СТОХАСТИЧЕСКАЯ МОДЕЛЬ ЦЕНООБРАЗОВАНИЯ НА РЫНКЕ НЕДВИЖИМОСТИ: ФОРМИРОВАНИЕ ЛОГНОРМАЛЬНОЙ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

УДК 339.13.017

Олег Витальевич Русаков,
к.ф-мат.н., доцент кафедры теории вероятностей и математической статистики Санкт-Петербургского государственного университета
Тел.: (812) 428-42-12
Эл. почта: OViRusakov@yahoo.co.uk

Михаил Борисович Ласкин,
к.ф-мат.н., доцент, директор ООО «Инвест-Строй»
Тел.: (812) 315-50-92
Эл. почта: laskinmb@yahoo.com

Ольга Ильинична Джаксумбаева,
к.э.н., Санкт-Петербургский государственный университет, Экономический факультет
Тел.: (812) 273-02-76
Эл. почта: olgadzh@rambler.ru

Мы строим стохастическую модель ценообразования на рынке вторичной недвижимости. Метод построения модели основывается на последовательных сравнениях цен. Мы доказываем, что при стандартных условиях на коэффициенты сравнения единственный невырожденный предел по распределению для цены имеет закон логнормального распределения вероятностей. Соответствие эмпирических распределений цен теоретически доказанному логнормальному закону мы подтверждаем на многочисленных данных статистики по недвижимости Санкт-Петербурга. При этом мы в основном используем мощный и чувствительный статистический критерий согласия Колмогорова-Смирнова. Опираясь на «Федеральный стандарт оценки №2», мы делаем вывод, что рекомендуемая наиболее вероятная цена – мода распределения корректно и однозначно определяется при аппроксимации логнормальным законом. Так как у логнормального распределения среднее превосходит наиболее вероятное значение – моду, то цены, вычисленные по математическому ожиданию, систематически завышены.

Ключевые слова: *рыночная стоимость недвижимости, стохастическая модель ценообразования, геометрическое броуновское движение, мода логарифмически нормального закона распределения, применение критерия согласия Колмогорова-Смирнова.*

Введение

При определении рыночной стоимости объектов недвижимости методом сравнительного подхода возникают следующие проблемы как теоретического, так и практического характера: выбора объектов сравнения, формирования достаточно представительной выборки, построения оценки рыночной стоимости на основе статистических данных.

Однако, в практике оценивания, как правило, выборки не являются представительными, содержат малое количество объектов сравнения, а проверка статистических гипотез, как справедливо отмечалось в [1], часто не производится.

В Федеральном Стандарте Оценки № 2 [2] в п.6 указано следующее положение: «При определении рыночной стоимости определяется наиболее вероятная цена, по которой объект оценки может быть отчужден на дату оценки на открытом рынке в условиях конкуренции, когда стороны сделки действуют разумно, располагая всей необходимой информацией, а на величине сделки не отражаются какие-либо чрезвычайные обстоятельства».

Аналогичные формулировки, в основе которых фигурирует «наиболее вероятная цена», содержат и многие признанные зарубежные стандарты оценки, такие как: IVS (p.30 a) [3], TEGOVA (EVS p. 5.3.1) [4], USPAP (Standard rule 6-2, p. c.) [5], RICS (p. 3.2.1) [6].

С математической точки зрения это означает, что определение рыночной стоимости, заданное стандартами, опирается на факт вероятностной природы цены: цена – величина случайная. При этом сама рыночная стоимость является числовой характеристикой распределения случайной величины (цены), её *модой* и показывает наиболее вероятное значение случайной величины.

Абсолютное большинство используемых в современной практике методов оценивания рыночной стоимости при статистической обработке имеющегося материала ориентированы на нормальное распределение, что плохо согласуется с эмпирическими распределениями цен. Так, в частности, одной из особенностей нормального распределения являются его симметричность, а значит, в частности, и равенство характеристик: моды, математического ожидания и медианы. Между тем, цены на однородные активы никогда не распределены нормально. Это следует хотя бы из того, что цены положительны, а нормально распределенная случайная величина со строго положительной вероятностью принимает отрицательные значения. Более того, распределение цен в большинстве случаев асимметрично влево.

В настоящей статье, при достаточно простых и естественных предположениях, устанавливается теоретический факт слабой сходимости (сходимости по распределению: в одномерном случае – слабая сходимость равносильна сходимости по распределению; однако, в данной статье в некоторых местах мы выходим за рамки одномерного случая, например, – в ключевой Лемме 1) процесса последовательных сравнений цен к логарифмически нормальному распределению. Получены статистические результаты, подтверждающие данный факт. Ряд ключевых примеров – приводится. Устанавливается ряд особенностей рыночной стоимости, которые вытекают из логарифмически нормального распределения цен. Рассмотрены примеры, основанные на статистическом материале по ценам предложений вторичного рынка жилой недвижимости в г. Санкт-Петербурге (СПб) по состоянию на 31.03.2014 г.

Oleg V. Rusakov,
PhD in physic and mathematic, Associate professor, Theory Probability and Mathematical Statistics department, Saint-Petersburg State University
Tel.: (812) 428-42-12
E-mail: OViRusakov@yahoo.co.uk

Michael B. Laskin,
PhD in physic and mathematic, Associate-professor, Director OOO "Invest-Stroy", Saint-Petersburg
Tel.: (812) 315-50-92
E-mail laskinmb@yahoo.com

Olga I. Jaksumbaeva,
PhD in economics, Economics department, Saint-Petersburg State University
Tel.: (812) 273-02-76
E-mail: olgadzh@rambler.ru

STOCHASTIC PRICING MODEL FOR THE REAL ESTATE MARKET: FORMATION OF LOG-NORMAL GENERAL POPULATION

We construct a stochastic model of real estate pricing. The method of the pricing construction is based on a sequential comparison of the supply prices. We proof that under standard assumptions imposed upon the comparison coefficients there exists an unique non-degenerated limit in distribution and this limit has the log-normal law of distribution. The accordance of empirical distributions of prices to the theoretically obtained log-normal distribution we verify by numerous statistical data of real estate prices from Saint-Petersburg (Russia). For establishing this accordance we essentially apply the efficient and sensitive test of fit of Kolmogorov-Smirnov. Basing on "The Russian Federal Estimation Standard N2", we conclude that the most probable price, i.e. mode of distribution, is correctly and uniquely defined under the log-normal approximation. Since the mean value of log-normal distribution exceeds the mode – most probable value, it follows that the prices valued by the mathematical expectation are systematically overstated.

Keywords: *real estate market value, stochastic model of pricing, geometric Brownian motion, mode of the logarithmically normal law distribution, applications of the Kolmogorov-Smirnov test of fit.*

1. Стохастическая модель ценообразования

Мы строим стохастическую модель ценообразования на рынке жилой недвижимости, основанную на последовательном сравнении цен объектов недвижимости. Коэффициенты сравнения мультипликативны, малы, случайны, имеют простое распределение (элементарны и однородны в определенном смысле). Мы показываем, что при достаточно большом числе сравнений N (начиная с 500, примерно, для ошибки аппроксимации порядка 5%, т.к. таковая ошибка обратно пропорциональна корню из N) при стандартных предположениях о малости и однородности коэффициентов сравнения с высокой точностью формируется логнормальное распределение цены (имеется в виду удельная цена – за 1 кв. м.). Мы полагаем, что выполнены следующие простые и стандартные предположения:

а) об однородности коэффициентов сравнения то есть величин, показывающих во сколько раз цена следующего (ближайшего) объекта сравнения больше или меньше цены предыдущего;

б) о независимости от предыдущих шагов сравнения увеличения/уменьшения цены. При выполнении данных предположений мы получаем приближение для распределения удельной цены объекта недвижимости к распределению вероятностей *логарифмически нормального закона.* Полученное приближение следует понимать в смысле предела по распределению (слабой сходимости) цен при последовательном случайном сравнении, когда число объектов сравнения $N \to \infty$.

Определение 1. *Мы говорим, что случайная величина X имеет логнормальное распределение (lognormal distribution), если ее логарифм нормально распределен, то есть она допускает следующее представление (в виде равенства по распределению)*

$$X \stackrel{d}{=} \exp\{\mu + \sigma\, \mathrm{N}(0,1)\}, \qquad (1)$$

здесь и далее:

вещественный параметр μ – математическое ожидание случайной величины $\ln X$; положительный параметр σ^2 – волатильность (volatility) – дисперсия распределения $\ln X$, σ – соответствующее стандартное отклонение; и в то же время: $\exp\{\mu\}$ *– медиана распределения величины X;* $\exp\{\mu + \sigma^2/2\}$ *– математическое ожидание величины X;* $\exp\{\mu - \sigma^2\}$ *– мода распределения величины X.*

Обозначение $N(\mu, \sigma^2)$ мы используем как для закона нормального распределения со средним μ, дисперсией σ^2, так и для соответствующей случайной величины. Значок $\stackrel{d}{=}$ обозначает равенство по распределению.

Введем неслучайную величину $S_0 > 0$, которой мы придаем смысл цены (цена за один квадратный метр) начального объекта сравнения. Запишем случайную величину X в виде

$$X = S_0 \exp\{(\mu - \ln S_0) + \sigma N(0, 1)\}. \qquad (2)$$

Введем параметры «отношение «сигнал/шум», которые также будем называть *относительным риском.* По сути – это принятые модификации известного *коэффициента Шарпа (Sharpe parameters)* [7].

$$R = R(\mu, S_0, \sigma) \stackrel{\Delta}{=} \frac{\mu - \ln S_0}{\sigma},$$

$$R_2 = R_2(\mu, S_0, \sigma^2) \stackrel{\Delta}{=} \frac{\mu - \ln S_0}{\sigma^2}. \qquad (3)$$

Величину $\mu - S_0$ мы интерпретируем, как полезный сигнал.

2. Процесс последовательных сравнений и его пределы по распределению

Процесс последовательных сравнений реализуется по следующей схеме. Выбирается некоторый начальный объект сравнения с ценой S_0. Выбор предполагается возможно случайным, а если неслучайным, то не из крайних значений спектра цен совокупности объектов сравнения. Переход к следующему объекту сравнения предполагается таким, что цена может увеличиться или уменьшится незначительно. Мы рассматриваем бинарную схему, лежащую в основе последовательности испытаний Бернулли. Цена следующего объекта сравнения определяется по формуле

$$S_1 \overset{\Delta}{=} S_0 Y_1, \qquad (4)$$

где Y_1 – случайная величина, принимающая два значения $U > 1$ (Up) и $0 < D < 1$ (Down) с соответствующими вероятностями $0 < p < 1$ и $q = 1 - p$. При этом предполагается, что величины U и D незначительно отличаются от единицы (в большую и меньшую стороны, соответственно). Переход к следующим объектам происходит по рекуррентному правилу, причем каждый раз умножение происходит на независимый коэффициент, и все коэффициенты одинаково распределены. Для того, чтобы доказать логнормальность предельного распределения данной рекуррентной процедуры, мы будем доказывать нормальность предела сумм логарифмов независимых одинаково распределенных коэффициентов сравнения (Y). Вследствие свойства непрерывности экспоненты и ограниченности траекторий броуновского движения на замкнутом интервале [8] будет достаточно этих оснований для установления необходимого факта логнормальности предельного закона распределения цен. Важно отметить, что вероятности p и q зависят от выбора начальной цены сравнения S_0. Мы изначально предполагаем, что существует некоторый спектр (интервал) цен объектов сравнения, который зависит от количества объектов N (вовлеченных в процесс последовательных сравнений). Число объектов, вовлеченных в процесс сравнения, должно быть соизмеримо с числом объектов всей существующей, или предполагаемой совокупности и они должны быть распределены однородно по совокупности.

В случае, когда цена S_0 находится в начале спектра цен объектов сравнения, то вероятность взять следующий объект более дорогим (то есть вероятность p события $\{S_1 > S_0\}$) превосходит вероятность q взять следующий объект более дешевым (события $\{S_1 < S_0\}$). При этом – чем ниже (левее в спектре) цена S_0, тем больше вероятность p превосходит вероятность q. Таким образом, мы предполагаем зависимость $p = p(S_0)$, $q = q(S_0)$. Также заметим, что в случае, когда выбор начальной цены сравнения S_0 случаен, возникает ее зависимость от количества объектов N, и, следовательно, появляется зависимость вероятностей p и q от N. Следующее предположение для упрощения построения конструкции процесса сравнения – это «логарифмически симметричные» коэффициенты сравнения, то есть

$$D \overset{\Delta}{=} \frac{1}{U}. \qquad (5)$$

Для дальнейших построений будем опираться на следующую стохастическую лемму «общего характера» [9]. Здесь эта лемма приводится в более слабом виде, чем в первоисточнике, но в достаточном для поставленных в настоящей работе целей. Приводится также и необходимое доказательство.

Используем следующие стандартные обозначения математического анализа: при $n \to \infty$ функция $\overset{=}{o}$ – «о малое», функция $\underset{=}{O}$ – «О большое».

Пусть случайные величины ξ, $\{\xi_j\}$, $j \in \mathbf{N}$, независимы, одинаково распределены и имеют распределение, зависящее от некоторого натурального n, произвольного, но фиксированного. Случайная величина ξ вводится с целью задания закона распределения.

Дальнейшая интерпретация – следующая: случайные величины вида (ξ) представляют собой соответствующие центрированные (то есть, после вычитания соответствующих математических ожиданий) логарифмы коэффициентов сравнения вида $\ln Y$, и этих коэффициентов сравнения рассматривается n штук. Константа C_n, введенная в следующей ниже формуле (6) – есть сумма математических ожиданий логарифмов (накопленное математическое ожидание логарифмов) коэффициентов сравнения вида $\ln Y$.

По последовательности $\xi_1, \xi_2, \ldots, \xi_n$ построим непрерывную случайную ломаную следующего вида:

$$g_n(t) \overset{\Delta}{=} \sum_{j=1}^{[nt]} \xi_j + \left(nt - [nt]\right)\xi_{[nt]+1} + tC_n, \qquad (6)$$

где $t \in [0, 1]$, а C_n – некоторая константа, зависящая только от n. Разумеется, что существует взаимно-однозначное соответствие между ломаными (g_n) и последовательностями $\xi_1, \xi_2, \ldots, \xi_n$. Это соответствие задается формулой

$$\xi_{k+1} = g_n\left(\frac{k+1}{n}\right) - g_n\left(\frac{k}{n}\right) - C_n\frac{1}{n}, \quad k = 0, 1, \ldots, n-1. \qquad (7)$$

Лемма 1. *Пусть для всех натуральных n и для некоторого $\delta > 0$ конечно математическое ожидание $E\left|\xi_1(n)\right|^{2+\delta} < \infty$. Пусть при $n \to \infty$ случайные величины $\xi_j \equiv \xi_j(n)$, $j = 1, \ldots, n$; и константы C_n, $\sigma_n > 0$ удовлетворяют следующим условиям:*

(*i*) $E\xi_1(n) = 0$;

(*ii*) $E\xi_1^2(n) = \dfrac{\sigma_n^2}{n} = \dfrac{\sigma^2}{n}(1 + \overset{=}{o}(1))$;

(*iii*) $E\left|\xi_1(n)\right|^r = \underset{=}{O}(n^{-r/2})$, $r > 0$;

(*iv*) $C_n = C(1 + \overset{=}{o}(1))$,

где C и $\sigma > 0$ – некоторые постоянные, не зависящие от n. Тогда конечномерные распределения случайных ломаных $g_n(t)$ при $n \to \infty$ слабо сходятся к конечномерным распределениям случайного процесса

$$\sigma W_t + Ct,\ t \in [0, 1], \qquad (8)$$

где W_t – стандартное броуновское движение. В пункте (iii) не предполагается существования момента произвольного порядка. Следует понимать так, что для всякого $r > 1$, такого что $E|\xi_1(n)|^r < \infty$ имеет место условие (iii). Естественно, здесь подойдут все $r \le 2 + \delta$.

О слабой сходимости более подробно смотри напр.: [10] для случайных величин, [8] для векторов и функций.

Доказательство Леммы 1. Обозначим $f(x)$ характеристическую функцию (преобразование Фурье меры) для распределения случайной величины ξ. Прежде всего заметим, что в силу условий (*i*) – (*iii*) при всяком фиксированном x справедливо разложение

$$\ln f(x) = -\frac{1}{2}\frac{\sigma^2}{n}x^2 + \overset{=}{o}\left(\frac{1}{n}\right), \quad n \to \infty.$$

Отметим, что подобного рода разложение используется, как правило, при доказательстве методом характеристических функций центральной предельной теоремы в форме Леви (см. [10]). Рассмотрим некоторое произвольное (не обязательно равномерное) разбиение временного промежутка [0, 1], заданное посредством точек $0 = t_0 < \cdots < t_m \le 1$. Вычислим распределение случайного вектора, составленного из значений в этих точках случайной ломаной g_n:

$$\varphi \equiv \varphi_n(m) \stackrel{\Delta}{=}$$
$$\stackrel{\Delta}{=} \left(g_n\left(t_1\right), g_n\left(t_2\right), \ldots, g_n\left(t_m\right)\right), \quad (10)$$

где $g_n(t)$ определены посредством равенства (6). В дальнейшем, при доказательстве настоящей леммы, опустим индекс n при обозначении случайной функции (случайной ломаной) $g_n(t) \equiv g(t)$. Обозначим через t_j^* ближайший слева к точке t_j узел равномерного разбиения на n частей интервала времени [0, 1]:

$$t_j^* \stackrel{\Delta}{=} \left[n t_j\right] \frac{1}{n} = t_j - \left\{n t_j\right\} \frac{1}{n}, \quad (11)$$

здесь $\{a\}$ обозначает дробную часть числа a. Из (6) нетрудно вывести равенство

$$g(t_j) - g(t_j^*) =$$
$$= \left(nt_j - \left[nt_j\right]\right)\left(\xi_{\left[nt_j\right]+1} + C_n \frac{1}{n}\right). \quad (12)$$

Так как первый сомножитель в правой части этого равенства равномерно по n ограничен единицей, а второй сомножитель имеет стремящиеся к нулю (при $n \to \infty$) дисперсию и среднее, то использование классического неравенства Чебышева легко устанавливает следующую сходимость по вероятности

$$\left|g_n(t_j) - g_n\left(t_j^*\right)\right| \xrightarrow{P} 0,$$
$$n \to \infty. \quad (13)$$

Таким образом, если у вектора φ существует предельное распределение, то оно совпадает с предельным распределением вектора

$$\varphi^* \equiv \varphi_n^*(m) \stackrel{\Delta}{=} \left(g(t_1^*), \ldots, g(t_m^*)\right) \stackrel{\Delta}{=}$$
$$\stackrel{\Delta}{=} \left(g\left(k_1 \frac{1}{n}\right), \ldots, g\left(k_m \frac{1}{n}\right)\right), \quad (14)$$

где $k_j = [nt_j]$, $j = 1, \ldots, m$. Рассмотрим случайный вектор ψ, составленный из приращений вектора φ^*:

$$\psi(n) \equiv \psi = (\psi_1, \ldots, \psi_m) \quad (15)$$

с компонентами

$$\psi_j \stackrel{\Delta}{=} g\left(k_j \frac{1}{n}\right) - g\left(k_{j-1} \frac{1}{n}\right). \quad (16)$$

Очевидно, что

$$\psi_j = \sum_{i=k_{j-1}+1}^{k_j} \xi_i + C_n\left(k_j - k_{j-1}\right)\frac{1}{n}. \quad (17)$$

Кроме того, $(\psi_1, \ldots, \psi_m)$ – независимые случайные величины, и даже более – каждая компонента вектора ψ состоит из суммы независимых случайных величин. Поэтому для всякого $j = 1, \ldots, m$ характеристическая функция f_j случайной величины ψ_j допускает представление

$$f_j(x; n) \equiv f_j(x) =$$
$$= \exp\left\{ixC_n(k_j - k_{j-1})\frac{1}{n}\right\} f^{k_j - k_{j-1}}(x),$$
$$x \in \mathbf{R}, \quad (18)$$

где $f(x)$ – характеристическая функция случайной величины ξ_1. Так как

$$\lim_{n\to\infty}\left(k_j - k_{j-1}\right)\frac{1}{n} = t_j - t_{j-1},$$
$$j = 1, \ldots, m \quad (19)$$

то при каждом фиксированном вещественном x для характеристической функции $f_j(x)$ при каждом $j = 1, \ldots, m$ справедливо предельное соотношение

$$\lim_{n\to\infty} f_j(x;n) =$$
$$= \exp\left\{ixC(t_j - t_{j-1})\right\} \times$$
$$\times \exp\left\{-\frac{\sigma^2}{2}(t_j - t_{j-1})x^2\right\}. \quad (20)$$

Здесь мы использовали предельное соотношение для последовательности $\{C_n\}$ из условия Леммы 1, а также представление для характеристической функции $f(x)$ случайной величины ξ_1. Таким образом, по теореме о непрерывном соответствии между характеристическими функциями и распределениями следует, что распределения случайного вектора ψ сходятся к m-мерному гауссовскому распределению с вектором средних

$$(t_1 - t_0, t_2 - t_1, \ldots, t_j - t_{j-1})C \quad (21)$$

и ковариационной матрицей

$$\delta_{ij}\sigma^2(t_j - t_{j-1}),\ i, j, \ldots, m, \quad (22)$$

где $\delta_{ij} = 0$ при $i \ne j$ и $\delta_{ij} = 1$ при $i = j$ – символ Кронекера.

Что же касается векторов φ и φ^*, то их распределения совпадают с точностью до бесконечно малого члена (при $n \to \infty$). В свою очередь, вектор φ^* получается из вектора ψ суммированием независимых компонент последнего, поэтому нетрудно вычислить общее предельное распределение векторов φ и φ^*:

гауссовское распределение с вектором средних $(t_1, \ldots, t_m)C$ и ковариационной матрицей $\sigma^2\min\{t_i, t_j\}$, $i, j, \ldots, m$.

Эти характеристики полностью соответствуют известным характеристикам конечномерных распределений винеровского процесса с коэффициентом масштаба σ и со сдвигом Ct. Таким образом, мы установили необходимый нам факт сходимости конечномерных распределений:

Конечномерные распределения процесса, заданного случайными ломаными $g_n(t)$, сходятся при $n \to \infty$ к конечномерным распределениям процесса

$$\sigma W_t + Ct,\ t \in [0, 1], \quad (23)$$

где W_t – стандартный винеровский процесс. Лемма доказана.

Теперь заметим, что распределение винеровского процесса в единице – стандартно нормально $N(0, 1)$. Таким образом, нам *достаточно* в формулировке леммы рассмотреть момент времени $t = 1$, когда количество наших объектов сравнения N совпадает с параметром n из предыдущей леммы. Тогда мы получим сходимость (слабую сходимость) распределений логарифмов цен к распределению нормального закона $N(C, \sigma^2)$, так как распределение случайной величины σW_1 + C нормально со средним C и дисперсией σ^2. Лемма 1 устанавливает факт сходимости по распределению к нормальному закону для логарифмов цен. Однако, нам нужна сходимость распределений собственно цен к логнормальному закону. Следующее Следствие 1 доказывает, что экспоненты от сумм (ξ_j) сходятся по распределению к логнормальному закону. Рассмотрим последовательность при натуральном n случайных величин $(g_n(1))$, которые являются концами случайных ломаных, что определены равенством (6).

Следствие 1 к Лемме 1. *Последовательность* $(\exp\{g_n(1)\})$ *сходится по распределению при* $n \to \infty$ *к логнормальной случайной величине* $\exp\{C + \sigma N(0, 1)\}$.

Доказательство Следствия 1. Введем обозначения $G_n = g_n(1)$, при всех $n \in \mathbf{N}$; $\Phi(u) = \int_{-\infty}^{u} \exp\left\{-\frac{1}{2}\upsilon^2\right\} d\upsilon$ – функция распределения стандартного нормального закона, $u \in \mathbf{R}$. Заметим, что сходимость по распределению равносильна сходимости функций распределений во всех точках непрерывности функции распределения предельного закона (см. [8]). При этом отметим, что функции распределения как нормального закона, так и логнормального, непрерывны на всей оси и на правой полуоси, соответственно. Из Леммы 1 имеем для любого вещественного a при $n \to \infty$, $P(G_n < a) \to \Phi(a)$.

Так как экспонента строго монотонно растущая функция и применение ее к обеим частям сохранит неравенство, а вероятность эквивалентных неравенств равна, то $P(G_n < a) = P(\exp(G_n) < \exp(a)) \to \Phi(a)$. Обозначим $x_0 = \exp(a)$.

Тогда при $n \to \infty$, $P(\exp(G_n) < x_0) \to \Phi(\ln x_0)$. При повторном использовании свойства экспоненты строго монотонно расти, легко видеть, что $\Phi(\ln x_0) = P(N(0,1) < \ln x_0) = P(\exp(N(0,1)) < x_0)$. В силу произвольности $a \in \mathbf{R}$ и взаимно-однозначности логарифма, как функции из правой полуоси в $\mathbf{R}$, окончательно имеем требуемое в Следствии 1 утверждение. Следствие 1 доказано.

Замечание к Лемме 1. *Параметр* $t \in [0, 1]$ *в формулировке Леммы естественно интерпретируется, как (непрерывное) время действия процесса сравнения. Этот параметр легко обобщается на случай* $t \in [0, T]$ *для произвольного* $T > 0$. *При вовлечении всё большего числа объектов сравнения в процесс сравнения при выполнении условия однородности мы получаем процесс Броуновского движения для логарифма цены с параметром: процессом линейно накапливаемой волатильности* $\sigma^2 t = D\{\sigma W_t\}$, $t \in [0, T]$. *В случае же неоднородности, зависимости волатильности от времени* $\sigma^2 = \sigma^2(t)$ *(но ограниченности в смысле среднего-квадратичного), и зависимости среднего от времени* $C = C(t)$ *мы получаем сходимость логарифмов цены к процессу*

$$\int_0^t C(\upsilon)d\upsilon + \int_0^t \sigma^2(u)dW_u, \quad 0 \le t \le T, \tag{24}$$

где dW_u *– стандартный гауссовский «белый шум». Процессом накопленной волатильности в этом случае является интеграл* $\int_0^t \sigma^2(u)du$.

3. Процесс последовательных сравнений и стохастическая модель ценообразования

Запишем схему сравнения по шагам

$$S_{j+1} \stackrel{\Delta}{=} S_j Y_{j+1}, \ j \le N, \tag{25}$$

– в рекуррентном виде;

$$S_l \stackrel{\Delta}{=} S_0 \prod_{j=1}^{l} Y_j, \ l \le N, \tag{26}$$

– в явном виде. Здесь $\{Y_j\}$, $j \le N$ – последовательность независимых одинаково распределенных случайных величин с распределением (задаваемым случайной величиной Y)

$$Y \stackrel{\Delta}{=} \begin{cases} U, \ p \\ D, \ q \end{cases} = \begin{cases} U, & p(S_0) \\ \dfrac{1}{U}, & q(S_0) \end{cases} \tag{27}$$

Для применения Леммы 1 перейдём к логарифмам

$$\xi'_j \stackrel{\Delta}{=} \ln Y_j, \ j \le N, \tag{28}$$

и, затем, проведём центрирование

$$\xi_j \stackrel{\Delta}{=} \xi'_j - E\xi'_j = \ln Y_j - E\ln Y_j, \quad j \le N. \tag{29}$$

Положим (в обозначениях Леммы 1) $n = N$. Положим первый и второй центральный момент равными следующим выражениям ($N \to \infty$)

$$\begin{cases} E\xi'_j = (\mu - \ln S_0)\dfrac{1}{N} + \overset{=}{o}(1) - C_j = \\ = C_j(N); \\ D\xi'_j = D\xi_j = \dfrac{\sigma^2}{N} + \overset{=}{o}(1) \end{cases} \tag{30}$$

(дисперсия от сдвига не зависит, а случайные величины ξ_j, $j \le N$, получены из «штрихованных» величин ξ'_j путем сдвига). Видим, что условия Леммы 1 выполнены; в формулировке Леммы 1 мы полагает $t = 1$ и утверждаем необходимый нам факт сходимости распределения логарифма цены к нормальному закону распределения при $N \to \infty$

$$\ln\left\{S_0 \prod_{j=1}^{N} Y_j\right\} = = \ln S_0 + \sum_{j=1}^{N} \xi_j + \sum_{j=1}^{N} C_j \Rightarrow \Rightarrow \sigma\, \mathrm{N}(0,1) + \mu \tag{31}$$

(Логарифмы $\ln S_0$ сократились!). Обозначение «$\Rightarrow$» мы полагаем общепринятым для сходимости по распределению, слабой сходимости распределений.

Замечание 1. *Отметим, что в предельном выражении (31) распределение коэффициента сравнения вида Y вовсе не обязано иметь бинарное распределение. Единственное требование – условия Леммы 1, связанные с независимостью и моментами. Поэтому мы можем говорить об инвариантности метода доказательства относительно выбора закона распределения коэффициентов сравнения.*

Замечание 2. *Вследствие факта сходимости в Лемме 1 накопленных сумм к процессу броуновского движения мы сможем получить сформированное логнормальное распределение и в случае, когда в процесс сравнения будет вовлечена подвыборка или увеличенная выборка объема* $[Nt]$, $0 < t < T < \infty$, $[\bullet]$ *– целая часть. Важно лишь, чтоб коэффициенты сравнения были независимыми и однородно малыми в определенном смысле. В таких случаях у нас лишь будет меняться дисперсия нормального распределения от логарифма в соответствии со свойством броуновского движения* $DW(t) = t$, $t \ge 0$. *В случае неоднородности у нас возникнет нетривиальный процесс накопленной волатильности (Смотри Замечание к Лемме 1.)*

Замечание 3. *Известно, что сходимость по распределению еще* *не влечет сходимость моментов (математического ожидания, дисперсии) и других характеристик распределений. Однако, характеристики будут сходиться при слабой сходимости распределений в случае, когда они равномерно ограничены. Данное условие в нашей модели автоматически выполняется, т.к. логарифмы всех цен ограничены одной константой, возможно, и весьма большой.*

4. Инвариантность распределения относительно выбора начального объекта сравнения и анализ коэффициентов

Рассмотрим вопрос зависимости вероятностей у коэффициентов сравнения от начального значения S_0. Применим известные элементарные формулы для дисперсии и математического ожидания бинарных случайных величин к случайным величинам вида ξ' с распределением, заданным (29), $j = 1, ..., N$,

$$E\xi_j' = p\ln U + d\ln D = (2p-1)\ln U, \qquad (32)$$

$$D\xi_j' = D\xi_j = (\ln U - \ln D)^2 pq = (2\ln U)^2 pq = 4p(1-p)(\ln U)^2 \qquad (33)$$

Приравниваем вычисленные математическое ожидание (32) и дисперсию (33) к выражениям в (30) и получаем, пренебрегая членами малого порядка, систему из двух уравнений (относительно неизвестных p и S_0)

$$\left.\begin{aligned}(2p-1)\ln U &= (\mu - \ln S_0)\frac{1}{N}\\ \sqrt{p(1-p)}\,2\ln U &= \frac{\sigma}{\sqrt{N}}\end{aligned}\right\}. \qquad (34)$$

Поделим в (34) первое уравнение на второе уравнение и получим следующее Утверждение 1, устанавливающее соотношение между вероятностью p и коэффициентом относительного риска R, определенного (3), (а, следовательно, – между p и S_0 при известных значениях параметров μ и σ^2).

Утверждение 1. Вероятность подъема цены p в процессе последовательных сравнений зависит только от коэффициента относительного риска и количеством объектов сравнения следующим образом:

$$\frac{1}{2}\frac{p-(1-p)}{\sqrt{p(1-p)}} = \frac{1}{\sqrt{N}}\frac{\mu - \ln S_0}{\sigma} = \frac{R}{\sqrt{N}}. \qquad (35)$$

Комментарий к Утверждению 1. *Уравнение (35) имеет в правой части член, убывающий к нулю со скорость порядка $\sqrt{N}$. Это является следствием того факта, что в следствие механизма работы Центральной предельной теоремы для последовательности испытаний Бернулли отклонение количества успехов от ожидаемого количества успехов попадает в интервал, пропорциональный $\sqrt{N}$. Другими словами, при случайном выборе логарифма начального значения $\ln S_0$ из N данных возможных значений мы с вероятностью, стремящейся к единице при $N \to \infty$, выберем случайное число $\ln S_0$, отличающееся от среднего μ на величину пропорциональную $\sqrt{N}$, когда логарифм коэффициента (одного) сравнения имеет порядок малости $1/\sqrt{N}$.*

Введем обозначение $l(p)$ для функции в левой части (35), заданной на интервале (0, 1) и затем рассмотрим поведение этой функции,

$$l(p) \overset{\Delta}{=} \frac{1}{2}\left(\sqrt{\frac{p}{1-p}} - \sqrt{\frac{1-p}{p}}\right). \qquad (36)$$

Нетрудно убедиться, что функция $l(p)$ непрерывная (даже гладкая – бесконечно-дифференцируема), строго монотонно возрастающая, стремится к минус бесконечности при $p \to 0$, равна нулю при $p = 1/2$ (то есть при симметричном случайном блуждании) и стремится к плюс бесконечности при $p \to 1$. Из этих свойств следует, что: *в уравнении* (36) *существует и единственное решение* $p \in (0, 1)$ *при всякой вещественной левой части.* Более того, при известных значениях параметров μ и σ в объеме выборки N мы по заданному значению S_0 можем вычислить $p = p(S_0)$, – и наоборот. Так как мы работаем с выборкой (цен на объекты недвижимости), имеем объем этой выборки N, то мы всегда можем вычислить выборочные среднее $\hat{\mu}$ и дисперсию $\hat{\sigma}^2$, подставить их в (35) вместо μ и σ, соответственно, и получить прямое (вычисляемое) соотношение между S_0 и $p = p(S_0)$. Простой анализ построенной сто-

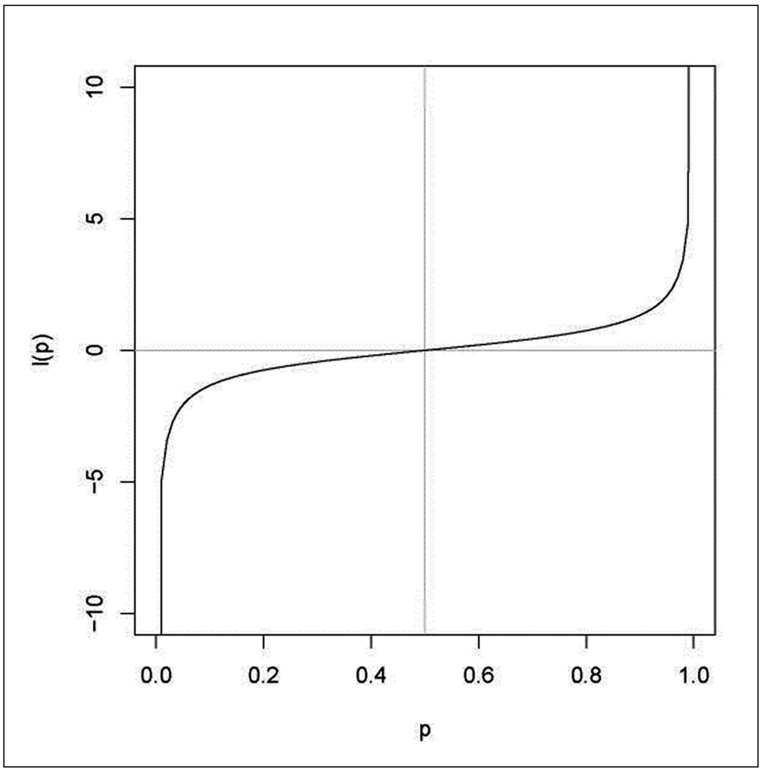

Рис. 1. График функции $l(p)$.

хастической модели и полученных формул позволяет сделать следующие важные практические выводы.

Вывод 1. *Так как выполняется равенство по распределению (1), то среднее логнормального закона распределения равно* $\exp\{\mu + \sigma^2 / 2\}$, *а мода равна* $\exp\{\mu - \sigma^2\}$, *и легко видеть, что отношение среднего к моде не зависит от* μ, *всегда больше единицы и равно* $\exp\{(3 / 2)\sigma^2\}$. *Разлагая экспоненту, получаем при малых* σ, *что это отношение приблизительно равно* $1 + (3 / 2)\sigma^2$. *Это означает, что применение средних арифметических для оценки рыночной стоимости (в качестве рекомендуемого наиболее вероятного значения – моды) приводит к систематической ошибке в сторону завышения, выраженную в относительной погрешности* $\exp\{(3 / 2)\sigma^2\} - 1(\times 100\%)$. *Например, при* $\sigma = 0{,}258$ *систематическая ошибка составляет 10,5%.*

Вывод 2. *Иногда в практике оценки вместо средних арифметических (оценка математического ожидания) применяют средние геометрические (оценка медианы). Так как медиана логнормального закона равна* $\exp\{\mu\}$, *то отношение медианы к моде не зависит от* μ, *всегда больше единицы и равно* $\exp\{\sigma^2\}$. *Разлагая экспоненту, получаем при малых* σ, *что это значение приблизительно равно* $1 + \sigma^2$. Это *означает, что применение средних геометрических для оценки рыночной стоимости (наиболее вероятного значения – моды) приводит к систематической ошибке в сторону превышения, выраженную в относительной погрешности* $\exp\{\sigma^2\} - 1(\times 100\%)$. *В частности, при* $\sigma = 0{,}258$ *систематическая ошибка составляет 6,9%.*

Вывод 3. *Из свойств логнормального распределения следует, что при постоянном* μ, *с ростом* σ^2 *увеличивается разница между математическим ожиданием, медианой и модой. Таким образом, при постоянном* μ *и растущей* σ^2 *среднее арифметическое растет, а рыночная стоимость (мода – наиболее вероятное значение) уменьшается.*

Итак, некоторый начальный спектр (интервал) цен достаточно общего вида (такого, чтобы величины p и $1 - p$, зависящие от N, не вырождались с ростом N) определяет параметры μ и σ формирующегося с помощью последовательных сравнений логнормального закона цены. Мы полагаем, что участники рынка действуют разумно и выбирают объекты сравнения, руководствуясь определенными принципами однородности – объекты имеют примерно одинаковый набор ценообразующих факторов. Тогда для разных групп объектов, обладающих одинаковой однородностью, будут формироваться логнормальные законы распределения с различными параметрами μ и σ. Такая модель автоматически приведет к смеси логнормальных распределений, что мы кратко и обсудим в следующем разделе.

5. Множество выборок объектов сравнения – формирование смеси распределений

Из обычной практики следует, что объекты сравнения далеко не всегда однородны. В этом случае элементы для сравнения мы берем из выборок, сформированных по тому или иному принципу однородности. Для каждой такой выборки формируется в соответствии с изложенным выше процессом сравнения свое логнормальное распределение, со своими параметрами μ и σ^2. При этом совокупное распределение получается путем смешивания (о смеси распределений см. [11]) полученных логнормальных распределений, когда смешивающие коэффициенты (веса) получаются пропорциональными количеству единиц сравнения в выборке, сформированной по данному принципу однородности. Отметим, что может быть принят альтернативный принцип формирования смешивающих коэффициентов – пропорционально совокупной денежной массе по каждой зоне ценовых предпочтений (по каждому району). Случай, когда веса пропорциональны денежной массе согласуется с тем, что логнормальное распределение получается для цены – денежной единицы.

Определение смеси. *Дадим определение для конечной смеси, то есть, когда смешивается конечное число распределений. Пусть имеется* m *функций распределения* $F_1, \ldots, F_m$ *– «элементов смеси», и* m *неотрицательных чисел, называемых «весами»,* $\alpha_1, \ldots, \alpha_m$, *сумма которых равна единице. Тогда функция* H, *определенная посредством равенства*

$$H(x) = \alpha_1 F_1(x) + \ldots + \alpha_m F_m(x), \quad x \in \mathbf{R} \tag{37}$$

является функцией распределения, которое называется смесью распределений $F_1, \ldots, F_m$ *с соответствующими весами* $\alpha_1, \ldots, \alpha_m$.

Необходимо отметить, что если у распределений существуют плотности $f_1, \ldots, f_m$, *то и у распределения* H *существует плотность* h, *равная*

$$h(x) = \alpha_1 f_1(x) + \ldots + \alpha_m f_m(x), \quad x \in \mathbf{R} \tag{38}$$

Обращаем внимание читателя на то, что определение смеси исходно никак не связано со случайными величинами (можно было бы построить аналог (37) для любых функций), но в прикладном смысле распределения без случайных величин практически не рассматриваются. Введем случайные величины $X_1, \ldots, X_m$, имеющие соответствующие функции распределения $F_1, \ldots, F_m$, и, так называемую, «смешивающую» случайную величину v, принимающую значения $1, \ldots, m$ с соответствующими вероятностями $\alpha_1, \ldots, \alpha_m$ (а числа α по построению естественно трактуются, как вероятности). Важно, чтоб случайная величина v не зависела от $X_1, \ldots, X_m$, при этом случайные величины $(X_1, \ldots, X_m)$ между собой могут как угодно зависеть. Тогда смесь реализуется по следующему алгоритму: сначала реализуем v, а затем реализуем X_v с тем номером, на который указала v. Полученную в результате такого алгоритма случайную величину обозначим Y. С помощью формулы полной вероятности легко видеть, что функция распределения для Y есть H. Отметим, что моменты распределений «смешиваются» при формировании смеси, то есть

$$EY^k = \alpha_1 EX_1^k + \ldots + \alpha_m EX_m^k \tag{39}$$

для любого натурального k. Также заметим, что ничего принципиального не изменится, если рассматривать счетный набор весов. Говоря языком чистой математики, смесь распределений – это выпуклая аффинная комбинация распределений.

Для дальнейших применений нам потребуется следующий факт, который, образно говоря, заключается в том, что «смесь нормальных распределений после потенцирования превращается в смесь логнормальных распределений» с теми же весами.

Утверждение 2:

***A**. Пусть функция распределения H случайной величины Y задается в виде, соответствующем (37), причем функции распределения $F_1, \ldots, F_m$ являются функциями распределения нормального закона с соответствующими параметрами $N(\mu_j, \sigma_j^2)$, $j = 1, \ldots, m$. Тогда случайная величина $\exp(Y)$ имеет функцию распределения вида $\alpha_1 G_1 + \ldots + \alpha_m G_m$, где G_j – функция распределения логнормального закона с параметрами: μ_j – медиана, σ_j^2 – волатильность; $j = 1, \ldots, m$.*

***B**. Обратно, пусть распределение случайной величины Z представимо в виде смеси логнормальных распределений с соответствующими параметрами: μ_j – медиана, σ_j^2 – волатильность; $j = 1, \ldots, m$. Тогда распределение случайной величины $Y = \ln(Z)$ является смесью с теми же весами нормальных распределений $N(\mu_j, \sigma_j^2)$, $j = 1, \ldots, m$.*

Доказательство Утверждения 2. Докажем сначала первую часть **A** Утверждения 2. Имеем для любого положительного *y,* используя определение функции распределения для случайных величин,

$$P(\exp(Y) < y) = P(Y < \ln(y)) =$$
$$= \sum_j \alpha_j P(X_j < \ln(y)) = .$$
$$= \sum_j \alpha_j P(e^{X_j} < y)$$

Мы использовали тот факт, что при применении к обеим частям неравенства монотонно растущей функции – неравенство не меняется. Логарифм и экспонента очевидно таковые. Часть **B** доказывается аналогично при эксплуатации тех же свойств логарифма и экспоненты. Утверждение 2 доказано.

Задача выделения из смеси распределений ее компонент относится к классу некорректных задач. Для решения таких задач часто используются факты и приемы, относящиеся к предметной области. В нашем случае – это разбиение общей выборки на естественные зоны, кластеры, районы и т.п.

Доказанное Утверждение 2. позволяет нам переходить от единичного нормального распределения к смеси нормальных распределений, и, в конечном итоге, к смеси логнормальных распределений, используя универсальную Лемму 1 и метод последовательных сравнений для каждой зоны, кластера и т.п.

Примеры в Приложении 2 иллюстрируют тот факт, что в зонах ценовых предпочтений (приведены только некоторые из 12 зон по СПб), в кластерах «тип дома» и даже в районах наблюдается хорошее соответствие логнормальному закону, как элементу общей смеси по СПб логнормальных распределений.

Важно заметить, что даже несмотря на тот факт, что смесь логнормальных распределений не является логнормальным распределением, в случае, когда множество совокупных значений параметров μ и σ^2 (взятых по всем зонам СПб) имеет небольшой разброс, компактно, полученное в результате смешивания [теоретическое] распределение может остаться унимодальным (то есть иметь одну моду – одно значение с максимальной вероятностью). Этот факт имеет свое подтверждение на реальных данных, см. Приложение 1, Рис. 2.

В любом случае, когда значения μ и σ^2 в смешивающих распределениях имеют небольшой разброс, и смешивающие распределения имеют веса примерно одинакового порядка, то итоговое распределение имеет, по меньшей мере, хорошо выделяемый, узкий интервал «максимальной вероятности». Данный тезис находит свою иллюстрацию даже в статистике, взятой по всему городу СПб (см. Приложение 1, Рис. 7).

6. Критерий согласия, статистическая оценка и калибровка параметров логнормального распределения цены

Все приведенные в настоящей статье диаграммы и аппроксимации получены с помощью статистического пакета ***R***. Модельным параметрам μ и σ^2 присваиваются значения оценочных. Оцениваются (стандартным методом вычисления выборочного среднего и среднеквадратичного отклонения (ско)) параметры среднего и масштаба для логарифмов цены. Модельные параметры μ и σ^2 мы полагаем равными тем значениям, что получены в результате оценивания.

Оценки параметров μ и σ^2 ($\hat{\mu}$ и $\hat{\sigma}^2$, соответственно) могут проводиться отдельно, но могут и быть взяты из работы тестов, представленных ниже. Статистические выводы начинаются с оценки согласия [*fit*] и качества согласия с логнормальным распределением цены по отдельно взятой выборке. Проверяется согласие с логнормальным распределением универсальным методом, предлагаемым пакетом ***R***, – функция *fitdistr* пакета ***R*** [12]. Алгоритм работы функции *fitdistr* основан на применении классического метода максимального правдоподобия для оценки параметра предлагаемого распределения (предлагается логнормальное распределение) с использованием меры «качества», основанной на среднеквадратичной метрике. Очевидно, что среднеквадратичная метрика чересчур чувствительна к «выбросам» и мало чувствительна к тяжести хвостов распределения, поэтому представляется оправданной возможность дальнейшей «корректировки» параметров, в том числе и визуальным способом. Далее проводится визуальный анализ полученной теоретической плотности [соответствующего выборке логнормального закона] с эмпирической плотностью, представленной в виде гистограммы. На этом шаге параметры μ и σ^2 калибруются путем «подгонки», с тем, чтобы добиться максимального визуального соответствия с эмпирической плотностью. Относительное изменение происходит с порядком не более 10%. Пусть полученные после подгонки параметры обозначаются μ_* и σ_*^2, соответственно. Для установления факта значимого (или не значимого) статистического соответствия анализируемой выборки закону логнормального распределения используется мощный критерий согласия Колмогорова-Смирнова (КС-тест) (см., [13], [14], [15], [16], [17]). При удовлетворении наперед заданному значению *p-value*на основе КС-теста принимается гипотеза о логнормальности с параметрами μ_* и σ_*^2 для распределения цены случайно взятого объекта недвижимости. После того, как распределение прошло КС-тест наиболее вероятной ценой берется мода (как и предписывается, собственно, самим определением понятия «моды») логнормального распределения с параметрами μ_* и σ_*^2, то есть

$$\exp(\mu_* \text{ и } \sigma_*^2). \qquad (40)$$

Для согласия выборки с законом распределения смеси логнормальных распределений применялись функции пакета *mixtools* в рамках пакета ***R***. С помощью этих функций формально выражалась смесь, построенная на основе разделения (разбиения) на районы, зоны ценовых предпочтений и т.п. Затем формировалась гипотеза о логнормальности элементов построенной смеси и пакетом *mixtools* оценивались параметры логнормальных распределений – элементов смеси. После этого исследуемая выборка проходила КС-тест на согласие с построенной смесью. Обработанные этой функцией данные показали (см. Приложение 1, Рис. 2), что выборка «масс-маркет» по СПб дает хорошо обусловленную моду для смеси, когда элементы смеси – районы города. Более того, все районы СПб, включая премиум сегмент, дают эмпирическое распределение, вполне удовлетворительно согласующееся посредством применения КС-теста со смесью логнормальных распределений. Представляется перспективным использовать в дальнейшем пакет *fitdistrplus* в рамках пакета ***R***. В пакете *fitdistrplus* есть опции, позволяющие улучшать оценки параметров, в частности, снижать чувствительность к «выбросам», лучше обрабатывать тяжелые хвосты и прочее.

Приложение 1

Рассмотрим ряд примеров. Обработан статистический материал – объявления о продажах объектов вторичной жилой недвижимости, опубликованных в [18]. Из выборки исключены некорректные объявления: нет цены или площади, другие ошибки. Рассматривается следующий масштаб распределения цен: тыс. руб. за 1 кв. м.

Разделение смеси по административным районам города. Масс-маркет

Пример статистического анализа выборки: «масс-маркет» Санкт-Петербурга. На приведенном ниже рис. 2 показаны гистограмма, кривая плотности логарифмически нормального распределения, – при оценивании «одиночным» логнормальным распределением, и кривая смеси при оценивании смесью логнормальных распределений. Каждый элемент смеси представляет из себя «масс-маркет» одного района города. Представлены все 18 районов СПб. Объем всей выборки: 12772 объявления о продаже. Веса смеси пропорциональны количествам объявлений. Легенда для диаграммы: пунктир – логнормальное распределение, сплошная линия – смесь логнормальных распределений. Вертикальной линией обозначена мода смеси; вертикальным пунктиром обозначена мода одного логнормального распределения. Полученные результаты: наиболее вероятные значения оказались равными 88,0669 тыс. руб. за кв. м. и 88,2176 тыс. руб. за кв. м., соответственно для единичного логнормального распределения и для смеси. Как видно, разница весьма незначительная, и при этом, распределение, полученное в результате смеси, сохраняет свойство унимодальности. Полученное в результате КС-теста, примененного к данной смеси, значение $p - value = 1{,}258e^{-6}$.

Необходимо отметить, что несмотря на сравнительно небольшую величину σ значение *p-value* мало (специфичной особенностью КС-теста при больших выборках является чувствительность к небольшим компонентам смести, что чрезмерно снижает значение *p-value*), что свидетельствует о крайне малой значимости модели одного логнормального распределения для решения задачи аппроксимации распределения данной выборки: $p - value < 1{,}517e^{-12}$, – на несколько порядков меньше, чем при использовании значимости для всего города СПб одного логнормального распределения (Для модели смеси), – и это является важным статистическим аргументом для разбиения города на подвыборки, формирующие отдельно взятые логнормальные распределения.

Следующий пример показывает, что выделение некоторых отдельно взятых административных районов города СПб дает хорошо обусловленное с точки зрения *p-value* КС-теста логнормальное распределение.

Разделение смеси по административным районам города

Разделение смеси по административным районам города показывает,

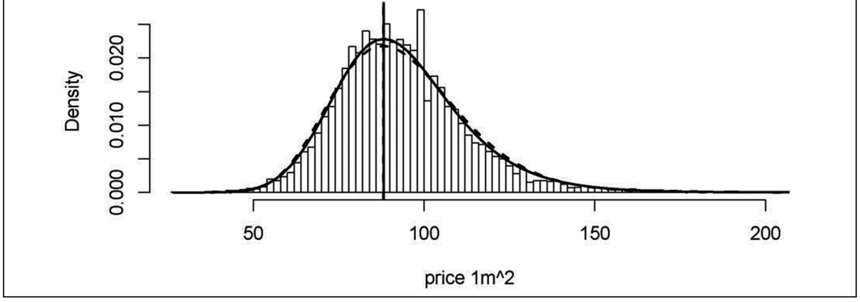

Рис. 2. Масс-маркет. Смесь

что большинство районов города (17 из 18) являются районами застройки, достаточно однородной, чтобы распределение цены 1 кв. м. в них могло быть представлено логарифмически нормальным распределением. Центральный район недостаточно хорошо аппроксимируется логнормальным распределением, но легко может быть разделен на зоны, аппроксимируемые логнормальным распределением. На приведенном ниже рис. 3 показаны гистограмма и кривая плотности логарифмически нормального распределения для Адмиралтейского района г. Санкт-Петербурга (за исключением премиум-сегмента – объектов на Конногвардейском бульваре).

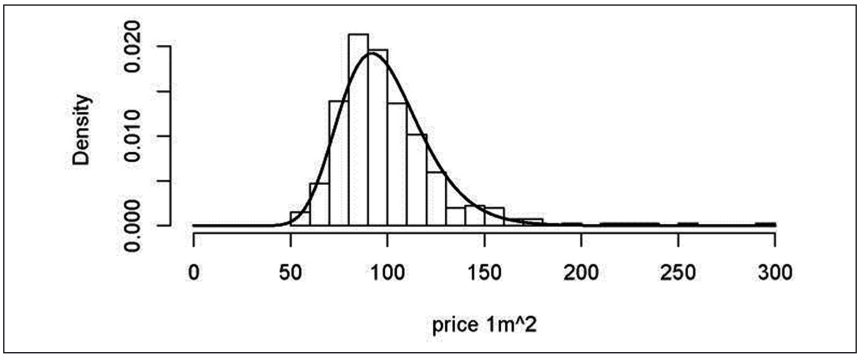

Рис. 3. Адмиралтейский район СПб

Выборка цен предложений вторичной жилой недвижимости в Адмиралтейском районе г. Санкт-Петербурга на 31.03.2014 г. и ее аппроксимация логнормальным распределением. Объем выборки: $N = 403$. Параметры логнормального закона: $\mu = 4{,}57$, $\sigma = 0{,}226$. Среднее арифметическое по выборке равно 99,799 тыс. руб. за кв. м. Оценка рыночной стоимости 1 кв. м., мода по логнормальной аппроксимации равна 91,983 тыс. руб. за кв. м. Результат КС-теста: *p-value* = 0,3245

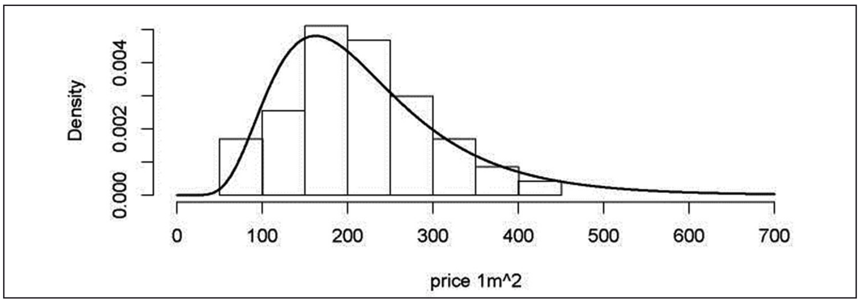

Рис. 4. Зона ценовых предпочтений «Золотой треугольник»

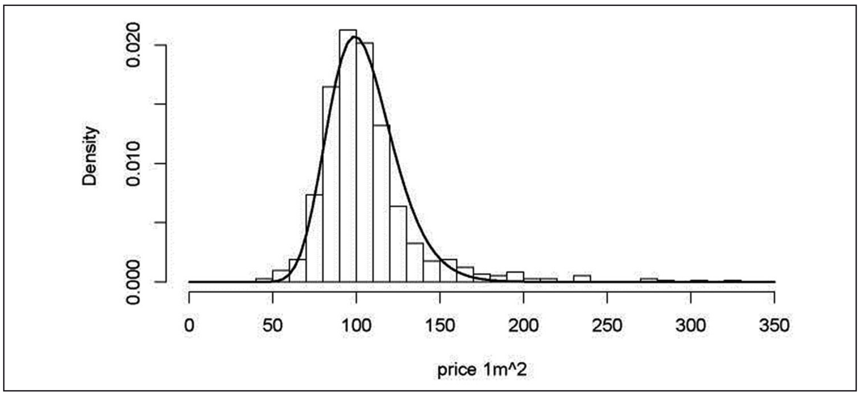

Рис. 5. Зона ценовых предпочтений № 4

Разделение исходной выборки по зонам ценовых предпочтений

Поскольку районы могут быть недостаточно однородными, в КУГИ г. Санкт-Петербурга в 2012 году был разработан принцип разбиения города по зонам ценовых предпочтений. Всего было выделено 12 зон (1 зона – наиболее дорогие адреса, увеличение номера зоны означает сдвиг к периферии) см. [19]. Разбиение исходной выборки по зонам ценовых предпочтений дает 12 подвыборок хорошо приближающихся логарифмически нормальным законом распределения. Например, на рис. 4 и рис. 5, представлены гистограммы и кривые плотности логнормального закона распределения для зоны 1 (Золотой треугольник) и зоны 4.

Выборка цен предложений вторичной жилой недвижимости в г. Санкт-Петербурге для зоны ценовых предпочтений №1 на 31.03.2014 г. и ее аппроксимация логнормальным распределением. Объем выборки: $N = 47$. Параметры логнормального закона: $\mu = 5{,}3$, $\sigma = 0{,}46$. Среднее арифметическое по выборке равно 213,358 тыс. руб. за кв. м. Оценка рыночной стоимости 1 кв. м., то есть мода по логнормальной аппроксимации равна 162,130 тыс. руб. за кв. м. Результат КС-теста: *p-value* = 0,713.

Выборка цен предложений вторичной жилой недвижимости в г. Санкт-Петербурге для зоны ценовых предпочтений №4 на 31.03.2014 г. и ее аппроксимация логнормальным распределением. Объем выборки: $N = 734$. Параметры логнормального закона: $\mu = 4{,}631$, $\sigma = 0{,}191$.Среднее арифметическое по выборке равно 106,339 тыс. руб. за кв. м. Оценка рыночной стоимости 1 кв. м., то есть мода по логнормальной аппроксимации равна 98,940 тыс. руб. за кв. м. Результат теста Колмогорова-Смирнова: *p-value* = 0,05332.

Разделение исходной выборки по типам домов

Разделение выборки по типам домов дает 12 подвыборок (с учетом объединения зданий разных номерных серий в группы панельной и блочной застройки). Все подвыборки хорошо аппроксимируются логнормальным распределением. На Рис. 6 представ-

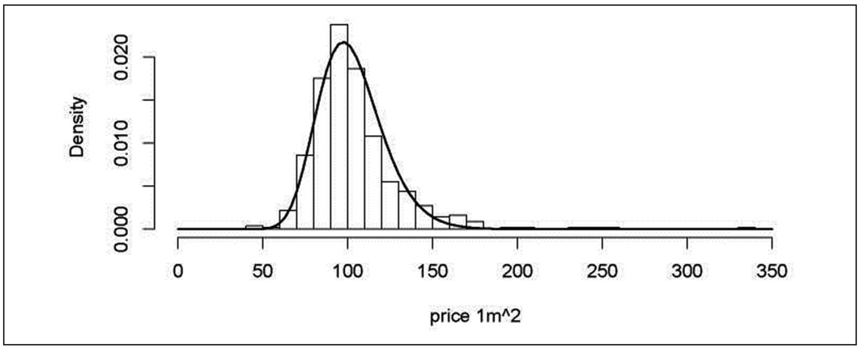

Рис. 6. Застройка сталинского периода

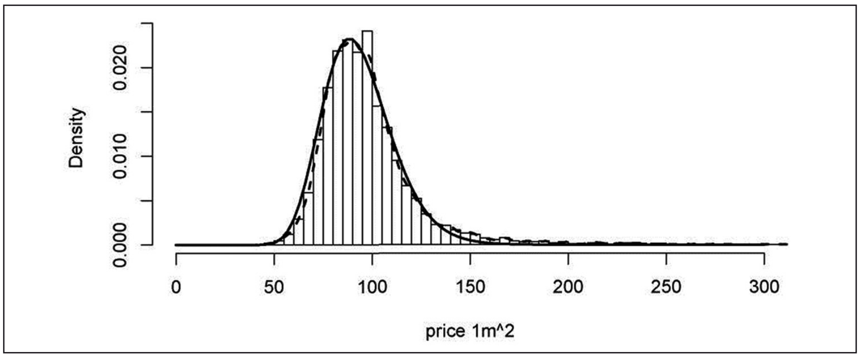

Рис. 7. Весь город СПб

лены гистограмма и кривая плотности логнормального распределения для «сталинских домов».

Выборка цен предложений вторичной жилой недвижимости в г. Санкт-Петербурге для застройки сталинского периода на 31.03.2014 г. и ее аппроксимация логнормальным распределением. Объем выборки: $N = 546$. Параметры логнормального закона μ = 4,612, σ = 0,185.Среднее арифметическое по выборке равно 104,035 тыс. руб. за кв. м. Оценка рыночной стоимости 1 кв. м., то есть мода по логнормальной аппроксимации, равна 97,298 тыс. руб. за кв. м. Результат теста Колмогорова-Смирнова: *p-value* = 0,1575 .

Весь город. Статистика, идентифицированная по зонам ценовых предпочтений

Рассмотрим выборку по всему городу СПб – статистику объявлений по всему городу, которая попала в идентификаторы зон ценовых предпочтений (см. [19]). Такая выборка, безусловно, является смесью не вполне однородных выборок, различающихся по целому спектру ценообразующих факторов, таких как место (административные, территориальные и т.п. деления), типам домов, этажности, типам квартир и т.д. На рис. 7 видно, что эмпирическое распределение имеет «тяжелый» правый «хвост», указывающий на необходимость дальнейшего разделения представленной смеси, видимо выделение такого/таких элементов смеси, как премиум сегмент/сегменты. Легенда для диаграммы: пунктир – ядерная оценка плотности, полученная с помощью пакета ***R***. Объем выборки: 10627. Параметры логнормального закона: μ = 4,539, σ = 0,20. Среднее арифметическое по выборке равно 97,462 тыс. руб. за кв. м. Оценка рыночной стоимости 1 кв. м., то есть мода по логнормальной аппроксимации равна 89,579 тыс. руб. за кв. м. Результат теста Колмогорова-Смирнова: *p-value* < 2,861e^{-8}.

Малое значение *p-value*, как и в случае Рис. 2 свидетельствует о малой значимости для всего города одного логнормального распределения, – и, как отмечалось выше, это является важным статистическим аргументом для разбиения города на подвыборки, формирующие отдельно взятые логнормальные распределения.

Литература

1. Грибовский С.В. «Математические методы оценки стоимости недвижимого имущества.» Учебное пособие. М.: Финансы и Статистика, 2008
2. Федеральный стандарт оценки №2 «Цель оценки и виды стоимости.» Приказ Минэкономразвития РФ № 255 от 20.07.2007. Зарегистрирован в Минюсте РФ 23.08.2007
3. International Valuation Standard Council, London, 2013. URL:http://www.valuersinstitute.com.au/docs/professional_practice/International%20Valuation%20Stand
4. European Valuation Standards, 7-th edition, Brussels, 2012. URL: http://www.tegova.org/en/p4fe1fcee0b1db
5. Uniform Standards of Professional Appraisal Practice, 2014-2015 edition, Annapolis, Maryland, USA, 2014. URL: http://www.uspap.org/
6. Royal Institution of Chartered Surveyors Valuation Professional Standard, London, 2014. URL:http://www.rics.org/uk/shop/RICS-Valuation-Professional-Standards-Red-Book-2013-19750.aspx
7. Sharpe, William R., "The Sharpe Ratio." Journal of Portfolio Management21(1): 49-58, 1994.
8. Биллингсли, П. «Сходимость вероятностных мер.» М.: Наука, 1977
9. Русаков О.В., Солев В.Н. «Предельный переход для классических моделей в стохастической финансовой математике (учебно-методическое пособие)» Издательство Санкт-Петербургского государственного университета, 1999.
10. Ширяев А.Н. «Вероятность» М.: Наука, 1989.
11. Феллер, В. «Введение в теорию вероятностей и ее приложения.» т. II, М.: Мир, 1984.
12. Venables, W.N. and Ripley, B.D. "Modern Applied Statistics with S." 4th edition. Springer, 2002.
13. Birnbaum, Z.W. and Fred, H.T. "One-sided confidence contours for probability distribution functions." The Annals of Mathematical Statistics, 22/4: 592-596, 1951.
14. William, J. Conover "Practical Nonparametric Statistics." New York: John Wiley & Sons. Pages 295-301 (one-sample Kolmogorov

test), 309-314 (two-sample Smirnov test), 1971.

15. Durbin, J. “Distribution theory for tests based on the sample distribution function.” SIAM, 1973.

16. Большев, Л.Н., Смирнов, Н.В., «Таблицы математической статистики.» М.: Наука, 1983.

17. Marsaglia, G., Tsang, W.W., and Wang, J. “Evaluating Kolmogorov´s Distribution.” Journal of Statistical Software, 8/18, 2003. URL: http://www.jstatsoft.org/v08/i18/.

18. Бюллетень Недвижимости №13/1605 Санкт-Петербург, 24.03.2014.

19. Отчет об определении кадастровой стоимости объектов недвижимости (за исключением земельных участков), расположенных на территории Санкт-Петербурга. КУГИ Правительства Санкт-Петербурга, том 2, раздел 2.3, СПб, 2012 http://rosreestr.ru/wps/portal/p/cc_ib_portal_services/cc_ib_ais_fdg/co/ Субъект РФ: С-Петербург. Виды объектов: Недвижимость, помещения. Отчет №32-1-0733/2012(2), Санкт-Петербург, 30.11.2012

References

1. Gribovsky, S. V. “Mathematical methods of survey (estimation) of real estate value” Textbook. M.: Finance and Statistics, (in Russian), 2008.

2. Federal Standard of Survey (Valuation) № 2 “The purpose of valuation and types of prices”. Decree of the Ministry of Economic Development of the Russian Federation №255 dated at 20.07.2007. Registered in the Ministry of Justice of RF 23.08.2007

3. International Valuation Standard Council, London, 2013. URL: http://www.valuersinstitute.com.au/docs/professional_practice/International%20Valuation%20Standards%202013.pdf

4. European Valuation Standards, 7th edition, Brussels, 2012. URL: http://www.tegova.org/en/p4fe1fcee0b1db

5. Uniform Standards of Professional Appraisal Practice, 2014-2015 edition, Annapolis, Maryland, USA, 2014. URL: http://www.uspap.org/

6. Royal Institution of Chartered Surveyors Valuation Professional Standard, London, 2014. URL: http://www.rics.org/uk/shop/RICS-Valuation-Professional-Standards-Red-Book-2013-19750.aspx

7. Sharpe, William R., “The Sharpe Ratio.”Journal of Portfolio Management 21(1):49-58, 1994.

8. Billingsley, P. “Convergence of Probability Measures” John Willey & Sons, New York, 1968.

9. Rusakov, O. V., Solev, V. N. “Passage to the Limit for Classical Models in Stochastic Financial Mathematics (study guide)” Saint-Petersburg State University Press, (in Russian) 1999.

10. Shiryaev, A. N. “Probability” M.: Nauka (in Russian), 1989.

11. Feller, W. “An Introduction to Probability Theory and Its Applications” Volume II, 3rd ed., John Willey & Sons, New York, 1971.

12. Venables, W.N. and Ripley, B.D. “Modern Applied Statistics with S.”4th ed., Springer, 2002.

13. Birnbaum, Z.W. and Fred, H.T. “One-sided confidence contours for probability distribution functions.” The Annals of Mathematical Statistics, 22/4: 592–596, 1951.

14. William, J. Conover “Practical Nonparametric Statistics.” New York: John Wiley & Sons. Pages 295–301 (one-sample Kolmogorov test), 309–314 (two-sample Smirnov test),1971.

15. Durbin, J. “Distribution theory for tests based on the sample distribution function.” SIAM, 1973.

16. Bolshev, L. N., Smirnov N. V., “Tables of Mathematical Statistics” M.: Nauka (in Russian), 1983 Real Estate Bulletin № 13/1605 Saint-Petersburg, 24.03.2014.

17. Marsaglia, G., Tsang, W.W., and Wang, J. “Evaluating Kolmogorov´s Distribution.” Journal of Statistical Software, 8/18, 2003. URL: http://www.jstatsoft.org/v08/i18/.

18. Real Estate Bulletin № 13/1605 Saint-Petersburg, 24.03.2014.

19. Report on determination of the real estate cadastre value (excluding land costs), situated in Saint-Petersburg. CASP Government of Saint Petersburg, edition 2, section 2.3, Saint-Petersburg, 2012. URL: http://rosreestr.ru/wps/portal/p/cc_ib_portal_services/cc_ib_ais_fdg/co/ Federal Subject of RF: Saint-Petersburg. Types of property: Real Estate, Buildings. Report №32-1-0733/2012(2), Saint-Petersburg, 30.11.2012.