

ОБНАРУЖЕНИЕ РАЗЛАДОК ВРЕМЕННЫХ РЯДОВ С ИСПОЛЬЗОВАНИЕМ КОПУЛА

В. А. МОРОЗОВ

*Белорусский государственный университет
Минск, Республика Беларусь
vmorozov@by.byelx.com:*

Рассматриваются копула модели зависимости случайных переменных. Предлагаются алгоритмы обнаружения разладки временного ряда на основе параметрической модели.

Ключевые слова: копула модель, обнаружение разладки.

ВВЕДЕНИЕ

Исследование зависимостей играет ключевую роль при принятии решений в условиях неопределенности и анализе риска. Естественным, что неадекватные модели могут привести к неточным оценкам риска и неоптимальным решениям. Традиционно для описания зависимости между случайными величинами используется корреляция Спирмена. Известно, однако, что такой подход удовлетворителен только при условии нормальности переменных или если их распределение принадлежит классу эллиптических. В тоже время известно, что распределения финансовых данных существенно ассиметричны. В последние годы наблюдается количественный рост статей по методам, использующим копула. В [1] копулы использовались в риск-анализе, в [2] в финансах, в [3]-[5] предложены тесты для сравнения копул и обнаружения разладок.

КОПУЛЫ: НЕКОТОРЫЕ ОПРЕДЕЛЕНИЯ И СВОЙСТВА

Копула – это функция, которая связывает одномерные функции распределения, чтобы получить многомерную функцию распределения. Если все величины имеют непрерывное распределение, то их копула – это просто многомерная функция распределения от своих равномерно на $(0, 1)$ маргинальных распределений. Рассмотрим вектор случайных переменных $x = (x_1, \dots, x_d)$ с совместной функцией распределения F и маргинальными распределениями $F_1, \dots, F_d$. По теореме Склера

$$F(x) = C(F_1(x_1), \dots, F_d(x_d)) \quad (1)$$

Традиционное использование корреляции для модели зависимости влечет гауссовскую копулу

$$C(u_1, \dots, u_d, \Sigma) = \Phi(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d), \Sigma)$$

где Σ – ковариационная матрица многомерного нормального распределения Φ . Близко связанной с гауссовской является копула студента с ν степенями свободы

$$C(u_1, \dots, u_d, \Sigma, \nu) = T(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d)), \Sigma, \nu)$$

Из любого многомерного распределения мы можем получить маргинальные распределения $F_1, \dots, F_d$ и копулу C . Но, что более важно, для моделирования временных рядов по любому множеству маргинальных распределений и копуле C , используя (1) мы можем получить совместное многомерное распределение. Очень важно, что маргинальные распределения могут быть разными, а выбор C не обусловлен выбором маргинальных распределений. Ясно, что копула C содержит всю информацию о зависимости переменных x_i . Поэтому, иногда, копулу называют “функцией зависимости”. Если многомерная функция распределения p раз дифференцируема, то можно показать:

$$f(x) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{i=1}^d f_i(x_i)$$

где

$$C(u_1, \dots, u_d) = \partial C(u_1, \dots, u_d) / \partial u_1 \partial u_2 \dots \partial u_d$$

- плотность d -мерной копулы $C(u_1, \dots, u_d)$.

Пусть дана выборка независимых одинаково распределенных случайных векторов

$$\{x^{(k)} = (x_1^{(k)}, \dots, x_d^{(k)})\}, k = 1, \dots, n.$$

Тогда из (4) получаем логарифм функции правдоподобия

$$L = L_C + \sum_{i=1}^d L_i$$

где

$$L_C = \sum_{k=1}^n \ln c(F_1(x_1^{(k)}), \dots, F_d(x_d^{(k)}))$$

- вклад из структуры зависимости, представленной копулой C и

$$L_i = \sum_{k=1}^n f_i(x_i^{(k)})$$

- вклад маргинальных плотностей. Предположим, что копула C принадлежит параметрическому семейству с параметром θ , а маргинальные распределения и одномерные плотности зависят от параметров

$$\alpha_i : F_i = F(x_i, \alpha_i), f_i = f(x_i, \alpha_i)$$

ОЦЕНИВАНИЕ ПАРАМЕТРОВ

Оценка максимального правдоподобия (MLE) находится из соотношения

$$(\hat{\alpha}, \hat{\theta}) = \arg \max_{\alpha, \theta} L(\alpha, \theta) = \arg \max_{\alpha, \theta} (L_C(\alpha, \theta) + \sum_{i=1}^d L_i(\alpha_i)), \alpha = (\alpha_1, \dots, \alpha_d)$$

Более привлекательным является 2-шаговый алгоритм (IFM), предложенный в [6]. На первом шаге оцениваются параметры маргинальных распределений, а затем

оцениваются параметры копулы, где в качестве α_i используются оценки, полученные на первом шаге. Хотя IFM алгоритм менее эффективен, он более привлекателен в вычислительном плане, позволяя большую гибкость при оценивании параметров маргинальных распределений. В частном случае многомерной нормальной плотности эти алгоритмы эквивалентны.

МЕТОДЫ ОБНАРУЖЕНИЯ РАЗЛАДКИ

Пусть дана выборка независимых одинаково распределенных случайных векторов

$$\{x^{(k)} = (x_1^{(k)}, \dots, x_d^{(k)})\}, k = 1, \dots, n$$

Простейший случай двух состояний, в котором структурные изменения происходят в определенные моменты времени t_0 , будет иметь плотность

$$f(x^{(k)}, \theta = \theta^1, \alpha = \alpha^1)$$

для первых t_0 наблюдений и

$$f(x^{(k)}, \theta = \theta^2, \alpha = \alpha^2)$$

для остальных $n - t_0$ наблюдений. Рассмотрим апостериорные и последовательные алгоритмы.

Апостериорный алгоритм. Логарифм функции правдоподобия имеет вид:

$$L(n_1) = \sum_{t < t_0} L(x^{(t)}, \alpha^1, \theta^1) + \sum_{t \geq t_0} L(x^{(t)}, \alpha^2, \theta^2)$$

А алгоритм обнаружения момента разладки

$$t_0 = \arg \max_{1 \leq k \leq n} L(k),$$

где оценки параметров находятся с помощью ранее рассмотренных алгоритмов.

Последовательные алгоритмы. Данный подход предполагает использование некоторой подходящей меры расстояния между распределениями и базируется на алгоритме кумулятивных сумм вида

$$S_t = \sum_{k=1}^t s_k,$$

где s_k - некоторая статистика. При этом правило подачи сигнала о разладке имеет вид

$$t_0 = \inf \{t \geq 1 : g_t \geq h\},$$

где

$$g_t = S_t - \min_{1 \leq k \leq t} S_k,$$

$h > 0$ - порог. Можно предложить следующие статистики.

1. Статистика, основанная на логарифме псевдоотношения правдоподобия, в которой истинные плотности заменяются копула плотностями

$$V_t = \sum_{k=1}^t v_k,$$

где

$$v_k = \ln[c(F_1(x_1^k, \alpha_1^{(1)}), \dots, c(F_d(x_d^k, \alpha_d^{(1)})) / c(F_1(x_1^k, \alpha_1^{(2)}), \dots, c(F_d(x_d^k, \alpha_d^{(2)})))] .$$

2. Статистика, основанная на расстоянии между распределениями, определяемом аналогом дивергенции Кульбака

$$D_t = \sum_{k=1}^t d_k ,$$

где

$$d_k = \int_X f(x, \alpha^1, \theta^1) v(x) dx + v_k, X \subset R^d ,$$

$$v(x) = \ln[c(F_1(x_1, \alpha_1^{(1)}), \dots, c(F_d(x_d, \alpha_d^{(1)})) / c(F_1(x_1, \alpha_1^{(2)}), \dots, c(F_d(x_d, \alpha_d^{(2)})))]$$

При использовании данных алгоритмов кумулятивных сумм необходимо выбрать расположение двух участков временного ряда для идентификации моделей. Оценки $\theta = \theta^1, \alpha = \alpha^1$ находят в фиксированном окне, например длины T . Оценки $\theta = \theta^2, \alpha = \alpha^2$ находим в скользящем окне такого же размера. Когда модели существенно различаются находим момент разладки.

ЛИТЕРАТУРА

1. Embrechts P.M.E., McNeil A.J., Straumann D. Correlation and dependency in risk management: Properties and pitfalls. Risk management : Value at Risk and Beyond, Cambridge University Press, 2002.
2. Frees E.W., Valdez E.A. Understanding relationships using copulas. North American Actuarial Journal, 2(1), 1998.
3. Remillard B., Scaillet O. Testing for equality between two copulas. Journal of Multivariate Analysis, vol.100, is.3, 2009.
4. Fermanian J.D., Scaillet O. Nonparametric estimation of copulas for time series. J.Risk, 5, 2003.
5. Xiaohong Chen.and Yanqin Fan. Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. The Canadian Journal of Statistics, Vol.33, 2005.
6. Joe H. Multivariate models and dependence concepts, *Monographs on Statistics and Applied Probability*, vol. 73. Chapman & Hall/CRC, New York, 2001.