МИНОБРНАУКИ РОССИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.
ЧЕРНЫШЕВСКОГО»

Кафедра математической экономики

АНАЛИЗ ДИНАМИКИ ФОНДОВЫХ ИНДЕКСОВ С ИСПОЛЬЗОВАНИЕМ НЕЧЕТКИХ МОДЕЛЕЙ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 451 группы
направления 38.03.05 Бизнес-информатика
механико-математического факультета
Лотарева Ильи Александровича

Научный руководитель
доцент, к.ф.–м.н., доцент ________________ В.В.Новиков

Заведующий кафедрой
д.ф.–м.н., профессор ________________ С.И.Дудов

Саратов 2020

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.Существуют и экономические приложения, хотя и не столь многочисленные, как в других областях. Иногда эти модели называют также моделями Такаги–Сугено–Канга. 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 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 помогают найти нужную форму нелинейности.

Целью настоящей дипломной работы является изучение подхода к построению модели Такаги-Сугено , которая применялась авторами для исследования динамики российских фондовых индексов. В рамках достижения данной цели были поставлены следующие задачи:

— изучение теории нечетких множеств;

— изучение основ классического и нечеткого регрессионного анализа;

— знакомство с методами построения нечетких регрессионных моделей Такаги-Сугено;

— на основе изученного теоретического материала написать на языке Python 3.6 программу, реализующую алгоритм Такаги-Сугено для моделирования динамики индексов Московской биржи.

Основное содержание работы

1. Необходимые сведения о нечётких множествах 2. Общее описание модели Такаги-Сугено 3. Определение функции принадлежности и параметров TS-модели 4. Анализ динамики фондовых индексов при помощи построения модели Такаги-Сугено для уменьшения ошибок регрессии на языке программирования Python 3.6

Необходимые сведения о нечетких множествах

Рассмотрим некоторое множество U, которое будем называть универсальным множеством.

Определение 1. *Любая функция* $\mu{:}U \to [0,\ 1]$ *называется функцией принадлежности.*

Определение 2. *График функции принадлежности* μ, *т.е. множество пар* $(u,\ \mu_A(u))$, $u \in U$, *называется нечётким множеством. Если нечёткое множество обозначить через* A, *то функция принадлежности этого нечёткого множества обычно обозначается* $\mu_A(u)$. *Иногда о числе* $\mu_A(u)$ *говорят, как о степени принадлежности элемента и нечёткому множеству* A.

Определение 3. *Пусть* A *– обычное подмножество множества* U. *Индикатором множества* A *называется функция* $I_A{:}U \to [0{,}1]$ *такая, что* $I_A(u){=}1$ *при* $u \in A$ *и* $I_A(u){=}0$ *при* $u \notin A$.

Индикатор является функцией принадлежности и, следовательно, график индикатора – это нечёткое множество;если обычное множество A, *то и нечёткое множество в этом случае будет обозначаться* A. *Такое двусмысленное использование обозначения требует определённого внимания. Формально обычные подмножества множества* U *не являются частным видом нечётких множеств. Нечётким множеством является график индикатора обычного подмножества множества* U.

Определение 4. *Носителем нечёткого множества* A *называется множество* $u \in U : \mu_A(u){>}0$.

Определение 5. *Нечёткое множество* A *называется нормализованным, если существует элемент* $u \in$ *такой, что* $\mu_A(u){=}1$.

Определение 6. *Пересечением нечётких множеств* A *и* B называется нечёткое множество с функцией принадлежности $\mu_A \cap_B(u){=}\min\{\mu_A(u),\ \mu_B(u)\}$, $u \in U$.

Определение 7. *Объединением нечётких A и B называется нечёткое множество с функцией принадлежности $\mu_A \cup_B(u)$=*$\max\{\mu_A(u), \mu_B(u)\}$*, $u \in U$.*
Определение 8. *Дополнением нечёткого множества A называется нечёткое множество с функцией принадлежности. $\mu_{\bar{A}}(u)$=1 – $\mu_A(u)$, $u \in U$.*
Определение 9. *η-срезом нечёткого множества A называется множество*

$$A_\eta = u \in U : \mu_A(u) \geq \eta,$$

$0<\eta\leq 1$.

При $U=\mathbb{R}$ вместо термина «нечёткое множество» в ряде случаев удобнее употреблять термин «нечёткая велечина». Тогда вместо обозначения A используется, например, обозначение ξ. Для каждого множества $B \subseteq \mathbb{R}$ определятеся

$$\sup_{u \in B} \mu_\xi(u).$$

Определение 10. Пусть $U=\mathbb{R}$. Нечёткое множества A называется выпуклым, если существует точка $\mu_A \in \mathbb{R}$ такая, что функция $\mu_A(u)$ является монотонно неубывающей на прямой$(-\infty,\mu_A]$ и является монотонно невозрастающей на полупрямой $[\mu_A,\infty)$.
Определение 11. Пусть $U=\mathbb{R}$, A – выпуклое нормализованное нечёткое множество с ограниченным носителем. Функции

$$z^L(\eta) = A_\eta, z^R(\eta) = \sup A_\eta, \eta \in (0, 1],$$

называются, соответственно, левым и правым индексами нечёткого множества A.

Можно было бы сказать, что если A – это выпуклое нормализованное нечёткое множество с ограниченным носителем, то будем называть A нечётким числом. Однако для определения нечётко-случайных велечин удобнее считать, что нечёткое число-это компактное подмножество плоскости, лежащее межде такими нечётким множеством A и осью абсцисс.

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=($x_{1,t}$,...,$x_{m,t}$), t=1,...,n, (объясняющие переменные) и одномерные наблюдения y_1,...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.

Определение степеней принадлежности. Этот шаг заключается в разделении m-мерных наблюдений на заданное число нечетких кластеров. Используется алгоритм нечеткой кластеризации *c-means*. Данный метод позволяет сопоставить одно наблюдение одновременно с несколькими кластерами с разной степенью принадлежности. Число кластеров равняется числу нечетких правил I.
Метод нечёткой кластеризации C-средних(c-means) позволяет разбить имеющееся множество элементов мощностью N на заданное число нечётких множеств k. Метод нечеткой кластеризации C-средних можно рассматривать как усовершенствованный метод k-средних, при котором для каждого элемента из рассматриваемого множества рассчитывается степень его принадлежности каждому из кластеров. Кластеризация — это разделение множества входных векторов на группы (кластеры) по степени «схожести» друг на друга. Классический пример c-means — т.н. «бабочка» (butterfly):

Алгоритм нечётких c-means – это алгоритм нечёткой кластеризации, находит компактные кластеры сферической формы. Расстояние между объектом и центром кластера расчитывается через стандартную Евклидову норму.
1. Установить параметры алгоритма: N - количество объектов, K-количество кластеров; m - экспоненциальный вес(обычно устанавливается m=2); ε - параметр остановки алгоритма.
2. Случайным образом сгенерировать матрицу нечёткого разбиения, удовлетворяющую условиям u_{ij}.

3. Рассчитать центры кластеров : $c_j = \frac{\sum_{i=1}^{N}(u_{ij})^m \cdot x_i}{\sum_{i=1}^{N}(u_{ij})^m}$.

4. Рассчитать расстояние между объектами из X и центрами кластеров: $D_{ij} = \sqrt{\|x_j - c_j\|^2}$.

5. Пересчитать элементы матрицы нечёткого разбиения:

если $D_{ij} > 0 : u_{ij} = \frac{1}{(D_{ij}^2 \sum_{k=1}^{K} \frac{1}{D_{ik}^2})^{\frac{1}{m-1}}}$, если D_{ij}=0: $u_{ij} = \begin{cases} 1, k = j \\ 0, k \neq \text{j}, \end{cases}$ k=$\overline{1, K}$.

6. Проверить условие $\|U - U^*\|^2 < \varepsilon$, где U^* - матрица нечёткого разбиения на предыдущей итерации алгоритма. Если «да» , то перейти к шагу 7, иначе - к шагу 3.

7. Конец.

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

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

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

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:ЕСЛИ (x_1=A_1),ТО ..., R^2:ЕСЛИ (x_1=A_2),ТО ....

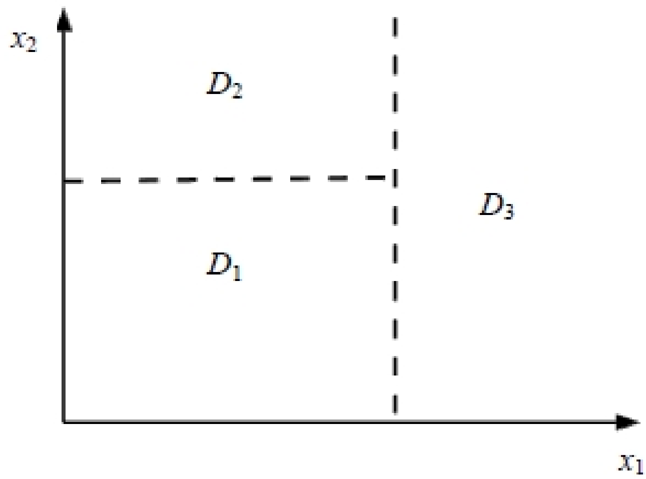
Данная процедура повторяется для каждой независимой переменной, для каждой модели считались MAPE и RMSE. Тогда x_i для которого модель давала наименьший MAPE, далее в тексте и на рисунках будет обозначаться как x_1.образом, на этапе 2 проводится m итераций.

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

Система нечетких правил 2:

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

Система нечетких правил 3:

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

В данной работе рассматривается только один временной ряд - индекс МосБиржи. Используются недельные данные с января 2018 г. по декабрь 2018 г.Длина временного ряда составляет 52 наблюдения.

Для построения эконометрической модели выбраны следующие показатели:

y_t – среднее значение индекса Московской Биржи за неделю t,t = 0,1,...,51;
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:

$$x_{3t} = \frac{y_{t-1}}{y_{t-2}} \tag{6}$$

Такой выбор экзогенных переменных позволяет отразить и положение на рынке акций, и положение на рынке облигаций. В качестве базовой используется следующая линейная модель:

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

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 нечёткими правилами является наилучшей. Система нечетких правил имеет следующий вид:

$$R^1\text{: ЕСЛИ } (x_{2,t}{=}A_1) \text{ И } (y_{t-1}{=}B_1) \text{ И } (x_{1,t}{=}C_1) \text{ И } (x_{3,t}{=}D_1)\text{, ТО}$$
$$y_t^1{=}\alpha_0^1{+}\beta_1^1 y_{t-1}{+}\beta_2^1 x_{1,t}{+}\beta_3^1 x_{2,t}{+}\beta_4^1 x_{3,t},$$

R^2: ЕСЛИ $(x_{2,t}{=}A_1)$ И $(y_{t-1}{=}B_1)$ И $(x_{1,t}{=}C_1)$ И $(x_{3,t}{=}D_1)$, 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: ЕСЛИ $(x_{2,t}{=}A_1)$ И $(y_{t-1}{=}B_1)$ И $(x_{1,t}{=}C_2)$, 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: ЕСЛИ $(x_{2,t}{=}A_1)$ И $(y_{t-1}{=}B_1)$ И $(x_{1,t}{=}C_3)$, 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: ЕСЛИ $(x_{2,t}{=}A_1)$ И $(y_{t-1}{=}B_2)$, ТО $y_t^5{=}\alpha_0^5{+}\beta_1^5 y_{t-1}{+}\beta_2^5 x_{1,t}{+}\beta 5_3 x_{2,t}{+}\beta_4^5 x_{3,t}$,
R^6: ЕСЛИ $(x_{2,t}{=}A_2)$, 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: ЕСЛИ $(x_{2,t}{=}A_3)$, 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 $A_1, A_2, A_3, B_1, B_2, C_1, C_2, C$ - некоторые нечеткие множества, функции принадлежности которых строятся при помощи алгоритма c-means,
$\alpha_0^i, \beta_1^i, \beta_2^i, \beta_3^i, \beta_4^i$, $i{=}1,...,7$ – рассчитанные коэффициенты уравнений.
Для подсчётов была написана программа для автоматизации больших подсчётов.

Заключение

В данной работе исследуется применение 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 переменные на объясняемые оказывается разным при различных условиях для разных нечётких правил.