Шведов А.С.[1], Свиязов В.А.[1]
Москва, НИУ ВШЭ[1]

Применение моделей, основанных на нечеткой логике, к финансовым временным рядам

Введенное в работе (Zadeh, 1965) понятие нечеткого множества стало одним из центральных в математике. Скажем, если X – это множество всех людей, то подмножество молодых людей по своему смыслу является нечетким. Степень принадлежности тридцатилетнего человека к этому нечеткому подмножеству может быть, например, 0.7, а степень принадлежности сорокалетнего человека – 0.3. При этом 0.7 и 0.3 не являются вероятностями, нечеткие множества – это принципиально другой подход к математическому моделированию неопределенности. И подход уже широко воспринятый практиками во многих областях. Не в последнюю очередь из-за того, что нечеткая теория математически проще, чем теория вероятностей. А также из-за того, что комбинирование подходов нечеткой теории и теории вероятностей во многих случаях оказывается полезным.

При рассмотрении некоторого универсального множества X любая функция $\mu : X \to [0,1]$ называется функцией принадлежности. График функции μ, то есть множество пар $\left(x, \mu(x)\right)$, $x \in X$, называется нечетким множеством. В основе раздела математики, называемого нечеткой теорией, обязательно должно лежать математически строгое определение нечеткого множества. И строгое определение именно такое, нечеткое множество – это график функции принадлежности. Распространенный в нечеткой теории оборот «обозначим через $\mu_A(x)$ функцию принадлежности нечеткого множества A» хотя и не содержит ничего неправильного, непривычен для математиков, работающих в других областях. График функции обозначается через A, а сама функция – через μ_A. А вот приведенная выше фраза «подмножество молодых людей по своему смыслу является нечетким» с точки зрения математической теории неправильная. Нечеткое множество – это не подмножество универсального множества. Хотя задачу описания множества молодых людей (понимаемого в обычном смысле) можно заменить на эквивалентную, но чуть более тяжеловесную задачу. Описать график индикатора множества молодых людей. (Напомним, что индикатором множества $E \subseteq X$ называется функция, принимающая значение 1 на множестве E и значение 0 – на множестве $X \setminus E$.) Тогда переход к нечеткому множеству молодых людей означает всего лишь замену индикатора на функцию принадлежности более общего вида.

С каждым нечетким множеством A может быть связана возможностная мера на X. Для любого множества $E \subseteq X$ значение возможностной меры определяется следующим образом: $Pos(E) = \sup\limits_{x \in E} \mu_A(x)$. Очевидно, что для любых двух множеств $E,F \subseteq X$ выполняется условие $Pos(E \cup F) = \max(Pos(E), Pos(F))$. Таким образом, обычное для вероятностной меры свойство аддитивности, что для непересекающихся множеств мера объединения равна сумме мер, для возможностной меры, вообще говоря, не имеет места. Название «возможностная мера» идет от рассмотрения функций принадлежности μ равных 1 в некоторой точке $x_0 \in X$ и равных 0 в других точках множества X. Тогда значение возможностной меры множества E дает ответ на вопрос, возможно x_0 или нет. Кроме возможностной меры в нечеткой теории часто используются еще две меры. Мера необходимости $Nec(E) = 1 - Pos(\bar{E})$, где $\bar{E} = X \setminus E$, и еще одна мера $Cr(E) = 0.5(Nec(E) + Pos(E))$. Нетрудно проверить, что для любого множества $E \subseteq X$ выполняется неравенство $Nec(E) \le Pos(E)$ и $Nec(E \cap F) = \min(Nec(E), Nec(F))$ для любых множеств $E,F \subseteq X$. Преимущество меры $Cr(\cdot)$ состоит в том, что $Cr(E) + Cr(\bar{E}) = 1$ для любого множества $E \subseteq X$.

В нечеткой логике важное значение имеет понятие *t*-нормы (или треугольной нормы). Функция $\Delta:[0,1]^2 \to [0,1]$ называется *t*-нормой, если $\Delta(x,y) = \Delta(y,x)$ для любых $x,y \in [0,1]$; $\Delta(x,1) = x$ для любого $x \in [0,1]$; $\Delta(x,\Delta(y,z)) = \Delta(\Delta(x,y),z)$ для любых $x,y,z \in [0,1]$; $\Delta(w,x) \le \Delta(y,z)$ для любых $w,x,y,z \in [0,1]$ таких, что $w \le y$, $x \le z$. Наиболее употребительными *t*-нормами являются $\Delta(x,y) = \min(x,y)$ и $\Delta(x,y) = xy$. С каждой *t*-нормой Δ связывается ассоциированная t-конорма ∇, также являющаяся отображением единичного квадрата $[0,1]^2$ в отрезок $[0,1]$ и определяемая при помощи формулы $\nabla(x,y) = 1 - \Delta(1-x, 1-y)$. Например, *t*-конорма $\nabla(x,y) = \max(x,y)$ является ассоциированной для *t*-нормы $\Delta(x,y) = \min(x,y)$.

В классической двузначной математической логике истинность каждого высказывания может равняться либо 1 (высказывание истинно), либо 0 (высказывание ложно). В нечеткой математической логике истинность высказывания A, обозначаемая $T(A)$, может быть любым действительным числом из отрезка $[0,1]$. После выбора *t*-нормы Δ (и, следовательно, ассоциированной с ней *t*-конормы ∇) истинность конъюнкции (то есть высказывания «A и B», где B – некоторое другое высказывание) определяется соотношением $T(A \wedge B) = \Delta(T(A), T(B))$. Истинность дизъюнкции (высказывания «A или B») определяется как $T(A \vee B) = \nabla\left(T(A), T(B)\right)$. Нечеткой импликацией называется любое отображение $[0,1]^2 \to [0,1]$, при котором точка (0,0) переходит в 1, точка (0,1) переходит в 1, точка (1,0) переходит в 0, точка (1,1) переходит в 1. Примером нечеткой импликации может быть отображение $(x,y) \to \nabla\left(1-x, \Delta(x,y)\right)$. Истинность высказывания $A \Rightarrow B$ (если A, то B) определяется как значение этого отображения в точке $\left(T(A), T(B)\right)$. Для определения истинности отрицания используется монотонно невозрастающая функция $N:[0,1] \to [0,1]$ такая, что $N(0)=1$, $N(1)=0$, $N(N(x))=x$ для любого $x \in [0,1]$. Примером такой функции является $N(x) = \left(1-x^a\right)^{1/a}$ при $a>0$. После выбора функции N истинность отрицания (то есть высказывания «не A») определяется соотношением $T(\neg A) = N\left(T(A)\right)$. Примеры с применением перечисленных операций нечеткой логики можно найти, например, в книгах (Пегат, 2013), (Рутковский, 2010).

Нечеткие системы Такаги–Сугено, предложенные в работе (Takagi, Sugeno, 1985), применяются и для статических, и для динамических задач. Для построения нечеткой системы берется некоторая выборка. После того, как нечеткая система построена, она используется следующим образом. По входному значению x (скалярному или векторному) строится выходное значение y. Нечеткая система состоит из нескольких нечетких правил. Например, нечеткое правило может быть таким. Если волатильность доходности фондового индекса низкая и доходность фондового индекса высокая, то использовать модель обобщенной авторегрессионной условной гетероскедастичности с такими-то параметрами. Антецедент другого нечеткого правила может иметь вид «волатильность доходности фондового индекса высокая и доходность фондового индекса средняя»; в этом случае параметры GARCH-модели будут другими. Степень активации каждого нечеткого правила определяется (для выбранного участка временного ряда) истинностью соответствующей конъюнкции (в смысле нечеткой логики). Выходное значение y рассчитывается как взвешенное среднее выходных значений всех нечетких правил с учетом степени активации каждого нечеткого правила. Заметим, что, хотя каждое нечеткое правило напоминает импликацию; если (…), то (…); истинность импликации в данном случае не рассчитывается.

В исследовании производится прогнозирование волатильности доходности фондовых индексов и отдельных акций российского финансового рынка. Модель, используемая для прогнозирования, является сочетанием системы нечёткого вывода Такаги–Сугено и модели GARCH. Ряд волатильности разбивается на C нечётких кластеров методом eTS (evolving Takagi–Sugeno), данный метод предложен в работе (Angelov, Filev, 2004). Каждому кластеру ставится в соответствие нечёткое правило, представляющее собой расчёт выходного значения при помощи локальной модели. Локальная модель в каждом нечетком правиле имеет вид модели GARCH. В совокупности результаты каждого нечеткого правила составляют нечёткое выходное значение, которое затем объединяется в скалярное выходное значение.

Волатильность моделируется следующим образом: $a_t = \sqrt{h_t}\,\varepsilon_t$. Здесь a_t – ряд волатильности; ε_t – процесс белого шума; $\sqrt{h_t}$ – гетероскедастичная составляющая волатильности. Каждое нечеткое правило представимо в следующем виде.

$$\text{ЕСЛИ } a_t \in F_k, \text{ ТО } h_t^{(k)} = \alpha_{k0} + \sum_{i=1}^{q} \alpha_{ki} a_{t-i}^2 + \sum_{j=1}^{p} \beta_{kj} h_{t-j}^{(k)},$$

$\alpha_{k0} > 0,\ \alpha_{ki} > 0,\ \beta_{kj} > 0\ \ \forall k, i, j$. Здесь k – номер нечеткого кластера, $k = 1,\ldots,C$; F_k – нечеткий кластер. Оцениваются параметры консеквента α_{ki} и β_{kj}.

Истинность высказывания $a_t \in F_k$ принимает значение из отрезка $[0,1]$ и вычисляется при помощи гауссовской функции принадлежности. Параметры функций принадлежности (параметры антецедента), а также структура кластеров, калибруются методом eTS. В модели eTS нечёткая кластеризация производится в режиме онлайн: как только в модель добавляется новая информация на новом шаге, запускается рекурсивный алгоритм обновления структуры нечётких кластеров и пересчёта параметров антецедента. В результате работы алгоритма принимается одно из следующих решений относительно структуры кластеров:

- структура кластеров остаётся неизменной;
- один из текущих кластеров заменяется новым;
- добавляется новый кластер.

Параметры консеквента оцениваются отдельно после оценки параметров антецедента численным методом наименьших квадратов.

Исследование было проведено на трех фондовых индексах российского рынка (в скобках приведены используемые далее обозначения): Индексе РТС (RTSI), Индексе нефти и газа (MOEXOG) и Индексе потребительского сектора (MOEXCN) – а также на пяти (обыкновенных) акциях фондового рынка РФ, входящих в топ 10 акций с наибольшим среднедневным оборотом на Московской бирже: акции Сбербанка (SBER), акции Газпрома (GAZP), акции Лукойла (LKOH), акции Полюса (PLZL) и акции Яндекса (YNDX).

В качестве исходных временных рядов были использованы дневные логарифмические доходности финансовых инструментов. Для каждого инструмента было взято по три временных диапазона:

1) Начало – 1 января 2018 года, выборка длиной 2 года;

2) Начало – 1 января 2015 года, выборка длиной 5 лет;
3) Начало – 1 января 2010 года, выборка длиной 10 лет (кроме акции Яндекса, которая начала торговаться на Московской Бирже в 2014 году).

Все рыночные данные, использованные в исследовании, были выгружены с сайта finam.ru.

Результаты работы моделей на различных выборках – среднеквадратичные ошибки прогноза – отображены в таблице 1.

Таблица 1

Среднеквадратичная ошибка (RMSE) нечёткой модели (Fuzzy GARCH) и классической модели (GARCH) на разных выборках.

Финансовый инструмент	**Начало выборки**	**Размер выборки**	**Fuzzy GARCH**	**GARCH**
RI.MOEXCN	2018-01-03	502	0.000178	0.000160
	2015-01-05	1258	0.000152	0.000140
	2010-01-11	2518	0.000644	0.008715
RI.MOEXOG	2018-01-03	502	0.000238	0.000202
	2015-01-05	1258	0.000284	0.000218
	2010-01-11	2518	0.000397	0.000331
LKOH	2018-01-03	502	0.000525	0.000485
	2015-01-05	1258	0.000603	0.000460
	2010-01-11	2518	0.000569	0.000481
GAZP	2018-01-03	502	0.001382	0.001177
	2015-01-05	1258	0.000956	0.000804
	2010-01-11	2518	0.000994	0.000838
YNDX	2018-01-03	502	0.003670	0.003136
	2015-01-05	1258	0.002657	0.002310

Продолжение Таблицы 1

Финансовый инструмент	**Начало выборки**	**Размер выборки**	**Fuzzy GARCH**	**GARCH**
RTSI	2018-01-03	502	0.000770	0.000682

	2015-01-05	1258	0.000873	0.004551
	2010-01-11	2518	0.001017	0.001509
PLZL	2018-01-03	502	0.002067	0.002523
	2015-01-05	1258	0.002235	0.001519
	2010-01-11	2518	0.002795	0.002356
SBER	2018-01-03	502	0.002686	0.001650
	2015-01-05	1258	0.001872	0.001411
	2010-01-11	2518	0.001689	0.001223

Из 23 рассмотренных временных рядов больше, чем вдвое различаются среднеквадратичные ошибки лишь для 2 временных рядов (0.000644 против 0.008715; 0.000873 против 0.004551), и в обоих случаях преимущество у нечеткой модели. Это показывает следующие направления исследований. Анализ тех свойств временного ряда, при которых применение для прогнозирования нечетких систем оказывается полезным, и использование более совершенных методов разбиения на нечеткие кластеры.

Исследование выполнено с использованием суперкомпьютерного комплекса НИУ ВШЭ (Kostenetskiy et al., 2021).

Список использованной литературы:

1. Пегат А. Нечеткое моделирование и управление. Москва: БИНОМ. Лаборатория знаний, 2013. – 798 с.
2. Рутковский Л. Методы и технологии искусственного интеллекта. Москва: Горячая линия – Телеком, 2010. – 520 с.
3. Angelov P.P., Filev D.P. An approach to online identification of Takagi-Sugeno fuzzy models // IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. 2004. Vol. 34. P. 484–498.
4. Kostenetskiy P.S., Chulkevich R.A., Kozyrev V.I. HPC resources of the Higher School of Economics // Journal of Physics: Conference Series. 2021. Vol. 1740(1). P. 12050.
5. Takagi T., Sugeno M. Fuzzy identification of systems and its applications to modeling and control // IEEE Transactions on Systems, Man, and Cybernetics. 1985. Vol. 15. P. 116–132.
6. Zadeh L.A. Fuzzy sets // Information and Control. 1965. Vol. 8. P. 338–353.