Модели для распределения приращений финансовых индексов

Болотин Д. А., Боровко Н. А.

18 апреля 2022 г.

Введение

Распределительные свойства приращений цены активов важны как с теоретической точки зрения, для понимания динамики рынка, так и для многочисленных приложений, таких как ценообразование производных продуктов и оценки стоимостного риска, в которых допущения о распределении играют решающую роль. Эти приложения, как и многие другие, предполагают различные временные горизонты, начиная от нескольких минут, типичных временных рамок, в которых совершаются рыночные сделки, до нескольких месяцев, которые являются временным горизонтом, о котором заботятся портфельные менеджеры. Это требует знания распределительных свойств приращений цен в различных временных масштабах.

Введение

Концепции масштабной инвариантности и масштабного поведения в настоящее время все чаще применяются за пределами их традиционных областей применения - физических наук. Их применение к финансовым рынкам, инициированное Мандельбротом в 1960-х годах, в последние годы вновь вызвало интерес, отчасти из-за обилия высокочастотных наборов данных и доступности компьютеров для анализа их статистических свойств.

Введение

Количественные исследования в основном сосредоточены на ''ликвидных'' финансовых рынках, то есть на организованных рынках, где транзакции происходят часто и число участников велико. Типичными примерами являются валютные рынки, организованные фьючерсные рынки и рынки фондовых индексов, а также рынок крупных акций. Цены на таких рынках регистрируются несколько раз в минуту, создавая множество данных для использования. Различные факторы, такие как государственная политика, процентные ставки и экономические условия, несомненно, влияют на поведение рынка. Однако точная природа их влияния не очень хорошо известна, и, учитывая сложный характер механизма ценообразования, простые детерминированные модели не в состоянии воспроизвести свойства, наблюдаемые в финансовых временных рядах.

Введение

Модель Башелье[Rama Cont, 1997], применяемая к логарифму цен (для обеспечения положительности цены), стала очень популярной в 1950-х годах и является одним из основных компонентов знаменитой формулы ценообразования опционов Блэка-Шоулза. Эта модель подразумевает, что приращения доходности активов (или цен на активы) независимы идентично распределенные (iid) гауссовы переменные. Действительно, если рассматривать каждое изменение цены как сумму множества небольших и независимых случайных вкладов различных рыночных факторов, Центральная предельная теорема предполагает Гауссово распределение как естественного кандидата.

Базовые модели

Прежде всего, существует Гауссова Гипотеза[Brada et al., 1965], гласящая, что распределение изменения цен на рынке является либо нормальным, либо логнормальным. В наиболее популярной форме эта гипотеза выглядит так:

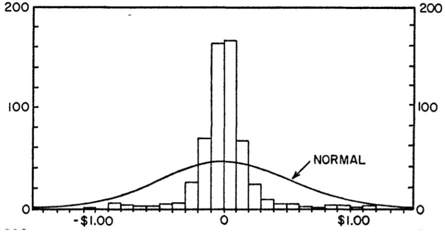
$$\Delta P_n = P(t + n\theta) - P(t + (n-1)\theta),$$

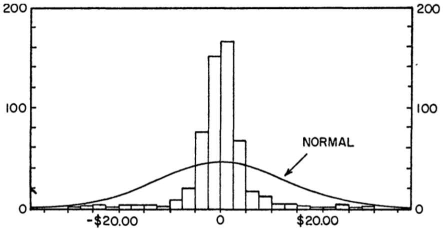
где θ - определенный временной интервал.

В силу предположения о независимости транзакций изменения цен $(\Delta P_1, ..., \Delta P_n)$ также будут независимыми случайными величинами. И эти приращения, либо их логарифмы будут нормально распределены. Это свойство позволяет нам использовать методы анализа, основанные на ЦПТ.

Базовые модели

Однако такое распределение подверглось сомнению в работе, исследовавшей дневные приращения индекса Dow Jones Industrial Average с 1957 по 1962, а также в статье Мандельброта, анализирующей логарифмы приращений цен на хлопок, шерсть, пшеницу и кукурузу.
Проблемы были в слишком толстых хвостах и слишком остроконечном распределении в центре графика.





Тогда была разработана более строгая форма закона:

$$\Delta P_n = P(t + nj) - P(t + (n-1)j),$$

где j - определенный интервал транзакций.

Разница в том, что беря интервал по времени, мы получаем неоднородное число транзакций - их может не быть вообще. Поэтому исходной формулой допускается пользоваться только в случае, если все временные интервалы содержат примерное одинаковое число сделок.
Рассматривается набор из 10 ценных бумаг, торговавшихся 102 дня в 1965 году. Такой набор обусловлен разной степенью их активности, стоимости и волатильности.

Базовые модели

Stock	Average daily volume (hundreds)	Volume range (hundreds)	Average daily trans-actions	Trans-action range	Average shares per trans-action	Price range
Anaconda	85	17–345	48	13–174	178	46–54
Chrysler	910	101–2536	440	50–1049	186	53–75
Eastman Kodak	56	23–269	30	13–115	186	107–122
IBM	91	25–256	64	21–174	142	426–478
Int'l Paper	124	45–536	45	18–164	276	28–35
National Lead	79	25–273	34	15–75	230	70–78
Proctor & Gamble	42	15–103	21	10–7	203	71–81
Union Carbide	58	17–171	26	10–53	224	102–112
United Aircraft	40	14–117	20	9–51	193	42–51
U. S. Steel	380	88–1055	173	45–468	220	45–57

Stock	Price changes				Standard deviations (Sx)	Longest runs unchanged prices
	Minus	Un-changed	Plus	Total		
Anaconda	1198	2350	1263	4811	$0.143	35
Chrysler	6872	24,891	6989	38,752	0.095	98
Eastman Kodak	995	990	1040	3025	0.260	20
IBM	1934	2219	2264	6417	0.516	18
Int'l Paper	1064	2397	1064	4525	0.105	33
National Lead	926	1519	991	3436	0.148	22
Proctor & Gamble	700	714	687	2101	0.199	13
Union Carbide	866	895	854	2615	0.202	9
United Aircraft	640	761	647	2048	0.174	16
U. S. Steel	3210	10,876	3221	17,307	0.081	63

Затем временной период был разделен на 3, соответствующих спаду, росту и стабильности.

Stock	Standard deviations		
	4/22–7/12	7/15–8/9	8/12–9/13
Anaconda	$0.143	$0.132	$.114
Chrysler	0.095	0.103	0.096
Eastman Kodak	0.260	0.285	0.270
IBM	0.516	0.461	0.547
International Paper	0.105	0.104	0.102
National Lead	0.148	0.139	0.175
Proctor & Gamble	0.199	0.233	0.229
Union Carbide	0.202	0.189	0.244
United Aircraft	0.174	0.160	0.159
U. S. Steel	0.081	0.084	0.082

Базовые модели

На основании тестов был сделан вывод, что в данном случае имеется конечная дисперсия. Затем было заключено, что ЦПТ работает в данном случае (на примере акций Chrysler), однако цены брались через 100, что превышает самый длинный период неизменности цен (98). Главный смысл приведенного исследования состоит в том, что классические методы, использующие моменты, могут быть применены к приращению цен акций, если производить разделение не по интервалам времени, а по интервалам транзакций.

Sample size	Actual standard deviation	Expected standard deviation
1	0.098	0.098
50	0.664	0.679
100	0.960	0.955

Однако данная гипотеза подверглась критике[FAMA, 1963], главным образом из-за хвостов: они оказались значительно выше, чем у нормального распределения. Поэтому была предложена Гипотеза Парето, базирующаяся на 2 важных утверждениях:
(1) дисперсии эмпирических распределений ведут себя так, как если бы они были бесконечными
(2) эмпирические распределения наилучшим образом соответствуют негауссовым членам семейства предельных распределений.
Если утверждение о бесконечной дисперсии верно, то статистические инструменты, основанные на предположении о конечной дисперсии, перестают показывать достойный результат.

Логарифм характеристической функции в такой модели будет иметь вид:

$$ln(f(t)) = ln \int_{-\infty}^{+\infty} exp(iut) dP(\bar{u} < u) = i\delta t - \gamma |t|^{\alpha} (1 + i\beta \frac{t}{|t|}) tg \frac{\alpha\pi}{2},$$

где $\alpha, \beta, \delta, \gamma$ - параметры распределения.

Здесь сложность в том, что плотности распределения известны только для некоторых частных случаев, одним из которых является Гауссово распределение, в этом случае логарифм характеристической функции равен:

$$ln(f(t)) = i\mu t - \frac{\sigma^2}{2} t^2$$

Базовые модели

Устойчивое распределение семейства Парето обладает 3 очень важными свойствами:
(1) Хвосты имеют асимптотическую природу Парето,
(2) имеет место инвариантность при сложении,
(3) эти распределения являются единственно возможными предельными распределениями для сумм независимых одинаково распределенных случайных величин.
Мандельброт проверил[Mandelbrot, 1963] гипотезу Парето, проанализировав логарифмы ежедневных приращений цен на хлопок. Он вычислил выборочные выборочные дисперсии на выборке до 1300 наблюдений и обнаружил, что по мере увеличения размера выборки дисперсия не устанавливается до какого-либо предельного значения, а скорее продолжает изменяться абсолютно беспорядочным образом, точно так, как и следовало ожидать при гипотезе Парето.

Базовые модели

Главная проблема гипотезы - когда $\alpha < 2$, дисперсия распределения будет бесконечной, поэтому выборочная диспресия будет неподходящей мерой. Более того, статистические методы, основанные на конечной дисперсии (например, метод наименьгих квадратов), также либо не будут работать вообще, либо будут работать в более слабой версии.

Однако это не означает, что нельзя оценить изменчивость распределения. Если $a > 1$, то оценки, включающие только первые степени, имеют конечное математическое ожидание. Это означает, что например, среднее абсолютное отклонение, которое тоже включает в себя только первые степени, тоже будет иметь конечное математическое ожидание, а значит, оно является лучшей мерой для данного распределения.

Базовые модели

Помимо двуз рассмотренных выше гипотез, существует альтернативная[Mandelbrot and Taylor, 1967], суть которой в том, что они обе не противоречат друг другу: приращения цен за фиксированное число транзакций могут иметь Гауссово распределение, в то время как при фиксированном интервале времени будет иметь место быть гипотеза Парето с бесконечной дисперсией. Противоречия не будет, поскольку количество транзакций за любой период времени является случайным.

Модели распределения индексов

В прикладной теории вероятностей обычно предполагается, что конкретная модель может быть достаточно обоснована только в том случае, если она является асимптотическим приближением, т.е. когда существует довольно простая предельная схема и соответствующая предельная теорема, в которой рассматриваемая модель действует в качестве предельное распределение[В. Ю. Королев, 2015]. Эти теоремы как бы позволяют нам разделить вклад внешних и внутренних факторов в случайность поведения анализируемого процесса.

Модели распределения индексов

В данном исследовании особое внимание уделено механизму формирования GG-распределений общего вида как асимптотических аппроксимаций. В нем же предложен аналог закона больших чисел для отрицательных биномиальных случайных сумм независимых неодинаково распределенных случайных величин.

Модели распределения индексов

В том числе были приведены примеры подгонки GG-распределений(обобщенного гаммараспределения) к реальным данным об интенсивности потоков событий на финансовых рынках, а также результаты сравнения таких моделей с GIG-распределениями(Generalized inverse Gaussian distribution). Параметры для этих моделей оцениваются и подгоняются при помощи использования сеточного метода разделения дисперсионно-сдвиговых смесей нормальных законов. Эффективность этого метода является достаточно важным обстоятельством при его применении для анализа реальных процессов с целью прогнозирования сопутствующих рисков, возможности и результаты которого на примере индекса Shanghai Composite будут показаны далее.

Почему распределения GG работает лучше в данной модели чем GIG?

В отличие от GIG-распределений, класс GG-распределений содержит законы, хвосты которых убывают вейбулловским образом. В частности, при параметре $0 < \nu < 1$ GG-распределения занимают промежуточное место между распределениями с экспоненциальным убыванием хвостов (показательное распределение, гамма-распределение) и «тяжелохвостыми» распределениями со степенным убыванием хвостов типа Ципфа–Парето.

Отличие GVG от GH

Асимптотическое поведение хвостов смесей нормальных законов в определенном смысле совпадает с аналогичным поведением хвостов смешивающих законов. В виду этого, в отличие от обобщенных гиперболических законов, класс обобщенных дисперсионных гамма-распределений(GVG distribution) содержит распределения с хвостами, убывающими экспоненциально-степенным образом. Такие особенности эмпирических распределений характерны для финансовых индексов с нерегулярным поведением, т.е. с кратковременными периодами повышенной волатильности или выбросами.

GVG распределения как асимптотические аппроксимации

В ходе исследования были изучены элементарные свойства дисперсионно-сдвиговых смесей нормальных законов, а также выведен в виде следствия критерий сходимости распределений случайных сумм к обобщенным дисперсионным гамма-распределениям:
Предположим, что существуют последовательность натуральных чисел $\{k_n\}_{n\geq 1}$ и число $\alpha \in \mathbb{R}$ такие, что имеет место сходимость

$$P(S_{n,k_n} < x) \Rightarrow \Phi(x-\alpha)$$

Предположим, что $N_m \to \infty$ по вероятности. Для того чтобы имела место сходимость распределений случайных сумм к обобщенным дисперсионным гаммараспределениям:

$$P(S_{n,N_n} < x) \Rightarrow P_{GVG}(x;\alpha,\sigma,\nu,\kappa,\delta)(1)$$

необходимо и достаточно, чтобы

$$P(N_n < xk_n) \Rightarrow P_{GG}(x;\nu,\kappa,\delta)(2)$$

В следствии 1 ключевым условием является соотношение (2). В дальнейшем будет приведен пример асимптотической конструкции, в рамках которой справедлива предельная теорема, влекущая сходимость (2). На основании трех вспомогательных лемм была получена предельная теорема:
Предположим, что случайные величины $M_j, j > 1$, удовлетворяют условию статистической устойчивости:
по вероятности при $n \to \infty$ с некоторым $\nu > 0$. Предположим, что случайная величина $\bar{N}_{r,p_n}$ имеет отрицательное биномиальное распределение (3) с параметрами $r > 0$ и $p_n \to 0$ и при каждом $n \in N$ независима от последовательности $\{M_j\}_{j \geq 1}$. Пусть случайные величины N_n имеют вид (4). Тогда

$$\lim_{n \to \infty} \sup_{x > 0} \left| P(p_n^{1/\nu} N_n < x) - \frac{\nu}{\Gamma(r)} \int\limits_0^x z^{\nu r - 1} e^{-z^\nu} dz \right| = 0.$$

GVG распределения как асимптотические аппроксимации

$$p_i^{(m+1)} = \frac{1}{n}\sum_{j=1}^{n} g_{ij}^{(m)}, j = 1...K(4).$$

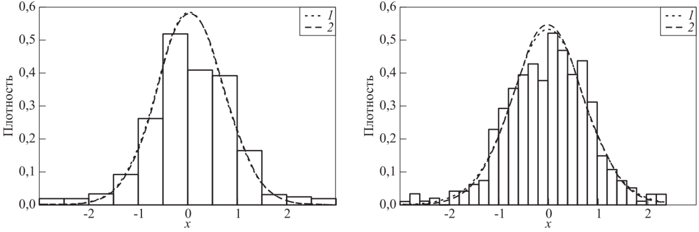
Таким образом, теорема 3, по сути, представляет собой закон больших чисел для отрицательных биномиальных случайных сумм необязательно одинаково распределенных независимых случайных величин и устанавливает условия сходимости распределений таких сумм к обобщенным гамма-распределениям.
При определенном подборе параметров в полученной теореме, можно получить пример, в котором асимптотическим распределением числа слагаемых в сумме, фигурирующий в следствии 1, является GG-распределение.

Сравнение GVG и GH распределений на примере индекса Shanghai Composite

В качестве данных были взяты котировки индекса Shanghai Composite. Будет рассматриваться изменение логарифмов значений индекса (цены) с частотой (тиком), равной 1 мин, на протяжении трех рабочих дней, начиная с открытия биржи 5 января 2015 г. Размер окна равен 3 ч: $w = 180$, сдвиг окна минимальный — одно наблюдение: $s = 1$ тик. Результаты проверки на целесообразность использования именно GVG-моделей для описания статистических закономерностей поведения приращений логарифмов индекса иллюстрируются на рис.2, где приведены результаты подгонки GVG и GH(Generalized hyperbolic) распределений к указанным приращениям при двух разных положениях скользящего окна.

Сравнение GVG и GH распределений на примере индекса Shanghai Composite

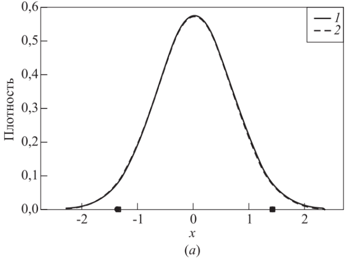
Оба распределения удовлетворительно описывают данные, однако, GVG распределение демонстрирует более высокое P-значение. Аналогичная проверкой было выявлено, что в 403 из 410 случаев P-значение для GVG-распределения оказалось выше, чем для GH-распределения.



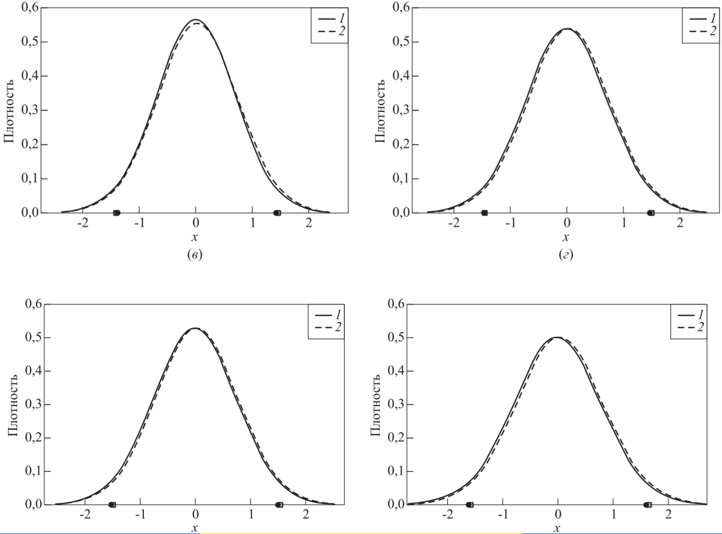
Результаты подгонки GVG (*1*) и GH (*2*) распределений к приращениям логарифмов индекса Shanghai Composite при двух разных положениях скользящего окна

Прогнозирование временного ряда при помощи GVG распределения

Далее применяется алгоритм прогнозирования. Результаты прогнозирования для разных временных дистанций представлены на рис. 3, где показано изменение истинного и прогнозируемого распределения с расширением горизонта прогнозирования.



Прогнозирование временного ряда при помощи GVG распределения



Прогнозирование временного ряда при помощи GVG распределения

На основании полученных результатов можно сделать вывод, что модель типа GVG-распределений достаточно эффективна для прогнозирования как минимум на 2 часа вперед. Также заметим, найденная модель достаточно точно оценивает хвосты распределений, что позволяет использовать полученные прогнозы для оценки рисков.

References I

Brada, J., Ernst, H., and Tessel, J. V. (1965).
The distribution of stock prices differences: Gaussian after all?

FAMA, E. F. (1963).
Mandelbrot and the stable paretian hypothesis.

Mandelbrot, B. (1963).
The variation of certain speculative prices.

Mandelbrot, B. and Taylor, H. (1967).
On the distribution of stock price differences.

Rama Cont, Marc Potters, J.-P. B. (1997).
Scaling in stock market data: stable laws and beyond.

В. Ю. Королев, А. Ю. Корчагин2, (2015).
ОБОБЩЕННЫЕ ДИСПЕРСИОННЫЕ ГАММА-РАСПРЕДЕЛЕНИЯ КАК МОДЕЛИ СТАТИСТИЧЕСКИХ ЗАКОНОМЕРНОСТЕЙ НА ФИНАНСОВЫХ РЫНКАХ.