

© О. О. Романова, 2017

Липецкий государственный технический  
университет (ЛГТУ), Россия

E-mail: romanova.olga.mail@gmail.com

## **К ВОПРОСУ ПРИМЕНИМОСТИ АППАРАТА МАРС К ЗАДАЧЕ ПРОГНОЗИРОВАНИЯ БАНКРОТСТВА: ЗАРУБЕЖНЫЙ ОПЫТ**

*В статье приводится краткое описание математического аппарата многомерных адаптивных регрессионных сплайнов (МАРС) и обсуждаются преимущества данного инструмента перед иными статистическими методами применительно к задаче прогнозирования банкротства организаций. В работе также приведен обзор существующих гибридных моделей прогнозирования банкротства с использованием МАРС, обсуждаются их достоинства и недостатки.*

**Ключевые слова:** многомерные адаптивные регрессионные сплайны, риск банкротства, прогнозирование банкротства.

### **Введение**

Начиная с 30-х годов XX столетия экономисты различных стран пытаются решить задачу прогнозирования финансового нездоровья организаций. Несмотря на многообразие подходов, используемых для решения прикладной задачи прогнозирования банкротства (как-то: статистические методы, методы искусственного интеллекта, гибридные методы) и огромное количество существующих моделей, на сегодняшний день говорить об универсальной модели крайне затруднительно. Вместе с тем последние исследования в данном направлении позволили выявить новые инструменты, которые в значительной степени превосходят разработанные в XX веке аналоги. Одним из таких инструментов можно считать аппарат многомерных адаптивных регрессионных сплайнов (далее – МАРС).

Аппарат МАРС был впервые детально описан Джеромом Фридменом в работе [1]. Основные особенности математического аппарата и опыт построения моделей на его основе были описаны нами в работах [2, 3].

Целью настоящего исследования стало изучение вопроса применимости аппарата МАРС для прогнозирования риска банкротст-

ва предприятий. Для достижения цели исследования были поставлены следующие задачи:

- изучение математических основ аппарата МАРС;
- выявление преимуществ и недостатков МАРС перед иными методами решения классической задачи регрессии;
- исследование опыта применения МАРС к прикладной задаче прогнозирования корпоративного банкротства.

Объект исследования – риски банкротства предпринимательских структур. Предмет исследования – применимость аппарата МАРС к прогнозированию рисков банкротства предприятий.

### ***Результаты исследования***

МАРС – это статистическая процедура, позволяющая решить классическую задачу регрессии: установить вид и параметры аппроксимирующей функции, описывающую функциональную зависимость отдельных наблюдений (исходные данные) с указанной точностью [1]. Пространство значений входных переменных разбивается на области со своими собственными уравнениями базисных функций. Это позволяет использовать МАРС даже в случае задач с «проклятием размерности», когда высокая размерность пространства значений входных переменных ограничивает применимость иных статистических процедур.

Метод МАР-сплайнов не имеет ограничения, характерного для иных статистических методов, в части наличия исходных предположений о типе зависимостей (линейных, степенных, экспоненциальных и проч.) между предикторными и выходными переменными [5]. Кроме того, метод МАРС чувствителен к изменению вида связи между предиктором и откликом, будь то: изменение формы связи (например, от линейной к степенной), добавление или вычитание некоторой константы для прогноза отклика справа от узловой точки предиктора, изменение наклона регрессионной функции [4].

Подобные особенности сплайнов достигаются за счет использования особого вида базисных функций:

$$\begin{aligned}(x-t)_+ &= \begin{cases} x-t, & \text{если } x > t, \\ 0, & \text{иначе,} \end{cases} \\ (t-x)_- &= \begin{cases} t-x, & \text{если } x < t, \\ 0, & \text{иначе.} \end{cases}\end{aligned}\tag{1}$$

Параметр  $t$  в формуле (1) – точка изгиба кусочной функции-галки (название объясняется видом графика подобной функции). Формулу вида (1) также можно представить как:

$$\begin{aligned}(x-t)_+ &= \max(0; x-t), \\ (t-x)_- &= \max(0; t-x).\end{aligned}\tag{2}$$

В многомерном случае для каждой компоненты  $x^j$  вектора предикторов  $X = (x^1, \dots, x^j, \dots, x^p)^T$  строятся базисные функции вида (1), (2) с узлами в каждой наблюдаемой переменной  $x_i^j$  ( $i=1, \dots, N, j=1, \dots, p$ ).

По исходным данным определяется набор базисных функций:

$$C = \{(x_j - t)_+, (t - x_j)_-\}_{t \in \{x_{1j}, \dots, x_{Nj}\}, j=1, \dots, p.}$$

Общее уравнение МАР-сплайнов для модели из  $M$  членов, отличных от константы, представляет собой взвешенную сумму базисных функций и их произведений и записывается в виде [1]:

$$y = f(X) = \beta_0 + \sum_{m=1}^M \beta_m h_m(X),\tag{3}$$

где:

$\beta_0$  – некоторый свободный член;

$\beta_m$  – параметр, определяемый для итоговой функции методом наименьших квадратов, некоторая константа;

$h_m(X)$  – базисная функция из множества  $C$  либо же произведение двух и более таких функций;

$X$  – вектор объясняющих переменных;

$M$  – число базисных функций.

В качестве базисной функции  $h_m(X)$  могут выступать:

– константа;

– базисная функция вида (1) или (2);

– произведение базисных функций (1), (2).

Основной принцип работы модели состоит в выборе нужной взвешенной суммы базисных функций из общего набора базисных функций, покрывающих все значения каждого предиктора (т.е. набор будет состоять из одной базисной функции и параметра  $t$  для каждого отдельного значения каждой предикторной переменной). Алгоритм МАР-сплайнов отыскивает в пространстве всех входных и предикторных переменных расположение узловых точек, а также взаимосвязи между переменными. В процессе поиска число добавленных к модели базисных функций из общего набора возрастает

до тех пор, пока не будет максимизирован общий критерий качества модели – обобщенное скользящее среднее (generalized cross-validation), который имеет вид [5]:

$$LOF(\hat{f}_m) = GCV(M) = \frac{1}{N} \frac{\sum_{i=1}^N (y_i - \hat{f}_m(x_i))^2}{\left(1 - \frac{C}{N}\right)^2}, \quad (4)$$

где:

$GCV(M)$  – критерий точности модели, отражающий рост дисперсии с ростом числа базисных функций;

$C$  – эффективное число параметров модели, которое в общем случае определяется как  $C = 2K - 1$  (где  $K$  – общее число параметров в модели).

Пошаговый алгоритм построения MAP-сплайна схож с алгоритмом кусочной линейной регрессии, только вместо регрессионных кусочных функций используются базисные функции из множества  $S$ . Метод MAP-сплайнов находит искомую зависимость за две стадии: «вперед» (forward stage) и «назад» (backward stage)[5]. Первая стадия заключается в добавлении базисных функций к набору, пока не будет достигнут либо заданный уровень точности модели (некоторый критерий невязки), либо же будет достигнуто максимальное число базисных функций. На каждом шаге работы алгоритма новая пара базисных функций является произведением функции  $B_m(X)$  на одну из базисных функций множества  $S$ , коэффициенты находятся методом наименьших квадратов. В результате получается модель из большого числа базисных функций, которая, как правило, чрезмерно подгоняет исходные данные. На второй стадии из набора удаляются функции, которые вносят наименьший вклад в ошибку  $GCV(M)$ . Пошаговый алгоритм построения MAP-сплайна описан в [1].

Главным свойством кусочно-линейных функций является их способность действовать локально, т. е. принимать ненулевые значения лишь на части их области определения. Результат умножения одной функции на другую отличен от нуля лишь в малой части пространства, где обе функции принимают ненулевые значения. Это и позволяет строить качественные модели, используя сплайны.

Еще одной особенностью аппарата MAP-сплайна является возможность моделирования как числовых, так и категориальных факторов (их кодов) без предварительной обработки.

Суммируя вышеизложенное, можно отметить следующие преимущества аппарата МАРС с точки зрения технического построения моделей перед иными статистическими инструментами решения задачи идентификации рисков банкротства:

- возможность моделирования нелинейных зависимостей между факторами и откликом;
- чувствительность к изменению вида зависимости между факторами и откликом;
- отсутствие первоначальных предположений о типе зависимости между факторами и откликом модели;
- скорость работы алгоритма, возможность обработки больших массивов данных;
- высокое качество моделей на большой выборке факторов при ограниченном числе наблюдений (так называемое «проклятие размерности»);
- удобство работы с категориальными переменными.

Вместе с тем выбор аппарата МАРС не является залогом успешной модели. Обладая рядом технических преимуществ перед классическими статистическими методами анализа, МАРС остаются чувствительными к первоначальным входным данным. Кроме того, для достижения высоких показателей статистического качества модель МАРС может стать громоздкой и трудной в интерпретации.

Нами был проведен обзор существующего опыта использования аппарата МАРС для решения прикладной задачи прогнозирования банкротства. Следует отметить, что отечественных исследований в данной тематике обнаружить не удалось.

*Гибридная модель на основе самоорганизующихся карт Кохонена и МАРС [6]*

Разработана модель для предприятий строительной отрасли Испании на основе аппарата МАРС и построения гибридной системы. Методика построения модели предполагает разделение массива предприятий на кластеры схожих по финансовым показателям субъектов, после чего каждый подобный кластер заменяется вектором. Идентичность векторов устанавливается на основании расстояния Махолонобиса. Процесс кластеризации происходит за счет особого вида нейронных сетей — самоорганизующихся карт Кохонена. Подобная кластеризация исходных данных позволяет решить проблему репрезентативности выборки для обучения нейронной сети, т.к. содержит достаточное количество как обанкротившихся, так и финансово здоровых предприятий. После решения задачи



кластеризации организаций производится математическое моделирование на основе аппарата МАРС. Выбор математического аппарата обусловлен его гибкостью, возможностью построения нелинейных зависимостей, возможностью сокращения факторов без ухудшения качества итоговой модели. В качестве факторов были отобраны показатели пятифакторной модели Альтмана по причине их доступности, простоты расчета, популярности данных предикторов в работах иных исследователей, в том числе при моделировании нейронных сетей. В результате построения модели 84,29% предприятий выборки были классифицированы корректно (при этом 88,7% — доля корректно классифицированных предприятий-банкротов, 60,4% — доля корректно классифицированных предприятий не-банкротов).

*Гибридная модель на основе генетического алгоритма, нечетких нейронных сетей и МАРС [7]*

В данном исследовании также предложена гибридная модель прогнозирования банкротства предприятия. Моделирование производится в два этапа: кластеризация на основе нейронных сетей и аппроксимация с использованием аппарата МАРС. В основе кластеризации исходных массивов предприятий лежит генетическая нечеткая нейронная сеть, позволяющая моделировать нелинейные связи, использующая принцип «черного» ящика, а также с улучшенными по сравнению с нечеткими нейронными сетями показателями обучения. Критерием кластеризации выбран метод нечетких С-средних. Особенностью данного метода является возможность принадлежности элемента нескольким кластерам одновременно, при этом для каждого кластера определяется вероятность такой принадлежности. Итоговая эффективность модели проверялась путем сравнения с откликами пятифакторной модели Альтмана, логит-модели Ольсона, модели Змеевского, модели Шумвея. Выборка была сформирована на основании финансовых данных предприятий Индии. Авторы не приводят статистических критериев качества модели, ограничиваясь графическим отображением проведенных кластеров.

*Гибридная модель на основе нечеткой нейронной сети и МАРС [4]*

В работе построена гибридная модель на основе нечеткой нейронной сети и аппарата МАРС. Исходные данные — финансовые показатели более 59 тысяч предприятий Испании. Первым этапом проводится кластеризация предприятий, критерием выбран метод

нечетких С-средних, как и в предыдущей описанной нами работе. Дальнейшая аппроксимация производится с использованием аппарата МАРС. Полученная модель сравнивалась с результатами классификации иных моделей, а именно: моделей дискриминантного анализа, модели МАРС, прямой нейронной сети – и показала наилучшие результаты.

#### *Гибридная модель на основе CART-алгоритма и МАРС [8]*

В работе построена гибридная модель на основе CART-алгоритма (classification and regression trees) и МАРС-моделирования. Базой для анализа стали финансовые показатели банков США с 2008 по 2013 годы, на основе которых были рассчитаны 10 финансовых коэффициентов. Также были построены отдельно модели только на основе CART-алгоритма и МАРС. В исследовании было показано, что в общем случае модель МАРС демонстрирует более высокую точность классификации, чем модель CART, а именно 94,37% корректно классифицированных банков против 94,07%. Гибридная модель, полученная в результате исследования, дает наиболее точные прогнозы как на обучающей, так и на контрольной выборке (98,84% и 96,92% соответственно).

#### **Заключение**

На основе проведенного анализа построенных зарубежными исследователями моделей МАРС можно сделать следующие выводы. Наиболее точными для целей прогнозирования банкротства предприятия оказались гибридные модели, совмещающие преимущества нейронных сетей и аппарата МАРС. Отдельные исследования демонстрируют преимущества подобных моделей над моделями с использованием только МАРС-аппарата. Вместе с тем все описанные выше существующие модели наукоемки и тяжелы в практической реализации в периметре предпринимательской структуры, т.к. требуют значительных ресурсов и предъявляют высокие требования к компетентности их конечного пользователя. Все это ограничивает сферу практического применения данных моделей. Таким образом, существующие гибридные модели на основе аппарата МАРС и нейронных сетей, несмотря на высокие статистические показатели качества, не могут являться доступным для подавляющего большинства предприятий инструментом идентификации рисков. Данное обстоятельство подчеркивает важность разработки универсальной модели для российских субъектов предпринимательства, отвечающей как требованиям качества, так и требованиям практической реализации.

В связи с указанными техническими преимуществами, а также исследованием опыта зарубежных исследователей аппарат MAP-сплайнов использовался нами для построения экспресс-моделей оценки риска банкротства предприятия, в том числе описанных в [3].

### **Литература**

1. Friedman J. H. Multivariate adaptive regression splines // *Annals of Statistics*. 1991. № 19. P. 1–141.

2. Романова О. О. Аппарат мар-сплайнов при оценке риска банкротства предприятий // *Концепт: электронный журнал*. 2016. Т. 11. С. 1401-1405. URL: <http://e-concept.tu/2016/86300.htm>.

3. Романова О. О., Иода Е. В. Инструменты идентификации риска банкротства предприятия: аппарат MAP-сплайнов // *Социально-экономические явления и процессы*. 2015. № 9. С. 140-145.

4. Bankruptcy forecasting: a hybrid approach using fuzzy c-means clustering and multivariate adaptive regression splines (MARS) / J. Andrus, P. Lorca, F. J. Cos Juez, F. Sanchez-Lasheras // *Expert Systems with Applications*. 2001. № 38. P. 1866–1875.

5. Hastie T., Tibshirani R., Friedman J. *The elements of statistical learning, data mining, inference and prediction*. NY: Springer, 2010. 544 p.

6. A hybrid device of self organizing maps (SOM) and multivariate adaptive regression splines (MARS) for the forecasting of firms' bankruptcy // *Accounting and Management Information Systems*. 2011. Vol. 10. № 3. P. 351–374.

7. A hybrid model for bankruptcy prediction using genetic algorithm, fuzzy c-means, and MARS / A. Martin, V. Gayathri, G. Saranya, P. Gayathri, P. Venkatesan // *International Journal on Soft Computing (IJSC)*. 2011. № 1. P. 213-245.

8. Zeineb Affes, Rania Hentati-Kaffel. Forecast bankruptcy using a blend of clustering and MARS model – Case of US banks. Documents de travail du Centre d'Economie de la Sorbonne 2016.26. URL: <https://halshs.archives-ouvertes.fr/halshs-01314553/document>.

*Статья поступила в редакцию 10.09.17 г.*

*Рекомендуется к опубликованию членом Экспертного совета  
канд. экон. наук, доцентом О. А. Горбуновой*