

Построение линейных комбинаций моделей временных рядов, оптимальных по оценке дисперсии прогноза

Н.В. Белецкая^{*,**}, Д.А. Петрусевич^{*}

^{*}Российский Технологический Университет (МИРЭА), Москва, Россия

^{**}Институт проблем передачи информации им. А.А. Харкевича, Москва, Россия

Поступила в редколлегию 07.09.2022

Аннотация—В рамках представленной работы рассматриваются вопросы, связанные с объединением или усреднением прогнозов моделей временных рядов. Объединение прогнозов нескольких моделей или объединение самих моделей (в беггинге моделей временных рядов) активно используется и исследуется при моделировании и прогнозировании временных процессов. В работе построены и анализируются функции оценки дисперсии прогноза модели $ARIMA(p, d, q)$ на 1, 2, 3 шага вперёд на основе представления в виде ψ -весов. При объединении или усреднении прогнозов нескольких моделей $ARIMA(p, d, q)$, $p < 4$, исследуется свойство выпуклости вниз функции дисперсии прогноза. Кроме этого, рассматривается объединение прогнозов двух моделей произвольной природы. Предполагается, что прогнозы участвуют во взвешенной линейной комбинации, причём, веса подбираются так, чтобы минимизировать оценку дисперсии прогноза. В рамках исследования показано, что при оптимальном выборе весов в комбинации двух моделей дисперсия прогноза не превышает дисперсии прогноза объединяемых моделей. В рамках практического эксперимента рассматривается взвешенная комбинация (с весами, оптимальными для минимизации оценки дисперсии прогноза) моделей $ARIMA(p, d, q)$ и моделей $ADL(p, q)$. Качество прогноза объединённых моделей не уступает качеству прогноза лучших моделей, участвующих в объединении. При исследовании объединения прогнозов трёх моделей дисперсия прогноза может как снижаться при объединении, так и превышать дисперсии прогноза отдельных моделей, поэтому пока сделать общие выводы при объединении больше, чем пары, моделей не представляется возможным.

КЛЮЧЕВЫЕ СЛОВА: $ARIMA(p, d, q)$, $ADL(p, q)$, информационный критерий Акаике, информационный критерий Байеса, оптимальное объединение, минимизация дисперсии прогноза, ψ -веса

DOI: 10.53921/18195822_2022_22_3_187

1. ВВЕДЕНИЕ

Временные ряды часто применяются для моделирования различных процессов. Они используются для прогнозирования поведения курсов валют, статистических экономических индексов, для прогнозирования уровня воды в реках и озерах, потока пациентов в больницах и других задачах. Чаще всего, у исследователя в наличии несколько моделей, применяемых для прогнозирования [1, 2]. В представленной работе исследуются способы объединения моделей и их прогнозов для повышения качества прогнозирования.

В задаче классификации часто применяются стратегии беггинга и бустинга. Они показывают прекрасные результаты, в подавляющем большинстве задач существенно превосходя обычные модели классификации по качеству [3–9]. Основная идея состоит в объединении слабых классификаторов (их точность не высока, но, по крайней мере, требуется, чтобы вероятность

правильного ответа классификатора превышала $1/2$). В некотором смысле случай временных рядов похож на задачу классификации: присутствует множество моделей, результаты которых можно объединить. Но прямой аналогии нет. Ситуация осложняется тем, что для обучения данные необходимо разделить на обучающую и тестовую выборку для того, чтобы оценить качество моделей, участвующих в объединении. По времени тестовая выборка всегда находится в конце исследуемого промежутка, поэтому математические модели, которые должны сделать прогноз на будущее, фактически обучаются по временному периоду, который заканчивается существенно раньше прогнозируемого периода. Объединение моделей на основе беггинга может потребовать ещё большего тестового периода. Это снижает качество моделей. Таким образом, модель, обученная на всем объёме данных, зачастую даёт лучший прогноз, чем объединяемые при усреднении или беггинге математические модели. Идея совмещения нескольких временных рядов, генерируемых при беггинге, близка к синтезу изображения на основе нескольких образов, но информация представлена в другом виде [10].

В представленной работе анализируются функции оценки дисперсии прогноза моделей ARIMA (p, d, q) : анализируются условия, при которых они являются выпуклыми вниз. Далее рассматривается оптимальное (с точки зрения оценки дисперсии прогноза) объединение двух моделей произвольной природы: дисперсия прогноза оптимального объединения моделей не превосходит дисперсии прогнозов объединяемых моделей. В серии практических экспериментов рассмотрено объединение прогнозов моделей ARIMA (p, d, q) и моделей ADL (p, q) . Совмещение прогнозов моделей различных типов является темой дальнейших исследований.

При оценке моделей ARIMA (p, d, q) используются значения информационных критериев Акаике (AIC) и Байеса / Шварца (BIC) [11, 12]. Следует подчеркнуть, что прямо в структуру критериев встроено требование низких порядков моделей (модели штрафуются за величину порядков). Это объясняется тем, что зачастую в моделях старших порядков ($p > 5$ или $q > 5$) появляются проблемы, связанные с линейной зависимостью между различными слагаемыми в модели. Вместе с тем, из-за того, что в минимизируемое значение критерия вносится дополнительное слагаемое, связанное с порядками модели, в рамках вычислительных экспериментов рассматривается несколько моделей, которые могут не обладать минимальным значением информационного критерия. В ходе экспериментов оценивается качество прогноза, которое модель делает на тестовый период времени.

2. ИССЛЕДУЕМЫЕ МОДЕЛИ ВРЕМЕННЫХ РЯДОВ

Следует сразу оговориться, что построение моделей ARIMA (p, d, q) [11], которые бы полностью описывали динамику рассматриваемого процесса, или построение моделей ADL (p, q) [12], обладающими коэффициентом детерминации $R^2 > 0.8$, встречается очень редко в связи с несколькими причинами. Во-первых, модели ARIMA предполагают зависимость значений временного ряда от значений того же ряда в прошлом, но зависимость от других параметров в такую систему не вводится, поэтому для того, чтобы смоделировать влияние других параметров на поведение ряда, приходится использовать другие модели (например, для оценки влияния макроэкономических индексов на поведение экономических временных рядов, в основном, используются модели, основанные на линейной регрессии). Модели ADL предполагают построение зависимости между значениями двух временных рядов, но ввод других параметров в модель проблематичен. Во-вторых, сами данные, по которым строятся временные ряды, могут включать ошибочные измерения (или данные, в которых пропущенные значения заменены специальным методом, но не совпадают с реальностью), измерения могут быть проведены с существенной дисперсией. Таким образом, данные включают в себя некоторый элемент неопределенности. Более подробно источники неопределенности моделей временных рядов обсуждаются в [13].

Вместе с тем, вопрос повышения качества прогнозирования стоит во многих сферах математического, в том числе в исследовании временных рядов. И объединение нескольких моделей временных рядов на основе некоторого принципа, например, на основе минимизации оценки дисперсии прогноза, может привести к новым результатам в моделировании временных рядов.

Основные вопросы, которые возникают при беггинге или усреднении прогнозов временных рядов, включают оценку точности прогноза (и её изменение при применении усреднения или беггинга) и основные преобразования, требуемые для проведения самого беггинга на временных рядах. Исследованию второго вопроса посвящены работы [13, 14]. Сейчас во многих работах беггинг применяется к временным рядам, отражающим процессы совершенно различной природы [1, 15, 16].

Для того, чтобы произвести оценку дисперсии прогноза модели временного ряда ARIMA (p, d, q) , необходимо произвести преобразование ряда к виду $MA(\infty)$ [17]. Модель скользящего среднего $MA(q)$ имеет вид:

$$\hat{X}_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (1)$$

где современное значение временного ряда X_t выражается с помощью суммы компонент белого шума ε_{t-q} [11, 12].

Вторая, не менее важная составляющая часть модели ARIMA (p, d, q) , – процесс авторегрессии $AR(p)$, который предполагает наличие взаимосвязи между современным значением ряда X_t и его предыдущими значениями X_{t-1}, \dots, X_{t-p} [11, 12]:

$$\hat{X}_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p}. \quad (2)$$

На практике пользуются моделями не выше 3-го или 5-го порядка: $p, q \leq 5$. Обычно моделей с таким количеством слагаемых достаточно для описания сложной динамики. Кроме этого, при повышении порядков появляются проблемы, связанные с мультиколлинеарностью (линейной зависимостью между слагаемыми), поэтому чаще всего исследователи ограничиваются моделями не выше 3-го порядка $p \leq 3$. Согласно теореме Волда [17], стационарный процесс может быть представлен в виде суммы процессов скользящего среднего (в общем случае, с бесконечным числом компонент). Множители перед элементами в такой сумме обычно называются ψ -весами. Используя такое представление, можно выразить дисперсию прогноза в виде суммы ψ -весов:

$$Var(\hat{x}_n - x_n) = \sigma^2 \sum_{j=0}^{n-1} \psi_j^2, \quad (3)$$

где \hat{x}_n – прогнозируемое значение ряда, x_n – реальное значение, σ – значение стандартной ошибки (получается в процессе обучения модели временного ряда). Для простоты предполагаем, что сезонность рядов совпадает (либо она отсутствует у всех исследуемых рядов, либо её структура одинакова у всех комбинируемых моделей). В целом, порядок d временной разности (для перехода к стационарному виду) также не влияет на представленные здесь рассуждения (подробнее эти аспекты моделирования временных рядов освещены в [11, 12]), поэтому далее опустим параметр d и будем говорить о стационарном процессе ARMA(p, q).

В рамках представленной работы также исследуются модели временных рядов ADL (p, q) (в англоязычной литературе чаще ArDL, модель авторегрессии и распределённого лага),

описывающая взаимосвязь двух рядов. Порядок авторегрессии p задаёт порядок зависимости исследуемого ряда y_t от своих предыдущих значений, а порядок распределённого лага q – зависимости от поведения второго ряда x_t :

$$\hat{X}_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + b_1 Y_{t-1} + \dots + b_q Y_{t-q}. \quad (4)$$

Таким образом, в отличие от моделей $AR(p)$ и $MA(q)$, рассмотренных выше, здесь присутствует зависимость не только от своих значений в предыдущие моменты времени, но и от поведения другого временного ряда. При этом, если порядок $q = 0$, то получим в качестве частного случая ряд $AR(p)$ (2).

3. ПРЕДСТАВЛЕНИЕ ДИСПЕРСИИ ПРОГНОЗА РЯДОВ $ARMA(p, q)$

Исследование рядов вида $AR(1)$ и $ARMA(1, q)$ дано согласно работе [18]. В общем случае модель $AR(p)$ представляется в виде ψ -весов следующим образом:

$$\begin{aligned} \psi_0^{AR} &= 1, \\ \psi_1^{AR} &= \varphi_1, \\ \psi_2^{AR} &= \varphi_1 \psi_1 + \varphi_2, \\ \psi_3^{AR} &= \varphi_1 \psi_2 + \varphi_2 \psi_1 + \varphi_3, \\ \psi_k^{AR} &= \sum_{i=1}^{k-1} \varphi_i \psi_{k-i}^{AR}. \end{aligned} \quad (5)$$

ψ -веса для модели $ARMA(p, q)$ отличаются наличием слагаемого θ_i :

$$\psi_k^{ARMA} = \sum_{i=1}^{k-1} \varphi_i \psi_{k-i}^{AR} + \theta_i. \quad (6)$$

Представление процесса $AR(1)$ в виде ψ -весов имеет вид геометрической прогрессии:

$$\begin{aligned} \psi_0^{AR(1)} &= 1, \\ \psi_1^{AR(1)} &= \varphi_1, \\ \psi_2^{AR(1)} &= \varphi_1 \psi_1 = \varphi_1^2, \\ \psi_k^{AR(1)} &= \varphi_1 \psi_{k-1}^{AR(1)} = \varphi_1^k. \end{aligned} \quad (7)$$

Подставляя ψ -веса для модели $AR(1)$ (7) в выражение для дисперсии прогноза (3) на n шагов, получим:

$$Var_{AR(1)}(\hat{x}_n - x_n) = \sum_{j=0}^{n-1} \psi_j^2 = \sigma^2 \left(1 + \sum_{j=1}^{n-1} \varphi_1^{2j} \right) = \sigma^2 \frac{1 - \varphi_1^{2n}}{1 - \varphi_1^2}. \quad (8)$$

Дисперсия прогноза (8) для модели $AR(1)$ остаётся конечной при прогнозе на любой период в будущем $n \rightarrow \infty$ при $|\varphi_1| < 1$:

$$\lim_{n \rightarrow \infty} Var_{AR(1)}(\hat{x}_n - x_n) = \sigma^2 \frac{1}{1 - \varphi_1^2}. \quad (9)$$

Представление процесса $ARMA(1, q)$ в виде ψ -весов отличается тем, что геометрическая прогрессия со знаменателем φ_1 начинается с ψ -веса с индексом $q + 1$:

$$\begin{aligned} \psi_0^{ARMA(1,q)} &= 1, \\ \psi_1^{ARMA(1,q)} &= \varphi_1 + \theta_1, \\ \psi_2^{ARMA(1,q)} &= \varphi_1 \psi_1 + \theta_2 = \varphi_1(\varphi_1 + \theta_1) + \theta_2, \\ \psi_{q+1}^{ARMA(1,q)} &= \varphi_1 \psi_q. \end{aligned} \quad (10)$$

В случае, если порядок скользящего среднего равен или ниже, чем период, на который делается прогноз ($q \leq n$), геометрической прогрессии вообще не наблюдается. Например, дисперсия прогноза модели $ARMA(1, 2)$ на 3 шага вперёд:

$$Var_{ARMA(1,2)}(\hat{x}_3 - x_3) = \sum_{j=0}^2 \psi_j^2 = \sigma^2 \left[1 + (\varphi_1 + \theta_1)^2 + (\varphi_1(\varphi_1 + \theta_1) + \theta_2)^2 \right]. \quad (11)$$

В случае, если порядок части скользящего среднего $MA(q)$ выше, чем период, на который делается прогноз ($q > n$), геометрическая прогрессия начинается с веса ψ_q . В выражении (12) последнее слагаемое как раз представляет её сумму:

$$\begin{aligned} Var_{ARMA(1,q)}(\hat{x}_n - x_n) &= \sum_{j=0}^2 \psi_j^2 = \\ &= \sigma^2 \left(1 + (\varphi_1 + \theta_1)^2 + (\varphi_1(\varphi_1 + \theta_1) + \theta_2)^2 + \dots + \psi_{q-1}^2 + \psi_q^2 \frac{1 - \varphi_1^{2(n-q)}}{1 - \varphi_1^2} \right). \end{aligned} \quad (12)$$

Продолжим выражение функции дисперсии прогноза для моделей $ARMA(2, q)$. Рассмотрим процесс $AR(2)$ (в сумме (2) присутствует два слагаемых X_{t-1} и X_{t-2}). Отличия в значениях ψ -весов модели $AR(1)$, представленных в (7), начинаются со второго веса:

$$\psi_2^{AR(2)} = \varphi_1 \psi_1 + \varphi_2 = \varphi_1^2 + \varphi_2. \quad (13)$$

Выражение для дисперсии уже не обладает такими замечательными свойствами, как в случае модели $AR(1)$. Приведём дисперсию прогноза на $k = 1, 2, 3$ шага вперёд (дисперсия прогноза на 1 шаг вперёд для всех моделей $ARMA(p, q)$ совпадает с σ^2):

$$\begin{aligned} Var_{AR(2)}(\hat{x}_1 - x_1) &= \sigma^2, \\ Var_{AR(2)}(\hat{x}_2 - x_2) &= \sigma^2 (1 + \varphi_1^2), \\ Var_{AR(2)}(\hat{x}_3 - x_3) &= \sigma^2 \left[1 + \varphi_1^2 + (\varphi_1^2 + \varphi_2)^2 \right]. \end{aligned} \quad (14)$$

Рассматривая ψ -веса для функций ARMA(2, q), нетрудно заметить, что отличия от выражений (13) и (14) появляются в связи с наличием параметров θ_i , характеризующих поведение части скользящего среднего MA(q). В частности, для модели ARMA(2, 1) ψ -веса имеют вид:

$$\psi_0^{ARMA(2,1)} = 1, \psi_1^{ARMA(2,1)} = \varphi_1 + \theta_1, \psi_2^{ARMA(2,1)} = \varphi_1(\varphi_1 + \theta_1) + \varphi_2. \quad (15)$$

Значение дисперсии прогноза на 2 и 3 шага вперёд меняется из-за наличия параметров θ_i :

$$\begin{aligned} Var_{ARMA(2,1)}(\hat{x}_2 - x_2) &= \sigma^2 \left[1 + (\varphi_1 + \theta_1)^2 \right], \\ Var_{ARMA(2,1)}(\hat{x}_3 - x_3) &= \sigma^2 \left[1 + (\varphi_1 + \theta_1)^2 + (\varphi_1(\varphi_1 + \theta_1) + \varphi_2)^2 \right]. \end{aligned} \quad (16)$$

Для модели ARMA(2, 2) ψ -веса имеют аналогичный вид. Отличие от весов для модели ARMA(2, 1), представленных в (15), заключается только в наличии зависимости ψ_2 от θ_2 :

$$\psi_2^{ARMA(2,2)} = \varphi_1(\varphi_1 + \theta_1) + \varphi_2 + \theta_2 = \psi_2^{ARMA(2,1)} + \theta_2. \quad (17)$$

Здесь с помощью верхнего индекса обозначена принадлежность ψ -веса к определённой модели. По сравнению с дисперсией прогноза модели ARMA(2,1) на 1-3 шага вперёд (16) меняется только прогноз на 3 шага вперёд:

$$Var_{ARMA(2,2)}(\hat{x}_3 - x_3) = \sigma^2 \left[1 + (\varphi_1 + \theta_1)^2 + (\varphi_1(\varphi_1 + \theta_1) + \varphi_2 + \theta_2)^2 \right]. \quad (18)$$

Обратим внимание на то, что, если ограничить дальность прогноза на будущее 3 временными промежутками, дисперсии моделей с большими значениями порядков скользящего среднего ARMA(2, q), $q > 2$, не отличаются от выражений, полученных для ARMA(2,1) и ARMA(2,2). Более того, при том же ограничении прогноза на 3 временных промежутка вперёд дисперсии прогнозов моделей ARMA(p , q), $p > 2$, будут иметь тот же вид, что и дисперсии прогнозов моделей ARMA(p , q), $p \leq 2$, с соответствующими значениями порядка части скользящего среднего q . Так происходит потому, что различия в ψ -весах наблюдаются только для весов ψ_i с индексами $i \geq 3$. Вместе с ними различия и в выражениях для дисперсии прогноза у моделей ARMA(p , q), $p > 2$, по сравнению с ARMA(p , q), $p \leq 2$, наблюдаются только в прогнозах на период в будущем дальше, чем на 3 шага вперёд. Например, дисперсии прогноза для модели AR(3) на 3 шага вперёд совпадают с дисперсиями прогноза модели AR(2), различия появляются в дисперсии прогноза на 4 шага вперёд.

4. ИССЛЕДОВАНИЕ ФУНКЦИИ ДИСПЕРСИИ ПРОГНОЗА РЯДОВ ARMA(p , q)

Теорема.

Теорема. *Функции дисперсий прогнозов для моделей ARMA(1, q) на 1, 2, 3 шага вперёд являются выпуклыми вниз. При условиях на параметры моделей ϕ_i , θ_i , указанных ниже, дисперсии прогнозов для моделей ARMA(p , q), $p = 2, 3$ тоже выпуклы вниз. Это утверждение важно для анализа взвешенной комбинации или усреднения прогнозов моделей.*

Доказательство.

Доказательство. Для моделей AR(1) функции дисперсий прогнозов на 1, 2, 3 шага вперёд (8) представляют собой комбинации 0-й, 2-й и 4-й степени от единственного параметра модели – φ_1 (графики функций дисперсии прогноза на 2 и 3 шага вперёд представлены на рисунке 1 в приложении. Для прогноза на 2 периода вперёд речь идёт о параболе с вершиной в $\varphi_1 = 0$ с ветвями, направленными вверх. Для прогноза на 3 шага вперёд функция дисперсии прогноза, её первая и вторая производная выписаны в (19). Очевидно, присутствует единственный минимум при $\varphi_1 = 0$:

$$\begin{aligned} \text{Var}_{AR(1)}(\hat{x}_3 - x_3) &= \sigma^2 (1 + \varphi_1^2 + \varphi_1^4), \\ (\text{Var}_{AR(1)}(\hat{x}_3 - x_3))' &= 2\sigma^2 \varphi_1 (1 + 2\varphi_1^2), \\ (\text{Var}_{AR(1)}(\hat{x}_3 - x_3))'' &= 2\sigma^2 (1 + 6\varphi_1^2). \end{aligned} \quad (19)$$

Для моделей AR(2) функции дисперсий прогнозов на 1, 2 шага вперёд совпадают с функциями для моделей AR(1). Гессиан для функции дисперсии прогноза модели AR(2) на 3 шага вперёд (14) имеет вид:

$$H(\text{Var}_{AR(2)}(\hat{x}_3 - x_3) / \sigma^2) = \begin{vmatrix} 2(6\varphi_1^2 + 2\varphi_2 + 1) & 4\varphi_1 \\ 4\varphi_1 & 2 \end{vmatrix} = 4(2\varphi_1^2 + 2\varphi_2 + 1). \quad (20)$$

Функция (20) положительна и положителен минор $\Delta_1 = 2(6\varphi_1^2 + 2\varphi_2 + 1)$ при условии $2\varphi_1^2 + 1 > -2\varphi_2$. График функции дисперсии прогноза модели AR(2) на 3 шага вперёд 2 представлен в приложении.

Для модели AR(3) функции дисперсии прогноза на 1, 2, 3 шага вперёд совпадают с функциями (14) для модели AR(2). То же касается и моделей больших порядков AR(p), $p > 3$. Различия в функциях дисперсии прогноза наблюдаются при прогнозировании на число шагов, большее 3.

Рассмотрим модели ARMA(1, q). Дисперсия прогнозов (11), (12) для модели ARMA(1, 1) выглядит следующим образом:

$$\begin{aligned} \text{Var}_{ARMA(1,1)}(\hat{x}_1 - x_1) &= \sigma^2, \\ \text{Var}_{ARMA(1,1)}(\hat{x}_2 - x_2) &= \sigma^2 [1 + (\varphi_1 + \theta_1)^2], \\ \text{Var}_{ARMA(1,1)}(\hat{x}_3 - x_3) &= \sigma^2 [1 + (\varphi_1 + \theta_1)^2 + \varphi_1^2(\varphi_1 + \theta_1)^2]. \end{aligned} \quad (21)$$

Графики функций дисперсии прогноза на 2 и 3 шага вперёд для модели ARMA(1, 1) представлены на рисунке 3 в приложении. Рассмотрим прогноз на 2 шага вперёд. Как видим, в сумме (21) геометрическая прогрессия начинается со второго слагаемого и имеет знаменатель φ_1^2 . Для определения положения экстремума рассмотрим частные производные первого порядка, они совпадают:

$$\left(\frac{\text{Var}_{ARMA(1,1)}(\hat{x}_2 - x_2)}{\sigma^2} \right)' \Big|_{\varphi_1} = \left(\frac{\text{Var}_{ARMA(1,1)}(\hat{x}_2 - x_2)}{\sigma^2} \right)' \Big|_{\theta_1} = 2(\varphi_1 + \theta_1) = 0. \quad (22)$$

Для модели, соответствующей экстремуму, $\theta_1 = -\varphi_1$, построим гессиан. Частные производные второго порядка имеют вид:

$$\begin{aligned} \left(\frac{Var_{ARMA(1,1)}(\hat{x}_2 - x_2)}{\sigma^2} \right)'' \Big|_{\varphi_1^2} &= 2, \\ \left(\frac{Var_{ARMA(1,1)}(\hat{x}_2 - x_2)}{\sigma^2} \right)'' \Big|_{\theta_1^2} &= 2, \\ \left(\frac{Var_{ARMA(1,1)}(\hat{x}_2 - x_2)}{\sigma^2} \right)'' \Big|_{\varphi_1\theta_1} &= 2. \end{aligned} \quad (23)$$

Гессиан функции дисперсии прогноза на 2 шага вперёд представлен в (24):

$$H \left(\frac{Var_{ARMA(1,1)}(\hat{x}_2 - x_2)}{\sigma^2} \right) = \begin{vmatrix} 2 & 2 \\ 2 & 2 \end{vmatrix} = 0. \quad (24)$$

Определитель (24) равен нулю, минор $\Delta_1 = 2 > 0$, поэтому для исследования функции построим её график. Он представлен в приложении 3. Функция достигает минимальных значений на прямой: $\theta_1 = -\varphi_1$. При этом, функция дисперсии прогноза обладает свойствами выпуклой вниз.

Рассмотрим дисперсию прогноза на 3 шага вперёд (21). Экстремум дисперсии достигается при условии равенства нулю первых частных производных:

$$\begin{aligned} \left(\frac{Var_{ARMA(1,1)}(\hat{x}_3 - x_3)}{\sigma^2} \right)' \Big|_{\varphi_1} &= 2(\varphi_1 + \theta_1)(2\varphi_1^2 + \varphi_1\theta_1 + 1) = 0, \\ \left(\frac{Var_{ARMA(1,1)}(\hat{x}_3 - x_3)}{\sigma^2} \right)' \Big|_{\theta_1} &= 2(\varphi_1 + \theta_1)(\varphi_1^2 + 1) = 0. \end{aligned} \quad (25)$$

Так же, как и в случае модели ARMA(1, 1), экстремум достигается при условии $\theta_1 = -\varphi_1$. Первое уравнение в системе (25) решается и при других условиях, но решение второго уравнения одновременно с первым возможно только при выполнении условия $\theta_1 = -\varphi_1$. Вычислим частные производные второго порядка для проверки критерия Сильвестра:

$$\begin{aligned} \left(\frac{Var_{ARMA(1,1)}(\hat{x}_3 - x_3)}{\sigma^2} \right)'' \Big|_{\varphi_1^2} &= 12\varphi_1(\varphi_1 + \theta_1) + 2(\theta_1^2 + 1), \\ \left(\frac{Var_{ARMA(1,1)}(\hat{x}_3 - x_3)}{\sigma^2} \right)'' \Big|_{\theta_1^2} &= 2(\varphi_1^2 + 1), \\ \left(\frac{Var_{ARMA(1,1)}(\hat{x}_3 - x_3)}{\sigma^2} \right)'' \Big|_{\varphi_1\theta_1} &= 4\varphi_1(\varphi_1 + \theta_1)(\varphi_1^2 + 1). \end{aligned} \quad (26)$$

При условии $\theta_1 = -\varphi_1$ частные производные второго порядка совпадают:

$$\left(Var_{ARMA(1,1)}(\hat{x}_3 - x_3) / \sigma^2 \right)'' \Big|_{\varphi_1\theta_1} = \left(Var_{ARMA(1,1)}(\hat{x}_3 - x_3) / \sigma^2 \right)'' \Big|_{\theta_1^2},$$

поэтому в гессиане присутствуют совпадающие строки и столбцы. При этом, миноры первого порядка $\left(Var_{ARMA(1,1)}(\hat{x}_3 - x_3) / \sigma^2 \right)'' \Big|_{\theta_1^2}$ или $\left(Var_{ARMA(1,1)}(\hat{x}_3 - x_3) / \sigma^2 \right)'' \Big|_{\varphi_1^2}$ положительны. Т.к. на основе критерия Сильвестра нельзя сделать выводов о характере функции, построим её график 3 и убедимся в том, что функция обладает свойствами выпуклой вниз и на прямой $\theta_1 = -\varphi_1$ достигается минимум.

При рассмотрении модели ARMA(1, 2) дисперсии прогнозов на 1, 2 шага вперёд остаются теми же, что и для модели ARMA(1, 1), поэтому выводы, сделанные выше, остаются в силе и для этих моделей. Рассмотрим прогноз на 3 шага вперёд. Отличие от выражения (21) заключается в наличии слагаемого θ_2 :

$$\text{Var}_{ARMA(1,2)}(\hat{x}_3 - x_3) = \sigma^2 \left[1 + (\varphi_1 + \theta_1)^2 + (\varphi_1(\varphi_1 + \theta_1) + \theta_2)^2 \right]. \quad (27)$$

Рассмотрим положение экстремума, приравняв все частные производные первого порядка нулю:

$$\begin{aligned} \left(\frac{\text{Var}_{ARMA(1,2)}(\hat{x}_3 - x_3)}{\sigma^2} \right)' \Big|_{\varphi_1} &= 2(\varphi_1 + \theta_1) + 2(2\varphi_1 + \theta_1)(\varphi_1(\varphi_1 + \theta_1) + \theta_2) = 0, \\ \left(\frac{\text{Var}_{ARMA(1,2)}(\hat{x}_3 - x_3)}{\sigma^2} \right)' \Big|_{\theta_1} &= 2(\varphi_1 + \theta_1) + 2\varphi_1(\varphi_1(\varphi_1 + \theta_1) + \theta_2) = 0, \\ \left(\frac{\text{Var}_{ARMA(1,2)}(\hat{x}_3 - x_3)}{\sigma^2} \right)' \Big|_{\theta_2} &= 2(\varphi_1(\varphi_1 + \theta_1) + \theta_2) = 0. \end{aligned} \quad (28)$$

Подставляя третье уравнение в системе (28) во второе, получаем:

$$\begin{aligned} \varphi_1 + \theta_1 &= 0, \\ \varphi_1(\varphi_1 + \theta_1) + \theta_2 &= 0. \end{aligned} \quad (29)$$

Или, ещё проще: $\varphi_1 + \theta_1 = 0$ и $\theta_2 = 0$. Для упрощения выкладок построим гессиан уже с учётом этих условий:

$$H \left(\frac{\text{Var}_{ARMA(1,2)}(\hat{x}_3 - x_3)}{\sigma^2} \right) = \begin{vmatrix} 2 + 2\varphi_1^2 & 2 + 2\varphi_1^2 & 2\varphi_1^2 \\ 2 + 2\varphi_1^2 & 2 + 2\varphi_1^2 & 2\varphi_1 \\ 2\varphi_1 & 2\varphi_1 & 2 \end{vmatrix}. \quad (30)$$

Первый и второй столбец определителя совпадают, поэтому он равен 0. Миноры имеют знаки: $\Delta_1 > 0$, $\Delta_2 = 0$, поэтому характер экстремума определяем по графикам сечений функции, которые по своей структуре повторяют то, что получено для модели ARMA(1, 1).

Рассмотрим модель ARMA(2, 1). Отличие в ψ -весах от случая модели ARMA(1, 2) заключается в замене $\theta_2 \Rightarrow \varphi_2$, поэтому с точностью до неё рассуждения, приведённые выше, и выкладки (27)-(30) сохраняются.

Рассмотрим модель ARMA(2, 2). Отличие в ψ -весах от случая модели ARMA(1, 2) заключается в появлении параметра φ_2 во втором весе ψ_2 (17). Параметры φ_2 и θ_2 равноправны, поэтому все результаты, полученные для модели ARMA(2, 1), сохраняют свою силу.

Следствие.

Следствие. При комбинировании прогнозов нескольких моделей дисперсия прогноза усреднённой модели окажется меньшей, чем у комбинируемых моделей, если в них присутствует равное количество моделей, предсказывающих противоположную динамику описываемого процесса (равное число моделей, у которых знаки φ_i и θ_i противоположны). В общем случае, дисперсия прогноза усреднённой модели не больше, чем у модели с наибольшей дисперсией прогноза и не меньше, чем у модели с наименьшей дисперсией прогноза, входящей в комбинацию. Это утверждение основано на том, что при усреднении происходит движение вниз по склону функции, выпуклой вниз. С геометрической точки зрения, дисперсия прогноза усреднённой модели будет наименьшей, если комбинируемым моделям соответствует равное количество точек на противоположных склонах.

5. МИНИМИЗАЦИЯ ДИСПЕРСИИ ПРОГНОЗА ПРИ ОБЪЕДИНЕНИИ МОДЕЛЕЙ

Функция объединения нескольких прогнозов выглядит следующим образом:

$$F = w_1 f_1 + w_2 f_2 + \dots + w_n f_n = \sum_{i=1}^n w_i f_i, \sum_{i=1}^n w_i = 1. \quad (31)$$

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. Основная идея концепции беггинга состоит в том, что на обучающей выборке удаётся оценить, насколько часто каждая из используемых моделей даёт правильный прогноз или ошибается. В соответствии с этим меняются веса w_i : вес модели, которая делает больше ошибок, снижается, а той модели, которая чаще даёт верный прогноз, – повышается. Если модели имеют один тип и обладают свойством линейности, то, в целом, можно сказать, что они объединяются в новую модель того же типа и с её помощью делается прогноз. Но по тому же принципу могут объединяться прогнозы моделей, не обладающих одинаковой природой. В рамках исследования нас интересует такое объединение прогнозов двух и большего количества моделей, при котором оценка дисперсии прогноза минимизируется. При объединении прогнозов двух моделей объединённый прогноз имеет вид (32):

$$F = w f_1 + (1 - w) f_2. \quad (32)$$

Т.к. веса всего два и между ними есть взаимосвязь (31), произведена замена: $w = w_1, w_2 = 1 - w_1$. Дисперсия прогноза, объединяющего предсказания двух моделей (32):

$$\text{Var}(\sigma_1, \sigma_2, \rho, w) = (w\sigma_1)^2 + ((1 - w)\sigma_2)^2 + 2w(1 - w)\rho\sigma_1\sigma_2. \quad (33)$$

В этой комбинации σ_i – стандартные ошибки объединяемых моделей, ρ – корреляция прогнозов моделей на тестовый период. В первую очередь нас интересует минимизация дисперсии по распределению весов.

Теорема.

Теорема. При оптимальном выборе веса w (минимизирующего величину дисперсии объединённого прогноза (33)) оценка дисперсии объединённого прогноза двух моделей не превосходит оценки дисперсии объединяемых моделей.

Доказательство.

Доказательство. При рассмотрении второй производной дисперсии (33) по весу w получим:

$$(\text{Var}(\sigma_1, \sigma_2, \rho, w))'' \Big|_w = 2(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2). \quad (34)$$

В связи с тем, что $|\rho| \leq 1$, значение функции (34) колеблется между положительными величинами, соответствующими точкам $\rho = -1, \rho = 1$ (более подробно в этом можно убедиться, взяв производную выражения (34) по ρ и проверив, что она отрицательная, следовательно, функция (34) монотонно убывает по ρ), можем произвести оценку:

$$(\sigma_1 - \sigma_2)^2 \leq (Var(\sigma_1, \sigma_2, \rho, w))'' \Big|_w \leq (\sigma_1 + \sigma_2)^2. \quad (35)$$

Вторая производная является неотрицательной величиной, поэтому функция дисперсии (33) выпукла вниз за исключением положений экстремумов. При этом, при дифференцировании функции объединённого прогноза получим единственный экстремум, который является минимумом. $(Var(\sigma_1, \sigma_2, \rho, w))' \Big|_w = 0$ при

$$\begin{aligned} w &= \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}, \\ 1 - w &= \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}. \end{aligned} \quad (36)$$

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. В силу однородности этих функций введём параметры отношения между σ_1 и σ_2 : $\alpha = \sigma_2/\sigma_1$, $\beta = \sigma_1/\sigma_2 = 1/\alpha$ – и разделим их на σ_1^2 . Тогда оптимальные веса (36) приобретают вид:

$$\begin{aligned} w &= \frac{1 - \alpha\rho}{1 + \alpha^2 - 2\alpha\rho}, \\ 1 - w &= \frac{1 - \beta\rho}{1 + \beta^2 - 2\beta\rho}. \end{aligned} \quad (37)$$

При таком значении веса дисперсия объединённого прогноза приобретает вид:

$$\frac{Var(\sigma_1, \sigma_2, \rho)}{(\sigma_1\sigma_2)^2} = \frac{1 - \rho^2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \quad (38)$$

или при использовании параметра α :

$$\frac{Var(\alpha, \rho)}{\sigma_2^2} = \frac{1 - \rho^2}{1 + \alpha^2 - 2\alpha\rho}. \quad (39)$$

Параметры ρ , σ_1 и σ_2 в условиях практического эксперимента являются константами. Для упрощения полученного выражения произведём замену:

$$\rho = k\alpha = k\frac{\sigma_2}{\sigma_1}.$$

Для определённости предположим, что $\sigma_2 \leq \sigma_1$, $\alpha \leq 1$. Дисперсия прогноза в новых переменных выражается следующим образом:

$$\frac{Var(\alpha, k)}{\sigma_2^2} = \frac{1 - k^2\alpha^2}{1 - (2k - 1)\alpha^2}. \quad (40)$$

Выражение в левой части равенства (40) неотрицательно. Т.к. коэффициент корреляции заключён в пределах $|\rho| = |k\alpha| \leq 1$, числитель выражения (40) тоже неотрицателен. Для того, чтобы выражение в знаменателе было неотрицательным, необходимо выполнение условия на коэффициент пропорциональности:

$$k = \rho/\alpha \leq \frac{1}{2} \left(\frac{1}{\alpha^2} + 1 \right),$$

избавляясь от k и переходя к параметрам ρ , α , получим неравенство:

$$\rho \leq \frac{1}{2} \left(\frac{1}{\alpha} + \alpha \right),$$

которое выполняется всюду (имеем ввиду, что $\alpha = \sigma_2/\sigma_1 \geq 0$).

Рассмотрим условия, при которых дисперсия объединённого прогноза ниже дисперсии σ_2^2 (в целом, не важно, дисперсию какой именно модели из двух выбирать, т.к. выражение для дисперсии (38) можно было сократить на σ_2^2 и получить результат, аналогичный (40), но дисперсия была бы выражена с помощью параметров σ_1^2 и β). Итак, $Var(\alpha, k)/\sigma_2^2 \leq 1$ при условии, что $1 - k^2\alpha^2 \leq 1 - (2k - 1)\alpha^2$ или

$$k^2 - 2k + 1 = (k - 1)^2 \geq 0.$$

Дисперсия объединённого прогноза $Var(\alpha, k)$ совпадает с дисперсией модели σ_2^2 при $\rho = \alpha = \sigma_2/\sigma_1$. Выкладки, связанные с сокращением и делением на σ_i^2 не корректны при условии, что у одной из моделей $\sigma_i = 0$, что бессмысленно с практической точки зрения. Аналогичные рассуждения можно привести и относительно пары параметров σ_1 , β . Таким образом, теорема в случае объединения прогнозов двух моделей доказана. Графики функции дисперсии объединённого (40) прогноза для двух моделей представлены в приложении на рисунке 6.

В случае объединения трёх или большего количества моделей такого замечательного эффекта, который указан в предыдущей теореме, не наблюдается. Дисперсия объединённого прогноза:

$$F = w_1 f_1 + w_2 f_2 + (1 - w_1 - w_2) f_3. \quad (41)$$

в случае трёх моделей имеет вид:

$$\begin{aligned} Var(\sigma_1, \sigma_2, \sigma_3, \rho_{12}, \rho_{13}, \rho_{23}, w_1, w_2) &= (w_1 \sigma_1)^2 + (w_2 \sigma_2)^2 + 2w_1 w_2 \rho_{12} \sigma_1 \sigma_2 \\ &+ 2w_1(1 - w_1 - w_2) \rho_{13} \sigma_1 \sigma_3 + 2w_2(1 - w_1 - w_2) \rho_{23} \sigma_2 \sigma_3. \end{aligned} \quad (42)$$

Здесь комбинации σ_i – стандартные ошибки объединяемых моделей, ρ_{ij} – парные корреляции прогнозов моделей с индексами i и j на тестовый период. Рассмотрим гессиан функции для того, чтобы было возможно охарактеризовать экстремумы, которые мы собираемся найти далее. Частные производные второго порядка имеют вид:

$$\begin{aligned} (Var(\sigma_1, \sigma_2, \sigma_3, \rho_{12}, \rho_{13}, \rho_{23}, w_1, w_2))'' \Big|_{w_1} &= 2(\sigma_1^2 + \sigma_3^2), \\ (Var(\sigma_1, \sigma_2, \sigma_3, \rho_{12}, \rho_{13}, \rho_{23}, w_1, w_2))'' \Big|_{w_2} &= 2(\sigma_1^2 + \sigma_3^2), \\ (Var(\sigma_1, \sigma_2, \sigma_3, \rho_{12}, \rho_{13}, \rho_{23}, w_1, w_2))'' \Big|_{w_1 w_2} &= 2(\sigma_3^2 + \rho_{12} \sigma_1 \sigma_2 - \rho_{13} \sigma_1 \sigma_3 - \rho_{23} \sigma_2 \sigma_3). \end{aligned} \quad (43)$$

Миноры первого порядка положительны. Гессииан имеет вид:

$$H(\text{Var}(\sigma_1, \sigma_2, \sigma_3, \rho_{12}, \rho_{13}, \rho_{23}, w_1, w_2))/4 = (\sigma_1^2 + \sigma_3^2)(\sigma_2^2 + \sigma_3^2) - (\sigma_3^2 + \rho_{12}\sigma_1\sigma_2 - \rho_{13}\sigma_1\sigma_3 - \rho_{23}\sigma_2\sigma_3)^2. \quad (44)$$

Это выражение имеет сложную форму, зависит от дисперсий объединяемых моделей σ_i^2 и взаимных корреляций их прогнозов на тестовый период ρ_{ij} . Для того, чтобы гарантировать наличие минимума в точке, где первые производные равны 0, необходимо потребовать, чтобы выражение (44) было положительным. В рамках вычислительных экспериментов было проверено, что для некоторых комбинаций параметров оно может быть нулевым или отрицательным, поэтому при объединении прогнозов необходимо его рассчитывать.

Рассмотрим положение экстремума по весам объединяемых моделей w_i . Приравняв частные производные по w_1, w_2 нулю, получим систему линейных по весам уравнений. Оптимальные веса для объединения трёх моделей имеют вид:

$$\begin{aligned} w_1 \cdot d &= (\sigma_3^2 - \rho_{23}\sigma_2\sigma_3)(\sigma_3^2 + \rho_{12}\sigma_1\sigma_2 - \rho_{23}\sigma_2\sigma_3) - (\sigma_3^2 - \rho_{13}\sigma_1\sigma_3)(\sigma_2^2 + \sigma_3^2 - 2\rho_{23}\sigma_2\sigma_3), \\ w_2 \cdot d &= (\sigma_3^2 - \rho_{23}\sigma_2\sigma_3)(\sigma_1^2 + \sigma_3^2 - 2\rho_{13}\sigma_1\sigma_3) - (\sigma_3^2 - \rho_{13}\sigma_1\sigma_3)(\sigma_3^2 + \rho_{12}\sigma_1\sigma_2 - \rho_{13}\sigma_1\sigma_3), \\ d &= (\sigma_3^2 + \rho_{12}\sigma_1\sigma_2 - \rho_{13}\sigma_1\sigma_3)(\sigma_3^2 + \rho_{12}\sigma_1\sigma_2 - \rho_{23}\sigma_2\sigma_3) - (\sigma_1^2 + \sigma_3^2 - 2\rho_{13}\sigma_1\sigma_3)(\sigma_2^2 + \sigma_3^2 - 2\rho_{23}\sigma_2\sigma_3). \end{aligned} \quad (45)$$

В выражениях (45) специально введён знаменатель d для упрощения выражений. Третий вес w_3 линейно зависит от рассчитанных выше: $w_3 = 1 - w_1 - w_2$.

Таким образом, при объединении трёх и большего количества моделей, необходимо индивидуально для каждого возможного объединения проверять гессииан дисперсии объединённого прогноза (44) и значение самой функции дисперсии для оптимального распределения весов (45).

Можно сделать вывод о том, что при наличии некоторого набора моделей временных рядов, использующихся как базовые инструменты для прогнозирования, объединение двух моделей является перспективным подходом. Объединение большего числа моделей должно опираться на математические расчёты, представленные выше. Можно предположить, что при наличии большого количества моделей процесс объединения (если моделей больше двух, то имеем ввиду объединение, не увеличивающее дисперсию прогноза) ведёт к постепенному уменьшению дисперсии прогноза, пока она не будет сходиться к некоторому предельному значению, в идеале равному нулю. Чем больше моделей для прогнозирования есть в наличии, тем более точным будет совмещённый прогноз. При этом, учитывая результаты, полученные в работе, рекомендуется использовать либо объединение пар моделей, выбранных из используемого набора, либо возможно объединение большего числа моделей при выполнении условия о том, что дисперсия объединённого прогноза снижается. К сожалению, на практике количество моделей не может быть настолько большим, чтобы было возможно говорить о сходимости оценки дисперсии прогноза к определённому пределу, поэтому мы приводим это рассуждение, схожее с действиями, применяемыми в статистике для большого числа наблюдений, для объяснения общей тенденции.

6. ОБЪЕДИНЕНИЕ ПРОГНОЗОВ ДВУХ МОДЕЛЕЙ ВРЕМЕННЫХ РЯДОВ AR(p)

Рассмотрим объединение двух прогнозов моделей типа AR(p), оптимальное по дисперсии прогноза. Функция дисперсии прогноза оптимального объединения двух моделей (38) имеет вид:

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

Рассмотрим дисперсию прогноза двух моделей $AR(p)$ на два шага вперёд (7):

$$\begin{aligned}\sigma_1 &= se_1 \sqrt{1 + \varphi_1^2}, \\ \sigma_2 &= se_2 \sqrt{1 + \varphi_2^2}.\end{aligned}\tag{46}$$

Здесь se_i – это стандартные ошибки моделей, а σ_i^2 – дисперсии объединяемых моделей вида $AR(p)$, индексы коэффициентов φ_i совпадают с номером модели ($i = 1, 2$), т.к. будет использован только один коэффициент. Эти значения можно подставить в дисперсии прогноза оптимального объединения двух моделей (38), но для того, чтобы проверить экстремумы этой функции, проще использовать производную сложной функции:

$$Var(\sigma_1, \sigma_2, \rho)' \Big|_{\varphi_i} = Var(\sigma_1, \sigma_2, \rho)' \Big|_{\sigma_i} \sigma_i' \Big|_{\varphi_i}.$$

Производная функции $Var(\sigma_1, \sigma_2, \rho)$ по φ_i имеет вид:

$$\begin{aligned}Var(\sigma_1, \sigma_2, \rho)' \Big|_{\varphi_1} &= \frac{2(1 - \rho^2)\sigma_1\sigma_2^3}{(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^2} (\sigma_2 - \rho\sigma_1) \frac{se_1\varphi_1}{\sqrt{1 + \varphi_1^2}}, \\ Var(\sigma_1, \sigma_2, \rho)' \Big|_{\varphi_2} &= \frac{2(1 - \rho^2)\sigma_1^3\sigma_2}{(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^2} (\sigma_1 - \rho\sigma_2) \frac{se_2\varphi_2}{\sqrt{1 + \varphi_2^2}}.\end{aligned}\tag{47}$$

Рассмотрим производную $Var(\sigma_1, \sigma_2, \rho)' \Big|_{\varphi_1}$, подставив значение σ_1 и $\sigma_1' \Big|_{\varphi_1}$. Первая функция в (47) преобразуется к виду:

$$Var(\sigma_1, \sigma_2, \rho)' \Big|_{\varphi_1} = \frac{2(1 - \rho^2)se_1^2\varphi_1\sigma_2^3}{(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^2} (\sigma_2 - \rho\sigma_1).\tag{48}$$

В (48) не везде подставлены выражения σ_i для упрощения. Для одинаковых значений ошибок $se_1 = se_2 = se$ и для значения корреляции прогнозов $\rho = 1/2$ график функции $Var(\varphi_1, \varphi_2)/se^2$ представлен в приложении на рисунке 4. Производная по φ_2 аналогична, требуется только поменять местами $\sigma_1 \leftrightarrow \sigma_2$, $\varphi_1 \leftrightarrow \varphi_2$. Комбинируя эти уравнения, видим, что производные равны нулю при $\rho = \pm 1$, $\varphi_i = 0$. Таким образом, следует рассмотреть прямые $\varphi_i = 0$, а также есть особые случаи, когда коэффициент корреляции между прогнозами $|\rho| = 1$, что нас не интересует с практической точки зрения. Рассмотрим график функции $Var(\sigma_1, \sigma_2, \rho) / ((1 - \rho^2)(\sigma_1\sigma_2)^2)$, чтобы определить характер экстремумов. График функции для $\rho = 1/2$ представлен на рисунке 4. Функция имеет характер выпуклой вниз и минимумы достигаются на прямых $\varphi_i = 0$. Рассматривая прогнозы на большее количество шагов вперёд, для модели $AR(1)$ продолжим получать схожие результаты, т.к. производные $Var(\sigma_1, \sigma_2, \rho)' \Big|_{\sigma_i}$

сохраняют свой вид, а в производных $\sigma_i' \Big|_{\varphi_j}$ (7) начинают появляться положительные множители, не влияющие на характер и расположение экстремумов (здесь j уже задаёт нумерацию коэффициентов моделей). Для моделей $AR(p)$ порядков $p > 1$ при прогнозировании более, чем на 2 шага вперёд, требуется дополнительное исследование.

Для дисперсии объединённого прогноза двух моделей $ARMA(p, q)$ на два шага вперёд (10), результат сохраняется при замене $\varphi_i \rightarrow \varphi_i + \theta_i$. Поэтому прямые, на которых достигается минимум дисперсии, перемещаются в $\varphi_i = -\theta_i$, но характер функции остаётся прежним (здесь i – номер модели). Для моделей $ARMA(p, q)$ порядков $p > 1$ или $q > 1$ при прогнозировании более, чем на 2 шага вперёд, требуется дополнительное исследование.

7. ПРАКТИЧЕСКИЙ ЭКСПЕРИМЕНТ

7.1. Исследование объединения прогнозов моделей ARIMA(p, d, q)

Рассмотрим месячный временной ряд индекса реальной заработной платы WAG [19] в период между 1993 и 2015 гг. Данные последнего года используем как тестовую выборку, на которую построенные модели сделают прогноз. Предыдущий период 1993 – 2014 гг. используется для настройки объединяемых моделей. В рамках первого эксперимента рассматривается только усреднение прогнозов нескольких моделей. Для настройки моделей воспользуемся функцией `auto.arima()` из языка R. Её результат используется для определения вида сезонности. Он устанавливается одинаковым для всех тестируемых моделей для определённости. Подробно этот эксперимент представлен в работе [18]. В рамках эксперимента подсчитано значение информационных критериев [11, 12] для моделей ARIMA(p, d, q), $p < 6$, $q < 6$, в последних столбцах таблицы 1 приведены оценки дисперсии прогноза и отклонения прогноза модели от реальных значений с помощью метрик RMSE и MAE:

$$RMSE = \sqrt{\frac{\sum_t (\tau(t) - ts(t))^2}{N}} \quad (49)$$

$$MAE = \frac{\sum_t |\tau(t) - ts(t)|}{N}.$$

Здесь N – длина тестового периода, $\tau(t)$ – реальные значения временного ряда, $ts(t)$ – значения, предсказанные тестируемой моделью.

Таблица 1. Модели временных рядов ARIMA(p, d, q) для индекса реальной зарплаты в РФ за период 1999 – 2015 гг.

№	Модель	AIC	BIC	RMSE	MAE	Оценка дисперсии прогноза
1	ARIMA(1, 0, 1)	864.56	873.73	3.56	2.38	85.94
2	ARIMA(2, 0, 0)	863.78	863.94	3.55	2.37	88.92
3	ARIMA(2, 0, 4)	869.55	872.95	3.52	2.36	91.77
4	ARIMA(3, 0, 5)	862.43	889.93	3.35	2.23	112.76
5	ARIMA(5, 0, 4)	865.30	866.81	3.31	2.21	100.52
6	Взвешенное среднее моделей 1, 5	-	-	3.56	2.38	85.94
7	Взвешенное среднее моделей 2, 3	-	-	3.55	2.37	88.92
8	Взвешенное среднее моделей 3, 4	-	-	3.52	2.36	91.77

В этом эксперименте корреляция прогнозов моделей, участвующих в объединении, превышает 0.99. Из-за этого фактически происходит выбор одной из объединяемых моделей. Модель №6 крайне близка к ARIMA(1, 0, 1), модель №7 – к ARIMA(2, 0, 0), а модель №8 – к ARIMA(2, 0, 4). В случае, если одна из моделей существенно превышает другую по качеству прогноза или, наоборот, модели предсказывают одинаковый результат, происходит выбор одной из них: веса моделей в комбинации преобразуются к виду (1, 0) или (0, 1).

Рассмотрим месячный временной ряд курса доллара к рублю RDEXRO [19]. Период между 1996 и 2020 гг. использовался для настройки и отбора наилучших моделей по качеству прогноза на 2021 г. По оценке дисперсии прогноза на 2021 г. пары моделей объединялись. Весовые коэффициенты при построении объединения 37 подбирались так, чтобы оценка дисперсии прогноза была минимальной (её величина указана в последнем столбце). Исходные модели

представлены в начале таблицы 2, объединённые пары моделей – в конце. Все модели не обладают сезонной составляющей, порядок $d = 1$.

Таблица 2. Модели временных рядов ARIMA(p, d, q) для процесса курса доллара к рублю

№	Модель	AIC	BIC	RMSE	MAE	Оценка дисперсии прогноза
1	ARIMA(1, 1, 0)	3352.75	3360.08	424.78	402.89	7779.38
2	ARIMA(2, 1, 1)	3353.69	3353.83	433.48	412.51	7754.75
3	ARIMA(3, 1, 3)	3357.95	3358.35	415.05	392.77	7775.61
4	ARIMA(4, 1, 0)	3354.81	3373.13	406.09	382.88	7751.54
5	ARIMA(5, 1, 5)	3353.63	3393.92	379.28	347.99	7385.71
6	Взвешенное среднее моделей 2, 3	-	-	415.05	392.77	-
7	Взвешенное среднее моделей 1, 5	-	-	379.28	347.99	-
8	Взвешенное среднее моделей 4, 5	-	-	379.28	347.99	-

В таблице 2 представлены комбинации моделей под номерами 6 – 8 (коэффициенты даны с точностью до 0.01, дополнительно указан коэффициент корреляции r прогнозов объединяемых моделей):

$$\begin{aligned}
 Model_6 &= 0.5Model_2 + 0.5Model_3, r = -0.16, \\
 Model_7 &= 0.49Model_1 + 0.51Model_5, r = -0.45, \\
 Model_8 &= 0.44Model_4 + 0.56Model_5, r = 0.79.
 \end{aligned}
 \tag{50}$$

Обратим внимание на то, что в некоторых случаях модели прогнозируют противоположное поведение временного ряда (коэффициенты корреляции прогнозов отрицательные), но, при этом, оценки по метрикам RMSE и MAE для объединённых моделей совпадают с оценками прогнозов наилучших моделей, участвующих в комбинации.

7.2. Исследование объединения прогнозов рядов ADL(p, q)

В следующем эксперименте исследуется взаимосвязь инфляции и безработицы на примере российской статистики [19] в период 2000–2020 гг. Соответствующая взаимосвязь на примере данных других стран известна достаточно давно [20], периодически появляются публикации, посвящённые исследованию этой взаимосвязи на примере отдельных стран или с учётом местных особенностей [21], [22], [23]. Для моделирования этой взаимосвязи построено несколько моделей ADL(p, q): независимая переменная – ряд безработицы, объясняемая переменная – ряд инфляции. Произведено объединение нескольких пар моделей. Весовой коэффициент в линейной комбинации подбирается в зависимости от того, насколько часто модель даёт наилучший прогноз среди списка конкурирующих моделей. Изначально все коэффициенты равны – речь идёт о усреднении прогнозов моделей. Затем модели делают прогноз. У модели, которая сделала самый близкий прогноз к реальным значениям в смысле евклидова расстояния (у модели-победителя), вес увеличивается на некоторую величину. У других моделей веса снижаются. Сумма весов остаётся равной единице. Дополнительно производится проверка весов на равенство нулю. Если он равен 0, то модель перестаёт участвовать в комбинации, её вес не получает отрицательное значение. В качестве данных для настройки моделей использовался

период 2000 – 2009 гг., на данных 2009–2019 гг. происходило оценивание прогнозов на следующий месяц и корректировка весов в объединении прогнозов моделей в зависимости от того, какая модель дала прогноз, самый близкий к реальному положению дел. Произведена оценка прогнозов на тестовый – 2020-й год. Результаты эксперимента представлены в таблице 3. В первой строке представлена модель ADL(5, 5). Некоторые из слагаемых не являются значимыми, поэтому в следующих двух моделях удалены незначимые слагаемые: во второй модели лаги безработицы 1-го, 2-го, 3-го порядка и лаг инфляции 4-го порядка. В третьей модели дополнительно удалён лаг инфляции 3-го порядка. Модель №5 – это модель с теми же старшими порядками, что и модель №4, но из неё удалены 1-й лаг безработицы и 4-й лаг инфляции. Так получено пять моделей, которые участвуют в экспериментах по объединению. Модель №6 – это среднее арифметическое прогнозов всех предыдущих моделей (или модель со средними значениями коэффициентов в силу линейности модели ADL). Модель №7 – это среднее арифметическое моделей 1–3, модель №8 – среднее арифметическое моделей 4–5. Модель №9 – это результат взвешенного объединения всех моделей, модель №10 – взвешенная комбинация моделей 1–3, модель №11 – взвешенная комбинация моделей 4–5. В столбцах приведены значения информационных критериев, коэффициента детерминации R^2 и меры ошибок RMSE, MAE для тестового года – 2020 г.

Таблица 3. Модели временных рядов $ADL(p, d, q)$ для процессов инфляции и безработицы российской макроэкономической статистики [19]

№	Модель	R^2	AIC	BIC	RMSE	MAE
1	ADL(5, 5)	0.397	-831.55	-796.56	0.058	0.045
2	$ADL(5, 5)_2$	0.371	-835.04	-810.81	0.058	0.044
3	$ADL(5, 5)_3$	0.353	-833.92	-812.39	0.057	0.043
4	$ADL(2, 5)$	0.285	-819.06	-792.14	0.058	0.045
5	$ADL(2, 5)_2$	0.281	-822.47	-800.94	0.058	0.046
6	Средняя модель	0.397	-831.55	-796.56	0.028	0.021
7	Среднее моделей 1, 2, 3	0.397	-831.55	-796.56	0.028	0.021
8	Среднее моделей 4, 5	0.285	-819.06	-792.14	0.028	0.021
9	Взвешенное среднее моделей 1–5	0.397	-831.55	-796.56	0.028	0.020
10	Взвешенное среднее моделей 1, 2, 3	0.397	-831.55	-796.56	0.028	0.021
11	Взвешенное среднее моделей 4, 5	0.285	-819.06	-792.14	0.028	0.021

В результате процесса объединения моделей модель 9 в таблице 3 имеет вид:

$$Model_9 = 0.6Model_5 + 0.24Model_1 + 0.16Model_3.$$

Здесь использована нумерация моделей из таблицы 3. Результаты даны с точностью до 0.01. При этом, в ходе эксперимента веса моделей 2 и 4 снизились до величины < 0.01 . Аналогично комбинации моделей с номерами 10, 11 имеют вид:

$$Model_{10} = 1.00Model_1,$$

$$Model_{11} = 0.89Model_4 + 0.11Model_5.$$

При построении модели 10 (она является взвешенной комбинацией моделей ADL(5, 5)) в большинстве случаев наилучший прогноз давала модель под номером 1, поэтому её вес в итоге оказался равен 1, а веса остальных моделей, участвующих в комбинации, снизились до величин

< 0.01 . Это один из возможных исходов построения комбинации, при которой редактируется в положительную сторону только вес модели, сделавшей лучший прогноз (победитель забирает всё). В дальнейших исследованиях планируется провести вычислительный эксперимент, в котором положительная добавка к весу распределяется между несколькими лучшими моделями.

Как видно, лучшие значения информационных критериев (AIC – критерий Акаике, BIC – критерий Шварца / Байеса) у моделей ADL(5, 5) с исключенными слагаемыми (модели 2 и 3), у моделей 6, 7 и 9, 10 (это взвешенные и невзвешенные комбинации всех моделей или моделей 1, 2, 3). Значения информационных критериев растут с ростом порядков модели, поэтому фактически модели высоких порядков штрафуются по сравнению с моделями низких порядков p и q . Интересно, что наилучшие значения критериев получены у моделей высоких порядков. Это означает, что модели "хорошо" подстроены под данные. С точки зрения качества прогноза на тестовый период, наилучшие результаты показала взвешенная комбинация всех моделей (модель 9), а также все усреднённые модели и модели, основанные на взвешенной комбинации. Их прогнозы лучше по критериям RMSE и MAE чем прогнозы моделей, построенных традиционным путём. Подход к изменению весов по ходу эксперимента может быть предметом дальнейших исследований в области объединения прогнозов временных рядов.

8. ЗАКЛЮЧЕНИЕ

В ходе исследований моделей временных рядов ARIMA (p, d, q) удалось построить функции оценки дисперсии прогноза на 1, 2 и 3 шага вперёд. Они выражаются с помощью ψ -весов (представление ряда в виде MA(∞)). Функции дисперсий прогнозов для моделей ARMA(1, q) на 1, 2, 3 шага вперёд являются выпуклыми вниз. При некоторых условиях на параметры моделей ϕ_i, θ_i , указанных в исследовании, дисперсии прогнозов для моделей ARMA(p, q), $p = 2, 3$ тоже выпуклы вниз. Таким образом, при объединении нескольких моделей оценка дисперсии прогноза снижается.

Оценка дисперсии прогноза для моделей ADL(p, q) основана на использовании критерия Стьюдента, поэтому вопросу объединения нескольких моделей ADL(p, q) будут посвящены дальнейшие исследования. Вместе с тем, проведено исследование оценки функции дисперсии объединения двух моделей произвольной природы. Показано, что при оптимальном выборе весовых коэффициентов в объединении оценка дисперсии прогноза объединённой модели (или оптимальной комбинации прогнозов двух моделей) не превосходит дисперсии объединяемых моделей.

В ходе практического эксперимента рассмотрено несколько моделей, полученных на основе взвешенной комбинации моделей временных рядов ARIMA (p, d, q) (эксперименты №1, №2) и ADL(p, q) (эксперимент №3). Практические эксперименты по усреднению моделей временных рядов ARIMA (p, d, q) приведены в работе [18].

Применение беггинга к временным рядам (речь фактически о введении взвешенной комбинации моделей на основе некоторого принципа) является популярной темой исследований в области моделирования временных рядов ([1], [13], [15], [16]). Очевидно, в ближайшие годы идея переноса этой технологии на временные ряды приведёт к введению новых методов прогнозирования.

9. ПРИЛОЖЕНИЯ

Графики функций дисперсии прогнозов для модели AR(1) (19) на 2 и 3 временных шага вперёд представлены на рисунке 1.

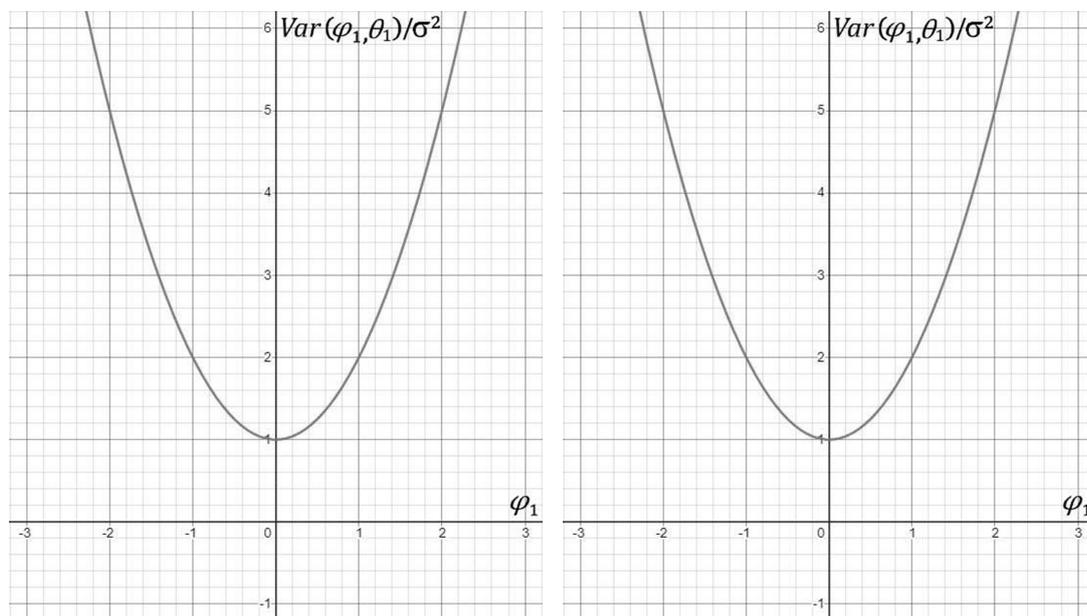


Рис. 1. Графики зависимости функций дисперсии прогноза модели AR(1) $Var(\varphi_1, \theta_1)/\sigma^2$ на 2 шага вперёд (слева) и 3 шага вперёд (справа) от параметра модели φ_1

График функции дисперсии прогноза для модели AR(2) (14) на 3 шага вперёд представлен на рисунке 2.

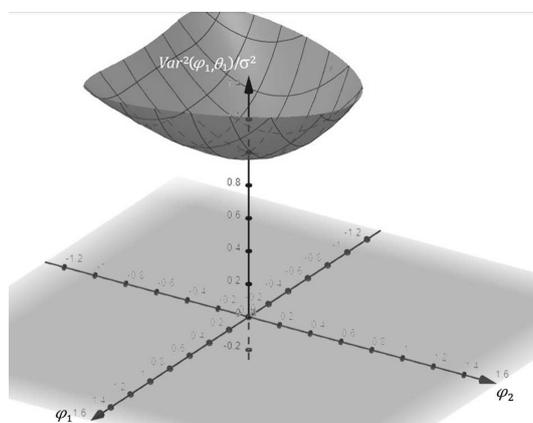


Рис. 2. График зависимости функции дисперсии прогноза для модели AR(2) на 3 шага вперёд от параметров модели φ_1, φ_2

Графики функций дисперсии прогнозов для модели ARMA(1, 1) $Var(\varphi_1, \varphi_2)/\sigma^2$ (21) на 2 и 3 временных шага вперёд представлены на рисунке 3.

Для случая совпадающих ошибок $se_1 = se_2 = se$ график зависимости функции дисперсии объединённого прогноза для двух моделей $Var(\varphi_1, \varphi_2)/se^2$ (48) на 2 шага вперёд от коэффициентов моделей AR(1) φ_1, φ_2 при $\rho = 1/2$ представлен на рисунке 4.

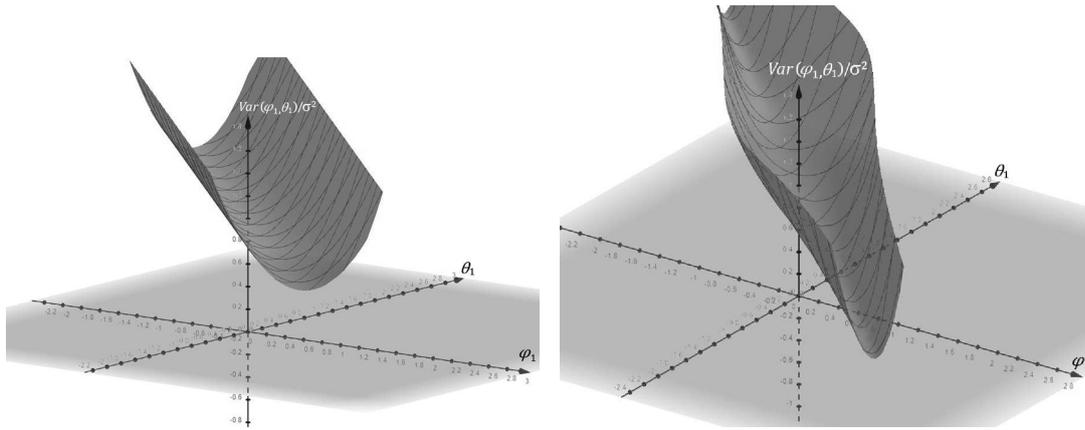


Рис. 3. Графики зависимости функций дисперсии прогноза модели ARMA(1, 1) $Var(\varphi_1, \theta_1)/\sigma^2$ на 2 шага вперёд (слева) и 3 шага вперёд (справа) от параметров модели φ_1, θ_1

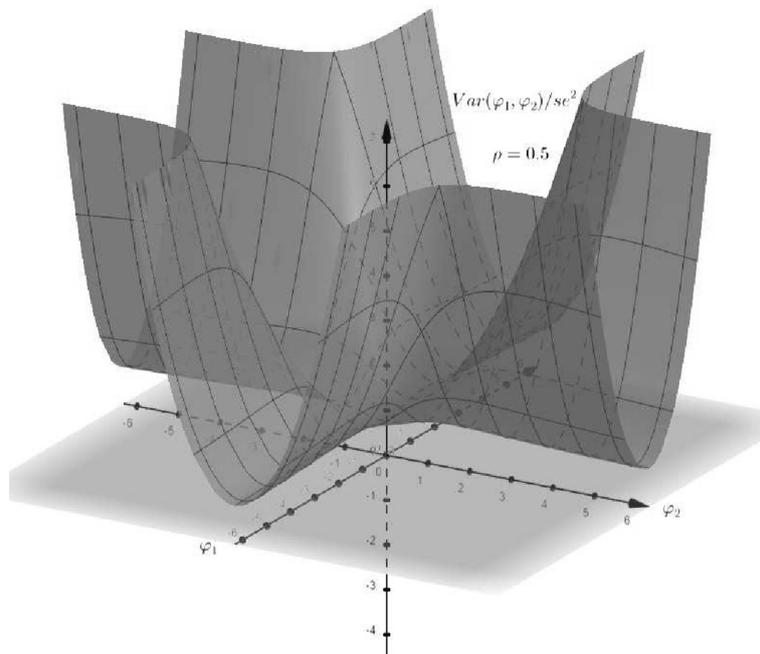


Рис. 4. График зависимости функции дисперсии объединённого прогноза двух моделей AR(1) на 2 шага вперёд $Var(\varphi_1, \varphi_2)/se^2$ от параметров объединяемых моделей φ_1, φ_2

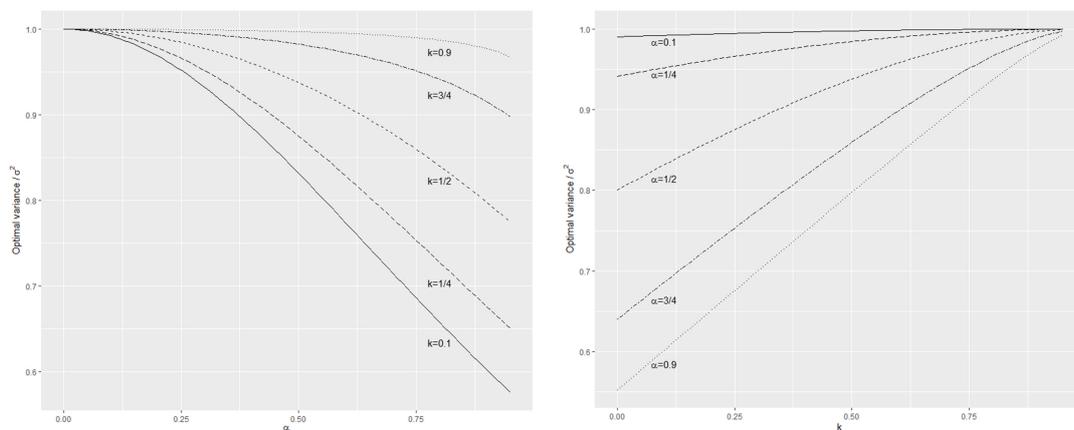


Рис. 5. Графики функций дисперсии объединённого прогноза для двух моделей в зависимости от α для некоторых значений $k = \rho/\alpha$ (слева) и в зависимости от $k = \rho/\alpha$ для некоторых значений α (справа)

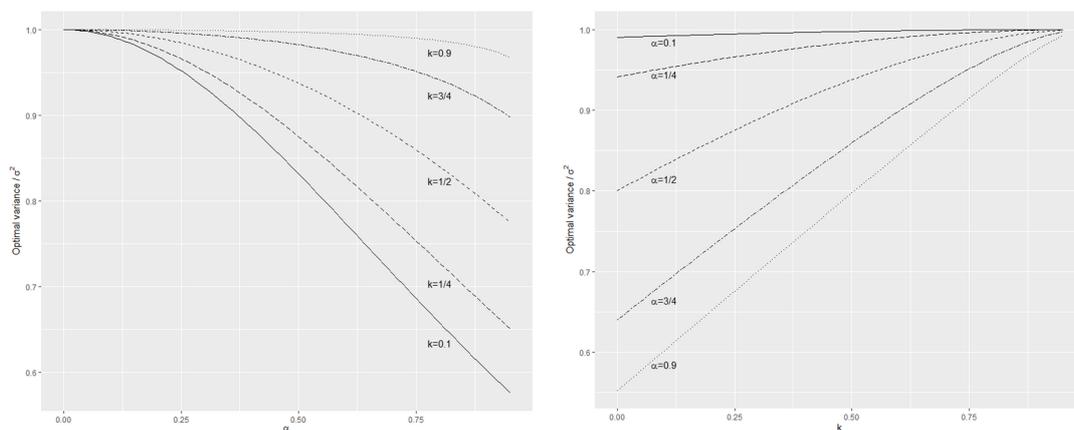


Рис. 6. Графики функций дисперсии объединённого прогноза для двух моделей в зависимости от α для некоторых значений $k = \rho/\alpha$ (слева) и в зависимости от $k = \rho/\alpha$ для некоторых значений α (справа)

График функции дисперсии объединённого прогноза для двух моделей вида AR(1) (40) при различных значениях параметра представлены на рисунке 6.

СПИСОК ЛИТЕРАТУРЫ

1. Stefenon S.F., Dal Molin Ribeiro M.H., Nied A., Yow K.C., Mariani V.C., dos Santos Coelho L., Seaman L.O. Time series forecasting using ensemble learning methods for emergency prevention in hydroelectric power plants with dam // *Electric Power Systems Research*, 2022, Vol.202, P.107584. doi: <https://doi.org/10.1016/j.epsr.2021.107584>
2. Shishmanova P.D. A comparative analysis of the main characteristics of human capital in Bulgaria and Russia // *Russian Technological Journal*, 2020, Vol.8(4), pp.141-151. doi: <https://doi.org/10.32362/2500-316X-2020-8-4-141-151>
3. Moral-Garcia S., Mantas C.J., Castellano J.G., Benitez M.D., Abellan J. Bagging of credal decision trees for imprecise classification // *Expert Systems with Applications*, 2020, Vol.141, P.112944. doi: <https://doi.org/10.1016/j.eswa.2019.112944>
4. Gu H., Cui Y.F., Xu L., Tu M.Y., Fu Y.J., Fu H.Y., Zhou Y.P. Bagging classification tree-based robust variable selection for radial basis function network modeling in metabonomics data analysis // *Chemometrics and Intelligent Laboratory Systems*, 2018, Vol.174, pp.76-84. doi: <https://doi.org/10.1016/j.chemolab.2017.12.014>
5. Subasi A., Kadasa B., Kremic E. Classification of the Cardiotocogram Data for Anticipation of Fetal Risks using Bagging Ensemble Classifier // *Procedia Computer Science*, 2019, Vol.168, pp.34-39. doi: <https://doi.org/10.1016/j.procs.2020.02.248>
6. Golov V.A., Petrushevich D.A. Data analysis methods in astronomic objects classification (Sloan Digital Sky Survey DR14) // *Russian Technological Journal*, 2021, Vol.9(3), pp.66-77. doi: <https://doi.org/10.32362/2500-316X-2021-9-3-66-77>
7. Pooja S.B., Siva Balan R.V., Anisha M., Muthukumar M.S., Jothikumar R. Techniques Tanimoto correlated feature selection system and hybridization of clustering and boosting ensemble classification of remote sensed big data for weather forecasting // *Computer Communications*, 2020, Vol.151, pp.266-274. doi: <https://doi.org/10.1016/j.comcom.2019.12.06>
8. Deepalakshmi S., Arulmurugan R. Gaussian bilateral filtered discrete Hartley feature transformation based infomax boosting for hyperspectral image classification // *International Journal of Intelligent Networks*, 2021, Vol.2, pp.195-203. doi: <https://doi.org/10.1016/j.ijin.2021.11.001>
9. Sigrist F. Gradient and Newton boosting for classification and regression // *Expert Systems with Applications*, 2021, Vol.167, P.114080. doi: <https://doi.org/10.1016/j.eswa.2020.114080>
10. Savinykh V.P., Gospodinov S.G., Kudzh S.A., Tsvetkov V.Y., Deshko I.P. Semantics of visual models in space research // *Russian Technological Journal*, 2022, Vol.10(2), pp. 51-58. <https://doi.org/10.32362/2500-316X-2022-10-2-51-58>
11. Hyndman R.J., Athanasopoulos G. *Forecasting: principles and practice*. Melbourne: OTexts, 2018 (2nd edition)
12. Stock J.H., Watson M.W. *Introduction to Econometrics*. Pearson, 2019
13. 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.268.2, pp.545-554. doi: <https://doi.org/10.1016/j.ejor.2018.01.045>
14. 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>
15. Larrea S.F., Porto A., Irigoyen E., Barragan A.J., Andujar J.M. Extreme learning machine ensemble model for time series forecasting boosted by PSO: Application to an electric consumption problem // *Neurocomputing*, 2021, Vol.452, pp.465-472. doi: <https://doi.org/10.1016/j.neucom.2019.12.140>

16. Godahewa R., Bandara K., Webb G.I., Smyl S., Bergmeir C. Ensembles of localised models for time series forecasting // Knowledge-Based Systems, 2021, Vol.233, P.107518. doi: <https://doi.org/10.1016/j.knosys.2021.107518>
17. Wald H. A study in the analysis of stationary time series. Uppsala: Almqvist and Wiksell Book Co., 1954 (2nd edition)
18. Musatov D., Petrushevich D. Modeling of forecasts variance reduction at multiple time series prediction averaging with ARMA(1, q) functions // Proceedings of MIP Computing-V 2022: V International Scientific Workshop on Modeling, Information Processing and Computing. 2022, Vol. 3091, pp. 60-65. doi: <https://doi.org/10.47813/dnit-mip5/2022-3091-1-11>
19. Dynamic series of macroeconomic statistics of the Russian Federation. Wage Index, Dollar rate. Retrieved from: <http://sophist.hse.ru/hse/nindex.shtml>, 04-04-2022
20. Philips A.W. The relation between unemployment and the Rate of change of money wage rates in the United Kingdom, 1861 – 1957 // *Economica*, 1958, Vol.25(100), pp.283-299
21. Karanassou M., Sala H. Snower D.J. Long-run inflation-unemployment dynamics: The Spanish Phillips curve and economic policy// *Journal of Policy Modeling*, 2008, Vol.30(2), pp.279-300. doi: <https://doi.org/10.1016/j.jpolmod.2007.04.002>
22. Karanassou M., Sala H. The US inflation-unemployment trade-off revisited: New evidence for policy-making // *Journal of Policy Modeling*, 2010, Vol.32(6), pp.758-777. doi: <https://doi.org/10.1016/j.jpolmod.2010.08.004>
23. Bhattarai K. Unemployment-inflation trade-offs in OECD countries // *Economic Modelling*, 2016, Vol.58, pp.93-103. doi: <https://doi.org/10.1016/j.econmod.2016.05.007>

Linear combination of time series models with minimal forecast variance

Beletskaya N., Petrushevich D.

In this paper construction of optimal combination of time series forecasts (by quality of prediction or forecast variance evaluation) is considered. Also averaging of multiple models' forecasts is in scope of this research as a part of weighted model combination. These approaches are widely used in time series modelling and forecasting. In the theoretical part functions evaluating forecast variance of ARIMA(p, d, q) models over 1, 2, 3 steps ahead are considered with use of ψ -weights. Property of convex downward is treated for averaged or weighted combination of several ARIMA(p, d, q), $p < 4$ models' forecasts. Also, forecasts combination for two models of arbitrary type is considered. Forecasts tak part in weighted combination and weights are counted in the way to minimize evaluation of forecast variance. In the experimental part weighted combinations (optimal by forecast variance) of ARIMA(p, d, q) models and ADL(p, q) models are built. Quality of combined models' forecasts isn't worse than accuracy of treated models' forecasts.

KEYWORDS: ARIMA(p, d, q), ADL(p, q), Akaike information criterion (AIC), Bayes information criterion (BIC), optimal combination, forecast variance minimization.