Вестник ТвГУ. Серия "Экономика и управление". 2015. №3. С. 268–274.

МАТЕМАТИЧЕСКИЕ И ИНСТРУМЕНТАЛЬНЫЕ МЕТОДЫ ЭКОНОМИКИ

УДК 338.27:519.862.6

ОБЪЕДИНЕНИЕ ПРОГНОЗОВ НА ОСНОВЕ ОДНОШАГОВОЙ М-ОЦЕНКИ ЭНДРЮСА

А.А. Васильев

Тверской государственный университет, г. Тверь

Предложено использовать для объединения прогнозов индивидуальных моделей прогнозирования одношаговую М-оценку Эндрюса вместо взвешенного арифметического среднего. Результаты исследования подтвердили гипотезу о том, что комбинированные модели прогнозирования на основе М-оценки Эндрюса в большинстве случаев превосходят по точности модель на основе взвешенного арифметического среднего значения.

Ключевые слова: *взвешенное арифметическое среднее, комбинированная модель, М-оценка Эндрюса, объединение прогнозов, одношаговая оценка, прогноз, простое арифметическое среднее.*

Наиболее часто в практике прогнозирования экономических показателей используются методы объединения индивидуальных прогнозов, основанные на их простом или взвешенном усреднении.

Результаты исследований комбинированных моделей на основе простого среднего арифметического показали, что простое среднее арифметическое чувствительно к аномальным прогнозам. Поэтому в ряде исследований предлагается для объединения прогнозов использовать усеченное среднее для отсеивания аномальных прогнозов [1, с. 47]. Предметом настоящего исследования является метод объединения индивидуальных прогнозов на основе другой робастной статистической оценки, а именно на основе одношаговой М-оценки Эндрюса.

М-оценкой называется любая оценка θ_n, определяемая как решение экстремальной задачи на минимум вида $\sum_{i=1}^{n} \rho(x_i, \theta_n) = min$ или как решение неявного уравнения вида $\sum_{i=1}^{n} \psi(x_i, \theta_n) = 0$, где ρ - произвольная функция; $\psi(x, \theta) = \frac{\partial}{\partial \theta} \rho(x, \theta)$ [2, с. 51]. Применительно к оценке параметра положения М-оценка имеет вид $\sum_{i=1}^{n} \psi(x_i - \theta_n) = 0$ [2, с. 52].

Один из видов функции ψ предложен в 1972 г. Дж. Эндрюсом. Функция ψ Эндрюса является немонотонной функцией, представляющей собой синусоиду. Она имеет вид [3, с. 185]

$$\psi_{sin(a)} = \begin{cases} sin\left(\dfrac{x}{a}\right) & при \quad -\pi a \le x \le \pi a, \\ 0 \quad при \quad |x| > \pi a. \end{cases}$$

Дж. Эндрюс рекомендовал выбирать значение a , равным 2,1.

Так как М-оценки параметра положения не инвариантны относительно параметра рассеивания, то целесообразно находить данные оценки как решение уравнения вида $\sum_{i=1}^{n} \psi\left(\dfrac{x_i - \theta_n}{S_n}\right) = 0$, где S_n - оценка среднего квадратического отклонения [3, с. 138].

Вычисление М-оценок производится на основе итерационной процедуры. Вариант итерационного алгоритма на основе метода Ньютона заключается в последовательном вычислении М-оценки по формуле [2, с. 149]

$$\theta_n^{(j+1)} = \theta_n^{(j)} + S_n^{(0)} \sum_{i=1}^{n} \psi\left(\frac{x_i - \theta_n^{(j)}}{S_n^{(0)}}\right) \bigg/ \sum_{i=1}^{n} \psi'\left(\frac{x_i - \theta_n^{(j)}}{S_n^{(0)}}\right),$$

где j , $j = 0, 1, \dots,$ - номер итерации.

К основной проблеме использования М-оценок параметра положения относится чувствительность к неверной оценке параметра рассеивания [3, с. 187]. Для решения этой проблемы рекомендуется использовать одношаговые М-оценки с робастными начальными значениями параметров положения и рассеивания [3, с. 187].

Одношаговые М-оценки, являющиеся результатом первого шага итерационной процедуры вычисления М-оценок, определяются по формуле [3, с. 139]

$$\theta_n^1 = \theta_n^{(0)} + S_n^{(0)} \sum_{i=1}^{n} \psi\left(\frac{x_i - \theta_n^{(0)}}{S_n^{(0)}}\right) \bigg/ \sum_{i=1}^{n} \psi'\left(\frac{x_i - \theta_n^{(0)}}{S_n^{(0)}}\right),$$

где $\theta_n^{(0)}$ и $S_n^{(0)}$ - начальные оценки параметров положения и рассеивания.

В качестве начальных оценок параметров положения и рассеивания рекомендуется использовать следующие робастные оценки [3, с. 139]: медиану $\theta_n^{(0)} = med(x_i)$ и нормированную медиану абсолютных отклонений $S_n^{(0)} = M'_{AO} = 1{,}483\, MAD(x_i) = 1{,}483\, med_i\, \{|x_i - med_j(x_j)|\}$ соответственно.

Основная проблема применения одношаговых М-оценок заключается в том, что при некоторых выборках знаменатель, равный $\sum_{i=1}^{n} \psi'\left(\dfrac{x_i - \theta_n^{(0)}}{S_n^{(0)}}\right)$, может оказаться равным нулю [3, с. 188]. Один их методов решения данной проблемы состоит в переходе к одношаговой W-оценке Тьюки вида $W_n^1 = \sum_{i=1}^{n} x_i w(u_i) \Big/ \sum_{i=1}^{n} w(u_i)$, где $u_i = \dfrac{x_i - med(x_i)}{1{,}483\, MAD(x_i)}$, а неотрицательная весовая функция w связана с ψ соотношением $w(u) = \psi(u)/u$ [3, с. 188].

Алгоритм вычисления одношаговой М-оценки Эндрюса включает следующие этапы.

1. Вычисление робастных начальных оценок параметров положения и рассеивания (медианы выборки данных и нормированной медианы абсолютных отклонений).

2. Задание значения параметра a в функции ψ Эндрюса.

3. Вычисление значений производной функции $\psi_{sin(a)}$ Эндрюса $\psi'_{sin(a)}$ при $\theta_n^{(0)} = M_0$ и $S_n^0 = M'_{AO}$ для каждого значения выборки по формуле

$$\psi'_{sin(a)}\left(\frac{x_i - M_0}{M'_{AO}}\right) = \begin{cases} -\dfrac{1}{a M'_{AO}} cos\left(\dfrac{x_i - M_0}{a M'_{AO}}\right) & при \quad -\pi a \le \dfrac{x_i - M_0}{M'_{AO}} \le \pi a, \\ 0 \quad при & \left|\dfrac{x_i - M_0}{M'_{AO}}\right| > \pi a. \end{cases}$$

4. Вычисление суммы производных функции $\psi_{sin(a)}$ Эндрюса $\psi'_{sin(a)}$ для всех значений выборки $\sum_{i=1}^{n} \psi'_{sin(a)}\left(\frac{x_i - M_0}{M'_{AO}}\right)$.

5. Если $\sum_{i=1}^{n} \psi'_{sin(a)}\left(\frac{x_i - \theta_n^{(0)}}{S_n^{(0)}}\right) \ne 0$, то вычисление одношаговой М-оценки Эндрюса, которая при $\theta_n^{(0)} = M_0$ и $S_n^0 = M'_{AO}$ имеет вид

$$\theta_n^1 = M_0 + M'_{AO} \sum_{i=1}^{n} \psi_{sin(a)}\left(\frac{x_i - M_0}{M'_{AO}}\right) \bigg/ \sum_{i=1}^{n} \psi'_{sin(a)}\left(\frac{x_i - M_0}{M'_{AO}}\right),$$

где $\psi_{sin(a)} = \begin{cases} sin\left(\dfrac{x_i - M_0}{a M'_{AO}}\right) & при \quad -\pi a \le \dfrac{x_i - M_0}{M'_{AO}} \le \pi a, \\ 0 \quad при & \left|\dfrac{x_i - M_0}{M'_{AO}}\right| > \pi a. \end{cases}$

Если $\sum_{i=1}^{n} \psi'_{sin(a)}\left(\frac{x_i - \theta_n^{(0)}}{S_n^{(0)}}\right) = 0$, то вместо одношаговой М-оценки Эндрюса вычисляется одношаговая W-оценка Тьюки вида

$$W_n^1 = \sum_{i=1}^{n} x_i\, w\left(\frac{x_i - M_0}{M'_{AO}}\right) \bigg/ \sum_{i=1}^{n} w\left(\frac{x_i - M_0}{M'_{AO}}\right),$$

где $w\left(\frac{x_i - M_0}{M'_{AO}}\right) = \left|\dfrac{\psi_{sin(a)}\left(\frac{x_i - M_0}{M'_{AO}}\right)}{\frac{x_i - M_0}{M'_{AO}}}\right| =$

$$
= \begin{cases} \left| \sin\left(\dfrac{x_i - M_0}{a M'_{AO}}\right) \Big/ \left(\dfrac{x_i - M_0}{M'_{AO}}\right) \right| & npu \quad -\pi a \le \dfrac{x_i - M_0}{M'_{AO}} \le \pi a, \\ 0 \quad npu \quad \left| \dfrac{x_i - M_0}{M'_{AO}} \right| > \pi a. \end{cases}
$$

В исследовании фиксированный базовый набор гибридной модели формировался с использованием индивидуальных моделей на один интервал времени вперед, которые могут применяться на начальных этапах прогнозирования: 1) модель на основе предыдущего значения показателя; 2) модель на основе абсолютного прироста за предыдущий интервал времени; 3) модель на основе коэффициента роста за предыдущий интервал времени; 4) модель на основе простого среднего значения; 5) модель на основе среднего абсолютного прироста; 6) модель на основе среднего коэффициента роста; 7) однопараметрическая модель Брауна на основе экспоненциального среднего нулевого порядка; 8) двухпараметрическая модель Хольта.

Для оценки точности комбинированных прогнозов, полученных при объединении индивидуальных прогнозов на основе М-оценки Эндрюса, были использованы фрагменты временных рядов объемов продаж легковых автомобилей, компьютеров, бензина, хлеба, мяса и мороженого, приведенные в [4, с. 79-80].

Для исследования точности прогнозов были использованы следующие показатели: максимальное значение модуля относительной ошибки прогноза (δ_{max}); средняя квадратическая ошибка (RMSE) прогноза; средняя абсолютная ошибка в процентах (MAPE). Значения перечисленных показателей были нормированы значениями соответствующих показателей для гибридной модели на основе объединения прогнозов с использованием взвешенного арифметического среднего значения.

Результаты исследований для временного ряда объема продажи хлеба представлены в табл. 1.

Таблица 1

Показатели точности прогноза объема продажи хлеба

($\delta_{max} \approx$ 58,91%; RMSE $\approx$ 19676,94 млн. руб.; MAPE=20,57%)

Оценка для объединения прогнозов	Показатели точности прогноза		
	δ_{max}	RMSE	MAPE
Взвешенное среднее	1,00	1,00	1,00
Простое среднее	0,70	0,94	1,00
Усеченное среднее (α =0,250)	0,68	0,87	0,91
Оценка Диксона в виде среднего из двух наилучших наблюдений	0,61	0,87	0,89
Оптимальная комплексная оценка по четырем квантилям	0,64	0,91	0,96
Одношаговая W-оценка с ψ -функцией Эндрюса (a=2,10)	0,70	0,90	0,94

Для сравнения в данной таблице приведены показатели точности прогноза комбинированных моделей на основе оценки Диксона в виде среднего из двух наилучших наблюдений и на основе оптимальной комплексной оценки по четырем квантилям [4, с. 76-77], которые являются наиболее точными при прогнозировании рассматриваемых временных рядов. В таблице приведены значения показателей точности комбинированного прогноза, рассчитанного с использованием одношаговой бивес-оценки Тьюки с весами на основе функции ψ Эндрюса, так как знаменатель одношаговой М-оценки с функцией ψ Эндрюса для всех уровней рассмотренных временных рядов был близок к нулю.

Для наглядности анализа полученных результатов комбинированные модели прогнозирования на основе оценок, перечисленных в табл. 1, упорядочены в табл. 2 по количеству временных рядов, для которых все показатели точности прогноза не хуже, чем при использовании комбинированной модели на основе взвешенного арифметического среднего значения.

Таблица 2

Упорядочивание комбинированных моделей по всем показателям

Оценка для объединения прогнозов	Не хуже, чем при использовании взвешенного среднего	В том числе лучше
Оценка Диксона в виде среднего из двух наилучших наблюдений	6 из 6	6 из 6
Оптимальная комплексная оценка по четырем квантилям	5 из 6	5 из 6
Усеченное среднее (α =0,250)	4 из 6	4 из 6
Одношаговая W-оценка с ψ -функцией Эндрюса (a=2,10)	4 из 6	4 из 6
Простое среднее	3 из 6	2 из 6

Результаты проведенных исследований (в том числе анализ табл. 1 и 2) позволяют сделать следующие выводы.

1. Точность пошагового прогноза на основе комбинированной модели с использованием одношаговой М-оценки Эндрюса на множестве рассмотренных временных рядов выше точности комбинированных моделей на основе простого и взвешенного арифметических средних.

2. Точность прогноза на основе комбинированной модели с использованием одношаговой М-оценки Эндрюса на множестве этих рядов несколько хуже точности комбинированных моделей на основе оценки Диксона в виде среднего из двух наилучших наблюдений и на основе оптимальной комплексной оценки по четырем квантилям.

3. Точность прогноза на основе комбинированной модели с использованием одношаговой М-оценки Эндрюса примерно соответствует точности комбинированных моделей на основе усеченного среднего.

4. Точность прогноза на основе комбинированной модели с использованием одношаговой М-оценки Эндрюса на множестве рассмотренных временных рядов практически не зависит от выбора значения параметра a из диапазона от 1 до 6.

Список литературы

1. Френкель А.А., Сурков А.А. Объединение прогнозов – эффективный инструмент повышения точности прогнозирования // Экономист. – 2015. - №1. – С. 44-56.
2. Хьюбер Дж.П. Робастность в статистике: научное изд. / Пер. с англ. – М.: Мир, 1984. – 304 с.
3. Робастность в статистике. Подход на основе функций влияния: научное изд. / Ф. Хампель, Э. Рончетти, П. Рауссеу, В. Штаэль; пер. с англ. – М.: Мир, 1989. – 512 с.
4. Васильев А.А. Гибридные модели прогнозирования объема продаж нового товара с использованием оценок на основе порядковых статистик // Современные научные исследования и инновации. - 2014. № 8 [Электронный ресурс]. - URL: http://web.snauka.ru/issues/2014/08/37268 (дата обращения 10.04.2015).

COMBINATION OF FORECASTS BASEDON THE ANDREWS' ONE STEP M-ESTIMATE

A.A. Vasil'ev

Tver State University, Tver

The author proposes to use the Andrews' one-step M-estimate instead of the weighted average mean for the combination of individual models forecasts. The results of the research has proved the hypothesis that the accuracy of combined forecasting models based on the Andrews' M-estimate exceeds the accuracy of the models based on the weighted average mean.
***Keywords**: the Andrews' M-estimate, combination of forecasts, combined model, forecast, one-step estimate, simple arithmetic mean, the weighted average mean.*

Об авторе:

ВАСИЛЬЕВ Александр Анатольевич – кандидат технических наук, доцент, заведующий кафедрой математики, статистики и информатики в экономике, Тверской государственный университет, (170000, г. Тверь, ул. Желябова, д. 33), e-mail: vasiljev-tvgu@yandex.ru

About the authors:

VASIL'EV Aleksandr Anatol'evich – Philosophy Doctor in Engineering Science, Associate Professor, Head of Department of Mathematics, Statistics and Informatics in Economics, Tver State University, (33, Zhelaybova St., Tver, 170000), e-mail: vasiljev-tvgu@yandex.ru

References

1. Frenkel' A.A., Surkov A.A. Ob#edinenie prognozov – jeffektivnyj instrument povyshenija tochnosti prognozirovanija. Jekonomist. 2015. №1. S. 44-56.
2. H'juber Dzh.P. Robastnost' v statistike: nauchnoe izd. Per. s angl. M.: Mir, 1984. 304 s.
3. Hampel' F., Ronchetti Je., Rausseu P., Shtajel' V. Robastnost' v statistike. Podhod na osnove funkcij vlijanija: nauchnoe izd. Per. s angl. M.: Mir, 1989. 512 s.
4. Vasil'ev A.A. Gibridnye modeli prognozirovanija ob#ema prodazh novogo tovara s ispol'zovaniem ocenok na osnove porjadkovyh statistik. Sovremennye nauchnye issledovanija i innovacii. 2014. № 8 [Jelektronnyj resurs]. - URL: http://web.snauka.ru/issues/2014/08/37268 (data obrashhenija 10.04.2015).