МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра дифференциальных уравнений и
математической экономики

АНАЛИЗ ЭКОНОМИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ МОДЕЛИ ТАКАГИ-СУГЕНО

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 451 группы
направления 38.03.05 — Бизнес-информатика

механико-математического факультета
Байгушевой Светланы Валерьевны

Научный руководитель
доцент, д. ф.-м. н., доцент ______________ В. В. Новиков

Заведующий кафедрой
зав. каф., д. ф.-м. н., профессор ______________ С. И. Дудов

Саратов 2022

ВВЕДЕНИЕ

Актуальность. Проблема моделирования фондовых показателей актуальна, так как прибыль инвесторов напрямую зависит от точности прогноза. Поэтому в данной области непрерывно ведутся исследования и применяются новые модели.

Целью данной работы является проведение оценки параметров моделей Такаги-Сугено для описания динамики фондовых индикаторов на примере индекса МосБиржи.

Задачи данной работы — изучить теоретическую базу, разработать код программы на языке Python алгоритма Fuzzy C-Means и алгоритма Гюстафсона-Кесселя, осуществляющие нечеткую кластеризацию, построить нечеткие регрессионные модели Такаги-Сугено для индекса МосБиржи, сравнить результаты построенных моделей, а также изучить их прогностические свойства.

Структура и содержание бакалаврской работы. Работа состоит из введения, трех разделов, которые включают в себя подразделы, заключения, списка использованных источников в количестве 20 штук.

Основное содержание работы

Во введении приводится краткое содержание работы и постановка задачи.

В первом разделе рассматривается четкая линейная регрессионная модель.

Второй раздел посвящен нечетким моделям Такаги-Сугено, который включает в себя алгоритмов Fuzzy C-Means и Гюстафсона-Кесселя, основы теории нечетких множествах и основные идеи, лежащие в основе TS-моделей.

В третьем разделе построены модели Такаги-Сугено для индекса МосБиржи, рассмотрено сравнение моделей и изучены их прогностические свойства.

В заключении описываются основные результаты работы и краткие выводы по ним.

Приложение содержит исходный код программы.

Алгоритм Fuzzy C-Means. Алгоритм представляет собой итерационную процедуру следующего вида:

Шаг1. Выбрать количество кластеров $2 \le c \le d$.

Шаг2. Выбрать скалярную метрику для отображения векторов данных на вещественную ось.

Шаг3. Выбрать параметр остановки δ.

Шаг4. Выбрать коэффициент нечеткости $w \in (1, \infty)$.

Шаг5. Проинициализировать матрицу разбиения (например, случайными значениями).

Шаг6. Вычислить прототипы(центры) кластеров по формуле:

$$c_l^{(i)} = \frac{\sum\limits_{j=1}^{d} (u_{ij}^{(l-1)})^w m_j}{\sum\limits_{j=1}^{d} (u_{ij}^{(l-1)})^w}, \quad 1 \le i \le c.$$

Шаг7. Для всех элементов данных высчитать квадраты расстояний до всех (центров) кластеров по формуле:

$$d_A^2(m_j, c^{(i)}) = ||m_j - c^{(i)}||_A^2 = (m_j - c^{(i)})^t A (m_j - c^{(i)}).$$

Шаг8. Обновить матрицу разбиения по формуле:

$$u_{ij}^{(l)} = \frac{1}{\sum\limits_{k=1}^{c} \left(\frac{d_A^2(m_j, c^{(i)})}{d_A^2(m_j, c^{(k)})} \right)^{\frac{1}{w-1}}}, \quad \text{для всех } 1 \le i \le c,\ 1 \le j \le d,$$

учитывая ограничения:

$$u_{ij} \in [0, 1]; \quad \sum_{i=1}^{c} u_{ij} = 1; \quad 0 < \sum_{j=1}^{d} u_{ij} < d$$

.

Шаг9. Проверить условие $||U^{(l)} - U^{(l-1)}|| < \delta$. Если условие выполняется, завершить процесс, если нет — перейти к шагу 7 с номером итерации $l = l + 1$.

Алгоритм Гюстафсона-Кесселя. Конструктивно алгоритм выглядит следующим образом:

Шаг1. Выбрать количество кластеров $2 \leq c \leq d$.

Шаг2. Выбрать критерий остановки $\delta > 0$.

Шаг3. Определить параметр нечеткости $w \in (1, \infty)$, например 2.

Шаг4. Проинициализировать матрицу разбиения (например, случайными значениями).

Шаг5. Рассчитать прототипы кластеров по формуле:

$$c_l^{(i)} = \frac{\sum\limits_{j=1}^{d} (u_{ij}^{(l-1)})^w m_j}{\sum\limits_{j=1}^{d} (u_{ij}^{(l-1)})^w}, \quad 1 \leq i \leq c.$$

Шаг6. Рассчитать ковариационные матрицы кластеров по формуле:

$$F^{(i)} = \frac{\sum\limits_{j=1}^{d} (u_{ij}^{(l-1)})^w (m_j - c^{(i)})(m_j - c^{(i)})^t}{\sum\limits_{j=1}^{d} (u_{ij}^{(l-1)})^w}$$

Шаг7. Рассчитать расстояния по формуле:

$$d^2_{F^{(i)}}(c_l^{(i)}, m_j) = (c^{(i)} - m_j)^t \left[|F^{(i)}|^{\frac{1}{r+1}} (F^{(i)})^{-1} \right] (c^{(i)} - m_j).$$

Шаг8. Обновить матрицу разбиения по формуле:

$$u_{ij}^{(l)} = \frac{1}{\sum\limits_{k=1}^{c} \left(\frac{d^2_{F^{(i)}}(m_j, c^{(i)})}{d^2_{F^{(i)}}(m_j, c^{(k)})} \right)^{\frac{1}{w-1}}}, \quad \text{для всех } 1 \leq i \leq c,\ 1 \leq j \leq d,$$

учитывая ограничения:

$$u_{ij} \in [0, 1]; \quad \sum_{i=1}^{c} u_{ij} = 1; \quad 0 < \sum_{j=1}^{d} u_{ij} < d$$

.

Шаг9. Проверить условие $||U^{(l)} - U^{(l-1)}|| < \delta$. Если условие выполняется, завершить процесс, если нет — перейти к шагу 5 с номером итерации $l = l + 1$.

Применение **нечетких множеств** позволяет описывать нечеткие понятия и знания, оперировать этими знаниями и делать нечеткие выводы. К преимуществам нечетких систем следует отнести их универсальность. Любая непрерывная функция может быть представлена нечеткой моделью с любой заданной точностью. Особые качества систем с нечеткой логикой позволяют синтезировать модель объекта на основании эвристической информации, полученной от эксперта или в результате эксперимента. Вместе с тем, нечетким системам присущи такие недостатки, как отсутствие алгоритмов синтеза устойчивых моделей и низкая скорость работы последних при большом числе управляющих правил.

С введением нечетких чисел оказалось возможным прогнозировать будущие значения параметров, которые меняются в установленном расчетном диапазоне. Вводится набор операций над нечеткими числами, которые сводятся к алгебраическим операциям с обычными числами при задании определенного интервала достоверности (уровня принадлежности). Применение нечетких чисел позволяет задавать расчетный коридор значений прогнозируемых параметров. Тогда ожидаемый эффект оценивается экспертом также как нечеткое число со своим расчетным разбросом (степенью нечеткости).

Описание TS-модели. При разработке моделей ВР приходится принимать решение о принадлежности рассматриваемых рядов классам, имеющим только детерминированный тренд (TS-модель) или стохастический тренд, возможно, вместе с детерминированным, (DS-модель). Проблема отнесения ВР к одному из указанных классов привлекает повышенное внимание специалистов-аналитиков в экономической сфере. Поскольку траектории TS- и DS- рядов существенно различаются между собой, вынесенное суждение о принадлежности анализируемого ряда выделенным классам важно при построении долгосрочных моделей прогнозирования. Линия тренда у TS-рядов представляет собой определенную осевую линию, относительно которой реальная траектория ряда совершает флуктуации с достаточно высокой частотой.В TS-рядах влияние предыдущих воздействий ослабевает с течением времени.

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

$$\text{ЕСЛИ } (x_1 = A^1), \text{ТО } y^1 = \alpha_0^1 + \beta_1^1 x_1 + \beta_2^1 x_2,$$
$$\text{ЕСЛИ } (x_1 = A^2), \text{ТО } y^2 = \alpha_0^2 + \beta_1^2 x_1 + \beta_2^2 x_2,$$
$$\text{ЕСЛИ } (x_1 = A^3), \text{ТО } y^3 = \alpha_0^3 + \beta_1^3 x_1 + \beta_2^3 x_2.$$

Здесь ЕСЛИ ($x_1 = A^1$) можно понимать как ЕСЛИ (x_1 малое). Это нечеткое правило действует не только для $x_1 \leq 0$, а для всех x_1, для которых $\mu_{A^1}(x_1) > 0$. Вклад первого нечеткого правила зависит от величины $\mu_{A^1}(x_1)$.

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

$$\hat{y} = \frac{\sum\limits_{i=1}^{I} G^i y^i}{\sum\limits_{i=1}^{I} G^i},$$

где G^i — действительное число, обозначающее степень правдивости i-го правила. I — число нечетких правил. В общем случае условие для i-го нечеткого правила будет иметь следующий вид:

$$\text{ЕСЛИ } (x_1 = A^1) \text{ И} \ldots \text{И } (x_m = A_m^i),$$

где $A_1^i, \ldots, A_m^i$ —нечеткие множества, $i = 1, \ldots, I$.

В данной работе G^i рассчитывается по формуле:

$$G^i = \mu_{A_1^i}(x_1) \times \ldots \times \mu_{A_m^i}(x_m).$$

Определение функций принадлежности и нечетких правил TS-модели. Рассмотрим основные характеристики функции принадлежности:

- Функция принадлежности нечеткого подмножества A приписывает каждому элементу $x \in X$ степень принадлежности к нечетному подмножеству A;

- Функция принадлежности определяет и указывает степень принадлежности элементов x подмножеству A, называемому в форме числовых значений в диапазоне [0,1];
- Функция принадлежности определяет характер термы. Численное значение функции принадлежности — степень принадлежности элемента $x \in X$;
- Функция принадлежности используется для представления лингвистических переменных в виде нечетких подмножеств.

Для определения оптимального числа нечетких правил и нахождения функций принадлежности существуют различные подходы. Часто предполагается, что функция принадлежности относится к некоторому семейству, и для определения параметров используется некоторая оптимизационная процедура. В качестве исходных данных для построения TS-модели берутся m-мерные наблюдения $x_t = (x_{1,t}, \ldots, x_{m,t}), t = 1, \ldots, n,$(объясняющие переменные) и одновременные наблюдения $y_1, \ldots, y_n$ (объясняемая переменная).

Процесс построения модели ведется итерационно. Перед началом каждой итерации формат системы нечетких правил считается выбранным. Определяются значения G^i 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. Данный метод позволяет сопоставить одно наблюдение одновременно с несколькими кластерами с разной степенью принадлежности. Число кластеров равняется числу нечетких правил L.

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

На этом шаге параметры определяются с помощью метода наименьших квадратов. При расчете $\hat{y}_t$, в качестве независимых переменных будут выступать:

$$\frac{G^i}{\sum\limits_{k=1}^{I} G^k}, \qquad (1)$$

$$\frac{G^i x_{j,t}}{\sum\limits_{k=1}^{I} G^k}, \tag{2}$$

где $t = 1, \ldots, n, i = 1, \ldots, I, j = 1, \ldots, m$. 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:

$$RMSE = \sqrt{\frac{\sum\limits_{t=1}^{n} (y_t - \hat{y}_t)^2}{n}}, \tag{3}$$

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|y_t - \hat{y}_t|}{y_t}. \tag{4}$$

Построение моделей Такаги-Сугено. В качестве примера применения алгоритма Такаги-Сугено рассматривается регрессионные модели с двумя нечеткими правилами для прогнозирования значения индекса МосБиржи. В качестве переменных выбраны следующие показатели:

- Скользящее среднее за последние 6 дней (5):

$$x_{1,k} = \frac{\sum\limits_{t=k-6}^{k} y_t}{6}, \tag{5}$$

  где y_t — значение индекса на момент закрытия t-го дня, t — количество наблюдений.
- Отклонение t-го наблюдения от скользящего среднего (6):

$$x_{2,t} = \frac{y_t - x_{1,k}}{x_{1,k}} 100\%. \tag{6}$$

Вычисления проводятся на языке программирования Python. В работе загружаются данные и преобразовываются, исходя из представленных формул независимых переменных. Далее в работе используется алгоритм Fuzzy C-Means и кластеризация по Гюстафсону-Кесселю. Затем определяются коэффициенты линейных уравнений в каждом нечетком правиле. Параметры определяются с помощью метода наименьших квадратов. Далее рассчитываются коэффициенты, исходя из результатов первое уравнение имеет вид:

$$y = 2778.97 + 0.714123x_1 + 234.309x_2,$$

а второе:

$$y = -7.85577 + 0.999259x_1 + 6.15260x_2.$$

Производится прогноз и оценка данного прогноза.

Сравним значения $MAPE$ и $RMSE$ результаты оценки построения TS-моделей для двух правил представлены в соответствии с таблицей:

Таблица 1 — Результат построения TS-моделей

	Модели	
	1	2
$MAPE$	12.952	1.0255
$RMSE$	441.22	40.334

Из полученных результатов можно сделать вывод о том, что TS модель построенная на основе кластеризации по Гюстафсону-Кесселю строит более точный прогноз.

ЗАКЛЮЧЕНИЕ

Один из наиболее популярных подходов построения моделей для прогнозов индексов фондовых бирж состоит в построении регрессии, где в качестве независимых переменных берутся технические индексы. По итогам проделанной работы можно сделать вывод, что модель Такаги-Сугено хорошо описывает данные. Можно также сделать вывод о том, что при использовании кластеризации по Гюстафсону-Кесселю модель строит прогноз лучше, чем с использованием алгоритма C-Means.

В соответствии с целью дипломной работы была проведена оценка параметров моделей Такаги-Сугено для описания динамики фондовых индикаторов на примере индекса МосБиржи. Для достижения целей были реализованы задачи: изучена теоретическая база, разработан код программы на языке Python алгоритма Fuzzy C-Means и алгоритма Гюстафсона-Кесселя, осуществляющие нечеткую кластеризацию, построены нечеткие регрессионные модели Такаги-Сугено для индекса МосБиржи, проведено сравнение построенных моделей, а также изучены их прогностические свойства.