

## Минимизация оценки дисперсии прогноза на примере моделей ETS

Н.В. Белецкая<sup>\*,\*\*</sup>, Д.А. Петрусевиц<sup>\*</sup>

<sup>\*</sup> Российский Технологический Университет (МИРЭА), Москва, Россия

<sup>\*\*</sup> Институт проблем передачи информации им. А.А. Харкевича, Москва, Россия

Поступила в редколлегию 20.02.2024 г. Принята 01.04.2024 г.

**Аннотация**—В рамках представленной работы рассматривается построение объединённой модели временных рядов (если две модели одинаковые по типу и обладают свойством аддитивности, например: ARIMA) или объединённого прогноза моделей (в случае отсутствия свойства аддитивности, например: для моделей ETS), при котором минимизируется оценка дисперсии прогноза. В отличие от других моделей временных рядов, где оценка дисперсии прогноза даётся на основе критерия Стьюдента, для моделей ARIMA и ETS можно построить функцию, которая связана с параметрами модели. Таким образом, есть возможность оценить величину доверительного интервала для прогноза и построить комбинации моделей с минимальной оценкой ширины интервала в зависимости от параметров комбинации. В теоретической части работы исследуются линейные комбинации прогнозов двух моделей, в которых минимизируется оценка дисперсии прогноза (вне зависимости от типа модели). Построен гессиян функции оценки дисперсии прогноза при построении линейной комбинации прогнозов. Он анализируется в условиях наличия экстремума (равенство нулю первых производных функции оценки дисперсии прогноза объединяемых моделей). Далее гессиян оценивается для некоторых групп моделей ETS, рассматриваются условия наличия минимума оценки дисперсии прогноза в стационарной точке в зависимости от значений параметров моделей.

**КЛЮЧЕВЫЕ СЛОВА:** ETS, экспоненциальное сглаживание, оптимальное объединение, минимизация дисперсии прогноза.

DOI: 10.53921/18195822\_2024\_24\_1\_16

### 1. ВВЕДЕНИЕ

При моделировании временных рядов исследователи применяют сразу несколько моделей. Отличаются они, как правило, и по своей природе: качественные зависимости, чаще всего, выражающиеся линейными регрессиями [1–3]; подобные модели, в которых строится линейная зависимость между параметрами, но учитывающие природу данных временных рядов (ArDL, VAR) [1]; модели временных рядов, в которых подразумевается допущение, что само поведение ряда вне других качественных зависимостей диктует его будущие значения, которые предсказываются на основе предыдущего поведения (ARIMA, ETS) [1, 2]. При прогнозировании величины может использоваться как множество моделей по отдельности, так и их комбинация. При объединении моделей разной природы, в основном, используется моделирование в несколько этапов [4–8]. В представленной работе рассматривается объединение однотипных моделей, для которых можно построить функцию оценки дисперсии прогноза, основанную на параметрах модели, а не на критерии Стьюдента. Таким образом, можно теоретически оценить эффект построения линейной комбинации прогнозов двух моделей в условиях минимизации оценки дисперсии прогноза. Представленное в теоретической части исследование применимо к комбинации моделей любой природы, в вычислительном эксперименте рассматриваются

примеры на моделях ETS [1]. Модели, основанные на нейронных сетях, чаще всего, хорошо справляются с моделированием временных рядов, не обладающих сезонностью, или работают с результатами преобразований, которые позволяют избавиться от сезонности. Работа сети LSTM представлена в исследованиях, например: [9–12], GRU – в [13], RNN – в [14, 15]. При этом, для сезонных временных рядов зачастую стандартные статистические модели обладают сравнимым качеством прогнозирования (такой пример и рассматривается в вычислительном эксперименте).

Описание построения взвешенной линейной комбинации, минимизирующей дисперсию прогноза, даётся в соответствии с работами [16, 17]. Функция объединения нескольких прогнозов выглядит следующим образом:

$$F = w_1 f_1 + w_2 f_2 + \dots + w_n f_n = \sum_{i=1}^n w_i f_i, \sum_{i=1}^n w_i = 1. \quad (1)$$

Здесь  $F$  – объединённый прогноз,  $f_i$  – прогнозы объединяемых моделей,  $w_i$  – весовые коэффициенты, с которыми происходит объединение. Если все коэффициенты совпадают:  $w_1 = w_2 = \dots = w_n = 1/n$ , – то получаем усреднение прогнозов  $n$  моделей. При объединении прогнозов двух моделей объединённый прогноз имеет вид (2):

$$F = w f_1 + (1 - w) f_2. \quad (2)$$

Т.к. веса всего два и между ними есть взаимосвязь (1), произведена замена:  $w = w_1, w_2 = 1 - w_1$ . Дисперсия прогноза, объединяющего предсказания двух моделей (2):

$$Var(\sigma_1, \sigma_2, \rho, w) = (w\sigma_1)^2 + ((1 - w)\sigma_2)^2 + 2w(1 - w)\rho\sigma_1\sigma_2.$$

При дифференцировании функции объединённого прогноза получим единственный экстремум:  $(Var(\sigma_1, \sigma_2, \rho, w))'_w = 0$  при

$$w = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2},$$

$$1 - w = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

Он является минимумом в связи с тем, что  $|\rho| \leq 1$ , вторая производная ограничена положительными значениями, функция оценки дисперсии прогноза выпукла вниз:

$$(Var(\sigma_1, \sigma_2, \rho, w))''_w = 2(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2),$$

$$(\sigma_1 - \sigma_2)^2 \leq (Var(\sigma_1, \sigma_2, \rho, w))''_w \leq (\sigma_1 + \sigma_2)^2.$$

Если вес  $w$  превышает 1, то его значение следует установить равным 1. Если он ниже 0, то его значение устанавливается равным 0 в рамках вычислительного эксперимента. При оптимальном значении веса дисперсия объединённого прогноза приобретает вид:

$$Var(\sigma_1, \sigma_2, \rho) = \frac{(\sigma_1\sigma_2)^2(1 - \rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}. \quad (3)$$

## 2. МИНИМИЗАЦИЯ ОЦЕНКИ ДИСПЕРСИИ ПРОГНОЗА ПРИ ОПТИМАЛЬНОМ ОБЪЕДИНЕНИИ МОДЕЛЕЙ

Для функции (3) найдём экстремумы по параметрам объединяемых моделей и проверим их на наличие минимумов или максимумов. Для этого вычислим первые производные выражения и составим условия на наличие экстремума, затем исследуя гессиан в этих точках.

Для определённости предположим, что дисперсии объединяемых моделей зависят от одного параметра (для первой модели обозначим как  $x$ ,  $\sigma_1 = \sigma_1(x)$ , для второй –  $y$ ,  $\sigma_2 = \sigma_2(y)$ ):

$$\text{Var}(\sigma_1, \sigma_2, \rho)'_x = \sigma_2^3(1 - \rho^2)(\sigma_1^2)'_x \frac{\sigma_2 - \rho\sigma_1(x)}{\sigma_1^2(x) + \sigma_2^2 - 2\rho\sigma_1(x)\sigma_2}. \quad (4)$$

Производную  $(\sigma_1^2)'_x$  не обрабатываем, т.к. она чаще всего проще выглядит проще именно в форме квадрата. Из-за того, что исходная функция симметрична относительно замены  $\sigma_1 \leftrightarrow \sigma_2$ , условия на наличие экстремума:

$$\begin{aligned} \text{Var}(\sigma_1, \sigma_2, \rho)'_x &= \sigma_2^3(1 - \rho^2)(\sigma_1^2)'_x \frac{\sigma_2 - \rho\sigma_1(x)}{[\sigma_1^2(x) + \sigma_2^2 - 2\rho\sigma_1(x)\sigma_2]^2} = 0, \\ \text{Var}(\sigma_1, \sigma_2, \rho)'_y &= \sigma_1^3(1 - \rho^2)(\sigma_2^2)'_y \frac{\sigma_1 - \rho\sigma_2(y)}{[\sigma_1^2 + \sigma_2^2(y) - 2\rho\sigma_1\sigma_2(y)]^2} = 0. \end{aligned}$$

Здесь  $x, y$  пробегает все параметры объединяемых моделей ( $x$  для первой,  $y$  для второй), поэтому уравнений получается в два раза больше, чем параметров модели. Например, для  $ETS(A, N, N)$  – два, по одному от каждой модели в объединении, для  $ETS(A, A, N)$  – четыре (два параметра в паре моделей) и т.д.

Сформулируем следующее утверждение о гессиане функции объединённой дисперсии (3).

**Теорема 1.** *Гессиан функции оценки дисперсии линейной взвешенной комбинации двух моделей, минимизированной по весу моделей (3), имеет диагональный блочный вид (блоки вокруг главной диагонали имеют ненулевые значения (в общем случае), зависящие от типа модели, а блоки вокруг побочной диагонали – нулевые. Он пропорционален определителю блочной матрицы, у которой на главной диагонали стоят функции  $(\sigma_1^2)'_x(\sigma_2(y) - \rho\sigma_1(x))$  и  $(\sigma_2^2)'_y(\sigma_1(x) - \rho\sigma_2(y))$  (здесь  $x$  пробегает параметры первой модели, участвующей в объединении, а  $y$  – второй).*

$$H \sim \begin{pmatrix} (\sigma_1^2)'_x(\sigma_2(y) - \rho\sigma_1(x)) & 0 \\ 0 & (\sigma_2^2)'_y(\sigma_1(x) - \rho\sigma_2(y)) \end{pmatrix}.$$

**Доказательство.** В точках, где выполняются соотношения (2), вычислим вторые производные. Для общности рассуждений предположим зависимость функции дисперсии объединяемых моделей от двух параметров, например:  $\sigma_1 = \sigma_1(x_1, x_2)$ , – при этом полагая, что  $\sigma_2 = \sigma_2(y)$  не зависит от  $x_1, x_2$ . Для этого заменим  $x$  на  $x_1$  в функции (4) и дифференцируем её по параметру  $x_2$ . Предварительно разделим выражение на две части: зависимую от  $x_1$  и  $x_2$  и не зависимую:

$$\begin{aligned} \text{Var}(\sigma_1, \sigma_2, \rho)'_{x_1} &= \sigma_2^3(1 - \rho^2) \times (\sigma_1^2)'_{x_1} \frac{\sigma_2 - \rho\sigma_1}{(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^2}, \\ f(x) &= (\sigma_1^2)'_{x_1} \frac{\sigma_2 - \rho\sigma_1}{(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^2}. \end{aligned}$$

Логарифмируем  $f(x)$ :

$$\ln f(x) = \ln [(\sigma_1^2)'_{x_1}] + \ln(\sigma_2 - \rho\sigma_1) - 2 \ln (\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2).$$

И дифференцируем по  $x_2$  ( $\sigma_2$  не зависит от неё):

$$f'_{x_2}(x)/f(x) = \frac{(\sigma_1^2)''_{x_1x_2}}{(\sigma_1^2)'_{x_1}} - (\sigma_1)'_{x_2} \left[ \frac{\rho}{\sigma_2 - \rho\sigma_1} + 4 \frac{\sigma_1 - \rho\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \right].$$

Приведём дроби в скобках к общему знаменателю и запишем ответ:

$$\begin{aligned} \text{Var}(\sigma_1, \sigma_2, \rho)''_{x_1x_2} &= \sigma_2^3(1 - \rho^2)(\sigma_1^2)'_{x_1} \frac{\sigma_2 - \rho\sigma_1}{[\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2]^2} \times \\ &\left[ \frac{(\sigma_1^2)''_{x_1x_2}}{(\sigma_1^2)'_{x_1}} - (\sigma_1)'_{x_2} \frac{-3\rho(\sigma_1^2 + \sigma_2^2) + 2(\rho^2 + 2)\sigma_1\sigma_2}{(\sigma_2 - \rho\sigma_1)(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)} \right]. \end{aligned}$$

Для того, чтобы найти смешанную производную по параметрам объединяемых моделей  $\sigma_1 = \sigma_1(x)$ ,  $\sigma_2 = \sigma_2(y)$ , аналогично дифференцируем (4) по  $y$ . Для этого разделим функцию на две части: не зависящие от  $y$  параметры и зависимую часть, которую обозначим  $g(y)$ :

$$\begin{aligned} \text{Var}(\sigma_1, \sigma_2, \rho)'_x &= (1 - \rho^2)(\sigma_1^2)'_x \times \sigma_2^3(y) \frac{\sigma_2(y) - \rho\sigma_1}{[\sigma_1^2(x) + \sigma_2^2(y) - 2\rho\sigma_1\sigma_2(y)]^2}, \\ g(y) &= \sigma_2^3(y) \frac{\sigma_2(y) - \rho\sigma_1}{[\sigma_1^2 + \sigma_2^2(y) - 2\rho\sigma_1\sigma_2(y)]^2}. \end{aligned}$$

Прологарифмируем  $g(y)$ :

$$\ln g(y) = 3 \ln(\sigma_2(y)) + \ln(\sigma_2(y) - \rho\sigma_1) - 2 \ln (\sigma_1^2 + \sigma_2^2(y) - 2\rho\sigma_1\sigma_2(y)).$$

Дифференцируем:

$$g'(y)/g(y) = 3 \frac{(\sigma_2(y))'_y}{\sigma_2(y)} + \frac{(\sigma_2)'_y}{\sigma_2(y) - \rho\sigma_1} - 2 \frac{(\sigma_2^2)'_y - 2\rho\sigma_1(\sigma_2)'_y}{\sigma_1^2 + \sigma_2^2(y) - 2\rho\sigma_1\sigma_2(y)}.$$

Здесь удобно в числителе последней дроби раскрыть производную квадрата, тогда вынесем общий множитель всех дробей  $(\sigma_2)'_y$ :

$$g'(y)/g(y) = (\sigma_2(y))'_y \left[ \frac{3}{\sigma_2(y)} + \frac{1}{\sigma_2(y) - \rho\sigma_1} - 4 \frac{\sigma_2(y) - \rho\sigma_1}{\sigma_1^2 + \sigma_2^2(y) - 2\rho\sigma_1\sigma_2(y)} \right].$$

Приведём вторую и третью дробь к общему знаменателю:

$$g'(y)/g(y) = (\sigma_2(y))'_y \left[ \frac{3}{\sigma_2(y)} + \frac{\sigma_1^2 + \sigma_2^2(y) - 2\rho\sigma_1\sigma_2(y) - 4[\sigma_2(y) - \rho\sigma_1]^2}{(\sigma_2(y) - \rho\sigma_1)(\sigma_1^2 + \sigma_2^2(y) - 2\rho\sigma_1\sigma_2(y))} \right].$$

Обратим внимание на то, что в числителе дроби получается квадрат. Раскроем скобки во второй дроби:

$$g'(y)/g(y) = (\sigma_2)'_y \left[ \frac{3}{\sigma_2} + \frac{-3\sigma_2^2 + (1 - 4\rho^2)\sigma_1^2 + 6\rho\sigma_1\sigma_2}{(\sigma_2 - \rho\sigma_1)(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)} \right].$$

Приведём дроби к общему знаменателю и приведём подобные:

$$g'(y)/g(y) = \frac{(\sigma_2)'_y \sigma_1}{\sigma_2(\sigma_2 - \rho\sigma_1)(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)} [-3\rho(\sigma_1^2 + \sigma_2^2) + 2(\rho^2 + 2)\sigma_1\sigma_2].$$

Подставим результат логарифмического дифференцирования в выражение для смешанной производной:

$$\begin{aligned} Var(\sigma_1, \sigma_2, \rho)''_{xy} &= \sigma_2^3(1 - \rho^2)(\sigma_1^2)'_x \frac{\sigma_2 - \rho\sigma_1}{[\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2]^2} \times \\ &\frac{(\sigma_2)'_y \sigma_1}{\sigma_2(\sigma_2 - \rho\sigma_1)(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)} [-3\rho(\sigma_1^2 + \sigma_2^2) + 2(\rho^2 + 2)\sigma_1\sigma_2]. \end{aligned}$$

Сокращаем:

$$Var(\sigma_1, \sigma_2, \rho)''_{xy} = \sigma_2^2(1 - \rho^2)(\sigma_1^2)'_x (\sigma_2)'_y \frac{-3\rho(\sigma_1^2 + \sigma_2^2) + 2(\rho^2 + 2)\sigma_1\sigma_2}{(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^3}.$$

Естественно, смешанная производная по нескольким параметрам второй модели будет выражаться аналогично (4) с учётом замены  $\sigma_1 \leftrightarrow \sigma_2$ . Рассмотрим вторые производные в местах экстремумов. В этих точках выполняются условия (2). За исключением особых случаев, рассмотренных ниже, условия на экстремум выглядят так:

$$(\sigma_i^2)' = 0.$$

В выражении второй производной по параметрам одной и той же модели (4) сокращаются  $(\sigma_1^2)'_{x_1}$  в первом слагаемом, а у второго слагаемого остаётся нулевой множитель  $(\sigma_1^2)'_{x_1} = 0$ . Смешанная производная состоит из единственного выражения, в котором участвует  $(\sigma_1^2)'_{x_1} = 0$ , поэтому:

$$\begin{aligned} Var(\sigma_1, \sigma_2, \rho)''_{x_1x_2} &= \sigma_2^3(1 - \rho^2) \frac{\sigma_2 - \rho\sigma_1}{(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^2} (\sigma_1^2)''_{x_1x_2}, \\ Var(\sigma_1, \sigma_2, \rho)''_{xy} &= 0. \end{aligned}$$

В такой ситуации гессиан распадается на части: левый верхний блок и правый нижний блок – функции, зависящие от параметров первой и второй моделей соответственно (если модели однотипные, то функции одинаковые и отличия между ними проявляются только в замене  $\sigma_1 \leftrightarrow \sigma_2$ ). Блоки вдоль побочной диагонали нулевые, потому что они относятся к смешанным производным по параметрам первой и второй модели  $(\sigma_i^2)''_{xy}$ , которые обращаются в ноль. Заметим, что каких-либо предположений о типах моделей не сделано, поэтому гессиан для функции прогноза взвешенной линейной комбинации двух моделей, минимальной по весу, диагонален и пропорционален определителю из вторых производных функций  $(\sigma_i^2)''_{x_1x_2}$ . Приведём пример простейшей модели  $ETS(A, N, N)$ :

$$H = (1 - \rho^2)^2 \frac{(\sigma_1\sigma_2)^3(\sigma_2 - \rho\sigma_1)(\sigma_1 - \rho\sigma_2)}{(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^4} \begin{pmatrix} (\sigma_1^2)''_x & 0 \\ 0 & (\sigma_2^2)''_y \end{pmatrix}$$

Рассмотрим особые случаи, при которых система (2) имеет решения, отличные от  $(\sigma_i^2)' = 0$ . Первая особая ситуация: числители дробей равны 0.

$$\begin{aligned} \sigma_2 - \rho\sigma_1 &= 0, \\ \sigma_1 - \rho\sigma_2 &= 0. \end{aligned}$$

Здесь подставим одно уравнение в другое и получим  $\rho = \pm 1$ . Это соответствует полностью согласующимся или полностью противоположным прогнозам объединяемых моделей. В первом случае объединять модели бессмысленно, одну надо исключить. Второй случай тоже на практике не возможен, т.к. не понятна трактовка моделей, предсказывающих полностью противоположное поведение рассматриваемого временного процесса. Другая особая ситуация:

$$\begin{aligned} \text{Var}(\sigma_1, \sigma_2, \rho)'_x &\sim \sigma_2 - \rho\sigma_1 = 0, \\ \text{Var}(\sigma_1, \sigma_2, \rho)'_y &\sim (\sigma_2^2)'_y = 0. \end{aligned}$$

В этой системе можно заменить  $\sigma_1 \leftrightarrow \sigma_2$  и получить аналогичный случай. Второе уравнение выполняется при условии  $\sigma_2 = 0$  или  $(\sigma_2^2)'_y = 0$ . В первом случае  $\sigma_1 = \sigma_2 = 0$ , речь идёт о том же самом экстремуме, что и для других моделей. Во втором случае  $w = 0, 1-w = 1$ , выбирается одна из двух моделей вместо построения их комбинации. В таком вырожденном случае один из блоков гессиана на главной диагонали обнуляется, гессиан остаётся диагональным, но равен 0.

### 3. ДИСПЕРСИЯ ПРОГНОЗА ОБЪЕДИНЁННОЙ МОДЕЛИ НА ПРИМЕРЕ ETS

В таблице 1 приведены основные типы моделей ETS. Для моделей, у которых достаточно просто выглядит выражение оценки дисперсии прогноза, приведём условия на параметры моделей для наличия экстремумов и гессианы для функции объединённого прогноза. Для моделей вида  $ETS(A, N, N)$  оценка дисперсии выглядит так:

$$\sigma_i^2 = se_i^2(1 + \alpha_i^2(n - 1)),$$

здесь  $se_i$  – стандартная ошибка модели,  $\alpha$  – параметр модели,  $n$  – номер шага по времени, на который делается прогноз (чем дальше прогноз, тем выше дисперсия). Оценка положения экстремума:

$$(\sigma_i^2)'_{\alpha_i} = 2se_i^2\alpha_i(n - 1) = 0, \alpha_i = 0.$$

Вторая производная везде положительная (за исключением прогноза на 1 шаг вперёд):

$$(\sigma_i^2)''_{\alpha_i} = 2se_i^2(n - 1).$$

Гессиан для оценки функции дисперсии прогноза имеет вид:

$$H = (1 - \rho^2)^2 4(n - 1)^2 \frac{(\sigma_1\sigma_2)^3(\sigma_2 - \rho\sigma_1)(\sigma_1 - \rho\sigma_2)}{(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^4} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Условие на наличие минимума в экстремуме функции:

$$\begin{aligned} \sigma_2 - \rho\sigma_1 &> 0, \\ \sigma_1 - \rho\sigma_2 &> 0. \end{aligned}$$

Минимум достигается при условии, что дисперсии объединяемых моделей не отличаются очень сильно, корреляция прогнозов не высока:  $\sigma_i > \rho\sigma_j, i \neq j$ .

Условие для наличия максимума:

$$\begin{aligned} \sigma_2 - \rho\sigma_1 &< 0, \\ \sigma_1 - \rho\sigma_2 &< 0. \end{aligned}$$

Это возможно при условии, что дисперсии объединяемых моделей не отличаются очень сильно, корреляция прогнозов высока:  $\sigma_i < \rho\sigma_j, i \neq j$ .

Экстремум отсутствует, если дисперсии моделей сильно отличаются (во втором случае знаки неравенств одновременно меняются на обратные):

$$\begin{aligned}\sigma_2 - \rho\sigma_1 &< 0, \\ \sigma_1 - \rho\sigma_2 &> 0.\end{aligned}$$

Рисунок 1 объединяет графики функции дисперсии прогноза (3) на  $n = 2$  шага вперёд для моделей типа  $ETS(A, N, N)$ :

$$Var(\sigma_1, \sigma_2, \rho) = \frac{se_1^2 se_2^2 (1 + \alpha_1^2)(1 + \alpha_2^2)(1 - \rho^2)}{se_1^2(1 + \alpha_1^2) + se_2^2(1 + \alpha_2^2) - 2\rho se_1 se_2 \sqrt{(1 + \alpha_1^2)(1 + \alpha_2^2)}} \quad (5)$$

при одинаковых среднеквадратических отклонениях моделей  $se_1 = se_2$  (случаи корреляции  $\rho = 0.1, 0.5, 0.9$ ). Графики для тех же уровней корреляции прогноза, но при соотношении СКО  $se_2/se_1 = 10$  представлены на рисунке 2.

Заметим на рисунке 1 наличие минимума вдоль осей  $\alpha_i$ . Интересно, что в случаях сильной и слабой корреляции  $\rho = 0.1$  и  $\rho = 0.9$  на этих прямых находится локальный максимум в нуле. При удалении от него функция уменьшается. Этому соответствует объединение моделей с сильно отличающимися коэффициентами  $\alpha_i/\alpha_j$ , а вместе с ними сильно отличаются дисперсии прогноза (в среднем случае  $\rho = 0.5$  такого эффекта нет). Обратим внимание на локальные максимумы, расположенные симметрично относительно 0, вдоль прямой, поворнутой относительно координатных осей. Им соответствует выбор одной из двух одинаковых моделей, получающихся при равных коэффициентах.

При рассмотрении объединения моделей с сильно отличающимися дисперсиями на рисунке 2 видно, что глобальный минимум достигается только вдоль той оси  $\alpha_1 = 0$ , которой соответствует параметр модели с меньшей дисперсией. Точкам вдоль второй оси  $\alpha_2$  соответствует локальный минимум, хотя при строгом анализе поведения гессиана в этом случае фиксируется отсутствие экстремума.

Для моделей вида  $ETS(A, A, N)$  оценка дисперсии выглядит так:

$$\sigma_i^2 = se_i^2(1 + (n-1)(\alpha_i^2 + \alpha_i\beta_i n + \beta_i^2 n(2n-1)/6)),$$

здесь  $\alpha, \beta$  – параметры модели. Оценка положения экстремума:

$$\begin{aligned}(\sigma_i^2)'_{\alpha_i} &= se_i^2(2\alpha_i + \beta_i n)(n-1) = 0, \\ (\sigma_i^2)'_{\beta_i} &= se_i^2(\alpha_i n + \beta_i n(2n-1)/3)(n-1) = 0,\end{aligned}$$

Решение системы достигается при  $\alpha_i = -\beta_i, n = 2$  или при  $\beta_i = 0, \alpha_i = 0$ , а также при  $n = 1$ . Вторые производные:

$$\begin{aligned}(\sigma_i^2)''_{\alpha_i} &= 2se_i^2(n-1), \\ (\sigma_i^2)''_{\beta_i} &= se_i^2 n(n-1)(2n-1)/3, \\ (\sigma_i^2)''_{\alpha_i\beta_i} &= 2se_i^2 n(n-1).\end{aligned}$$

Гессиан для оценки функции дисперсии прогноза имеет вид:

$$H \sim \begin{pmatrix} 2(\sigma_2 - \rho\sigma_1) & n(\sigma_2 - \rho\sigma_1) & 0 & 0 \\ n(\sigma_2 - \rho\sigma_1) & n(2n-1)(\sigma_2 - \rho\sigma_1)/3 & 0 & 0 \\ 0 & 0 & 2(\sigma_1 - \rho\sigma_2) & n(\sigma_1 - \rho\sigma_2) \\ 0 & 0 & n(\sigma_1 - \rho\sigma_2) & n(2n-1)(\sigma_1 - \rho\sigma_2)/3 \end{pmatrix},$$

коэффициент перед матрицей составим так, чтобы он был всегда положителен, и выпишем отдельно:

$$\frac{(\sigma_1\sigma_2)^6(1-\rho^2)^4(n-1)^4}{(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^8}.$$

Условие на наличие минимума в экстремуме функции:

$$\begin{aligned}\Delta_1 &\sim \sigma_2 - \rho\sigma_1 > 0, \\ \Delta_2 &\sim 2n(2n-1)/3 - n^2 > 0, \\ \Delta_3 &\sim 2(\sigma_1 - \rho\sigma_2)\Delta_2 > 0, \\ \Delta_4 &\sim [2n(2n-1)/3 - n^2]^2 (\sigma_1 - \rho\sigma_2)^2 (\sigma_2 - \rho\sigma_1)^2 > 0.\end{aligned}$$

В случае миноров  $\Delta_2, \Delta_4$  неуказанные множители – квадраты выражений  $\sigma_i - \rho\sigma_j$ , которые не влияют на знак. Таким образом, так же, как и при рассмотрении модели  $ETS(A, N, N)$ , для наличия минимума в стационарной точке нужно выполнение двух условий:

$$\begin{aligned}\sigma_2 - \rho\sigma_1 &> 0, \\ \sigma_1 - \rho\sigma_2 &> 0,\end{aligned}$$

а также  $\Delta_2 \sim n(n-2)/3 > 0$ , что выполняется при  $n > 2$ . Это возможно при низких значениях коэффициента корреляции и сопоставимых оценках дисперсий прогноза.

Условие на наличие минимума в экстремуме функции – это чередование знаков миноров, начиная с минуса. Учитывая то, что второй и четвёртый минор автоматически положительные (квадраты выражений, умноженные на положительное число) при  $n > 2$ , получим те же условия, что и для модели  $ETS(A, N, N)$ :

$$\begin{aligned}\sigma_2 - \rho\sigma_1 &< 0, \\ \sigma_1 - \rho\sigma_2 &< 0,\end{aligned}$$

что возможно при низких значениях коэффициента корреляции и сопоставимых оценках дисперсий прогноза. При отличающихся существенно дисперсиях прогнозов получим:

$$\text{sign}(\sigma_2 - \rho\sigma_1) \neq \text{sign}(\sigma_1 - \rho\sigma_2),$$

и экстремум в исследуемой точке отсутствует.

При  $n = 2$  нужно дополнительное исследование, т.к.:

$$\det H \sim \begin{vmatrix} 2 & 2 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \end{vmatrix} = 0.$$

В этой ситуации получаем частный случай модели  $ETS(A, N, N)$ .

Характер функций дисперсии прогноза остаётся тем же, что и в случае моделей типа  $ETS(A, N, N)$  представлен на рисунках (1, 2). Например, для прогнозов на  $n = 3$  и  $n = 7$  шагов вперёд:

$$\begin{aligned}n = 3 : \sigma_i^2 &= se_i^2(1 + 2(\alpha_i + 3/2\beta_i)^2 + \beta_i^2/2), \\ n = 7 : \sigma_i^2 &= se_i^2(1 + 6(\alpha_i + 7/2\beta_i)^2 + 35\beta_i^2/2)\end{aligned}$$

получим то же самое выражение, что и для модели  $ETS(A, N, N)$  после подстановок:  $x_i = 2(\alpha_i + 3/2\beta_i)^2 + \beta_i^2/2$  и  $x_i = 6(\alpha_i + 7/2\beta_i)^2 + 35\beta_i^2/2$  соответственно, где  $x_i$  – новые переменные. В общем виде выделение полного квадрата даёт:

$$\sigma_i^2 = se_i^2(1 + (n-1)((\alpha_i + \beta_i n/2) + \beta_i^2(n^2 - 2)/12)),$$

таким образом, замена  $((\alpha_i \beta_i n/2)^2 + \beta_i^2(n^2 - 2)/12) \rightarrow \alpha_i$  приводит к выражению (5) для модели  $ETS(A, A, N)$ . Анализ областей, которые не достигаются при таком переходе, сложен и выходит за рамки нашего исследования.

При объединении моделей вида  $ETS(A, A, A)$  оценка дисперсии выглядит так:

$$\sigma_i^2 = se_i^2(1 + \alpha_i^2(n-1) + \gamma_i k(2\alpha_i + \gamma_i), k = \lceil \frac{n-1}{m} \rceil).$$

здесь  $\alpha_i, \beta_i, \gamma_i$  – параметры модели, коэффициент  $k$  связан с наличием и периодом сезонности. Оценка положения экстремума:

$$\begin{aligned} (\sigma_i^2)'_{\alpha_i} &= se_i^2(2\alpha_i(n-1) + 2\gamma_i k) = 0, \\ (\sigma_i^2)'_{\gamma_i} &= se_i^2(2\alpha_i k + 2\gamma_i k) = 0. \end{aligned}$$

Решение системы достигается при  $\alpha_i = -\gamma_i, n = k+1 (m=1)$  (т.е. для несезонной модели) или при  $k=0, \alpha_i=0$ . Вторые производные:

$$\begin{aligned} (\sigma_i^2)''_{\alpha_i} &= 2se_i^2(n-1), \\ (\sigma_i^2)''_{\gamma_i} &= 2se_i^2 k, \\ (\sigma_i^2)''_{\alpha_i \gamma_i} &= 2se_i^2 k. \end{aligned}$$

Гессиан для оценки функции дисперсии прогноза имеет вид:

$$\det H \sim \begin{vmatrix} 2(n-1)(\sigma_2 - \rho\sigma_1) & 2k(\sigma_2 - \rho\sigma_1) & 0 & 0 \\ 2k(\sigma_2 - \rho\sigma_1) & 2k(\sigma_2 - \rho\sigma_1) & 0 & 0 \\ 0 & 0 & 2(n-1)(\sigma_1 - \rho\sigma_2) & 2k(\sigma_1 - \rho\sigma_2) \\ 0 & 0 & 2k(\sigma_1 - \rho\sigma_2) & 2k(\sigma_1 - \rho\sigma_2) \end{vmatrix}.$$

Миноры, которые начинаются в левом верхнем углу определителя, выглядят так:

$$\begin{aligned} \Delta_1 &\sim 2(n-1)(\sigma_2 - \rho\sigma_1), \\ \Delta_2 &\sim 8k(n-k-1)(\sigma_2 - \rho\sigma_1)^2, \\ \Delta_3 &\sim 2(n-1)(\sigma_1 - \rho\sigma_2)\Delta_2, \\ \Delta_4 &\sim 64k^2(n-k-1)^2(\sigma_2 - \rho\sigma_1)^2(\sigma_1 - \rho\sigma_2)^2. \end{aligned}$$

Здесь опущены положительные коэффициенты, не влияющие на знаки миноров. Итог получается таким же, как для предыдущих случаев: Для наличия максимума нужно, чтобы первый и третий минор были отрицательными:

$$\begin{aligned} \sigma_2 - \rho\sigma_1 &< 0, \\ \sigma_1 - \rho\sigma_2 &< 0, \end{aligned}$$

что может произойти при высокой корреляции между прогнозами объединяемых моделей  $\rho$  и сопоставимыми по масштабу оценками дисперсии прогноза объединяемых моделей  $\sigma_i \sim \sigma_j$ .

И наконец, экстремум в стационарной точке отсутствует при чередовании знаков в неравенствах:

$$\text{sign}(\sigma_2 - \rho\sigma_1) \neq \text{sign}(\sigma_1 - \rho\sigma_2)$$

в тех случаях, когда объединяются модели с сильно отличающимися оценками дисперсии прогноза  $\sigma_i \gg \sigma_j$  и тогда  $\sigma_i > \rho\sigma_j$ .

Т.к.  $k \leq n - 1$ , условия на наличие минимума или максимума сводятся к знакам соотношений вида  $\sigma_i - \rho\sigma_j$ . Второй и четвёртый определитель всегда неотрицательные. Для наличия минимума нужно, чтобы первый и третий миноры были положительными:

$$\begin{aligned}\sigma_2 - \rho\sigma_1 &> 0, \\ \sigma_1 - \rho\sigma_2 &> 0,\end{aligned}$$

что возможно при малых значениях корреляции прогнозов  $\rho$  и сопоставимых значениях оценок дисперсии прогноза объединяемых моделей  $\sigma_i \sim \sigma_j$ .

Случай  $k = 1$  – оценки дисперсии прогноза для несезонных моделей ведёт к рассмотрению указанных выше моделей. Выделяя полный квадрат, получим:

$$\sigma_i^2 = se_i^2(1 + k(\alpha_i + \gamma_i)^2 + \alpha_i^2(n - k - 1)).$$

что опять же после замены  $k(\alpha_i + \gamma_i)^2 + \alpha_i^2(n - k - 1) \rightarrow \alpha_i$  (с учётом того, что  $k \leq n - 1$ ) приводит к выражению (5). Таким образом, рассмотрение различных моделей *ETS*, обладающих относительно простым выражением для оценки функции дисперсии прогноза, фактически приводит к одной и той же ситуации, рассмотренной в приложении к простейшему типу модели *ETS(A, N, N)*.

#### 4. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

В качестве эксперимента рассмотрим временной ряд месячных показателей индекса потребительских цен по данным макроэкономической статистики РФ [18] в период между январём 1991 года и январём 2022 года. Часть данных, на которой обучаются сравниваемые модели: с 1991 по 2021 гг. Последние 12 месяцев используются в качестве тестовой выборки. В таблице 1 представлены основные сезонные и несезонные модели *ETS* [1], сравнение их прогнозов на тестовый период с реальными данными с помощью метрик *RMSE*, *MAE*, значение информационного критерия Акаике *AIC* [1, 2] и оценка дисперсии прогноза  $\sigma^2$  согласно выражениям, представленным в предыдущем пункте. Обозначения моделей *ETS* даны в соответствии с [1]: первая буква ответственна за форму представления остатков (аддитивная или мультипликативная), вторая – за наличие и вид тренда, третья – за вид сезонности. Для сравнения приведены результаты построения модели *ARIMA* на исследуемом временном ряде (значение информационного критерия Акаике вводится и для этой модели, но функциональная форма отличается [1, 19], поэтому не приводим его в таблице), а также результаты работы модели *ETS* на основе беггинга [19–24] *BaggedETS* (пакет доступен в библиотеках на языках R и Python). Основной идеей подхода на основе беггинга является снижение дисперсии остатка ряда при разложении ряда на тренд – сезонность – остаток путём порождения многих последовательностей на основе существующих данных и их усреднения. Остаток может обладать статистическими свойствами, на опознание и использование которых при прогнозировании и направлен подход.

Объединённые пары моделей представлены в таблице 2. В связи с тем, что взвешенная комбинация моделей не является моделью *ETS*, данные по информационным критериям не

**Таблица 1.** Модели временных рядов *ETS*, *ARIMA* и *BaggedETS* для процесса месячных показателей индекса потребительских цен РФ [18]

Модель	<i>RMSE</i>	<i>MAE</i>	<i>AIC</i>	$\sigma^2$
<i>ETS</i> ( <i>A</i> , <i>N</i> , <i>A</i> )	0.35	0.26	843.78	0.45
<i>ETS</i> ( <i>A</i> , <i>A</i> , <i>A</i> )	0.35	0.27	843.30	0.59
<i>ETS</i> ( <i>A</i> , <i>A<sub>d</sub></i> , <i>A</i> )	0.35	0.26	850.31	0.62
<i>ETS</i> ( <i>A</i> , <i>N</i> , <i>N</i> )	0.50	0.36	991.68	1.57
<i>ETS</i> ( <i>A</i> , <i>A</i> , <i>N</i> )	0.50	0.36	995.64	1.60
<i>ETS</i> ( <i>A</i> , <i>A<sub>d</sub></i> , <i>N</i> )	0.50	0.36	997.83	1.51
<i>BaggedETS</i>	0.29	0.24	–	–
<i>ARIMA</i>	0.34	0.28	–	–

приведены. Корреляция прогнозов объединяемых пар моделей на тестовом периоде отображена во втором столбце. В третьем и четвёртом столбцах приведены оценки близости прогноза комбинации моделей к реальным значениям, в последнем столбце – теоретическая оценка дисперсии прогноза.

**Таблица 2.** Объединения прогнозов моделей временных рядов *ETS* для процесса месячных показателей индекса потребительских цен РФ [18]

Модель	Корреляция прогнозов	<i>RMSE</i>	<i>MAE</i>	$\sigma^2$
<i>ETS</i> ( <i>A</i> , <i>A</i> , <i>A</i> ) + <i>ETS</i> ( <i>A</i> , <i>A<sub>d</sub></i> , <i>A</i> )	0.77	0.45	0.40	0.54
<i>ETS</i> ( <i>A</i> , <i>A<sub>d</sub></i> , <i>N</i> ) + <i>ETS</i> ( <i>A</i> , <i>A<sub>d</sub></i> , <i>A</i> )	-0.10	0.49	0.43	0.40
<i>ETS</i> ( <i>A</i> , <i>A<sub>d</sub></i> , <i>N</i> ) + <i>ETS</i> ( <i>A</i> , <i>A</i> , <i>A</i> )	0.20	0.28	0.22	0.50
<i>ETS</i> ( <i>A</i> , <i>A</i> , <i>N</i> ) + <i>ETS</i> ( <i>A</i> , <i>A<sub>d</sub></i> , <i>A</i> )	-0.11	0.51	0.45	0.40
<i>ETS</i> ( <i>A</i> , <i>A</i> , <i>N</i> ) + <i>ETS</i> ( <i>A</i> , <i>A</i> , <i>A</i> )	0.19	0.29	0.23	0.50

Следует отметить снижение, как теоретической оценки дисперсии прогноза, так и практическое уточнение прогнозов моделей. В таблице 2 все комбинации моделей обладают не только лучшими теоретическими оценками дисперсии прогноза (последний столбец таблиц 1, 2), но и лучшими оценками качества произведённого прогноза по метрикам *RMSE*, *MAE* по сравнению с исходными моделями *ETS*. Более того, характеристики прогноза модели *BaggedETS*, которая обладала существенным отрывом по сравнению со стандартными моделями *ETS*, оказались близки с оценками лучших взвешенных комбинаций моделей *ETS*, приведённых в таблице 2.

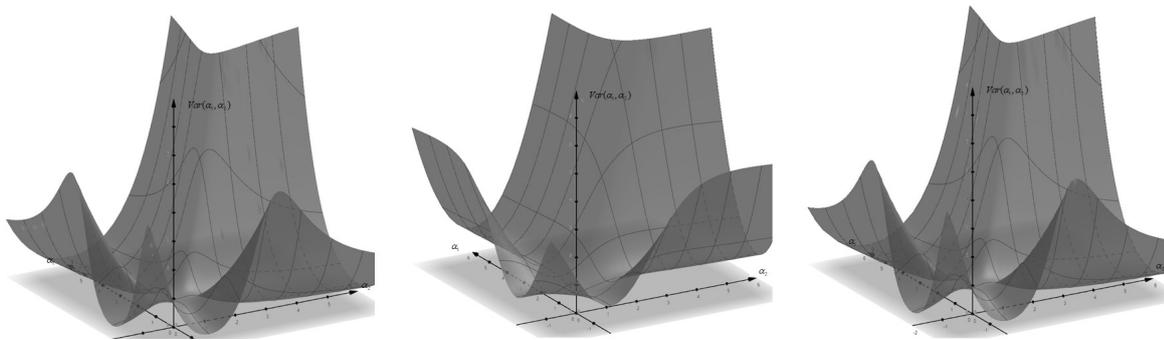
## 5. ЗАКЛЮЧЕНИЕ

В работах [16, 17] теоретически исследуются и в вычислительном эксперименте построены взвешенные линейные комбинации моделей, в которых веса подбираются на основе минимизации теоретической оценки дисперсии прогноза. В представленной работе сделан шаг вперёд и рассмотрен гессиан функции оценки дисперсии прогноза в стационарных точках при таком объединении. Показано, что минимум оценки дисперсии достигается в случае объединения моделей с дисперсиями, равными по порядку величины, и невысокими значениями корреляции прогнозов объединяемых моделей. При объединении моделей с высокой корреляцией прогнозов или с отличающимися на порядок дисперсиями минимум может не достигаться (может иметь место максимум или отсутствие экстремума). При этом, с практической точки зрения объединять модели, дающие сильно коррелирующие прогнозы, бессмысленно, а сильная отрицательная корреляция прогнозов различных моделей одного процесса встречается редко.

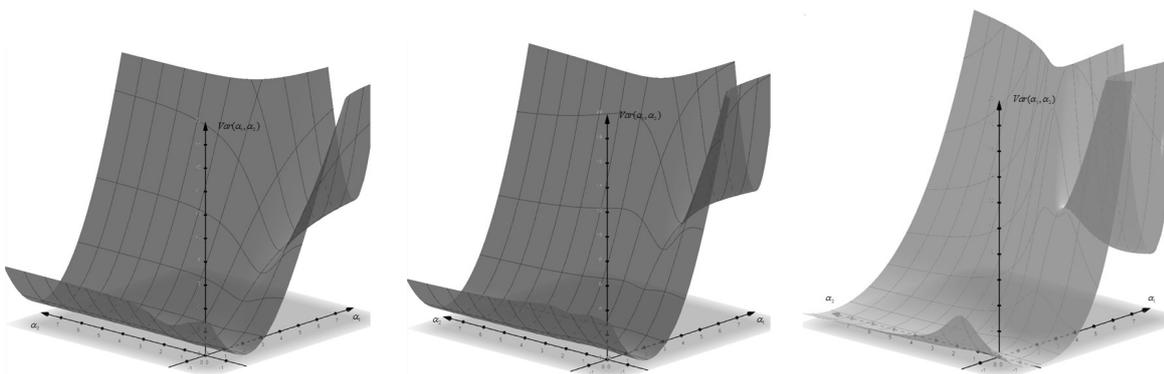
Рассмотрены функции дисперсии объединённого прогноза для некоторых моделей типа *ETS*. В случаях несезонной модели  $ETS(A, A, N)$  и сезонной модели  $ETS(A, A, A)$  удалось свести рассмотрению функции дисперсии к случаю более простой модели  $ETS(A, N, N)$ , которая подробно рассмотрена в работе. На практике случаи стационарных точек, в которых исследование гессиана не даёт утверждать наличие минимума, часто сводятся к локальным минимумам функции дисперсии прогноза. В ходе вычислительного эксперимента показано, что взвешенное объединение прогнозов привело к улучшению оценок  $RMSE$ ,  $MAE$ , характеризующих близость предсказываемых моделью значений к реальным данным по сравнению с исходными моделями *ETS*.

## ПРИЛОЖЕНИЕ

Рисунок 1 содержит графики функции дисперсии прогноза (5) на  $n = 2$  шага вперёд для моделей типа  $ETS(A, N, N)$  при одинаковых среднеквадратических отклонениях моделей  $se_1 = se_2$  (случаи корреляции  $\rho = 0.1, 0.5, 0.9$ ). Графики для тех же уровней корреляции прогноза, но при соотношении  $se_2/se_1 = 10$  представлены на рисунке 2.



**Рис. 1.** Графики функции дисперсии прогноза (5) на  $n = 2$  шага вперёд для моделей типа  $ETS(A, N, N)$  при одинаковых среднеквадратических отклонениях моделей  $se_1 = se_2$  (случаи корреляции прогнозов  $\rho = 0.1, 0.5, 0.9$  слева направо)



**Рис. 2.** Графики функции дисперсии прогноза (5) на  $n = 2$  шага вперёд для моделей типа  $ETS(A, N, N)$  при соотношении СКО моделей  $se_2/se_1 = 10$  (случаи корреляции прогнозов  $\rho = 0.1, 0.5, 0.9$  слева направо)

## СПИСОК ЛИТЕРАТУРЫ

1. Hyndman R.J., Athanasopoulos G. *Forecasting: principles and practice*. Melbourne: OTexts, 2021 (3rd edition)
2. Stock J.H., Watson M.W. *Introduction to Econometrics*. Pearson, 2018 (4th edition)
3. Аристов В.В., Строганов А.В., Ястребов А.Д. Моделирование пространственного распространения волн пандемии COVID-19 в России на основе кинетико-переносного описания. *Российский Технологический Журнал*, 2023, Том 11(4), с.59-71. doi: 10.32362/2500-316X-2023-11-4-59-71. <https://doi.org/10.32362/2500-316X-2023-11-4-59-71>
4. Abebe M., Noh Y., Kang Y.-J., Seo C., Kim D., Seo J. Ship trajectory planning for collision avoidance using hybrid ARIMA-LSTM models. *Ocean Engineering*, 2022, vol. 256, p. 111527. doi: 10.1016/j.oceaneng.2022.111527.
5. Wu Z., Yin H., He H., Li Y. Dynamic-LSTM hybrid models to improve seasonal drought predictions over China. *Journal of Hydrology*, 2022, vol. 615, p. 128706. doi: 10.1016/j.jhydrol.2022.128706.
6. Yan Y., Wang X., Ren F., Shao Z., Tian C. Wind speed prediction using a hybrid model of EEMD and LSTM considering seasonal features. *Energy Reports*, 2022, vol. 8, pp. 8965-8980. doi: 10.1016/j.egyр.2022.07.007.
7. Wang J., Wang P., Tian H., Tansey K., Liu J., Quan W. A deep learning framework combining CNN and GRU for improving wheat yield estimates using time series remotely sensed multi-variables. *Computers and Electronics in Agriculture*, 2023, vol. 206, p.107705. doi: 10.1016/j.compag.2023.107705.
8. Hua H., Liu M., Li Y., Deng S., Wang Q. An ensemble framework for short-term load forecasting based on parallel CNN and GRU with improved ResNet. *Electric Power Systems Research*, 2023, vol. 216, 109057. doi: 10.1016/j.epr.2022.109057.
9. Wang P., Zheng X., Ai G., Liu D., Zhu B. Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru and Iran. // *Chaos, Solitons and Fractals*, 2020, vol.140, p. 110214. doi: 10.1016/j.chaos.2020.110214
10. Cascone L., Sadiq S., Ullah S., Mirjalili S., Ur H., Siddiqui R., Umer M. Predicting Household Electric Power Consumption Using Multi-step Time Series with Convolutional LSTM. // *Big Data Research*, 2023, vol.31, p.100360. doi: 10.1016/j.bdr.2022.100360
11. Wang H., Zhang Y., Liang J., Liu L. DAFA-BiLSTM: Deep Autoregression Feature Augmented Bidirectional LSTM network for time series prediction. // *Neural Networks*, 2023, vol.157, pp.240-256. doi: 10.1016/j.neunet.2022.10.009.
12. Zhao L., Mo C., Ma J., Chen Z., Yao C. LSTM-MFCN: A time series classifier based on multi-scale spatial-temporal features. // *Computer Communications*, 2022, vol. 182, pp.52-59. doi: 10.1016/j.comcom.2021.10.036
13. Zhang D., Sun W., Dai Y., Liu K., Li W., Wang C. A hierarchical early kick detection method using a cascaded GRU network. // *Geoenergy Science and Engineering*, 2023, vol.222, p.211390 doi: 10.1016/j.geoen.2022.211390
14. Amalou I., Mouhni N., Abdali A. Multivariate time series prediction by RNN architectures for energy consumption forecasting. *Energy Reports*. 2022, vol. 8, pp.1084-1091. doi: 10.1016/j.egyр.2022.07.139.
15. Aseeri A. Effective RNN-Based forecasting methodology design for improving short-term power load forecasts: application to large-scale power-grid time series. *Journal of Computational Science*. 2023, vol.68, p.101984. doi: 10.1016/j.jocs.2023.101984.
16. Musatov D., Petrusевич D. Modeling of forecasts variance reduction at multiple time series prediction averaging with ARMA(1, q) functions. *Proc. of MIP Computing-V 2022: V International Scientific Workshop on Modeling, Information Processing and Computing*. Krasnoyarsk, Russia, 2022, vol. 3091, pp.60-65. doi: 10.47813/dnit-mip5/2022-3091-1-11.

17. Beletskaya N., Petrusевич D. Linear Combinations of time series models with minimal forecast variance. *Journal of Communications Technology and Electronics*, 2023, vol. 67, is. 1, pp.144-158. doi: 10.1134/S1064226922130022.
18. Единый архив экономических и социологических данных. Динамические ряды макроэкономической статистики РФ. Индекс потребительских цен. URL: <http://sophist.hse.ru/hse/nindex.shtml>, 01-02-2024
19. Грамович Я.В., Мусатов Д.Ю., Петрусевич Д.А. Применение беггинга в прогнозировании временных рядов. *Российский Технологический Журнал*. 2024, Том 12(1), с.101-110. doi: <https://doi.org/10.32362/2500-316X-2024-12-1-101-110>.
20. Petropoulos F., Hyndman R.J., Bergmeir C. Exploring the sources of uncertainty: Why does bagging for time series forecasting work? // *European Journal of Operational Research*. 2018, Vol.268, is.2, pp.545-554. doi: <https://doi.org/10.1016/j.ejor.2018.01.045>.
21. Bergmeir C., Hyndman R.J., Benitez J.M. Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation // *International Journal of Forecasting*. 2016, Vol.32, pp.303-312. doi: <https://doi.org/10.1016/j.ijforecast.2015.07.002>.
22. Shafik N., Tutz G. Boosting nonlinear additive autoregressive time series // *Statistics and Data Analysis*, 2009, vol.53, is.7, pp.2453-2464. doi: 10.1016/j.csda.2008.12.006.
23. Liu X., Liu A., Chen J.L., Li G. Impact of decomposition on time series bagging forecasting performance // *Tourism Management*, 2023, vol.97, p.104725. doi: 10.1016/j.tourman.2023.104725.
24. Matias J.M.,Febrero-Bande M., Gonzalez-Manteiga W., Reboredo J.C. Boosting GARCH and neural networks for the prediction of heteroskedastic time series. // *Mathematical and Computer Modelling*, 2010, Vol.51., is.3-4 pp.256-271. doi: 10.1016/j.mcm.2009.08.013

## Forecast variance minimization on the example of ETS models

N.V. Beletskaya, D.A. Petrusевич

In the research combined model over a few time series models is constructed (if models are additive, for example: ARIMA) or predictions of a few models are combined (in case if models aren't additive, for example: ETS). Evaluation of forecast variance is minimized in this construction. For majority of models forecast variance evaluation is based on Student's criterion. But in the special cases of ETS and ARIMA models (ETS models can be viewed as a special case of ARIMA models) forecast variance evaluation depends on model parameters. Thus, confidence interval of prediction can be expressed via model parameters and one can construct combinations of models with minimal evaluation of forecast variance. In the theoretical part of this research linear combination of two models with minimal forecast variance evaluation is constructed. Also, hessian of forecast variance evaluation is constructed and investigated. This function is also under investigation for some groups of ETS models: conditions under which extremal points are variance minima are described in terms of model parameters.

**KEYWORDS:** ETS, optimal combination, minimization of variance forecast evaluation.