Экономический журнал ВШЭ. 2017. Т. 21. № 3. С. 434–450.
HSE Economic Journal, 2017, vol. 21, no 3, pp. 434–450.

Анализ динамики фондовых индексов с использованием нечетких моделей Такаги – Сугено

Могилевич Е.О., Шведов А.С.

В данной работе проведены оценка параметров и исследование применимости моделей Такаги – Сугено для описания динамики фондовых индикаторов на примере основных индексов Московской биржи: ММВБ, РТС и отраслевого индекса нефти и газа. Дается обзор литературы по применению моделей Такаги – Сугено для прогнозирования некоторых зарубежных фондовых индексов и цен акций. Модели Такаги – Сугено представляют собой обобщение классических эконометрических подходов, это обобщение достигается за счет использования систем нечетких правил. Каждая модель Такаги – Сугено может рассматриваться как модификация некоторой линейной эконометрической модели. При этом существующие результаты из теории аппроксимации показывают, что при помощи модели Такаги – Сугено может быть приближенно представлена и произвольная нелинейная эконометрическая модель. В данной работе строятся модели Такаги – Сугено для российских фондовых индексов, для нахождения функций принадлежности используется метод нечеткой кластеризации. Коэффициенты линейной зависимости в каждом нечетком правиле находятся при помощи процедуры Сугено – Канга, основанной на применении метода наименьших квадратов. Проведенные расчеты показывают, что модель Такаги – Сугено дает уменьшение ошибки прогноза по сравнению с немодифицированной линейной моделью. В некоторых примерах ошибка прогноза уменьшается примерно в 4 раза.

Ключевые слова: нечеткие системы; рынок акций; регрессия, прогнозирование.

1. Введение

Проблема моделирования фондовых показателей является актуальной, поскольку от точности прогноза напрямую зависит прибыль инвесторов. Поэтому в данной области непрерывно ведутся исследования и применяются новые модели.

Могилевич Елена Олеговна – студентка 2 курса магистратуры факультета экономических наук Национального исследовательского университета «Высшая школа экономики». E-mail: gurjanova-elena@yandex.ru.

Шведов Алексей Сергеевич – доктор физ.-мат. наук, профессор департамента прикладной экономики факультета экономических наук Национального исследовательского университета «Высшая школа экономики». E-mail: ashvedov@hse.ru.

Статья поступила: 29.06.2017/Статья принята: 27.09.2017.

Модели Такаги – Сугено были предложены в работе [Takagi, Sugeno, 1985] и с тех пор получили большие применения в различных прикладных областях (см., например, [Пегат, 2009; Рутковская и др., 2013]). Существуют и экономические приложения, хотя и не столь многочисленные, как в других областях. Иногда эти модели называют также моделями Такаги – Сугено – Канга, поскольку в работе [Sugeno, Kang, 1988] предложен способ оценки параметров модели, основанный на применении метода наименьших квадратов и ставший широко распространенным. Также эти модели называют TS-моделями или TSK-моделями.

Особенностью этих моделей является то, что за счет использования систем нечетких правил оказывается возможным, не выходя за рамки линейных зависимостей, учитывать то, что влияние объясняющих переменных на объясняемые при различных условиях может быть разным. Успешность TS-моделей во многом объясняется тем, что конкретный вид функциональной зависимости (вообще говоря, нелинейной) одних показателей от других не предполагается. Если в классических нелинейных эконометрических моделях исследователь, как правило, более или менее произвольно с самого начала выбирает форму нелинейной зависимости, то нечеткие модели Такаги – Сугено сами «нащупывают» нужную форму нелинейности. Разумеется, мы не собираемся оспаривать важность применения нелинейных эконометрических моделей; эти модели «работают» и во многих случаях «работают» очень успешно.

На сегодняшний день существуют различные подходы к имплементации моделей Такаги – Сугено. Например, большую популярность имеет система ANFIS — Adaptive-Network-Based Fuzzy Inference System – нейро-нечеткая система на базе нечетких правил Такаги – Сугено [Jang, 1993]. В экономической сфере TS-модели применяются для построения прогнозов фондовых индексов и курсов акций различных компаний (подробнее об этих приложениях см. раздел 2). Но существуют и другие приложения: так, в работах [Olej, Kfupka, 2005; Olej, 2005] модель Такаги – Сугено применяется для предсказания значений ВВП; в работах [Rodriguez, Anders, 2004; Pingan, Xiaohong, 2000] моделируются цены на электроэнергию.

В данной работе используется модификация модели Такаги – Сугено, которая базируется на нечеткой кластеризации c-means, для изучения динамики российских фондовых индексов. Структура работы следующая. В разделе 2 дается обзор некоторых публикаций по построению TS-моделей для финансовых временных рядов. В разделе 3 излагаются основные идеи, лежащие в основе TS-моделей. В разделе 4 представлен один из существующих подходов к поиску оптимального числа и вида нечетких правил для моделей Такаги – Сугено. В разделах 5 и 6 построен ряд моделей Такаги – Сугено для основных индексов Московской биржи: индексов ММВБ, РТС и отраслевого индекса нефти и газа; изучаются прогностические свойства этих моделей.

2. Обзор литературы по некоторым применениям моделей Такаги – Сугено для изучения фондовых индексов

Традиционный подход к изучению динамики фондовых индикаторов и курсов акций можно разделить на следующие ветви. Первая ветвь – это прогнозирование на базе анализа макроэкономических показателей, таких как экспорт, импорт, обменный курс, уровень инфляции и безработицы, процентные ставки, персональные показатели деятельности компании и других фундаментальных величин. Вторая ветвь базируется на познании законов поведения фондовых рынков и основывается на гипотезе, что «история повторяет саму себя», тем самым подразумевая, что корреляция между значениями за прошлые промежутки времени и настоящие является определяющим звеном при формировании актуальной цены. Такой технический подход включает моделирование различных индексов и основывается на регрессионном анализе либо на анализе временных рядов.

Модель Такаги – Сугено, как правило, применяется с использованием экзогенных переменных и поэтому несет в себе черты обоих подходов. Остановимся на некоторых публикациях по моделям Такаги – Сугено. В работе [Sheta, 2006] анализ недельных значений индекса S&P 500 проводится на базе среднего показателя за прошлый промежуток времени, но при этом используются также и информация об облигациях казначейства США, о дивидендах и доходности для отдельных акций. По результатам исследования модель показала хорошие результаты как в оценке параметров модели (построенные оценки достаточно информативные), так и в качестве прогнозов.

В работе [Boyacioglu, Avci, 2010] используется система ANFIS для прогнозирования Turkey ISE National 100 Index. Опираясь на гипотезу о том, что доходность индекса может быть предсказана на основе макроэкономических данных и финансовых показателей компаний, в этом исследовании в качестве экзогенных переменных берутся цены на золото, обменный курс доллара, процент по депозитам, индексы потребительских цен и цен производителя, ставка по облигациям. Также авторы статьи включили в модель значения индексов ряда других фондовых бирж. В работе исследовано влияние числа нечетких правил и вида функции принадлежности на прогнозную силу модели.

Применяется также подход с использованием информации об историческом тренде технических индексов. В работе [Chang et al., 2004] предложено брать в качестве независимых переменных следующие технические индексы: скользящее среднее, отклонение от скользящего среднего, индекс относительной силы и другие. В работе [Chang, Liu, 2008] TS-модель с использованием данного подхода применяется к прогнозированию TSE Index, взвешенного по капитализации индекса Тайваньской фондовой биржи, и курса акций компании MediaTek Inc. По итогам расчетов TS-модель показала очень хорошее улучшение прогнозной силы – средняя абсолютная процентная ошибка упала почти в 3 раза. В работе [Chang, Fan, 2008] используется модифицированная TS-модель с вейвлет-преобразованием исходных данных. Примененная трансформация показала небольшое улучшение по сравнению с обычной TS-моделью, значительно меньшее, чем улучшение при использовании вейвлет-преобразования для обычной регрессии.

В работе [Esfahanipour, Aghamiri, 2010] TS-модель строится с использованием тех же индексов, что и в [Chang et al., 2004], применительно к некоторым индексам фондовой биржи Тегерана; TEPIX и другим. В статье используется метод c-means нечеткой кластеризации. Результаты по прогнозной силе модели получаются схожие с результатами статьи [Chang, Liu, 2008].

3. Общее описание TS-модели

При использовании классической импликации эконометрическая модель могла бы иметь, например, следующий вид:

ЕСЛИ (x_1 малое), ТО $y = \alpha_0^1 + \beta_1^1 x_1 + \beta_2^1 x_2$,

ЕСЛИ (x_1 среднее), ТО $y = \alpha_0^2 + \beta_1^2 x_1 + \beta_2^2 x_2$,

ЕСЛИ (x_1 большое), ТО $y = \alpha_0^3 + \beta_1^3 x_1 + \beta_2^3 x_2$,

где $\alpha_0^1, \alpha_0^2, \alpha_0^3, \beta_1^1, \beta_1^2, \beta_1^3, \beta_2^1, \beta_2^2, \beta_2^3$ – параметры модели (см. рис. 1). С одной стороны, это хорошо, что в модели допускаются различные коэффициенты для передачи одной и той же зависимости при различных условиях. Но возникают серьезные вопросы. Оправдано ли, например, что значение x_1 = 3,99 относится к средним, а значение x_1 = 4,01 – к большим? Полезно ли резкое изменение вида зависимости при переходе через точки x_1 = 0 и x_1 = 4? Или все же лучше взять какую-то нелинейную модель? Эти возражения снимаются, если вместо классических импликаций используются нечеткие импликации, как это делается в моделях Такаги – Сугено.

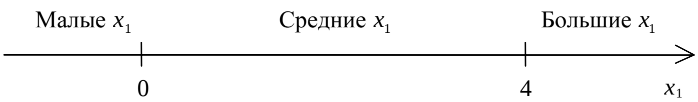


Рис. 1. Границы областей для переменной x_1

Пусть нечеткие множества A^1, A^2, A^3 означают соответственно малые, средние и большие значения x_1 (определение нечеткого множества и исторические ссылки см., например, в работе [Шведов, 2017]). Функции принадлежности $\mu_{A^1}(x_1), \mu_{A^2}(x_1), \mu_{A^3}(x_1)$ этих нечетких множеств изображены на рис. 2.

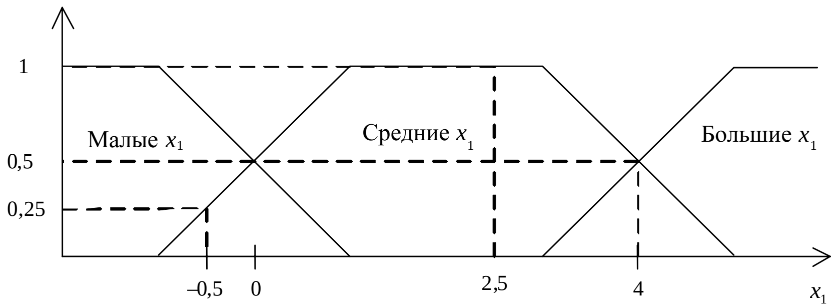


Рис. 2. Функции принадлежности нечетких множеств A^1, A^2, A^3

Например, степень принадлежности точки x_1 = –0,5 к множеству средних значений равняется 0,25. Степень принадлежности x_1 = 2,5 к множеству средних значений равняется единице. Степень принадлежности x_1 = 4 к множеству больших значений равняется 0,5.

В модели Такаги – Сугено нечеткие правила будут иметь следующий вид:

$$\text{ЕСЛИ } \left(x_1 = A^1\right), \text{ТО } y^1 = \alpha_0^1 + \beta_1^1 x_1 + \beta_2^1 x_2 ,$$

$$\text{ЕСЛИ } \left(x_1 = A^2\right), \text{ ТО } y^2 = \alpha_0^2 + \beta_1^2 x_1 + \beta_2^2 x_2 ,$$

$$\text{ЕСЛИ } \left(x_1 = A^3\right), \text{ ТО } y^3 = \alpha_0^3 + \beta_1^3 x_1 + \beta_2^3 x_2 .$$

Запись ЕСЛИ $\left(x_1 = A^1\right)$ можно понимать так же, т.е. ЕСЛИ (x_1 малое). Но теперь это нечеткое правило действует не только для $x_1 \le 0$, а для всех x_1, для которых $\mu_{A^1}\left(x_1\right) > 0$. Вклад первого нечеткого правила зависит от величины $\mu_{A^1}\left(x_1\right)$.

Прогнозное значение $\hat{y}$ будет определяться следующим выражением:

$$\hat{y} = \frac{\sum_{i=1}^{I} G^i y^i}{\sum_{i=1}^{I} G^i} , \qquad (1)$$

где G^i – действительное число, обозначающее степень правдивости *i*-го правила. Здесь I – число нечетких правил, в приведенном примере I = 3. В общем случае условие для *i*-го нечеткого правила будет иметь следующий вид:

$$\text{ЕСЛИ } \left(x_1 = A_1^i\right) \text{ И ... И } \left(x_m = A_m^i\right),$$

где $A_1^i, \dots, A_m^i$ – нечеткие множества, $i = 1, \dots, I$.

В работе [Takagi, Sugeno, 1985] G^i задается по следующей формуле:

$$G^i = \min\left(\mu_{A_1^i}\left(x_1\right), \dots, \mu_{A_m^i}\left(x_m\right)\right). \qquad (2)$$

В данной работе G^i рассчитывается по следующей формуле:

$$G^i = \mu_{A_1^i}\left(x_1\right) \times \dots \times \mu_{A_m^i}\left(x_m\right). \qquad (3)$$

Обе формулы (2) и (3) для нахождения G^i являются широко распространенными. Могут применяться и другие *t*-нормы. Определение треугольной нормы для случая двух переменных можно найти, например, в книге [Рутковская и др., 2013]. Аналогично определяется треугольная форма и для случая произвольного числа переменных.

В данном примере число нечетких правил и функции принадлежности заданы изначально. Во многих практических ситуациях (в том числе и в приводимых ниже расчетах) и то, и другое определяется исходя из данных. Нечеткими могут быть и исходные данные. Однако в настоящей работе мы ограничимся случаем, когда исходные данные четкие.

4. Определение функций принадлежности и параметров TS-модели

Вопросы идентификации TS-моделей рассматриваются во многих работах. Назовем, например, книги [Lilly, 2011; Zhang, Liu, 2006]. Для определения оптимального числа нечетких правил и нахождения функций принадлежности существуют различные подходы. Часто предполагается, что функция принадлежности относится к некоторому семейству (например, гауссовские функции принадлежности), и для определения параметров используется некоторая оптимизационная процедура. В этом разделе дается краткое описание того подхода, который применяется в настоящей работе. В качестве исходных данных для построения TS-модели нами берутся *m*-мерные наблюдения $x_t = \left(x_{1,t},\dots,x_{m,t}\right), t = 1,\dots,n$, (объясняющие переменные) и одномерные наблюдения $y_1,\dots,\mathrm{y}_n$ (объясняемая переменная).

Процесс построения модели ведется итерационно. Перед началом каждой итерации формат системы нечетких правил считается выбранным. Определяются значения G^i (см. формулу (1)) и коэффициенты линейного уравнения для каждого нечеткого правила.

Каждая итерация состоит из двух шагов.

Шаг 1. Определение степеней принадлежности.

Этот шаг заключается в разделении *m*-мерных наблюдений на заданное число нечетких кластеров. Используется алгоритм нечеткой кластеризации c-means (см., например, [Davé, Krishnapuram, 1997]). Данный метод позволяет сопоставить одно наблюдение одновременно с несколькими кластерами с разной степенью принадлежности. Число кластеров равняется числу нечетких правил *I*.

Шаг 2. Определение коэффициентов линейных уравнений в каждом нечетком правиле.

На этом шаге параметры определяются с помощью метода наименьших квадратов. Согласно формуле (1), при расчете $\hat{y}_t$ в качестве независимых переменных будут выступать $\dfrac{G^i}{\sum_{k=1}^{I} G^k}$ и $\dfrac{G^i x_{j,t}}{\sum_{k=1}^{I} G^k}$, где $t = 1,\dots,n, i = 1,\dots, I, j = 1,\dots,m$. Таким образом методом наименьших квадратов будет оцениваться линейная модель с $I \times (m+1)$ коэффициентами.

В качестве критериев при проведении итерационного процесса используются среднеквадратичная ошибка прогноза RMSE и средняя абсолютная процентная ошибка MAPE, которые вычисляются по следующим формулам:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}}, \tag{4}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{y_t}. \tag{5}$$

Цикл повторяется для итераций с различным числом нечетких правил и с различными наборами переменных в каждом нечетком правиле. В качестве оптимальной выбирается система нечетких правил, дающая минимальную абсолютную процентную ошибку MAPE.

Нечеткие правила подбирались на основе алгоритма, представленного в работе [Zhang, Liu, 2006]. Итерационный процесс разбивается на этапы. К каждому этапу относятся итерации, для которых количество нечетких правил одинаково. Но вид этих нечетких правил меняется от итерации к итерации.

На каждой итерации система нечетких правил строится путем некоторого изменения системы нечетких правил предыдущей итерации. При каждой итерации рассчитываются RMSE и MAPE, однако значение MAPE является решающим – по нему определяется, имеет ли смысл продолжать итерационный процесс.

Этап 1. Используется одно нечеткое правило. Соответствующее линейное уравнение будем называть базовой эконометрической моделью.

Этап 2. Сначала модель с двумя нечеткими правилами задается следующим образом:

$$R^1: \text{ЕСЛИ } (x_1 = A_1), \text{ ТО ...,}$$
$$R^2: \text{ЕСЛИ } (x_1 = A_2), \text{ ТО ....}$$

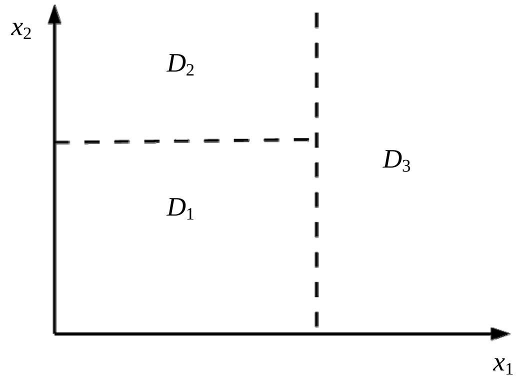
Данная процедура повторяется для каждой независимой переменной, для каждой модели считались MAPE и RMSE. Тогда x_i, для которого модель давала наименьший MAPE, далее в тексте и на рисунках будет обозначаться как x_1. Таким образом, на этапе 2 проводится *m* итераций.

Этап 3. На этом этапе при построении нечетких правил сначала используются только значения для x_1 (ср. пример из раздела 3), однако теперь наблюдения делятся на 3 кластера. Потом в нечеткие правила добавляется вторая переменная путем деления левого или правого кластера на 2 части. Нечеткие правила тогда выглядят следующим образом:

$$R^1: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_2 = B_1), \text{ ТО ...,}$$
$$R^2: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_2 = B_2), \text{ ТО ...,}$$
$$R^3: \text{ЕСЛИ } (x_1 = A_2), \text{ ТО ....}$$

В качестве второй переменной поочередно выступают x_j, где $j \neq 1$ из предыдущего этапа, т.е. все независимые переменные, кроме x_1. Модификация, дающая наименьшее значение MAPE, запоминается и используется на следующем этапе в качестве базовой.

Этап 4. На этом этапе в модель добавляется четвертое правило. Если на предыдущем этапе наименьший MAPE давала модификация только с одной переменной, то новое правило добавляется путем деления переменной на 4 кластера. Если на предыдущем этапе наименьший MAPE давала модификация с двумя переменными, то новое правило можно было добавить путем деления второй переменной на 3 кластера, либо с добавлением третьей переменной. Для определенности предположим, что на этапе 3 наименьшее значение MAPE давало следующее деление (см. рис. 3).



D_1 – область значений (x_1, x_2), для которых вклад первого нечеткого правила является наибольшим (т.е. $G^1 \geq G^2, G^1 \geq G^3$);
D_2 – область значений (x_1, x_2), для которых вклад второго нечеткого правила является наибольшим (т.е. $G^2 \geq G^1, G^2 \geq G^3$);
D_3 – область значений (x_1, x_2), для которых вклад третьего нечеткого правила является наибольшим (т.е. $G^3 \geq G^1, G^3 \geq G^2$).

Рис. 3. Деление на кластеры на этапе 3, дающее минимальное значение MAPE

Тогда на этапе 4 будут сравниваться следующие системы нечетких правил.

Система нечетких правил 1:

$$R^1: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_2 = B_1), \text{ ТО ...},$$
$$R^2: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_2 = B_2), \text{ ТО ...},$$
$$R^3: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_2 = B_3), \text{ ТО ...},$$
$$R^4: \text{ЕСЛИ } (x_1 = A_2), \text{ ТО ....}$$

Система нечетких правил 2:

$$R^1: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_2 = B_1) \text{ И } (x_3 = C_1), \text{ТО } ...,$$
$$R^2: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_2 = B_1) \text{ И } (x_3 = C_2), \text{ТО } ...,$$
$$R^3: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_2 = B_2), \text{ТО } ...,$$
$$R^4: \text{ЕСЛИ } (x_1 = A_2), \text{ТО } ....$$

Система нечетких правил 3:

$$R^1: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_2 = B_1), \text{ТО } ...,$$
$$R^2: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_2 = B_2) \text{ И } (x_3 = C_1), \text{ТО } ...,$$
$$R^3: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_2 = B_2) \text{ И } (x_3 = C_2), \text{ТО } ...,$$
$$R^4: \text{ЕСЛИ } (x_1 = A_2), \text{ТО } ....$$

Система нечетких правил 4:

$$R^1: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_2 = B_1), \text{ТО } ...,$$
$$R^2: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_2 = B_2), \text{ТО } ...,$$
$$R^3: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_3 = C_1), \text{ТО } ...,$$
$$R^4: \text{ЕСЛИ } (x_1 = A_1) \text{ И } (x_3 = C_2), \text{ТО } ....$$

В качестве x_3 поочередно выступают все переменные, кроме x_1 и x_2, лучшей признается модель с минимальным значением MAPE.

Этап 5. Последующие этапы строятся аналогично до тех пор, пока значение MAPE на этапе k не станет больше, чем MAPE на этапе $k-1$.

5. Пример построения модели Такаги – Сугено с экзогенными переменными

В данной работе модели Такаги – Сугено строились для основных индексов Московской биржи (ММВБ и РТС) и отраслевого индекса нефти и газа, который был выбран для исследования, поскольку экономика России напрямую зависит от экспорта углеводородов, что подчеркивает актуальность прогнозирования данного индекса.

В этом разделе рассматривается только один временной ряд – индекс ММВБ. Используются недельные данные с января 2010 г. по апрель 2017 г. Длина временного ряда составляет 188 наблюдений.

В данном примере не производится деления на тестовую и обучающую выборки. Целью расчета является построение прогноза на неделю вперед с учетом имеющейся информации.

Для построения эконометрической модели выбраны следующие показатели:

y_t – среднее значение индекса ММВБ за неделю $t, t = 0,1,\ldots,187$;

x_{1t} – количество акций, проданных за неделю t (млрд руб.);

x_{2t} – значение для недели t кривой бескупонной доходности государственных облигаций со сроком погашения 1 год (% годовых);

x_{3t} – средняя доходность индекса за неделю, которая рассчитывается по формуле:

(6) $$x_{3t} = \frac{y_{t-1}}{y_{t-2}}.$$

Такой выбор экзогенных переменных позволяет отразить и положение на рынке акций, и положение на рынке облигаций.

В качестве базовой используется следующая линейная модель:

(7) $$y_t = \alpha_0 + \beta_1 y_{t-1} + \beta_2 x_{1,t} + \beta_3 x_{2,t} + \beta_4 x_{3,t}.$$

Далее видно, что все переменные влияют на формирование нечетких правил.

Правила для модели Тагаки – Сугено строились согласно алгоритму, описанному выше. Анализ показал, что модель с 7 нечеткими правилами является наилучшей. Система нечетких правил имеет следующий вид:

$$R^1: \text{ЕСЛИ } \left(x_{2,t} = A_1\right) \text{ И } \left(y_{t-1} = B_1\right) \text{ И } \left(x_{1,t} = C_1\right) \text{ И } \left(x_{3,t} = D_1\right), \text{ТО}$$
$$y_t^1 = \alpha_0^1 + \beta_1^1 y_{t-1} + \beta_2^1 x_{1,t} + \beta_3^1 x_{2,t} + \beta_4^1 x_{3,t},$$
$$R^2: \text{ЕСЛИ } \left(x_{2,t} = A_1\right) \text{ И } \left(y_{t-1} = B_1\right) \text{ И } \left(x_{1,t} = C_1\right) \text{ И } \left(x_{3,t} = D_2\right), \text{ТО}$$
$$y_t^2 = \alpha_0^2 + \beta_1^2 y_{t-1} + \beta_2^2 x_{1,t} + \beta_3^2 x_{2,t} + \beta_4^2 x_{3,t},$$
$$R^3: \text{ЕСЛИ } \left(x_{2,t} = A_1\right) \text{ И } \left(y_{t-1} = B_1\right) \text{ И } \left(x_{1,t} = C_2\right), \text{ТО}$$
$$y_t^3 = \alpha_0^3 + \beta_1^3 y_{t-1} + \beta_2^3 x_{1,t} + \beta_3^3 x_{2,t} + \beta_4^3 x_{3,t},$$
$$R^4: \text{ЕСЛИ } \left(x_{2,t} = A_1\right) \text{ И } \left(y_{t-1} = B_1\right) \text{ И } \left(x_{1,t} = C_3\right), \text{ТО}$$
$$y_t^4 = \alpha_0^4 + \beta_1^4 y_{t-1} + \beta_2^4 x_{1,t} + \beta_3^4 x_{2,t} + \beta_4^4 x_{3,t},$$
$$R^5: \text{ЕСЛИ } \left(x_{2,t} = A_1\right) \text{ И } \left(y_{t-1} = B_2\right) \text{ ТО},$$
$$y_t^5 = \alpha_0^5 + \beta_1^5 y_{t-1} + \beta_2^5 x_{1,t} + \beta_3^5 x_{2,t} + \beta_4^5 x_{3,t},$$
$$R^6: \text{ЕСЛИ } \left(x_{2,t} = A_2\right) \text{ ТО},$$
$$y_t^6 = \alpha_0^6 + \beta_1^6 y_{t-1} + \beta_2^6 x_{1,t} + \beta_3^6 x_{2,t} + \beta_4^6 x_{3,t},$$
$$R^7: \text{ЕСЛИ } \left(x_{2,t} = A_3\right) \text{ ТО},$$
$$y_t^7 = \alpha_0^7 + \beta_1^7 y_{t-1} + \beta_2^7 x_{1,t} + \beta_3^7 x_{2,t} + \beta_4^7 x_{3,t}.$$

Здесь $A_1, A_2, A_3, B_1, B_2, C_1, C_2, C_3$ – некоторые нечеткие множества, функции принадлежности которых строятся при помощи алгоритма c-means, $\alpha_0^i, \beta_1^i, \beta_2^i, \beta_3^i, \beta_4^i$, $i = 1, \ldots, 7$ – рассчитанные коэффициенты уравнений.

В табл. 1 представлены результаты измерений, жирным шрифтом выделены результаты для оптимального числа правил. Столбец с одним нечетким правилом соответствует результатам, полученным по классической линейной модели.

Таблица 1.

Результаты построения TS-модели для разного числа правил

	Количество правил							
	1	2	3	4	5	6	7	8
MAPE	1,683	1,649	1,627	1,591	1,510	1,510	**1,394**	1,512
RMSE	33,904	33,216	32,370	31,912	30,104	29,836	**27,207**	30,665

Из табл. 1 видно, что модель Тагаки – Сугено дает видимое уменьшение ошибок регрессии: MAPE с 1,68 до 1,39%, RMSE с 33,9 до 27,2, причем уменьшение ошибки прогноза видно уже для первого правила. RMSE рассчитывается по формуле (4), MAPE рассчитывается по формуле (5), $n = 188$.

6. Пример построения модели Такаги – Сугено с техническими индексами

В данном примере используются ежедневные значения индексов ММВБ, РТС и отраслевого индекса нефти и газа с января 2014 г. по апрель 2017 г., всего $n_1 = 494$ наблюдения для обучающей выборки и $n_2 = 334$ наблюдения для тестовой выборки.

Один из наиболее популярных подходов построения моделей для прогнозов индексов фондовых бирж состоит в построении регрессии, где в качестве независимых переменных берутся технические индексы. Такой подход используется в работе [Chang, Liu, 2008], где представлена модель для взвешенного по капитализации индекса Тайваньской фондовой биржи TSE Index и курса акций компании MediaTek Inc. Оказывается, модель Тагаки – Сугено может существенно улучшить прогноз по сравнению с обычной регрессией. В качестве переменных выбраны следующие показатели:

- Скользящее среднее за последние 6 дней:

$$x_{1,\mathrm{k}} = \frac{\sum_{t=k-6}^{k} y_t}{6}, \tag{8}$$

где y_t – значение индекса на момент закрытия t-го дня, $t = 1, \ldots, 424$ для обучающей выборки и $t = 1, \ldots, 334$ для тестовой выборки.

- Отклонение *t*-го наблюдения от скользящего среднего:

(9) $$x_{2,t} = \frac{y_t - x_{1,t}}{x_{1,t}} \cdot 100\% .$$

В данном примере изучается зависимость среднеквадратичной ошибки RMSE и абсолютной средней процентной ошибки MAPE от числа правил. Поскольку переменных всего две, то нет необходимости проводить итерационный процесс, как это изложено в разделе 4, достаточно рассмотреть все возможные варианты.

Сначала описывается построение модели Такаги – Сугено, основанной на технических индексах, для самого простого случая – при делении каждой независимой переменной на два нечетких кластера.

Сначала нечеткие правила задаются только для переменной x_1:

$$R^1: \text{ЕСЛИ } \left(x_{1,t} = A_1\right), \text{ТО}$$
$$y_t^1 = \alpha_0^1 + \alpha_1^1 x_{1,t} + \alpha_2^1 x_{2,t} ,$$
$$R^2: \text{ЕСЛИ } \left(x_{1,t} = A_2\right), \text{ТО}$$
$$y_t^2 = \alpha_0^2 + \alpha_1^2 x_{1,t} + \alpha_2^2 x_{2,t} .$$

По результатам запоминается значение MAPE. Затем аналогичным образом деление на кластеры проводилось по переменной x_2, по результатам шага запоминалось значение MAPE.

Если выбрано C – количество нечетких кластеров для каждой переменной, то число нечетких правил в модели Такаги – Сугено $I = C^2$. Например, поэтому при $C = 2$ база состояла из 4 нечетких правил:

$$R^1: \text{ЕСЛИ } \left(x_{1,t} = A_1\right) \text{ И } \left(x_{2t} = B_1\right), \text{ТО } y_t^1 = \alpha_0^1 + \alpha_1^1 x_{1,t} + \alpha_2^1 x_{2,t} ,$$
$$R^2: \text{ЕСЛИ } \left(x_{1,t} = A_1\right) \text{ И } \left(x_{2t} = B_2\right), \text{ТО } y_t^2 = \alpha_0^2 + \alpha_1^2 x_{1,t} + \alpha_2^2 x_{2,t} ,$$
$$R^3: \text{ЕСЛИ } \left(x_{1,t} = A_2\right) \text{ И } \left(x_{2t} = B_1\right), \text{ТО } y_t^3 = \alpha_0^3 + \alpha_1^3 x_{1,t} + \alpha_2^3 x_{2,t} ,$$
$$R^4: \text{ЕСЛИ } \left(x_{1,t} = A_2\right) \text{ И } \left(x_{2t} = B_2\right), \text{ТО } y_t^4 = \alpha_0^4 + \alpha_1^4 x_{1,t} + \alpha_2^4 x_{2,t} .$$

Здесь A_1, A_2, B_1, B_2 – некоторые нечеткие множества, функции принадлежности которых строятся при помощи алгоритма c-means, $\alpha_0^i, \alpha_1^i, \alpha_2^i$, $i = 1,\dots,4$ – рассчитанные коэффициенты уравнений. Рассматриваются различные значения вплоть до 10, MAPE и RMSE рассчитываются по следующим формулам:

$$RMSE = \sqrt{\frac{\sum_{t=n_1}^{n_1+n_2} \left(\mathrm{y}_t - \hat{\mathrm{y}}_t\right)^2}{n_2}} ,$$

$$MAPE = \frac{1}{n_2} \sum_{t=n_1}^{n_1+n_2} \frac{|y_t - \hat{y}_t|}{y_t}.$$

В таблицах 2, 3 приведены результаты расчетов. Столбец с одним нечетким правилом соответствует результатам, полученным по классической линейной модели. Видно, что модель Тагаки – Сугено позволяет значительно уменьшить ошибку прогноза по сравнению с классической линейной моделью (примерно в 4 раза и даже больше).

Таблица 2.

Значения RMSE для тестовой выборки

	Модель Тагаки – Сугено									
	Количество правил									
	1	2	3	4	5	6	7	8	9	10
ММВБ	5,878	3,551	4,379	2,794	2,477	2,879	3,219	5,167	**1,202**	1,169
РТС	3,433	1,872	1,290	1,321	1,039	**0,997**	1,285	1,838	1,990	1,135
Oil&Gas	15,790	8,281	8,823	7,596	9,897	10,242	9,710	**4,272**	6,062	4,697

Таблица 3.

Значения MAPE для тестовой выборки

	Модель Тагаки – Сугено									
	Количество правил									
	1	2	3	4	5	6	7	8	9	10
ММВБ	0,226	0,126	0,163	0,091	0,079	0,101	0,105	0,144	**0,036**	0,053
РТС	0,237	0,151	0,108	0,095	**0,078**	0,078	0,100	0,133	0,115	0,088
Oil&Gas	0,237	0,099	0,115	0,094	0,132	0,137	0,120	**0,051**	0,070	0,051

Однако, как видно из рис. 4, зависимость между количеством правил и MAPE немонотонная.

Рис. 4. Зависимость МАРЕ от количества нечетких правил для каждого фондового индекса

7. Заключение

В данной работе исследуются применения моделей Такаги – Сугено к российским фондовым индексам.

По итогам проделанной работы можно сделать вывод, что модель Такаги – Сугено хорошо описывает данные и уменьшает ошибку прогноза. Как и следовало ожидать, влияние объясняющих переменных на объясняемые оказывается разным при различных условиях для разных нечетких правил. В целом, результаты исследования для Московской биржи согласуются с результатами предшествующих работ, полученными для некоторых зарубежных фондовых бирж.

* *
*

СПИСОК ЛИТЕРАТУРЫ

Пегат А. Нечеткое моделирование и управление. М.: Бином. Лаборатория знаний, 2013.

Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. 2-е изд. М.: Горячая линия – Телеком, 2013.

Шведов А.С. Нечеткое математическое программирование: краткий обзор // Проблемы управления. 2017. Вып. 3. С. 2–10.

Boyacioglu M.A., Avci D. An Adaptive Network-based Fuzzy Inference System (ANFIS) for the Prediction of Stock Market Return: The Case of the Istanbul Stock Exchange // Expert Systems with Applications. 2010. Vol. 37. № 12. P. 7908–7912.

Chang P.C., Fan C.Y. A Hybrid System Integrating a Wavelet and TSK Fuzzy Rules for Stock Price Forecasting // IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2008. Vol. 38. № 6. P. 802–815.

Chang P.C., Liu C.H. A TSK Type Fuzzy Rule Based System for Stock Price Prediction // Expert Systems with Applications. 2008. Vol. 34. № 1. P. 135–144.

Chang P.C., Wang Y.W., Yang W.N. An Investigation of the Hybrid Forecasting Models for Stock Price Variation in Taiwan // Journal of the Chinese Institute of Industrial Engineers. 2004. Vol. 21. № 4. P. 358–368.

Davé R.N., Krishnapuram R. Robust Clustering Methods: A Unified View // IEEE Transactions on Fuzzy Systems. 1997. Vol. 5. № 2. P. 270–293.

Esfahanipour A., Aghamiri W. Adapted Neuro-fuzzy Inference System on Indirect Approach TSK Fuzzy Rule Base for Stock Market Analysis // Expert Systems with Applications. 2010. Vol. 37. № 7. P. 4742–4748.

Jang J.S.R. ANFIS: Adaptive-Network-Based Fuzzy Inference System // IEEE Transactions on Systems, Man, and Cybernetics. 1993. Vol. 23. № 3. P. 665–685.

Lilly J.H. Fuzzy Control and Identification. John Wiley & Sons, 2011.

Olej V. Design of the Models of Neural Networks and the Takagi – Sugeno Fuzzy Inference System for Prediction of the Gross Domestic Product Development // WSEAS Transactions on Systems. 2005. Vol. 4. № 4. P. 314–319.

Olej V., Kfupka J. Prediction of Gross Domestic Product Development by Takagi–Sugeno Fuzzy Inference Systems // Intelligent Systems Design and Applications, 2005. ISDA'05. Proceedings. 5th International Conference on. IEEE, 2005. P. 186–191.

Pingan Z., Xiaohong G. Fuzzy Modeling for Electrical Market Price Forecasting // Intelligent Control and Automation, 2000. Proceedings of the 3rd World Congress on. IEEE, 2000. Vol. 3. P. 2262–2266.

Rodriguez C.P., Anders G.J. Energy Price Forecasting in the Ontario Competitive Power System Market // IEEE Transactions on Power Systems. 2004. Vol. 19. № 1. P. 366–374.

Sheta A. Software Effort Estimation and Stock Market Prediction Using Takagi – Sugeno Fuzzy Models // Fuzzy Systems, 2006 IEEE International Conference on. IEEE, 2006. P. 171–178.

Sugeno M., Kang G.T. Structure Identification of Fuzzy Model // Fuzzy Sets and Systems. 1988. Vol. 28. № 1. P. 15–33.

Takagi T., Sugeno M. Fuzzy Identification of Systems and its Applications to Modeling and Control // IEEE Transactions on Systems, Man, and Cybernetics. 1985. № 1. P. 116–132.

Zhang H., Liu D. Fuzzy Modeling and Fuzzy Control. Springer Science & Business Media, 2006.

Modeling the Stock Market Dynamics by Takagi – Sugeno Fuzzy Inference Systems

Elena Mogilevich[1], **Alexei Shvedov**[2]

[1] National Research University Higher School of Economics,
26, Shabolovka st., Moscow, 119049, Russian Federation.
E-mail: gurjanovaelena@yandex.ru

[2] National Research University Higher School of Economics,
20, Myasnitskaya ul., Moscow, 101000, Russian Federation.
E-mail: ashvedov@hse.ru

In this paper we evaluate the parameters and study the applicability of the Takagi – Sugeno models to describe the dynamics of the main indices of the Moscow stock market: MICEX, RTS indices and the oil and gas industry index. Review of TS-models for forecasting foreign stock indices and stock prices is given. TS-models represent a generalization of classical econometric approaches. It is achieved using systems of fuzzy rules. Any Takagi – Sugeno model can be considered as a modification of some linear econometric model. Results from the approximation theory show that a Takagi – Sugeno model can approximate any nonlinear econometric model. In this paper, we construct TS-models for Russian stock indices. The fuzzy clustering method is used to find the membership functions. The coefficients of the linear equation in each fuzzy rule are found using the Sugeno – Kang procedure based on the least squares method. Calculations show that the Takagi – Sugeno model in all cases gives a decrease in the forecast error in comparison with the unmodified linear model. In some examples, the forecast error decreases roughtly in 4 times.

Key words: fuzzy systems; stock market; regression, forecasting.

JEL Classification: C51.

* *
*

References

Pegat A. (2013) *Nechetkoe modelirovanie i upravlenie* [Fuzzy Modeling and Control]. Moscow: Binom. Knowledge lab.

Rutkovskaja D., Pilin'skij M., Rutkovskij L. (2013) *Nejronnye seti, geneticheskie algo-ritmy i nechetkie sistemy* [Neural Networks, Genetic Algorithms and Fuzzy Systems]. 2nd ed. Moscow: Hot Line – Telecom.

Shvedov A.S. (2017) Nechetkoe matematicheskoe programmirovanie: kratkij obzor [Fuzzy Mathematical Programming: An Overview]. *Control Sciences*, rel. 3, pp. 2–10.

Boyacioglu M.A., Avci D. (2010) An Adaptive Network-based Fuzzy Inference System (ANFIS) for the Prediction of Stock Market Return: The Case of the Istanbul Stock Exchange. *Expert Systems with Applications*, 37, 12, pp. 7908–7912.

Chang P.C., Fan C.Y. (2008) A Hybrid System Integrating a Wavelet and TSK Fuzzy Rules for Stock Price Forecasting. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38, 6, pp. 802–815.

Chang P.C., Liu C.H. (2008) A TSK Type Fuzzy Rule Based System for Stock Price Prediction. *Expert Systems with Applications*, 34, 1, pp. 135–144.

Chang P.C., Wang Y.W., Yang W.N. (2004) An Investigation of the Hybrid Forecasting Models for Stock Price Variation in Taiwan. *Journal of the Chinese Institute of Industrial Engineers*, 21, 4, pp. 358–368.

Davé R.N., Krishnapuram R. (1997) Robust Clustering Methods: A Unified View. *IEEE Transactions on Fuzzy Systems*, 5, 2, pp. 270–293.

Esfahanipour A., Aghamiri W. (2010) Adapted Neuro-fuzzy Inference System on Indirect Approach TSK Fuzzy Rule Base for Stock Market Analysis. *Expert Systems with Applications*, 37, 7, pp. 4742–4748.

Jang J.S.R. (1993) ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Systems, Man, and Cybernetics*, 23, 3, pp. 665–685.

Lilly J.H.(2011) *Fuzzy Control and Identification*. John Wiley & Sons.

Olej V. (2005) Design of the Models of Neural Networks and the Takagi–Sugeno Fuzzy Inference System for Prediction of the Gross Domestic Product Development. *WSEAS Transactions on Systems*, 4, 4, pp. 314–319.

Olej V., Kfupka J. (2005) Prediction of Gross Domestic Product Development by Takagi – Sugeno Fuzzy Inference Systems. *Intelligent Systems Design and Applications*, ISDA'05, Proceedings, 5th International Conference on, pp. 186–191.

Pingan Z., Xiaohong G. (2000) Fuzzy Modeling for Electrical Market Price Forecasting. *Intelligent Control and Automation*, Proceedings of the 3rd World Congress on, 3, pp. 2262–2266.

Rodriguez C.P., Anders G.J. (2004) Energy Price Forecasting in the Ontario Competitive Power System Market. *IEEE Transactions on Power Systems*, 19, 1, pp. 366–374.

Sheta A. (2006) Software Effort Estimation and Stock Market Prediction Using Takagi – Sugeno Fuzzy Models. *Fuzzy Systems*, *IEEE International Conference on*, pp. 171–178.

Sugeno M., Kang G.T. (1988) Structure Identification of Fuzzy Model. *Fuzzy Sets and Systems*, 28, 1, pp. 15–33.

Takagi T., Sugeno M. (1985) Fuzzy Identification of Systems and its Applications to Modeling and Control. *IEEE Transactions on Systems, Man, and Cybernetics*, 1, pp. 116–132.

Zhang H., Liu D. (2006) *Fuzzy Modeling and Fuzzy Control*. Springer Science & Business Media.