li Y., Xiang Z., Xiong T. (2020) The behavioral mechanism and forecasting of Beijing housing prices from multiscale perspective. *Discrete Dynamics in Nature and Society*, vol. 2020, article ID 5375206. <https://doi.org/10.1155/2020/5375206>
26. Ohnishi T., Mizuno T., Shimizu C., Watanabe T. (2011) On the evolution of the house price distribution. *Columbia Business School. Center of Japanese Economy and Business. Working Paper Series*, no. 296, pp. 1–20.
27. Raymond Y.C.T. (1997) An application of the ARIMA model to real estate prices in Hong-Kong. *Journal of Property Finance*, vol. 8, no. 2, p. 152–163. <https://doi.org/10.1108/09588689710167843>
28. Singh Y. (2005) *Model for forecasting price of houses in city of Stillwater*, M.S. dissertation, Oklahoma State University.
29. Sa'at N.F., Maimun N.H.A., Idris N.H. (2021) Enhancing the accuracy of Malaysian house price forecasting: a comparative analyses on the forecasting performance between the hedonic price model and artificial neural network model. *Planning Malaysia*, vol. 19, no. 3, p. 249–259.
30. Rusakov O., Laskin M., Jaksumbaeva O. (2016) Pricing in the real estate market as a stochastic limit. Log Normal approximation. *International Journal of the Mathematical Models and Methods in Applied Sciences*, vol.10, pp. 229–236.
31. Rusakov O., Laskin M., Jaksumbaeva O., Ivakina A. (2015) Pricing in real estate market as a stochastic limit. Lognormal approximation, 2015 *Second International Conference on Mathematics and Computers in Sciences and in Industry (MCSI)*. Institute of Electrical and Electronics Engineers Inc., pp. 235–239. <https://doi.org/10.1109/MCSI.2015.48>
32. Aitchinson J., Brown J.A.C. (1963) *The Lognormal distribution with special references to its uses in economics*. Cambridge: University Press.
33. Rusakov O.V. (2017) Pseudo-Poisson processes with stochastic intensity and the class of processes generalizing the Ornstein-Uhlenbeck process. *Vestnik of Saint Petersburg University. Mathematics. Mechanics. Astronomy*, vol. 4, no. 2, pp. 247–257 (in Russian).
34. Yakubovich Y., Rusakov O., Gushchin A. (2022) Functional limit theorem for the sums of PSI-processes with random intensities. *Mathematics*, vol. 10, no. 21, 3955. <https://doi.org/10.3390/math10213955>
35. Feller W. (1971) *An introduction to probability theory and its applications*. Wiley.
36. Laskin M.B. (2017) Adjustment of market value according to the pricing factor “object area”. *Property Relations in the Russian Federation*, no. 8(191), pp. 86–99 (in Russian).
37. Laskin M.B. (2018) Determination of the trading discount based on market data and cadastral value. *Business Informatics*, no. 3(45), pp. 53–61. <https://doi.org/10.17323/1998-0663.2018.3.53.61>
38. Korkmaz S., Goksuluk D., Zararsiz G. (2014) MVN: An R package for assessing multivariate normality. *The R Journal*, vol. 6/2, pp. 151–162.
39. Korkmaz S., Goksuluk D., Zararsiz G. (2022) *Package 'MVN'*. Available at: <https://cran.r-project.org/web/packages/MVN/MVN.pdf> (accessed 30 November 2023).

### About the authors

#### Mikhail B. Laskin

Cand. Sci. (Phys.-Math.), Associate Professor;

Senior Researcher, Laboratory of Intelligent Systems, Saint Petersburg Federal Research Center, Russian Academy of Sciences, 39, 14th Line, Vasilyevsky Island, St. Petersburg 199178, Russia;

E-mail: [laskinmb@yahoo.com](mailto:laskinmb@yahoo.com)

ORCID: 0000-0002-0143-4164

#### Oleg V. Rusakov

Cand. Sci. (Phys.-Math.);

Associate Professor, Department of Probability Theory and Mathematical Statistics, Saint Petersburg State University, 7/9, Universitetskaya Nab., St. Petersburg 199034, Russia;

E-mail: [ovirusakov@yahoo.co.uk](mailto:ovirusakov@yahoo.co.uk)

ORCID: 0000-0001-7444-9896