

А.Ю.Корчагин¹, В.Ю.Королев^{1,2}¹Московский государственный университет им. М.В. Ломоносова;²Институт проблем информатики РАН, Москва, Россия

(E-mail: vkorolev@cs.msu.ru)

Прогнозирование финансовых рисков с помощью модифицированного сеточного метода скользящего разделения дисперсионно-сдвиговых смесей нормальных законов

В статье описан метод прогнозирования финансовых рисков с использованием параметрических моделей из класса дисперсионно-сдвиговых смесей нормальных законов. Предложенный метод берет за основу результаты работы модифицированного двухэтапного сеточного метода оценивания параметров обобщенных гиперболических распределений в скользящем режиме. Подробно обсуждаются вопросы практического применения метода, скорость его работы. Описывается процесс настройки параметров (обучения модели). На реальных данных иллюстрируется точность прогнозирования при использовании разных метрик в зависимости от настроек метода, в том числе рассматриваются долгосрочные прогнозы на более чем один шаг вперед.

Ключевые слова: прогнозирование финансовых рисков, дисперсионно-сдвиговая смесь нормальных законов, обобщенное гиперболическое распределение, сеточный метод разделения смесей.

1 Введение

Одной из важнейших практических задач на финансовых рынках является задача прогнозирования рисков, связанных с превышением тех или иных показателей критических порогов. Статистический анализ финансовых данных, как правило, показывает, что многие исследуемые распределения отличаются наличием так называемых тяжелых хвостов. Точная оценка данных хвостов позволяет адекватно оценить соответствующие риски.

Помимо непосредственного исследования распределений, любая финансовая организация заинтересована в получении более-менее достоверных прогнозов на основе наблюдаемых данных. Прогнозирование несет в себе большой спекулятивный фактор, но некоторые жесткие требования к любому осмысленному методу прогнозирования известны заранее: метод должен работать достаточно быстро, чтобы прогноз оставлял время для принятия решения, и он должен показывать хорошие результаты на случайно выбранных исторических данных.

Для упрощения задачи оценки и прогнозирования распределений часто используется подход снижения размерности путем априорного сужения классов допустимых смесей. В частности, модели, основанные на дисперсионно-сдвиговых смесях нормальных законов, показали высочайшую адекватность при решении практических задач, связанных с описанием эволюции различных финансовых индексов. Более того, такие модели имеют серьезное теоретическое обоснование, так как они выступают в качестве предельных распределений для случайных блужданий со случайными интенсивностями скачков [1].

В работе предлагается алгоритм прогнозирования параметров дисперсионно-сдвиговых смесей (в частности, оценки рисков) на примере известной группы процессов — обобщенных гиперболических процессов (GH-processes). Класс таких смесей очень обширен и, в частности, включает в себя обобщенные гиперболические распределения, которые были введены О.Е.Барндорфф-Нильсеном в 1977–1978 гг. как класс специальных сдвиг-масштабных смесей нормальных законов [2–4].

Пусть $\alpha \in R$, $\beta \in R$. Если функцию распределения обобщенного гиперболического закона с параметрами α , β , ν , μ , λ обозначить через $P_{GH}(x; \alpha, \beta, \nu, \mu, \lambda)$, то по определению

$$P_{GH}(x; \alpha, \beta, \nu, \mu, \lambda) = \int_0^{\infty} \Phi\left(\frac{x - \beta - \alpha z}{\sqrt{z}}\right) p_{GIG}(z; \nu, \mu, \lambda) dz, \quad x \in R,$$

где $\Phi(x)$ — стандартная нормальная функция распределения.

$$\Phi(x) = \int_{-\infty}^x \varphi(z) dz, \quad \varphi(x) = \frac{1}{\sqrt{2x}} e^{-x^2/2}, \quad x \in R,$$

$p_{GIG}(z; \nu, \mu, \lambda)$ — плотность обобщенного обратного гауссовского распределения.

$$p_{GIG}(z; \nu, \mu, \lambda) = \frac{\lambda^{\nu/2}}{2\mu^{\nu/2} K_{\nu}(\sqrt{\mu\lambda})} \cdot x^{\nu-1} \cdot \exp\left\{-\frac{1}{2}\left(\frac{\mu}{x} + \lambda x\right)\right\}, \quad x > 0.$$

Здесь $\nu \in R$,

$$\mu > 0, \lambda \geq 0, \text{ если } \nu < 0;$$

$$\mu > 0, \lambda > 0, \text{ если } \nu = 0;$$

$$\mu \geq 0, \lambda > 0, \text{ если } \nu > 0;$$

$K_{\nu}(z)$ — модифицированная бесселева функция третьего рода порядка ν ,

$$K_{\nu}(z) = \frac{1}{2} \int_0^{\infty} y^{\nu-1} \exp\left\{-\frac{z}{2}\left(y + \frac{1}{y}\right)\right\} dy, \quad z \in C, \operatorname{Re} z > 0.$$

2 Описание метода прогнозирования финансовых рисков и его свойства

Приведем описание метода на примере пятипараметрического семейства обобщенных гиперболических процессов. Необходимо отметить, что метод использует только сами значения параметров распределений и никоим образом не учитывает вид самого семейства, поэтому данный метод без ограничения может быть применен к любому другому параметризованному семейству распределений.

Возьмем интересующий нас временной ряд и применим к нему стандартный подход разделения смесей нормальных законов с использованием скользящего окна. Для этого зафиксируем размер окна w и сдвиг окна $s \leq w$. Далее на каждом окне применим модифицированный двухэтапный сеточный метод разделения дисперсионно-сдвиговых смесей нормальных законов, предложенный в работе [5].

На первом этапе на положительной полупрямой выделим основную часть носителя смешивающего распределения, т.е. ограниченный интервал, вероятность которого, вычисленная в соответствии со смешивающим распределением, практически равна единице. На этот интервал накинem конечную сетку, содержащую, возможно, очень много известных узлов u_1, \dots, u_K . Считая параметр сдвига β равным нулю, приблизим искомое обобщенное гиперболическое распределение конечной смесью нормальных законов:

$$P_{GH}(x; \alpha, 0, \nu, \mu, \lambda) \approx \sum_{i=1}^K p_i \Phi\left(\frac{x - \alpha u_i}{\sqrt{u_i}}\right), \quad x \in R. \quad (2)$$

В смеси, стоящей в правой части соотношения (2), неизвестными являются только параметры $p_1, \dots, p_{K-1}, \alpha$. Пусть x_1, \dots, x_n — анализируемая выборка значений случайной величины с оцениваемым обобщенным гиперболическим распределением. Итерационный процесс, определяющий сеточный EM-алгоритм для данной задачи, задается следующим образом. Пусть $p_1^{(m)}, \dots, p_{K-1}^{(m)}, \alpha^{(m)}$ — оценки параметров p_1, \dots, p_{K-1} и α на m -й итерации, $p_K^{(m)} = 1 - p_1^{(m)} - \dots - p_{K-1}^{(m)}$. Обозначим

$$\varphi_{ij}^{(m)} = \frac{1}{\sqrt{u_i}} \varphi\left(\frac{x_j - \alpha^{(m)} u_i}{\sqrt{u_i}}\right), \quad g_{ij}^{(m)} = \frac{p_i^{(m)} \varphi_{ij}^{(m)}}{\sum_{r=1}^K p_r^{(m)} \varphi_{rj}^{(m)}}, \quad i = 1, \dots, K; \quad j = 1, \dots, n.$$

Тогда, используя стандартные рассуждения, определяющие вычислительные формулы EM-алгоритма для параметров конечной смеси нормальных законов (см., например, [6], разделы 5.3.7–5.3.8), следует положить

$$p_i^{(m+1)} = \frac{1}{n} \sum_{j=1}^n g_{ij}^{(m)}, \quad i = 1, \dots, K. \quad (3)$$

Обозначим $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$. Используя соотношение (5.3.24) в [6], с учетом очевидного равенства $\sum_{j=1}^n g_{ij}^{(m)} = 1$ можно заметить, что уточненная оценка параметра α имеет вид

$$\alpha^{(m+1)} = \frac{\bar{x}}{\sum_{i=1}^K u_i p_i^{(m+1)}}, \quad (4)$$

т.е. равна отношению генерального выборочного среднего и текущего эмпирического среднего смешивающего распределения, что вполне согласуется с тем, что в данном случае $EX = \alpha EU$.

Как известно, классический EM-алгоритм обладает свойством монотонности. Поэтому если узлы u_1, \dots, u_K сетки различны, неотрицательны и известны, то итерационный процесс (3)–(4) является монотонным, т.е. каждая его итерация не уменьшает целевую сеточную функцию правдоподобия

$$L(p_1, \dots, p_K, \alpha; x_1, \dots, x_n) = \prod_{j=1}^n \left[\sum_{i=1}^K \frac{p_i}{\sqrt{u_i}} \varphi \left(\frac{x_j - \alpha^{(m)} u_i}{\sqrt{u_i}} \right) \right].$$

В разделе 5.7.4 книги [6] показано, что при каждом фиксированном значении параметра α сеточная функция правдоподобия $L(p_1, \dots, p_{K-1}, \alpha; x_1, \dots, x_n)$ вогнута по аргументам p_1, \dots, p_{K-1} . Поэтому на каждом шаге итерационного процесса вместо соотношения (3) можно использовать любой более быстрый алгоритм максимизации функции $L(p_1, \dots, p_{K-1}, \alpha^{(m)}; x_1, \dots, x_n)$ по переменным p_1, \dots, p_{K-1} . Например, оценки весов p_1, \dots, p_{K-1} можно искать методом условного градиента [6,7].

Таким образом, на первом этапе получают оценки параметра α и весов всех узлов u_i конечной сетки, накинута на носитель смешивающего обобщенного обратного гауссовского распределения $p_{GIG}(z; \nu, \mu, \lambda)$.

На втором этапе остается применить какой-либо стандартный метод подгонки обобщенного обратного гауссовского распределения $p_{GIG}(z; \nu, \mu, \lambda)$ к эмпирическим данным типа гистограммы $(u_1; p_1), \dots, (u_K; p_K)$. Например, параметры ν, μ и λ можно оценить, решая задачу наименьших квадратов

$$(\nu^*, \mu^*, \lambda^*) = \arg \min_{\nu, \mu, \lambda} \sum_{i=1}^K \left[p_i - \int_{\frac{1}{2}(u_{i-1} + u_i)}^{\frac{1}{2}(u_i + u_{i+1})} p_{GIG}(z; \nu, \mu, \lambda) du \right]^2,$$

где $u_0 = 0, u_{K+1} = \infty$. Или же эти параметры можно найти из условия

$$(\nu^*, \mu^*, \lambda^*) = \arg \min_{\nu, \mu, \lambda} D_{KL} [p_{GIG}(z; \nu, \mu, \lambda), h(u)],$$

где

$$h(u) = \begin{cases} 0, & u \leq \frac{1}{2}u_1 \\ \frac{u_{i+1} - u_{i-1}}{2p_i}, & \frac{1}{2}(u_{i-1} + u_i) < u \leq \frac{1}{2}(u_i + u_{i+1}) \\ 0, & u > \frac{1}{2}(3u_K - u_{K-1}) \end{cases}$$

— гистограмма, построенная по значениям $(u_1; p_1), \dots, (u_K; p_K)$, а $D_{KL} [p_{GIG}(z; \nu, \mu, \lambda), h(u)]$ — расстояние (дивергенция) Кульбака–Лейблера.

Конкретный алгоритм выбора сетки на первом этапе описан в работе [5].

В качестве входных данных для метода прогнозирования мы будем использовать результат работы описанного выше двухэтапного сеточного метода разделения дисперсионно-сдвиговых смесей нормальных законов, т.е. ряд оценок параметров распределений, полученных на \hat{N} известных окнах с историческими наблюдениями, p_1, \dots, p_N , где каждое $p_i = (\alpha_i, \beta_i, \nu_i, \mu_i, \lambda_i)^T$.

Конечная задача прогнозирования — получить оценки p_{n+1}, p_{n+2}, \dots для окон, которые будут частично или полностью состоять из будущих наблюдений.

Для этого рассмотрим соотношение

$$\tilde{p}_{i+1} = F_1 p_i + F_2 p_{i-1} + \dots + F_r p_{i-r+1},$$

где $r \in \mathbb{N}$ — заранее фиксированный параметр, имеющий смысл порядка прогноза; $F_j \in R^{5 \times 5}$ — матрицы-регрессоры.

Процедура поиска параметров модели имеет вид

$$(F_1, \dots, F_r) = \arg \min \sum_{i=r+1}^{N-1} (p_{i+1} - \tilde{p}_{i+1})^2. \quad (5)$$

По сути, мы имеем регрессионную модель, где поиск матриц F_j (обучение модели) производится с использованием минимизации суммарного RSS по $\tilde{N} - r$ предсказаниям модели на известных данных. Для простоты обозначений примем $N = \tilde{N} - r$.

Соотношение (5) представляет из себя разновидность линейной регрессии и ее программная реализация не представляется сложной. Более того, во многих статистических пакетах есть встроенные функции расчета матриц F_j для случаев $r = 1, 2$.

3 О точности прогнозов

Остаточная сумма квадратов RSS, полученная в результате поиска матриц F_j , может служить критерием качества модели (будучи деленной на N), но эта метрика не учитывает самого вида распределений. При этом некоторые параметры распределений дают гораздо больший вклад в вид итогового распределения по сравнению с остальными.

Чтобы точнее оценить качество прогноза, рассмотрим вопрос близости непосредственно самого ГН-распределения с полученными параметрами \tilde{p}_{i+1} к истинному ГН-распределению с параметрами p_{i+1} для конкретных данных. Для этого будем использовать несколько метрик, описываемых ниже.

Символами $\tilde{F}(x)$ и $F(x)$ обозначим функции распределения P_{GH} обобщенного гиперболического распределения с параметрами \tilde{p}_{i+1} и p_{i+1} соответственно. По аналогии соответствующие плотности распределений обозначим $\tilde{f}(x)$ и $f(x)$.

1. *Метрики C, L_1, L_2 .* Эти часто применяемые метрики близости распределений имеют вид

$$C(\tilde{f}, f) = \sup_{x \in R} |\tilde{f}(x) - f(x)|, \quad L_1(\tilde{f}, f) = \int_{-\infty}^{+\infty} |\tilde{f}(x) - f(x)| dx, \quad L_2(\tilde{f}, f) = \left[\int_{-\infty}^{+\infty} (\tilde{f}(x) - f(x))^2 dx \right]^{1/2}.$$

Чем меньше значения метрик C, L_1 или L_2 , тем ближе распределения друг к другу.

2. *Метрика Intersect.* Эта метрика имеет смысл графического пересечения плотностей

$$I(\tilde{f}, f) = 1 - \int_{-\infty}^{+\infty} \min(\tilde{f}(x), f(x)) dx.$$

Чем ближе Intersect к нулю, тем ближе распределения друг к другу. При этом для наглядности весьма удобно представлять Intersect в процентах.

3. *Оценки хвостов.* Для оценки хвостов распределений, определяющих соответствующие риски, целесообразно рассмотреть следующие характеристики качества прогнозирования хвостов:

$$W_{0,05} = \tilde{F}(x_{0,05}), \quad W_{0,95} = \tilde{F}(x_{0,95}),$$

где x_α — квантиль порядка α распределения $F(x)$. Из-за того, что ГН-распределения является абсолютно непрерывными, эти характеристики можно переписать в более удобной форме:

$$W_{0,05} = \tilde{F}(F^{-1}(0,05)), \quad W_{0,95} = \tilde{F}(F^{-1}(0,95)).$$

Чем ближе $T_{0,05}$ к значению 0,05, тем точнее мы оценили левый хвост. Аналогично для правого хвоста распределения — чем ближе $T_{0,95}$ к значению 0,95, тем точнее оценка.

Помимо оценки веса левого или правого хвоста, имеет смысл рассмотреть, насколько точно соотносятся абсолютные значения самих квантилей распределений, т.е. насколько точен наш прогноз границ хвостов:

$$S_{0,05} = |\tilde{x}_{0,05} - x_{0,05}|, S_{0,95} = |\tilde{x}_{0,95} - x_{0,95}|$$

или в терминах обратной функции распределения

$$S_{0,05} = |\tilde{F}^{-1}(0,05) - F^{-1}(0,05)|, S_{0,95} = |\tilde{F}^{-1}(0,95) - F^{-1}(0,95)|.$$

Чем меньше значение S_α , тем точнее оценка хвостов.

Применяя описанные метрики при разных параметрах прогнозирования (порядок регрессии, количество исторических наблюдений для обучения модели), мы можем найти оптимальную для каждой конкретной задачи модель прогнозирования.

4 Результаты практического применения метода на реальных данных

В [8] детально описаны особенности использования модифицированного сеточного алгоритма и приведены практические советы для получения входного ряда p_1, \dots, p_n .

В качестве исходных данных возьмем индекс KOSPI (Korea Composite Stock Price Index) — основной индикатор корейской биржи, включающий в себя все торгующиеся на этой бирже компании. Индекс начал рассчитываться в 1983 г. и до сих пор является ключевым для корейского рынка, поэтому он достаточно часто подвергается анализу и попыткам предсказания своей динамики. Подробную информацию о KOSPI можно найти на сайте агентства Блумберг: www.bloomberg.com/quote/KOSPI:IND.

Мы будем рассматривать изменение логарифмов значений индекса (цены) с частотой (тиком), равной 1 минуте, на протяжении трех рабочих дней, начиная с открытия биржи 8-го декабря 2014 г. Размер окна установим равным 3 ч: $w = 180$, сдвиг окна минимальный — одно наблюдение: $s = 1$ тик.

Применяя модифицированный сеточный алгоритм к исходным данным, мы получим ряд оценок параметров $p_i = (\alpha_i, \beta_i, \nu_i, \mu_i, \lambda_i)^T$. В конкретной модели без ущерба для точности мы положили $\beta_i = 0$, оставив 4 параметра. Так как алгоритм устойчив к входным данным, а сдвиг окна — минимальный, ряд p_i дает нам сглаженную картину. На рисунке 1 представлено изменение восстановленного обобщенного гиперболического распределения с течением времени.

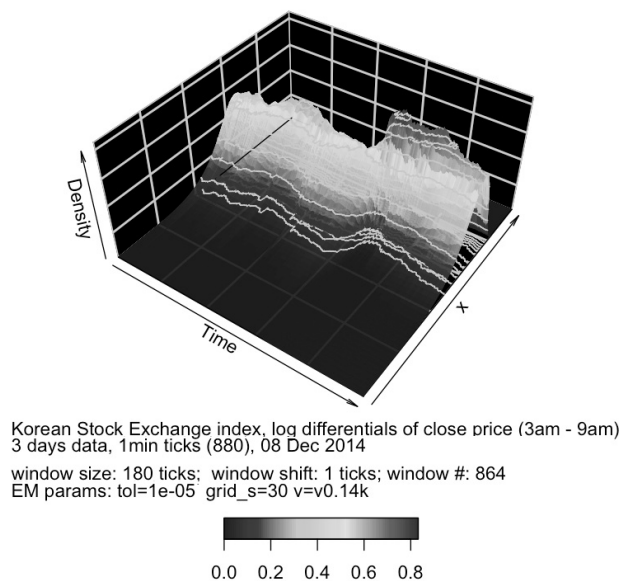


Рисунок 1. Изменение GH-аппроксимации для распределения индекса KOSPI во времени

4.1 Описание процедуры прогнозирования, исходные выбранные модели

Среди всех окон (864) мы выделим точку во времени ($T = 500$), которую будем считать «текущим моментом времени», границей известных данных. Данные до этой точки будут использоваться для обучения модели, данные после нее — для получения прогнозов и сравнения полученных результатов с историческими значениями.

Для анализа качества прогнозирования будем использовать значения метрик, вычисленные для следующих окон в будущем:

- $T + 1$ — окно, сдвинутое на 1 тик, минимальный прогноз в 1 мин, 99% данных известны;
- $T + 10$ — прогноз в 10 тиков (10 мин), 94% наблюдений известны в момент T ;
- $T + 60$ — прогноз в 60 тиков (1 ч), 66% наблюдений известны в момент T ;
- $T + 120$ — прогноз в 120 тиков (2 ч), 33% наблюдений известны в момент T ;
- $T + 180$ — прогноз в 180 тиков (3 ч), 0% наблюдений известны в момент T .

На рисунке 2 показана иллюстрация рассматриваемой временной линии в данном эксперименте. На линии отмечен момент T , который мы считаем текущим моментом времени. Все события левее от точки T являются для нас доступными историческими данными, события правее будут прогнозироваться (а позднее сравниваться с истинными значениями).

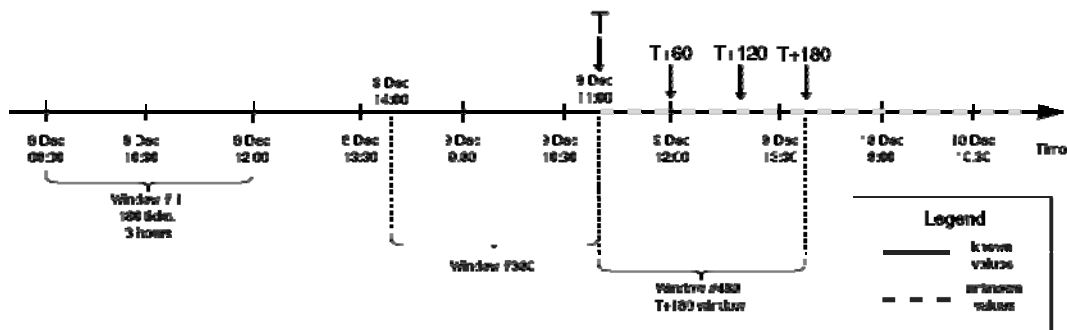


Рисунок 2. Временная линия в эксперименте, Kogea Exchange местное время

В качестве моделей мы будем использовать разные модели с $r = 1, 2$ и 3 , т.е. матричную регрессию первого, второго и третьего порядков. Помимо использования регрессий разного порядка, мы также будем изменять число последних известных исторических наблюдений N , используемых для поиска параметров регрессии. Модели, использующие меньше исторических данных, с большей вероятностью отразят недавние тренды в прогнозе, но при этом они менее точны, так как используют меньше входных данных.

Задача поиска матриц-регрессоров в данном случае является задачей отыскания соответственно 16, 32 и 48 численных параметров для случаев $r = 1, 2$ и 3 . Учитывая число неизвестных параметров, зададим число используемых известных исторических наблюдений для обучения модели в таблице 1.

Т а б л и ц а 1

Рассматриваемые модели, параметры r и N

	$N = 10$	$N = 20$	$N = 50$	$N = 100$	$N = 200$
$r = 1$ (16 коэфф.)					
$r = 2$ (32 коэфф.)					
$r = 3$ (48 коэфф.)					

4.2 Выбор лучшей модели с использованием фиксированного горизонта прогнозирования

После того, как были вычислены необходимые коэффициенты для выбранного семейства моделей, зафиксируем наиболее актуальный для нас горизонт прогнозирования: допустим, нас больше всего интересуют прогнозы на 1 ч вперед, или, используя введенное ранее обозначение, нас интересует окно $T + 60$.

Проведем сравнительный анализ расчетов основных метрик для различных моделей:

Т а б л и ц а 2

Анализ качества прогноза на фиксированный горизонт в зависимости от модели, +1 ч (T+60).

Модель		Прогноз T + 60								
r	N	RSS	C	L ₁	L ₂	I	W _{0,05}	W _{0,95}	S _{0,05}	S _{0,95}
1	10	0,5838	0,0805	0,1067	0,0706	5,68%	0,0597	0,9350	0,1122	0,0512
1	20	0,3778	0,0919	0,1183	0,0793	6,23%	0,0556	0,9311	0,0658	0,0632
1	50	0,3508	0,0626	0,0820	0,0548	4,40%	0,0550	0,9383	0,0570	0,0419
1	100	0,3152	0,1236	0,1599	0,1066	8,36%	0,0584	0,9203	0,1021	0,0929
1	200	0,9201	0,2863	0,3551	0,2353	18,05%	0,0415	0,8662	0,1215	0,1938
2	50	0,2756	0,1309	0,1699	0,1128	8,81%	0,0521	0,9163	0,0263	0,1052
2	100	0,2918	0,0784	0,1162	0,0737	6,22%	0,0676	0,9340	0,2018	0,0549
2	200	0,9038	0,2727	0,3409	0,2254	17,33%	0,0417	0,8707	0,1171	0,1880
3	100	0,2717	0,0644	0,1029	0,0639	5,55%	0,0683	0,9371	0,2065	0,0457
3	200	0,7581	0,2692	0,3326	0,2215	16,94%	0,0437	0,8734	0,0890	0,1817

Для удобства анализа полученных результатов в таблице жирным шрифтом выделены лучшие значения метрик в каждой из групп $r = 1, 2$ и 3 .

Из полученных значений становится ясно, что в группах $r = 1$ и $r = 2$ модель с минимальным компонентным RSS не является наилучшей с точки зрения оценки интересующих нас показателей, и подтверждает разумность выбора наших метрик. Среди всех моделей выбор падает на две: $r = 1, N = 50$ и $r = 1, N = 100$ — соответствующие ряды в таблице подсвечены серым цветом. Исходя из полученных значений метрик, наилучшей моделью в данном случае является регрессия первого порядка: $r = 1, N = 50$.

4.3 Анализ точности прогнозирования и особенностей полученной модели

Рассмотрим вопрос о точности прогнозирования для выбранной модели $r = 1, N = 50$, в зависимости от горизонта прогнозирования. Для каждой из ранее выбранных точек в будущем ($T + 1, T + 60, T + 120, T + 180$) построим соответствующий прогноз и исследуем референсные метрики.

Т а б л и ц а 3

Анализ качества прогноза в зависимости от горизонта прогнозирования, $r = 1, N = 50$

r	N		C	L ₁	L ₂	I	W _{0,05}	W _{0,95}	S _{0,05}	S _{0,95}
1	50	T+1	0,0066	0,0071	0,0050	0,62%	0,0499	0,9494	0,0011	0,0021
1	50	T+10	0,0495	0,0570	0,0397	3,13%	0,0515	0,9434	0,0174	0,0224
1	50	T+60	0,0626	0,0821	0,0548	4,40%	0,0550	0,9383	0,0570	0,0419
1	50	T+120	0,0515	0,0765	0,0463	4,08%	0,0553	0,9620	0,0624	0,0534
1	50	T+180	0,1639	0,2385	0,1483	12,13%	0,0468	0,9827	0,0412	0,1984

Как и ожидалось, по метрикам мы наблюдаем рост несоответствия между прогнозом и истинными значениями с расширением горизонта прогнозирования. На рисунке 3 показано изменение истинного и прогнозируемого распределений с расширением горизонта прогнозирования, а также приведены значения параметров указанных распределений.

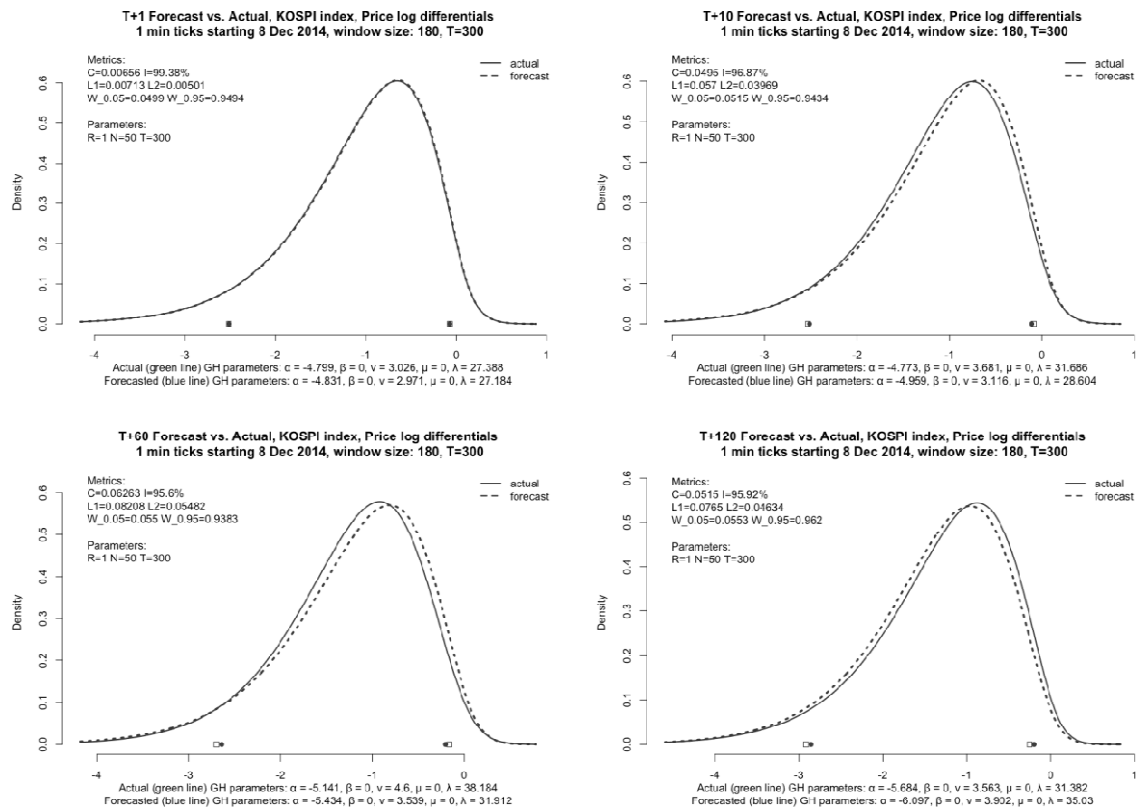


Рисунок 3. Прогнозируемое и истинное распределения, горизонты 1 мин, $T + 1$ (вверху слева); 10 мин, $T + 10$ (вверху справа); 1 ч, $T + 60$ (внизу слева); 2 ч, $T + 120$ (внизу справа)

Особый интерес представляет сравнение динамики отдельных параметров распределения прогноза и истинных исторических значений без использования самого вида распределения. На рисунке 4 отдельно показано соотношение между прогнозом параметра λ (после $T = 300$) и истинными значениями.

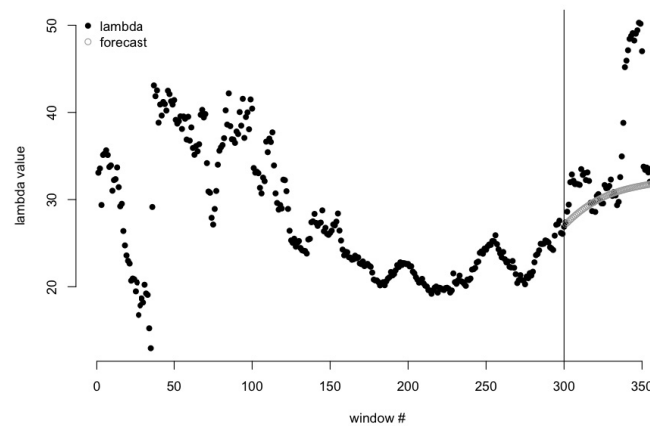


Рисунок 4. Прогнозируемое и истинные значения параметра λ

На основании полученных данных можно сделать вывод, что найденная описанным методом модель хорошо подходит для прогнозирования как минимум на 2 ч вперед. Важно отметить, что найденная модель достаточно точно оценивает хвосты распределений, позволяя нам полагаться на полученные прогнозы при оценке рисков.

Интересную картину можно получить, если непрерывно проследить за получаемым прогнозом, начиная с одной минуты и вплоть до 3 ч, постоянно сравнивая его с историческими распределениями. Удобно наблюдать за этой эволюцией в формате видео или ускоренного слайдшоу. Пример для найденной модели выше доступен по ссылке: git.io/b4jA

При практическом использовании модели на поступающих в реальном времени данных постепенно становятся доступными новые исходные параметры, что позволяет проводить повторное обучение модели (пересчитывая регрессионные матрицы). Это особенно удобно, поскольку поиск необходимых матриц происходит очень быстро и не представляет большой вычислительной сложности.

Работа выполнена при поддержке Российского научного фонда, грант 14-11-00364.

Список литературы

- 1 *Королев В.Ю.* Обобщенные гиперболические распределения как предельные для случайных сумм // Теория вероятностей и ее применения. — 2013. — Т. 58. — Вып. 1. — С. 117–132.
- 2 *Назаров А.Л.* Приближенные методы разделения смесей вероятностных распределений: Дис. ... канд. физ.-мат. наук. — М.: МГУ им. М.В. Ломоносова, 2013.
- 3 *Barndorff-Nielsen O.E.* Exponentially decreasing distributions for the logarithm of particle size // Proc. Roy. Soc. Lond., Ser. A. — 1977. — Vol. 353. — P. 401–419.
- 4 *Barndorff-Nielsen O.E.* Hyperbolic distributions and distributions of hyperbolae // Scand. J. Statist. — 1978. — Vol. 5. — P. 151–157.
- 5 *Королев В.Ю., Корчагин А.Ю.* Модифицированный сеточный метод разделения дисперсионно-сдвиговых смесей нормальных законов // Информатика и ее применения. — 2014. — Т. 8. — Вып. 4. — С. 11–19.
- 6 *Королев В.Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Моск. ун-та, 2011.
- 7 *Королев В.Ю., Назаров А.Л.* Разделение смесей вероятностных распределений при помощи сеточных методов моментов и максимального правдоподобия // Автоматика и телемеханика. — 2010. — Вып. 3. — С. 98–116.
- 8 *Корчагин А.Ю., Ярошенко И.И.* О практическом использовании модифицированного сеточного метода разделения дисперсионно-сдвиговых смесей нормальных законов // Информатика и ее применения. — 2015. — Т. 9. — Вып. 1. — С. 2–10.

А.Ю.Корчагин, В.Ю.Королев

Қалыпты заңдағы сырғымалы бөлінген дисперсиялық жылжымалы қоспаның модификацияланған торлық әдіс арқылы қаржылық тәуекелін болжау

Мақалада қалыпты заңдардың дисперсиялық жылжымалы қоспасы бар кластардан параметрлік модельдер арқылы қаржылық тәуекелді болжау әдісі сипатталған. Ұсынылып отырған әдіс негіз ретінде сырғымалы тәртіпте жалпыланған гиперболалық үлестірілген параметрлерді бағалауда екі кезеңді модификацияланған торлық әдістің жұмыс нәтижелері алынды. Әдістің қолданылуы мен жұмыстың жылдамдығы жан-жақты талданды. Параметрлерін келтіру мәселелері (модельді оқыту) нақты берілгендерде әдістің параметрлеріне сәйкес әр түрлі метрикаларды пайдаланып, болжаудың дәлдігі суреттелген, сонымен қатар бір қадамнан ұзақ мерзімді болжаулар қарастырылды.

A.Yu.Korchagin, V.Yu.Korolev

Prediction of financial risks using a modified grid method sliding separation precipitation shear mixtures of normal laws

Describes a method for predicting financial risks using parametric models from the class of precipitation-shear mixtures of normal laws. The proposed method takes as its basis the results of the modified two-step grid method for estimating the parameters of the generalized hyperbolic distributions in the sliding mode. Discusses in detail the practical application of the method, the speed of his work. Describes how to set parameters (training model). On real data illustrates the prediction accuracy when using different metrics depending on the setting method, including long-term forecasts are considered in more than one step ahead.

References

- 1 Korolev V.Yu. *The theory of probability and Applications*, 2013, 58, 1, p. 117–132.
- 2 Nazarov A.L. *Approximate methods for separating mixtures of probability distributions*: Dis. ... cand. phis.-math. sciences, Moscow: M.V.Lomonosov Moscow State University, 2013.
- 3 Barndorff-Nielsen O.E. *Proc. Roy. Soc. Lond., Ser. A*, 1977, 353, p. 401–419.
- 4 Barndorff-Nielsen O.E. *Hyperbolic distributions and distributions of hyperbolae* // *Scand. J. Statist.*, 1978, 5, p. 151–157.
- 5 Korolev V.Yu., Korchagin A.Yu. *Informatics and Applications*, 2014, 8, 4, p. 11–19.
- 6 Korolev V.Yu. *Probabilistic and statistical decomposition methods volatility chaotic processes*, Moscow: Publ. House of the Moscow University, 2011.
- 7 Korolev V.Yu., Nazarov A.L. *Automation and Remote Control*, 2010, 3, p. 98–116.
- 8 Korchagin A.Yu., Yaroshenko I.I. *Informatics and Applications*, 2015, 9, 1, p. 2–10.