



Национальный исследовательский университет  
«Высшая школа экономики»

Научная конференция  
«Соседи по науке»

Пермь, 2024

# ПРОГНОЗИРОВАНИЕ РЕГИОНАЛЬНОЙ ИНФЛЯЦИИ: ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ ИЛИ МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

**Авторы:**

**Татьяна Витальевна Букина,**

к.г.н., доцент, доцент департамента экономики и финансов НИУ ВШЭ – Пермь

**Дмитрий Викторович Кашин,**

к.э.н., доцент департамента экономики и финансов НИУ ВШЭ – Пермь

## Актуальность исследования

- Один из важнейших индикаторов состояния экономики – инфляция – рассматривается в работе на основе индекса потребительских цен (ИПЦ).
- Прогноз национальной инфляции важен не только для государственных структур, но и для бизнеса и домохозяйств – будь то решение об утверждении корпоративной стратегии или решение о плане крупных покупок.
- Прогнозирование региональной инфляции с одной стороны может повысить качество национального прогноза инфляции, с другой – поможет местным стейкхолдерам учесть сведения об инфляции в своей деятельности.

## Цель исследования

Определить модель, которая точнее остальных прогнозирует инфляцию в регионах ПФО.



Сравнить инструменты ML (метод опорных векторов, градиентный бустинг и случайный лес) с эконометрическими моделями временных рядов (AR, ARIMA).

## Эконометрика и прогнозирование инфляции

Взаимосвязь между уровнем безработицы и уровнем цен [Phillips, 1958; Samuelson, Solow, 1960]

Классификация моделей [Stock, Watson, 2008].  
4 группы: (1) ARMA, RW, USCV; (2) модели по кривым Филлипса; (3) модели инфл. ожиданий; (4) многофакторные модели

Объединение дезагрегированных прогнозов инфляции повышает качество общего прогноза инфляции [Hubrich, 2005]

Использование агрегированных и комбинированных прогнозов в России [Андреев, 2016; Байбуза, 2018].

Сложные модели прогнозирования инфляции имеют ошибку прогноза порой больше, чем модели, основанные только на временном ряде инфляции.

## ML и прогнозирование инфляции

Превосходство моделей машинного обучения над классическими эконометрическими моделями [Chakraborty, Joseph, 2017]:

- k-ближайших соседей (kNN)
- метод опорных векторов (SVM)
- метод дерева решений (DT)
- метод случайного леса (RF)
- гребневая регрессия (Ridge regression)
- модель нейронных сетей (NN)

На российских данных 2 статьи [Байбуза, 2018; Семитуркин, Шевелев, 2022]:

- метод опорных векторов (SVM)
- модель случайного леса (RF)
- модель градиентного бустинга (GB)
- модели с регуляризацией (LASSO и Elastic Net)
- модели RW и AR

**Методы машинного обучения позволяют, как минимум, строить прогнозы региональной инфляции, сопоставимые с прогнозами по эконометрическим моделям**

## Гипотезы исследования

1. Модели машинного обучения показывают превосходство над моделями временных рядов при построении прогнозов региональной инфляции.



2. Модели временных рядов позволяют точнее прогнозировать инфляцию на горизонте в 1 месяц, а на более длительных горизонтах планирования ошибка прогноза этих моделей больше, чем у прогнозов, построенных методами машинного обучения.



## Данные и методология

14 регионов Приволжского федерального округа.

Региональные и внешние макроэкономические показатели – ИПЦ, уровень безработицы, курс доллара США, ключевая ставка.

Выборка с января 2010 г. по декабрь 2022 г.

Источник данных – Единая межведомственная информационно-статистическая системы (ЕМИСС).

R Studio для прогнозирования временных рядов.

Python для моделей машинного обучения.

## Данные и методология

Методы семейства X-11 (X-12, X-12ARIMA и др.), а также TRAMO/SEATS [Пильник и др., 2015] не использовались, а брались фиктивные (дамми) переменные на номер месяца года. В случае с моделями машинного обучения использовался порядковый номер месяца года.

Прогнозирование в режиме псевдореального времени (т.к. одни показатели представлены на начало месяца, другие на конец месяца).

Уровень безработицы (квартальные данные) – интерполяция с сезонной корректировкой.

Натуральный логарифм к данным по инфляции, расширенный тест Дики – Фуллера => в первых разностях.

Горизонты прогнозирования в 1, 3, 6, 9, 12, 21 и 24 месяца, которые следуют за последним доступным месяцем в обучающей выборке.



## Данные и методология

Бенчмарк (базовая модель) – модель авторегрессии первого порядка AR(1).

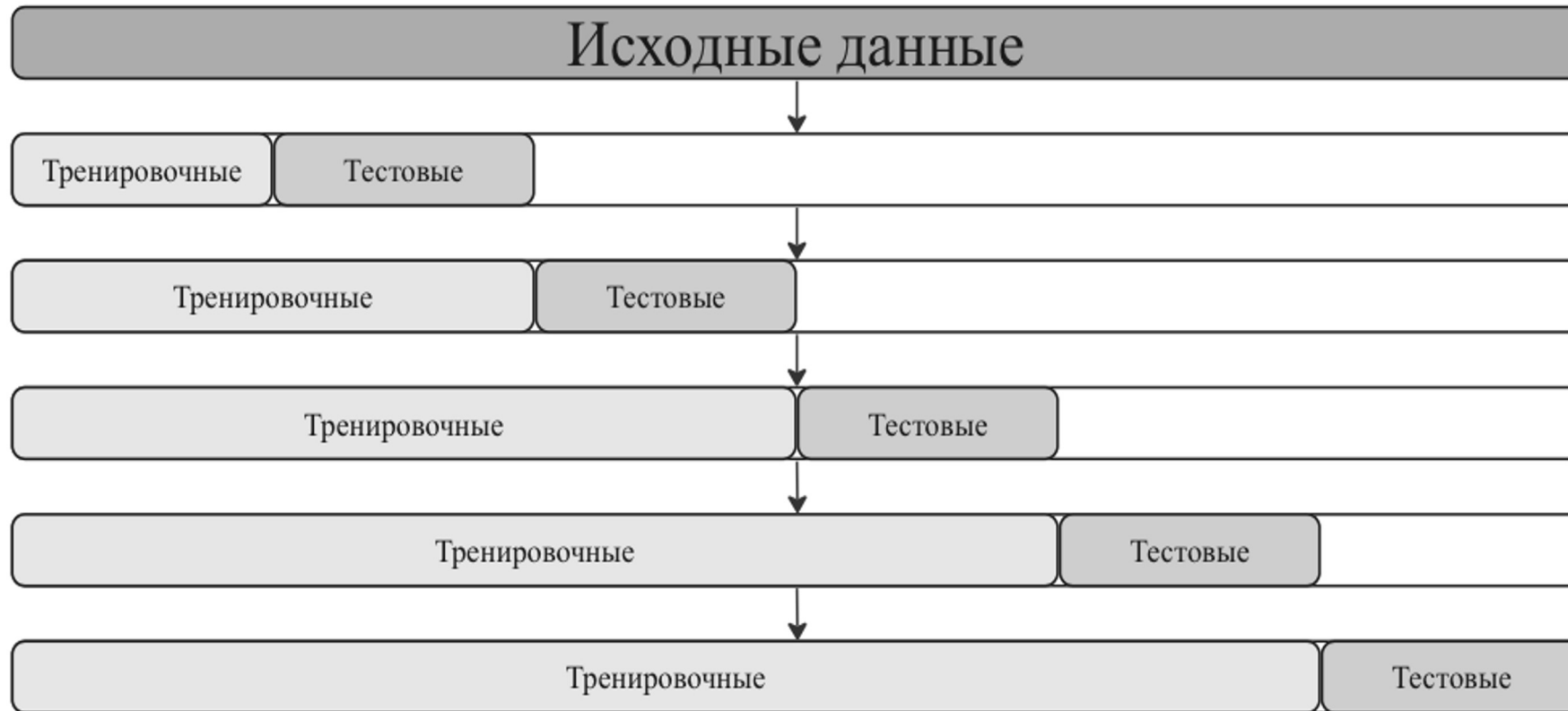
По каждой модели рассчитано отношение  $RMSE_{Model}$  к  $RMSE_{AR}$  (если  $> 1$ , то модель хуже AR).

Прогноз AR строится сразу на момент времени  $(t + k)$ . Также использован ARIMAX (с курсом доллара).

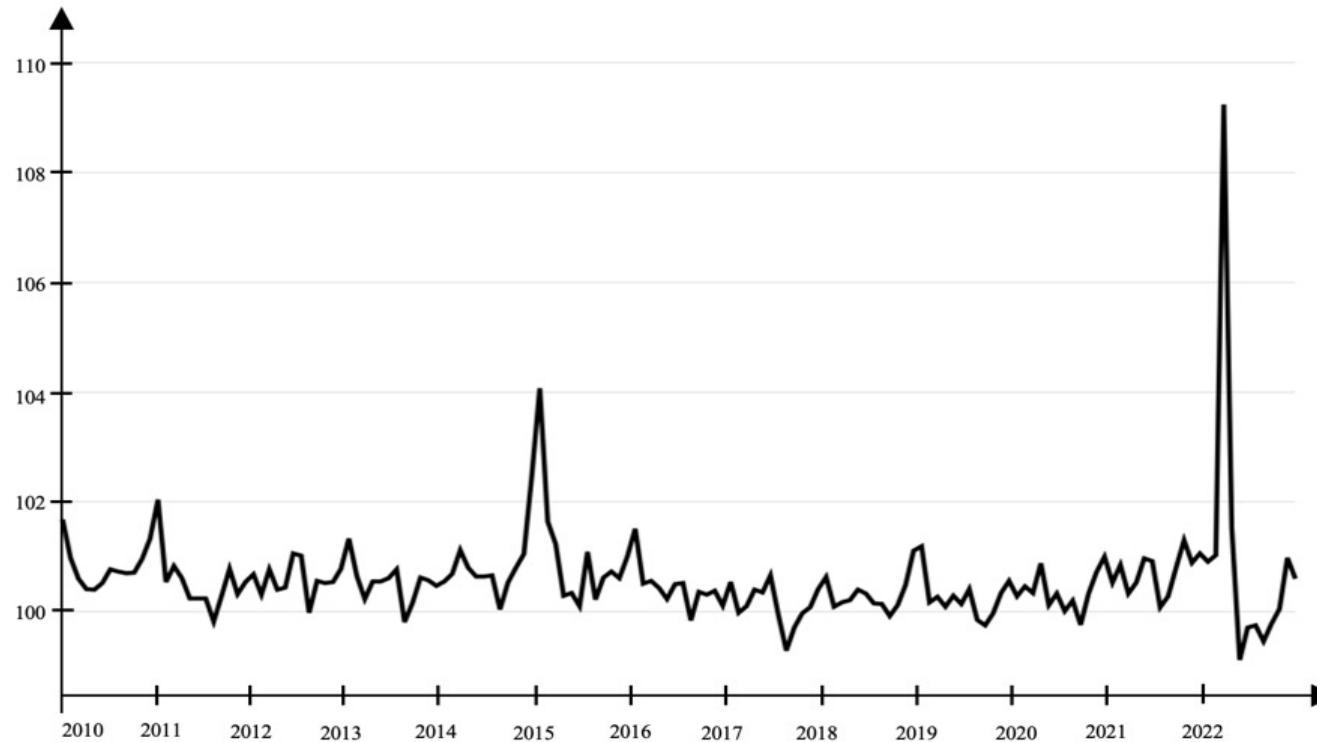
Модели ML – метод опорных векторов, градиентный бустинг и случайный лес.

Для подбора гиперпараметров использовали кросс-валидацию. Выбирали между двумя подходами: использовать тестовые выборки с одинаковым количеством наблюдений или использовать кросс-валидацию на скользящем окне.

## Кросс-валидация временных рядов



## Инфляция в Пермском крае (по месяцам)



Инфляция приведена к стационарному процессу через натуральный логарифм и первую разность в логарифмах для всех регионов в выборке.

## Модели ARIMAX по регионам ПФО

Регион	Спецификация ARIMAX(p,d, q)
Республика Башкортостан	(0,1,3)
Республика Марий Эл	(0,1,3)
Республика Мордовия	(5,1,0)
Республика Татарстан	(1,1,1)
Удмуртская Республика	(0,1,2)
Чувашская Республика	(1,1,1)
Пермский край	(5,1,0)
Кировская область	(1,1,1)
Нижегородская область	(1,1,1)
Оренбургская область	(0,1,2)
Пензенская область	(1,1,1)
Самарская область	(0,1,5)
Саратовская область	(1,1,1)
Ульяновская область	(3,1,2)

- Спецификация модели ARIMAX отличается от региона к региону.
- Можно выделить наиболее часто встречающуюся модификацию модели – ARIMAX (1, 1, 1).



## RMSEARIMAX / RMSEAR по регионам ПФО

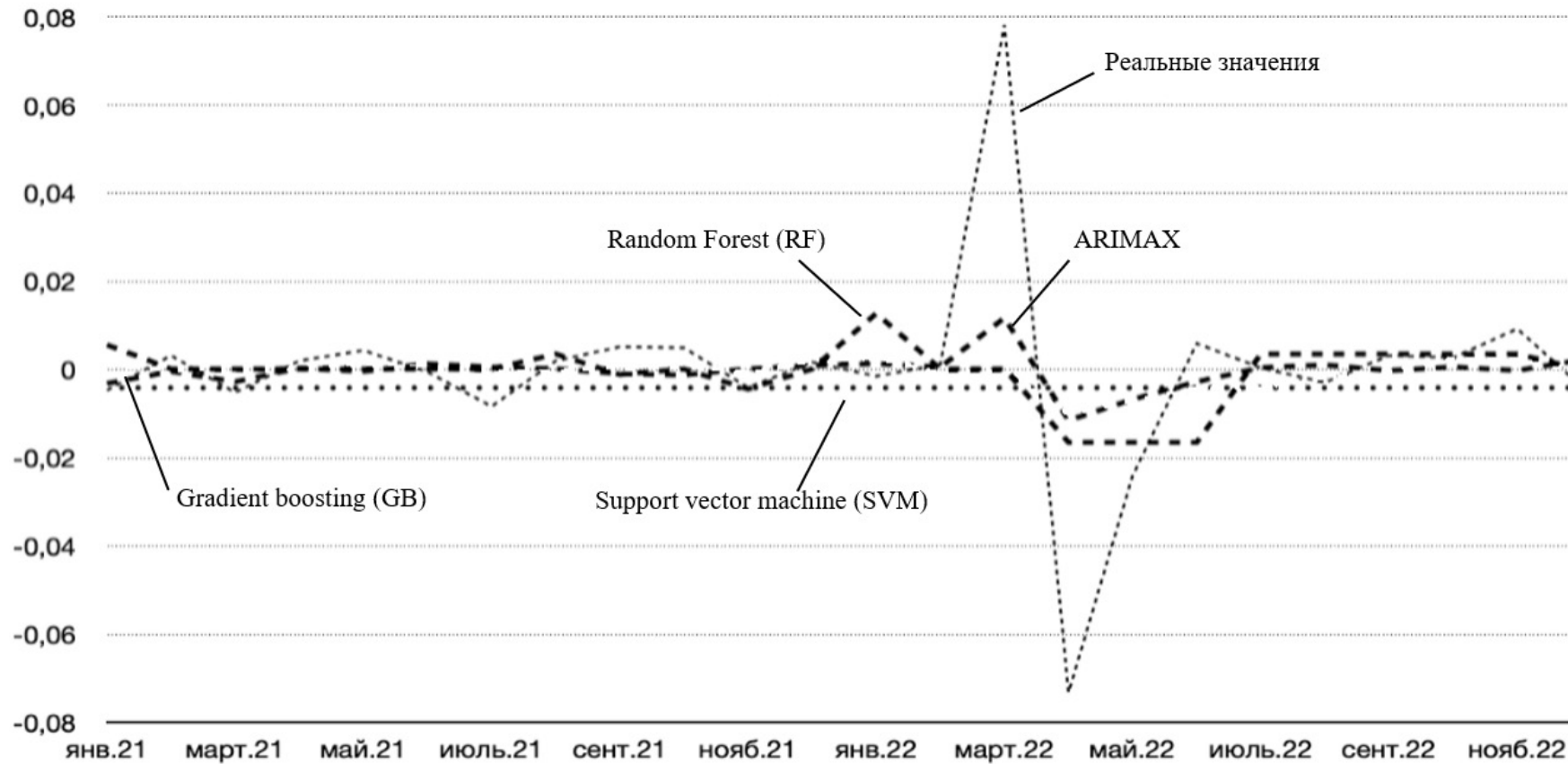
Регион \ горизонт прогноза (месяц)	1	3	6	9	12	21	24
Республика Башкортостан	0,036	0,891	0,868	0,994	0,991	0,903	0,903
Республика Марий Эл	0,334	0,839	0,874	0,985	1,008	0,869	0,869
Республика Мордовия	1,276	4,906	3,042	1,203	1,120	0,824	0,824
Республика Татарстан	0,222	1,003	1,038	1,010	1,010	0,948	0,948
Удмуртская Республика	0,038	0,775	0,778	0,946	1,042	0,851	0,852
Чувашская Республика	1,623	1,140	1,107	1,069	1,066	0,882	0,882
Пермский край	0,366	0,632	0,810	0,981	1,014	0,848	0,850
Кировская область	0,783	0,965	0,967	0,988	1,014	1,014	0,873
Нижегородская область	1,487	1,261	1,120	1,102	1,065	0,838	0,838
Оренбургская область	0,171	0,884	0,964	1,006	1,010	0,962	0,962
Пензенская область	0,434	0,811	0,991	1,006	1,030	0,882	0,882
Самарская область	0,016	0,778	0,788	0,982	1,013	0,925	0,925
Саратовская область	0,574	0,848	0,894	0,984	1,001	0,923	0,923
Ульяновская область	0,471	0,934	1,111	1,069	1,073	0,910	0,909
<b>Среднее значение</b>	<b>0,559</b>	<b>1,190</b>	<b>1,097</b>	<b>1,023</b>	<b>1,033</b>	<b>0,899</b>	<b>0,889</b>

## Выводы по $RMSE_{ARIMAX}$ / $RMSE_{AR}$ по регионам ПФО

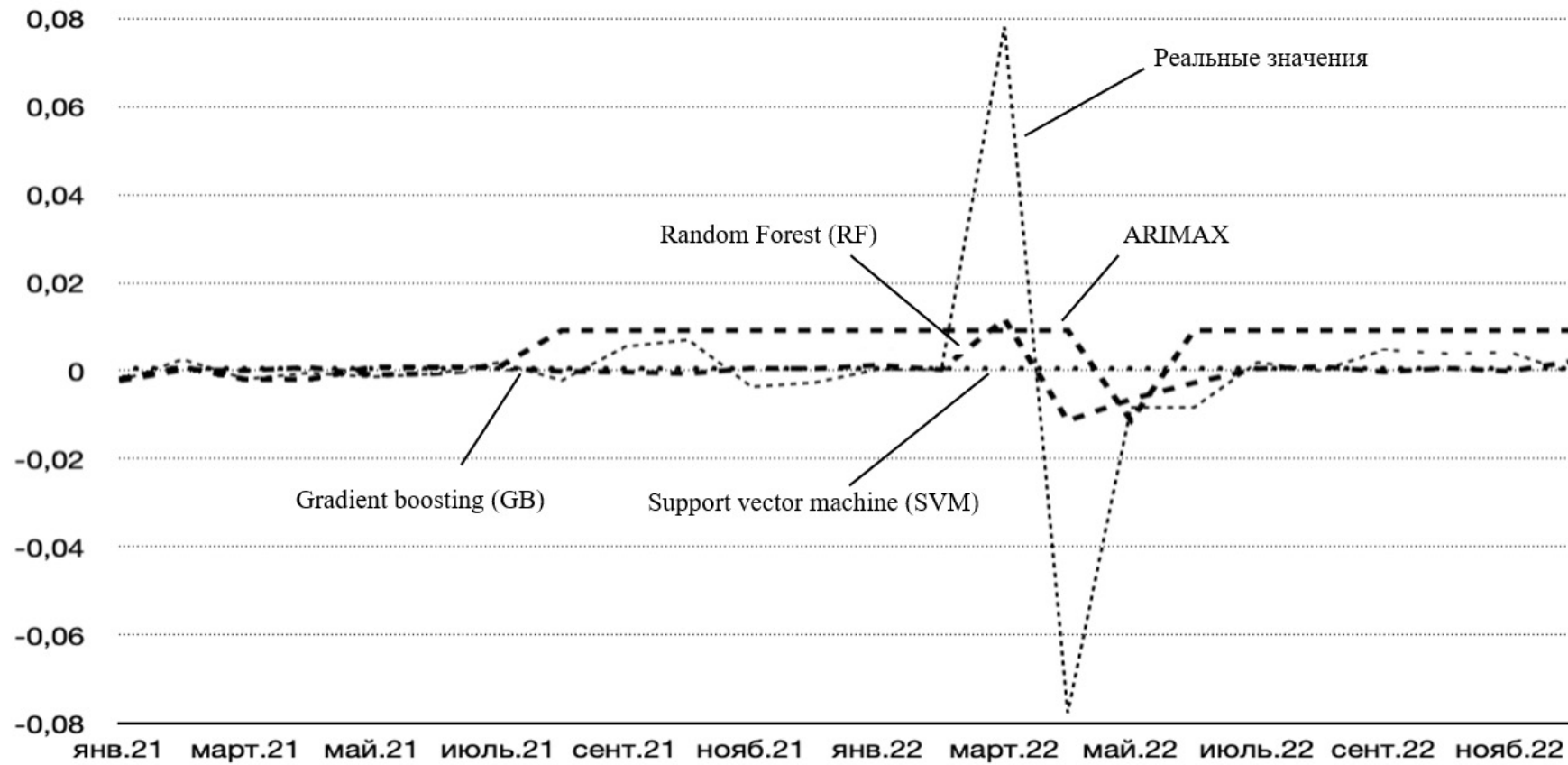
Модель ARIMAX в среднем превосходит модель-бенчмарк AR(1) как на достаточно коротких горизонтах планирования (1 месяц), так и на длинных временных промежутках в 21 и 24 месяца:

- На длинных временных горизонтах предыдущей инфляции в периодах  $t-24$  и  $t-21$  перестает оказывать влияние на текущую инфляцию и возрастает роль курса доллара.
- Курс доллара достаточно быстро реагирует на изменения в экономике, а значит, основываясь на его динамике, можно строить более точный прогноз краткосрочной региональной инфляции.

## Инфляция и ее прогноз для Пермского края

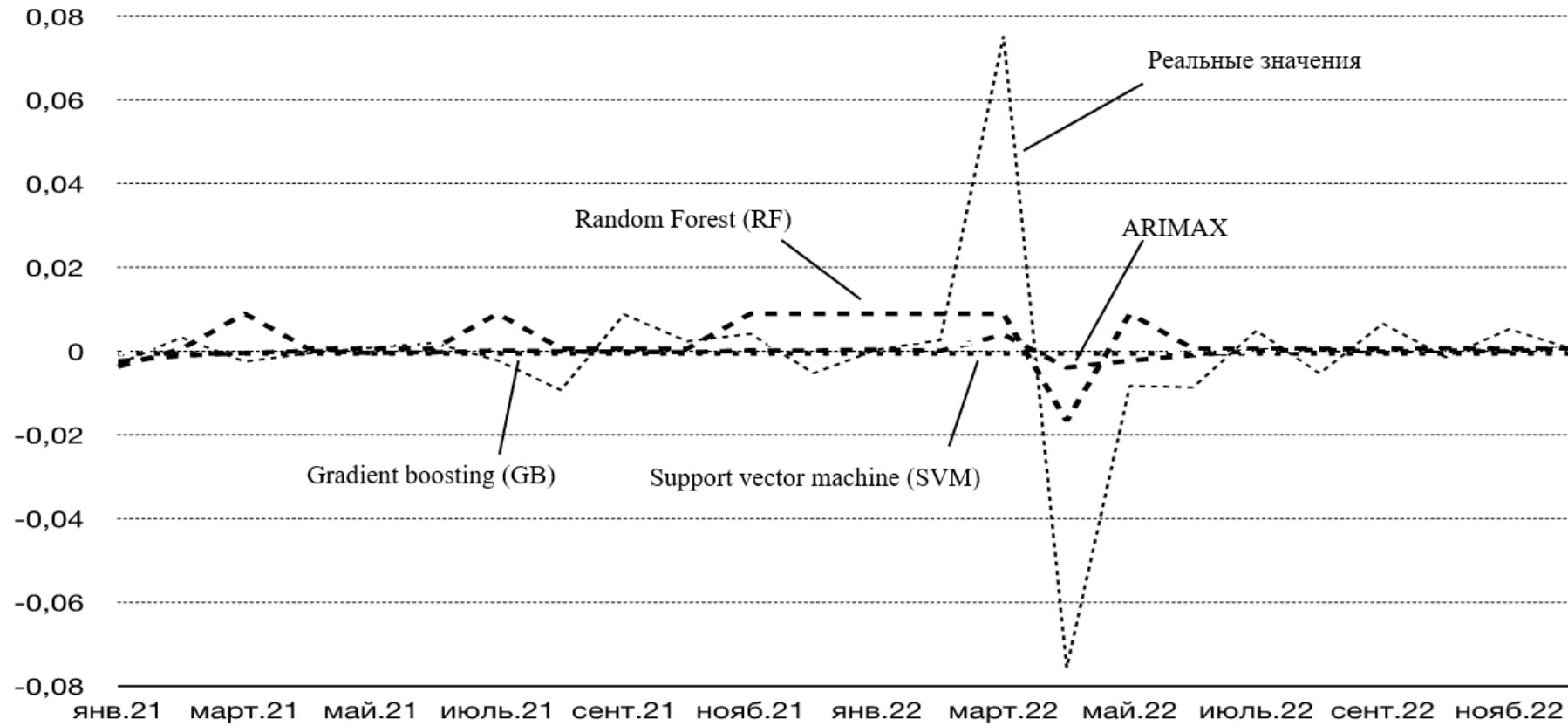


## Инфляция и ее прогноз для Удмуртской Республики

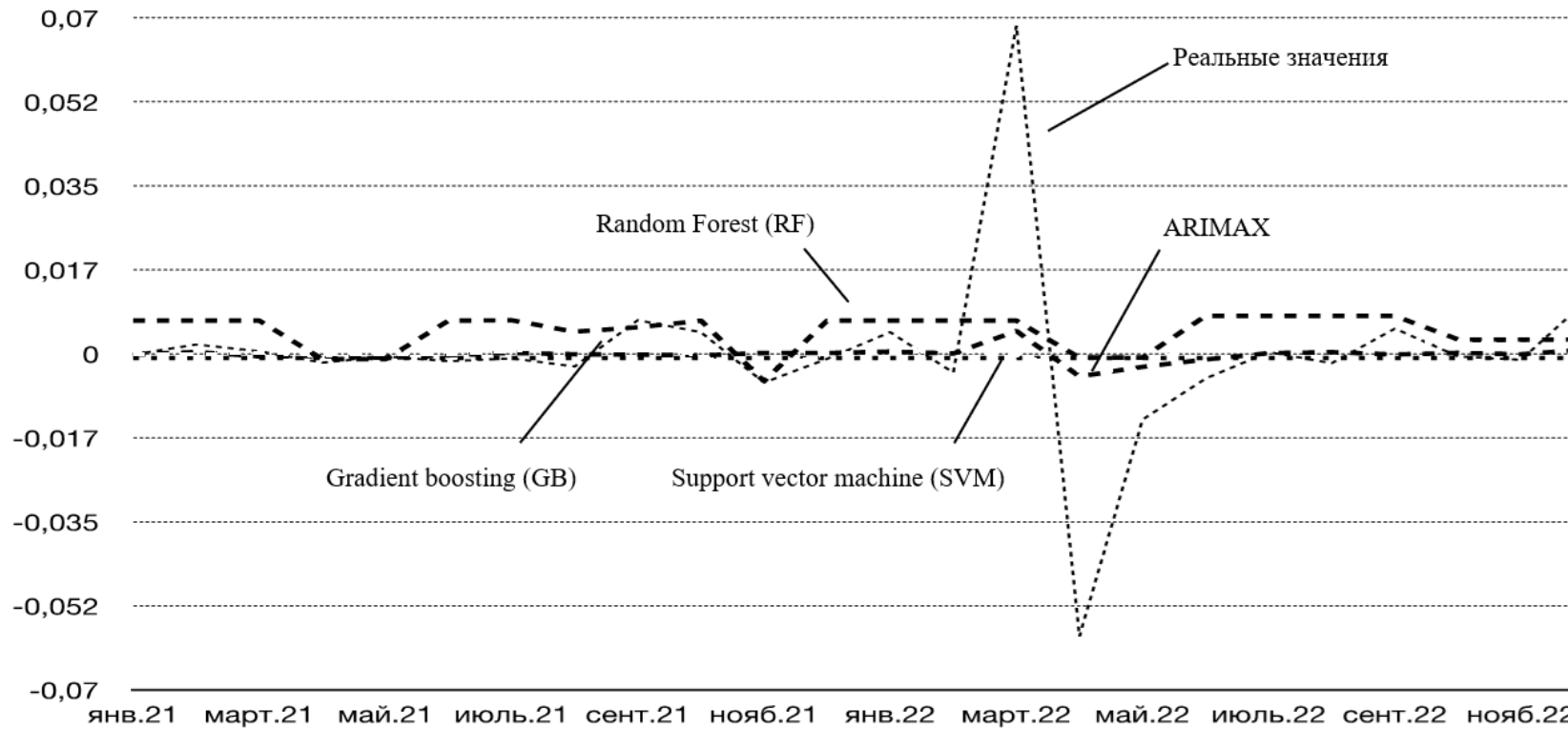




## Инфляция и ее прогноз для Республики Татарстан



## Инфляция и ее прогноз для Самарской области



## Результаты по прогнозам

1. По каждому региону, как и в случае со сравнением эконометрических моделей, рассчитаны относительные показатели  $RMSE_{Model} / RMSE_{AR}$  и среднее значение по всем регионам.
2. Модель градиентного бустинга демонстрирует более точные прогнозы как на коротких и средних периодах в 1, 3 и 6 месяцев, так и на длинных периодах в 21 и 24 месяца.
3. Метод опорных векторов выдает линейный прогноз с постоянно повторяющимся значением, однако прогнозы на 21 и 24 месяца оказались лучше, чем прогнозы по AR(1)

## Важность (частота) факторов при моделировании

Случайный лес		Градиентный бустинг	
Фактор	Частота	Фактор	Частота
Инфляция t-1	14	Инфляция t-1	14
Среднее значение инфляции t-3	4	Месяц	4
Месяц	4	Среднее значение инфляции t-3	3
Стандартное отклонение инфляции t-3	2	Инфляция t-2	2
Стандартное отклонение ключевой ставки t-9	2	Ключевая ставка t-12	2

В модели случайного леса сезонность инфляции характерна только для 4 регионов: Пермский край, Нижегородская область, Пензенская область и Саратовская область.

В модели градиентного бустинга номер месяца важен в Республике Татарстан.

Среднее значение инфляции за три прошедших месяца достаточно значимый фактор в моделях градиентного бустинга по: Республике Мордовия, Нижегородской области и Ульяновской области.

## Усредненные результаты оценки качества моделей по регионам ПФО

Средние значения $RMSE_{Model} / RMSE_{AR}$							
Период прогноза (месяц)	1	3	6	9	12	21	24
ARIMAX	0,559	1,190	1,097	1,023	1,033	0,899	0,889
RF	2,207	1,556	1,587	1,023	1,464	0,226	0,220
GB	0,999	0,797	0,838	1,069	1,035	0,275	0,223
SVM	2,940	1,680	1,656	2,060	1,046	0,389	0,271

Модель градиентного бустинга для прогнозов на 3, 6, 21 и 24 месяца превосходит бенчмарк, на горизонте в 1 месяц показывает качество сравнимое с бенчмарком.

Модель случайного леса показывает наиболее точные прогнозы (наименьшее относительное  $RMSE_{Model} / RMSE_{AR}$ ) инфляции на горизонтах в 21 и 24 месяца относительно всех остальных моделей и наихудшее качество в прогнозах на 12 месяцев.

## Выводы и ключевые результаты

- ✓ **Модели машинного обучения** могут эффективно прогнозировать инфляцию на более длительных временных периодах (21 и 24 месяца), в то время как **классические эконометрические модели AR и ARIMAX** показывают низкое качество прогнозов на таких дальних горизонтах планирования;
- ✓ Результаты соответствуют практике прогнозирования макроэкономических показателей, которая свидетельствует о том, что **для среднесрочных прогнозов** обычно **рекомендуется использовать либо структурные модели (например, DSGE), либо их полуструктурные аналоги (например, QPM)**;
- ✓ Наиболее точной оказалась **модель градиентного бустинга** (превосходит модель-бенчмарк AR(1) в 4-х из 7 рассматриваемых периодов), и немного уступает только при прогнозах на 9 и 12 месяцев.

## Наблюдения и ограничения

- ✓ Прогнозный период, 2021–2022 гг., является относительно нетипичным для российской и мировой экономики;
- ✓ Доступный набор макроэкономических показателей может быть расширен в дальнейших исследованиях;
- ✓ Качество данных может вызывать вопросы как из-за пересмотра расчета показателей в разные временные промежутки, обновления ранее представленных данных в конце года, так и специфики отдельно взятого региона;
- ✓ Ограничение в вычислительной мощности, подбор гиперпараметров в моделях машинного обучения может был выполнен с различными значениями самих гиперпараметров.
- ✓ Развитием исследования может стать использование нейронных сетей, так и комбинирование прогнозов по нескольким моделям в один прогноз.

Спасибо за внимание!