

Секция «Подсекция аспирантов. Региональная экономика, экономика промышленности и экономика инноваций: от стратегии к практике развития»

## ARIMA-профиль: определение и поиск на примере прогнозирования инфляции в Томской области

Научный руководитель – Богданов Александр Леонидович

*Шабанова Екатерина Юрьевна*

*Аспирант*

Национальный исследовательский Томский государственный университет, Томск, Россия

*E-mail: ekaterina.shabanova.tsu@mail.ru*

Одной из задач региональных отделений Центрального Банка РФ является прогнозирование индекса потребительских цен – официального измерителя уровня инфляции в России [1,2]. Сегодня, как никогда, Центробанк нуждается в инструменте, способном не просто предсказывать будущие значения инфляции, но и расширять знания о структуре исследуемого процесса. Зачастую в качестве такого инструмента используются модели ARIMA.

Под ARIMA-профилем в рамках настоящего исследования будет пониматься набор правил, разделяющих ARIMA-модели на "подходящие" и "не подходящие" для решения задачи прогнозирования.

Цель исследования состоит в том, чтобы экспериментально **продемонстрировать существование ARIMA-профиля** задачи прогнозирования ИПЦ в Томской области.

Модель ARIMA определяется параметрами  $p, d, q, P, D, Q, s$ . Параметры  $d$  и  $D$  отвечают за приведение временного ряда к стационарному виду. Ненулевые значения параметров  $p$  и  $q$  означают, что текущее значение ИПЦ определяется через  $p$  предыдущих значений ИПЦ и  $q$  предыдущих шоков. Аналогично с учетом сезонности интерпретируются параметры  $P$  и  $Q$ . Параметры сами по себе представляют исследовательский интерес, так как они «рассказывают» о структуре изучаемого явления.

Исходные данные представляли собой ежемесячные значения ИПЦ в Томской области за период с января 2002 года по август 2023 года. Данные получены с сайта ФСГС «Росстат» [3].

Для нахождения ARIMA-профиля задачи прогнозирования ИПЦ в Томской области было необходимо перебрать множество из 324-х различных ARIMA-моделей и оценить качество прогноза, которое они способны обеспечить.

Весь временной ряд несколько раз разбивался на обучающее, валидационное и тестовое множества (Рисунок 1). Обучающее и валидационное множества постепенно сдвигались по временному ряду слева направо, образуя «новые» наборы данных для оценки моделей. Всего было выполнено 69 разбиений.

Качество прогноза моделей оценивалось с помощью метрики RMSE. Чем меньше величина ошибки RMSE, тем лучше считается модель.

В результате эксперимента на каждом из 69-и разбиений были получены ранжированные списки из 324-х моделей ARIMA. Каждой модели присвоен ранг от 1 до 324 в соответствии с ее позицией в списке.

На тепловой карте (Рисунок 2) были отмечены случаи, когда  $ARIMA(p,d,q)(P,D,Q)s$  оказывалась в списке 10% лучших моделей. На оси абсцисс отложены 69 периодов прогнозирования, на оси ординат – 324 ARIMA-модели. Клетка на тепловой карте закрашена красным, если модель оказалась среди 10% лучших на конкретном разбиении. Если бы модели не улавливали никаких закономерностей в данных, то все они имели бы одинаковые

шансы на попадание в топ-список и тепловая карта соответствовала бы представленной на рисунке 3.

Заметно, что **реальная картина отличается** от тепловой карты, полученной путем случайного ранжирования моделей. Так можно заключить о **наличии структуры в инфляционном процессе** Томской области и существовании ARIMA-профиля.

ARIMA-профиль можно было бы описать простым перечислением «хороших» и «плохих» моделей, однако оно будет мало информативным. Хотелось бы иметь несколько лаконичных «правил», разделяющих модели на «хорошие» и «плохие». Набор разделяющих правил был получен как решение задачи классификации ARIMA-моделей на «плохие» и «хорошие» с помощью деревьев решений. «Хорошими» были отмечены модели, более 10 раз оказавшиеся в топ-списке, «плохими» – все остальные. Так, в выборке оказалось 20% «хороших» и 80% – «плохих» моделей. Предикторами в задаче классификации выступали параметры  $p, d, q, P, D, Q, s$ , а целевой переменной – метка класса «плохая модель» ("B") или «хорошая модель» ("G").

На рисунке 4 представлено полученное дерево решений. Среди всех разделяющих правил выделены два, пользующиеся наибольшим доверием (критерий Джини для них сильнее всего отличается от нежелательного значения 0,5). Правила, наилучшим образом отделяющие «хорошие» модели от «плохих»:

1. Если параметр  $Q = 0$  и параметр  $P = [0, 1]$ , то независимо от значений остальных параметров, с вероятностью 90% эта модель окажется **не подходящей** для прогнозирования ИПЦ в Томской области. Шаблон такой модели можно записать в виде  $ARIMA(*, *, *)([0, 1], *, 0)$ , где символ \* обозначает любое возможное значение соответствующего параметра.

2. Если параметр  $Q = 2$  и параметр  $P = 1$ , то с вероятностью 82% модель окажется **подходящей** для прогнозирования ИПЦ в Томской области. Шаблон модели:  $ARIMA(*, *, *) (2, *, 1)$ . Выносимая из этого шаблона закономерность – текущее значение инфляции в Томской области зависит от инфляции, которая наблюдалась год назад и два года назад, а также от значения прошлогоднего шока.

Далее для проверки полученных результатов с помощью группы «плохих» и группы «хороших» моделей были построены прогнозы на тестовый период – последние 12 месяцев, не использованные при поиске ARIMA-профиля (Рисунок 5).

Прогнозы двух групп моделей были усреднены (Рисунок 6). Для сравнения точности двух усредненных прогнозов в каждом из 12 месяцев были рассчитаны их ошибки RMSE и вычислена разность этих ошибок (Рисунок 7). Положительное значение разности означало, что ошибка усредненного прогноза «плохих» моделей выше ошибки «хороших». Как видно из рисунка, практически во всех периодах средний прогноз моделей, определяемых как "хорошие", действительно оказался точнее.

Таким образом, было показано, что **ARIMA-профиль** задачи прогнозирования ИПЦ в Томской области **существует** – все множество моделей можно разделить на те, которые определено «подходят» для решения поставленной задачи и те, которые определено «не подходят». ARIMA-профиль был описан набором правил, отделяющих одно подмножество от другого.

Полученный ARIMA-профиль можно использовать для выбора модели при решении задачи прогнозирования ИПЦ и лучшего понимания закономерностей инфляционного процесса в Томской области.

## Источники и литература

- 1) "Об утверждении Официальной статистической методологии организации статистического наблюдения за потребительскими ценами на товары и услуги и расче-

та индексов потребительских цен": Приказ Росстата от 30.12.2014 № 734 (ред. от 27.11.2019) // «КонсультантПлюс»: справочная правовая система. URL: <http://www.consultant.ru> (дата обращения: 15.08.2023).

- 2) Центральный банк Российской Федерации [Электронный ресурс] // Официальный сайт – Электрон. дан. – Москва – URL: <http://www.cbr.ru> (дата обращения: 01.08.2023).
- 3) Федеральная служба государственной статистики [Электронный ресурс] // Официальный сайт – Электрон. дан. – Москва – URL: <http://www.gks.ru> (дата обращения: 07.08.2023).

### Иллюстрации

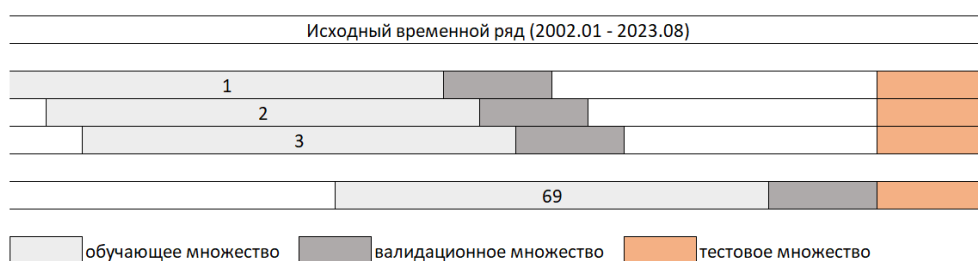


Рис. : Разбиение ряда на обучающее, валидационное и тестовое множества: на каждом из 69-и разбиений оценивалась точность прогноза 324-х ARIMA-моделей

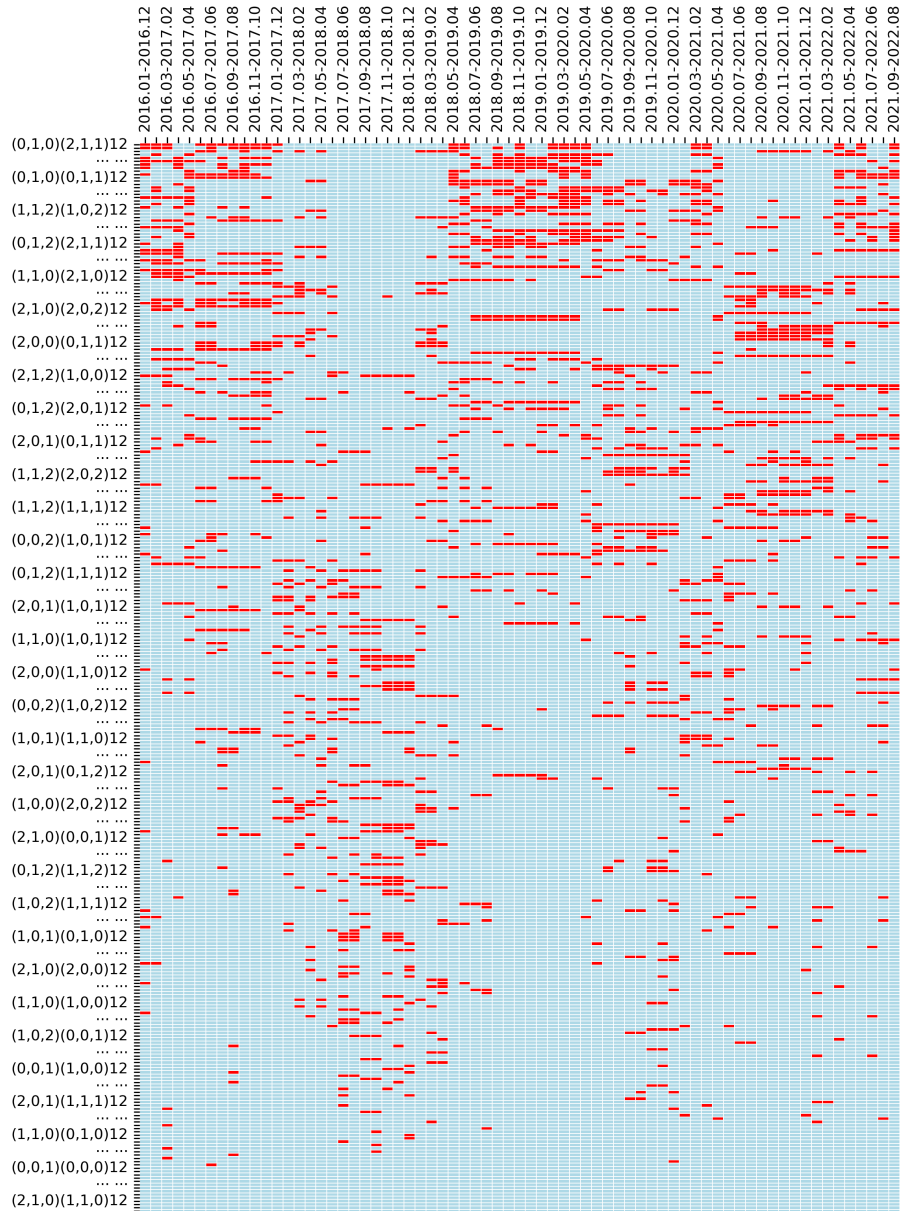


Рис. : Попадание ARIMA(p,d,q)(P,D,Q)s в 10% лучших моделей: ранжирование по RMSE

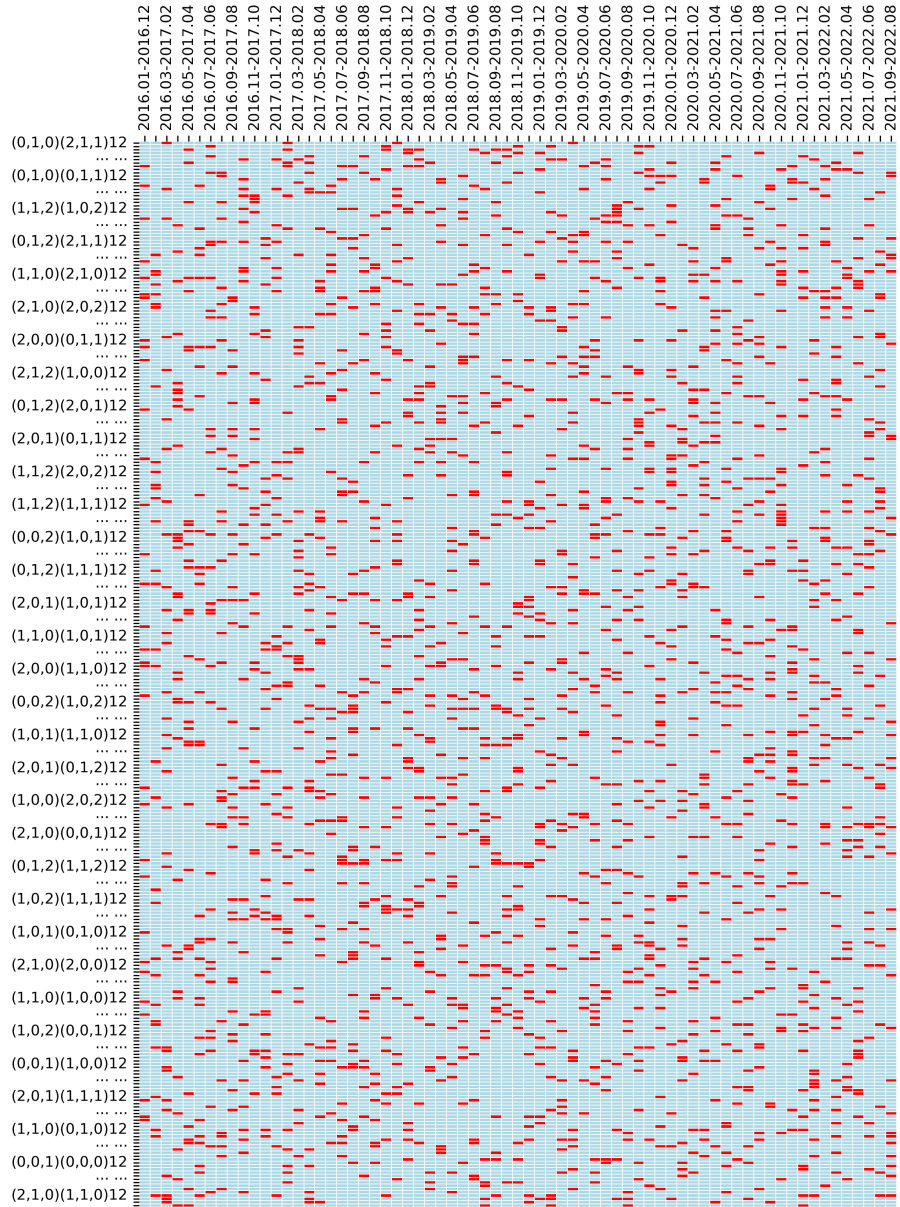


Рис. : Попадание ARIMA(p,d,q)(P,D,Q)s в 10% лучших моделей: случайное ранжирование

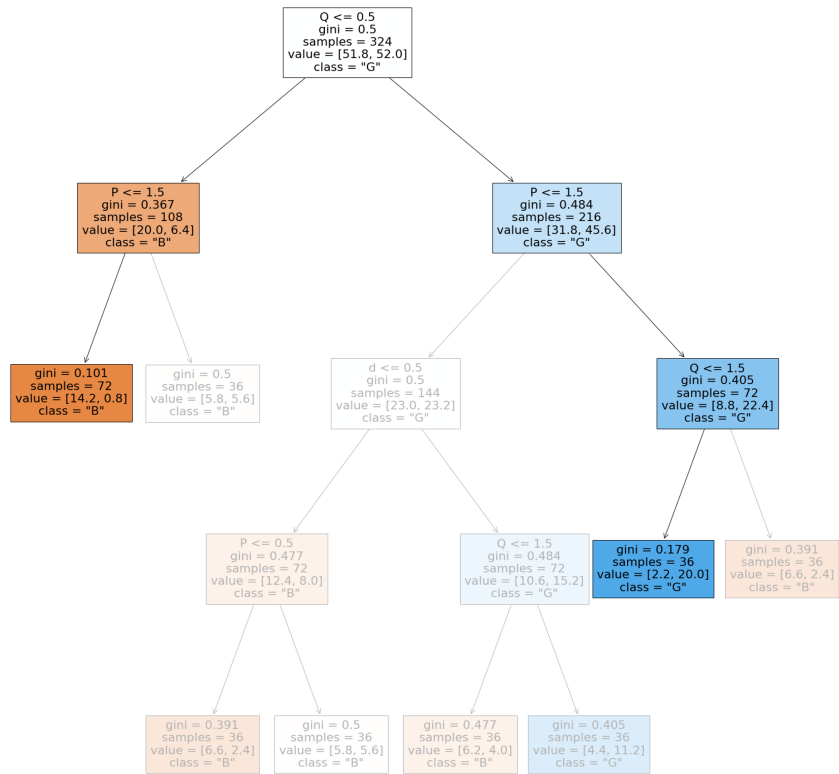


Рис. : Набор правил, разделяющих ARIMA-модели на две группы: «подходящие» и «не подходящие» для прогнозирования ИПЦ в Томской области

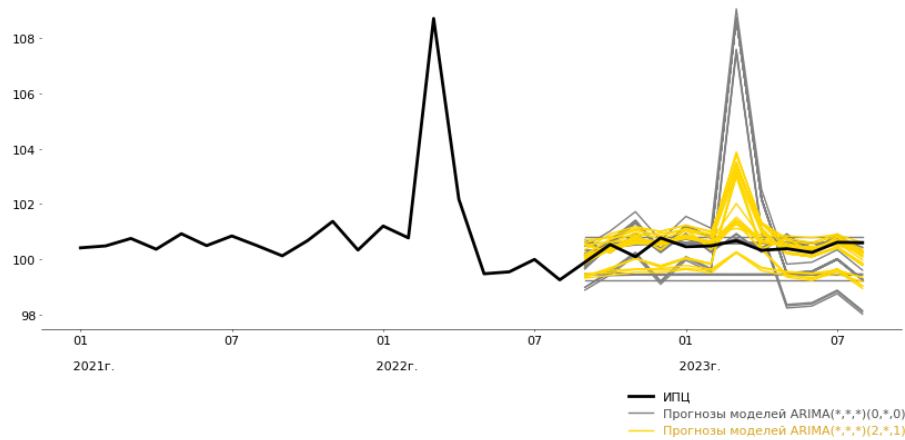


Рис. : Прогнозы «плохих» и «хороших» моделей на тестовый период: 36 моделей ARIMA(\*, \*,\*)(0,\*,0)12 и 36 моделей ARIMA(\*, \*,\*)(2,\*,1)12

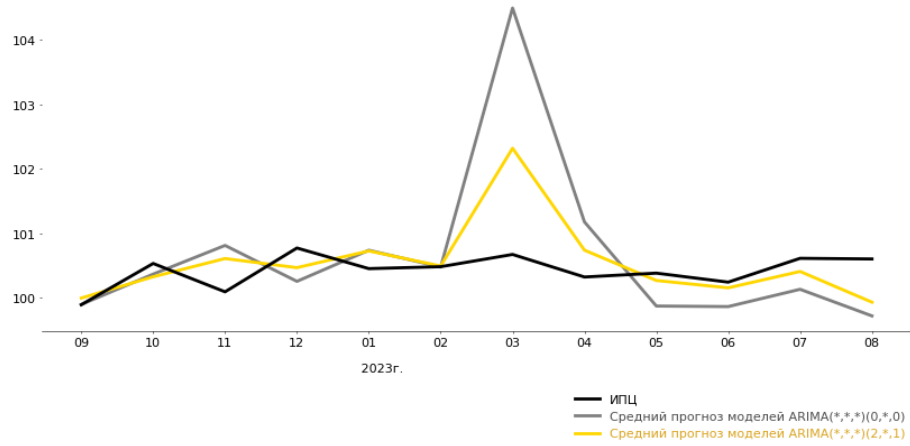


Рис. : Усредненные прогнозы «хороших» и «плохих» моделей

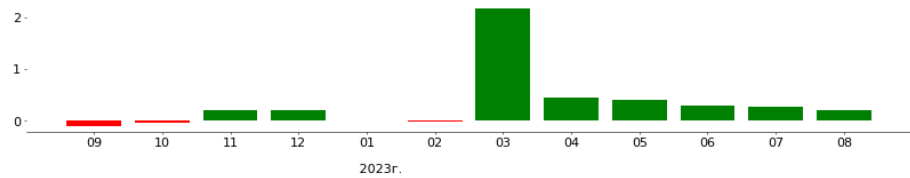


Рис. : Разности ошибок усредненных прогнозов: практически во всех периодах средний прогноз «хороших» моделей оказался более точным