УДК 519.246.8

ПРОЦЕССЫ ХОУКСА С ПЕРЕМЕННОЙ БАЗОВОЙ ИНТЕНСИВНОСТЬЮ, УПРАВЛЯЕМОЙ ЦЕПЬЮ МАРКОВА

Л.Г. Егорова
Национальный исследовательский университет «Высшая школа экономики»
Россия, 109028, Москва, Покровский бульвар, 11
E-mail: legorova@hse.ru
Институт проблем управления им. В.А. Трапезникова РАН
Россия, 117997, Москва, Профсоюзная ул., 65

Ключевые слова: временные ряды, кластерная структура, процесс Хоукса, цепь Маркова.

Аннотация: В данной работе предложена модель модифицированного процесса Хоукса для моделирования временных рядов, имеющих кластерную структуру и гетероскедастичность, в которой эти артефакты объясняются с содержательной точки зрения как следствие воздействия двух причин – внешнего мира, генерирующего события с некоторыми переменными уровнями системного риска, и внутренней рефлексивности процесса, склонного повышать вероятность появления следующего события в зависимости от предыстории развития процесса. Переменный уровень системного риска управляется марковской цепью и отражает возможность изменения характеристик внешней среды.

1. Введение

Временные ряды очень часто имеют сложную структуру, когда не выполняется предпосылка о постоянстве дисперсии и в таких рядах четко видна тенденция к кластеризации отклонений от средних значений. Такое поведение часто свойственно финансовым рядам (например, стоимостям акций), когда периоды с низким уровнем волатильности сменяются периодами повышенной волатильности, что соответствует периодам турбулентности на рынке и резких изменений стоимости финансового актива вследствие каких-то внешних либо внутренних потрясений.

Для моделирования таких рядов наряду с традиционными методами анализа временных рядов с помощью эконометрических моделей широко используется также модель процессов Хоукса, которые естественным образом описывают свойство так называемого «самовозбуждения», когда произошедшее событие оказывает влияние на будущее поведение процесса и увеличивает появление еще одного события. Интенсивность процесса Хоукса зависит как от функции базовой интенсивности, носящей в классической модели Хоукса неслучайный характер и моделирующей влияние на процесс экзогенных факторов (системный риск), так и от функции ядра, в котором реализована зависимость появления новых событий от предыстории, формирующая эндогенный характер возможных изменений.

Однако во временном ряде могут быть периоды разного поведения системы, в которых сама базовая интенсивность явным образом меняется, что формирует разные периоды временного ряда с похожей внутренней структурой внутри каждого периода – когда есть кризисные периоды с высокой волатильностью системы и периоды более спокойного поведения системы (с наличием, тем не менее, отдельных отклонений и возможным появлением кризисов малого масштаба). Поэтому для моделирования подобных временных рядов предложена модификация процессов Хоукса, в которых базовая интенсивность является случайной величиной с двумя возможными значениями, управляемой дискретной цепью Маркова. Вероятности нахождения цепи в том или ином состоянии будут оцениваться с помощью фильтра Гамильтона.

2. Модель

2.1. Процессы Хоукса

Пусть $(\Omega, \mathcal{F}, P)$ – вероятностное пространство, моменты времени $\{t_i: 0 = t_0 < t_1 < t_2 < \cdots\}$ составляют точечный процесс и обозначают моменты наступления событий (это могут быть как сами джокеры, так и моменты изменения состояния системы в ответ на появление джокеров). Процесс $N(t) = N_t = \sum_{i \in N^*} 1_{\{t_i \le t\}}$, показывающий сколько событий наступило к моменту времени t, называется считающим процессом.

Интенсивность процесса описывает среднее число событий в единицу времени

$$\lambda(t|\mathcal{F}_t) = \lim_{\Delta t \to 0} E\left[\frac{N(t+\Delta t) - N(t)}{\Delta t}|\mathcal{F}_t\right]$$

или, эквивалентным образом, вероятность появления события

$$\lambda(t|\mathcal{F}_t) = \lim_{\Delta t \to 0} \frac{P[N(t+\Delta t) - N(t) > 0|\mathcal{F}_t]}{\Delta t}$$

В процессах Хоукса функция интенсивности процесса записывается следующим образом:

$$\lambda(t) = \lambda_0(t) + \sum_{t_i < t} \nu(t - t_i),$$

где $\lambda_0: \mathbb{R} \to \mathbb{R}_+$ обозначает детерминированную «базовую» интенсивность, а функция $\nu: \mathbb{R}_+ \to \mathbb{R}_+$ отражает влияние произошедших событий в моменты времени t_i на текущее значение интенсивности.

В пионерской работе Хоукса [1], предложившего эту модель случайного процесса, в качестве ядра используется следующая функция

$$\nu(t) = \sum_{j=1}^{P} \alpha_j e^{-\beta_j t}.$$

В наиболее простом случае, когда есть только одна составляющая влияния прошлых событий на будущие и есть только постоянный системный риск (базовая интенсивность), интенсивность процесса Хоукса задается как

$$\lambda(t) = \lambda_0 + \sum_{t_i < t} \alpha e^{-\beta(t - t_i)}.$$

Здесь параметр α задает «скачок» интенсивности вероятности появления последующих событий при появлении очередного события, а параметр β определяет скорость снижения интенсивности обратно к базовому уровню с течением времени.

2.2. Цепь Маркова

Цепи Маркова служат одним из основных инструментов моделирования сложных систем, когда изучаемая система θ_t в момент времени t может находиться в одном из состояний из множества возможных состояний $s_t \in S = \{1, \dots, N\}$, а ее функционирование определяется последовательными случайными переходами между состояниями, причем вероятность выбора определенного состояния зависит только от текущего состояния цепи, но не от ее предыдущей истории:

$$\Pr[\theta_{t+1}=s_{t+1}|\theta_0=s_0,\theta_1=s_1,\ldots,\theta_t=s_t]=Pr[\theta_{t+1}=s_{t+1}|\theta_t=s_t].$$

Пусть матрица $P \in M_{NN}$ обозначает матрицу переходов между N состояниями в цепи Маркова и p_{ij} – это вероятность перехода из состояния i в состояние j. Ограничимся простым случаем, когда цепь Маркова однородна, то есть вероятности перехода не зависят от времени и являются константами. Благодаря марковскому свойству динамику системы описывать достаточно просто, если известны начальное распределение вероятностей p_0 состояния системы (в момент времени t_0) и матрица переходов:

$$Pr[\theta_{t+1}=s_{t+1}]=\sum_{s_t\in S}Pr[\theta_t=s_t]\,Pr[\theta_{t+1}=s_{t+1}|\theta_t=s_t]=\sum_{s_t\in S}p_i(s_t)p_{ij}.$$

Марковская цепь называется эргодической, если возможно перейти из любого состояния системы в любое другое ее состояние за $t>0$ шагов, так что $p_{ij}^t>0$. При этом марковская цепь называется регулярной, если существует число k, такое что переход из любого состояния в любое другое состояние может быть совершен за k шагов. В этом случае при любом начальном распределении вероятностей p^0 существует предел распределения вероятностей между значениями цепи $\lim_{t\to\infty}p_{ij}^t$, который равен $\lim_{t\to\infty}p_{ij}^t=w$, такой что $w=wP$, то есть существует стационарное распределение.

Пусть изучаемые процессы могут быть описаны двумя возможными состояниями системы $S=\{L,H\}$, которые отражают низкий и высокий уровень системного риска вследствие каких-либо внешних факторов. Такая модель соответствует наличию двух уровней базовой интенсивности в модели Хоукса: состояния λ_L и состояния λ_H, которые будут описывать функционирование системы в периоды отсутствия джокера и после его появления (аналогично поведению сейсмических датчиков при отсутствии землетрясения и после его начала).

2.3. Фильтр Гамильтона

Для оценки параметров процесса Хоукса с базовой интенсивностью, управляемой марковской цепью используется оптимизация функции правдоподобия с предварительной фильтрацией Гамильтона для предсказания режима функционирования системы, то есть уровня базовой интенсивности. Пусть $O_j, j=1,\ldots,l$ обозначают моменты наступления событий (резкие скачки показателей сейсмографа, например).

Обозначим как Θ набор параметров процесса Хоукса и цепи Маркова, которые мы хотим определить, то есть $\Theta=\{\lambda_L,\lambda_H,\alpha,\beta\}$. Пусть $f(O_k|\Theta,O_1,\ldots,O_{k-1})$– это функция плотности вероятности наступления события O_k при заданных параметрах Θ и предыстории процесса, т.е. моментов наступления прошлых событий $O_1,\ldots O_{k-1}$. Тогда логарифм функции правдоподобия будет равен

$$\begin{aligned}log\,L\\ &= log\,f(O_1|\Theta)+log\,f(O_2|\Theta,O_1)\\ &+log\,f(O_3|\Theta,O_1,O_2)+\ldots+log\,f(O_n|\Theta,O_1,\ldots,O_{n-1}).\end{aligned}$$

В фильтре Гамильтона условная вероятность новых событий по формуле Байеса равна

$$\begin{aligned}f(O_k|\Theta,O_1,\ldots,O_{k-1})\\ &=\sum_{i=1}^{l}\sum_{j=1}^{l}p_i(t_{k-1}|\Theta,O_1,\ldots O_{k-1})p_{ij}(t_{k-1},t_k|\Theta)\\ &\times f(O_k|\Theta,\theta_{k-1}=i,\theta_k=j),\end{aligned}$$

где $p_i(t_{k-1}|\Theta, O_1, \dots O_{k-1})$ – это вероятность нахождения в состоянии i в момент времени t_{k-1} при условии предыстории, $p_{ij}(t_{k-1}, t_k|\Theta)$ – это вероятность перехода из состояния i в момент t_{k-1} в состояние j в момент времени t_k при заданных параметрах Θ.

Вероятности $p_i(t_{k-1}|\Theta, O_1, \dots O_{k-1})$ могут быть найдены рекурсивно из $f(O_{k-1}|\Theta, O_1, \dots, O_{k-2})$ по правилу:

$$p_i(t_{k-1}|\Theta, O_1, \dots O_{k-1}) = \frac{\sum_{j=1}^{N} p_j\,(t_{k-2}|\Theta, O_1, \dots O_{k-2}) p_{ji}(t_{k-2}, t_{k-1}|\Theta)\, f(O_{k-1}|\Theta, \theta_{k-1})}{f(O_{k-1}|\Theta, O_1, \dots, O_{k-2})}.$$

Для начала рекурсивного определения вероятностей нам необходимо задать $f(O_1|\Theta)$. Воспользуемся наличием стационарного состояния цепи Маркова и определим $f(O_1|\Theta)$ следующим образом

$$f(O_1|\Theta) = \sum_{i=1}^{N} p_i\,(\Theta) f(O_1|\Theta, \theta_1 = i),$$

где $p_i\,(\Theta)$ являются вероятностями стационарного распределения марковской цепи.

3. Результаты расчетов

Предложенная модель и процедура оценки ее параметров с помощью фильтра Гамильтона была реализована в среде R и проведены расчеты для различных реализаций и наборов параметров системы. Для симуляции данных в виде процесса Хоукса, в [2] был предложен метод, основанный на «процедуре прореживания» [3]. Алгоритм с «прореживанием» был модифицирован нами так, чтобы при каждой попытке разыгрывания очередного момента появления события максимальный уровень интенсивности брался с учетом того, в каком состоянии на этот момент находится цепь Маркова.

Приведем далее результаты реализации процесса Хоукса, управляемого цепью Маркова и оценки его параметров (рис. 1). С помощью алгоритма прореживания был симулирован процесс на 500 шагов, базовая вероятность принимала два возможных значения $\{\lambda_H = 5, \lambda_L = 1\}$ с матрицей перехода между состояниями $P = \begin{pmatrix} 0.95 & 0.05 \\ 0.01 & 0.99 \end{pmatrix}$. Параметры ядра процесса Хоукса были равны $\alpha = 0.5, \beta = 0.7$.

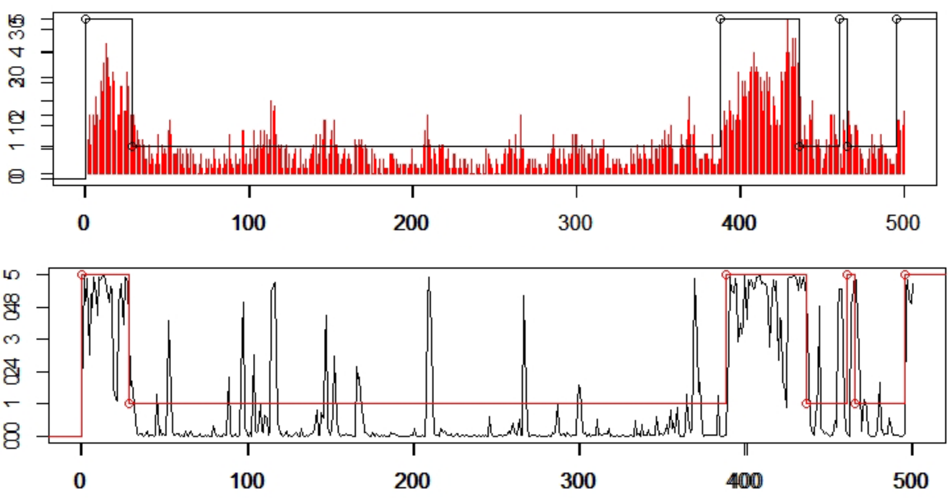

Рис. 1. Пример реализации процесса Хоукса с переменной базовой интенсивностью, управляемой цепью Маркова.

В верхней части рисунка ступенчатой функцией отмечены моменты изменения базовой интенсивности процесса Хоукса вследствие переключения в цепи Маркова между состояниями L и H, а также приведен сам процесс (по вертикальной оси частота наступления событий в день) – хорошо видно, как в периоды высокой λ_H базовой интенсивности частота появления событий в симулированном процессе с такими характеристиками намного выше, чем в периоды низкой λ_L базовой интенсивности, при этом всегда (и в периоды с высокой, и в периоды с низкой λ_0) наблюдается кластерная структура событий, когда есть «всплески» из-за наличия функции ядра в процессе Хоукса, повышающего вероятность появления нового события при появлении очередного события – только в периоды с λ_H такие кластеры становятся более концентрированными и случается больше событий. В нижней части рисунка приведены вновь моменты переключения базовой интенсивности процесса Хоукса между состояниями и вероятность перехода от состояния λ_L к состоянию λ_H, оцененная фильтром Гамильтона.

Видно, что хотя в периоды низкой базовой интенсивности при наступлении небольшого числа событий вследствие самовозбуждающейся природы процесса Хоукса фильтр периодически ошибается, повышая вероятность перехода в состояние системы с λ_H вместо λ_L, но все же быстро возвращается к корректному прогнозу состояния и в целом неплохо предсказывает поведение временного ряда и параметры системы – оценки параметров этой реализации равны $\hat{\lambda}_H = 6.74, \hat{\lambda}_L = 1.46, \hat{\alpha} = 0.44, \hat{\beta} = 0.69$ при истинных параметрах $\lambda_H = 5, \lambda_L = 1, \alpha = 0.5, \beta = 0.7$.

Таким образом, предложенная модель процесса Хоукса с переменной базовой интенсивностью, управляемой цепью Маркова, в целом неплохо описывает поведение системы, функционирующей в разных режимах, и алгоритм оценки параметров такой модели способен определять моменты переключения режимов и параметры самого процесса при определенных условиях, когда в данных есть достаточно много долгих периодов в режиме повышенной базовой интенсивности.

4. Заключение

В данной работе была рассмотрена модель временного ряда с кластерной структурой. Процессы Хоукса позволяют моделировать сложный характер процесса, когда будущие значения временного ряда могут зависеть от прошлых значений. Эта модель дополнена переменной базовой интенсивностью, управляемой дискретной цепью Маркова. Предложен алгоритм оценки параметров такого процесса с помощью фильтра Гамильтона.

Исследование осуществлено в рамках Программы фундаментальных исследований НИУ ВШЭ. Работа выполнена при поддержке Международного центра анализа и выбора решений НИУ ВШЭ. Автор выражает благодарность Donatien Hainaut, PhD, профессору Université Catholique de Louvain, за идею работы, плодотворные дискуссии и код для реализации фильтра Гамильтона.

Список литературы

1. Hawkes A.G. Spectra of Some Self-Exciting and Mutually Exciting Point Processes // Biometrika. 1971. Vol. 58, No. 1. P. 83-90.
2. Ogata Y. On Lewis' simulation method for point processes // IEEE Transactions on Information Theory. 1981. Vol. IT-27, No. 1. P. 23-31.
3. Lewis P. A. W., Shedler G. S. Simulation of Nonhomogeneous Poisson processes by thinning // Naval Research Logistics Quarterly. 1979. Vol. 26, No. 3. P. 403-413.