



Munich Personal RePEc Archive

Construction of short-term forecasts of socio-economic indicators of Russia based on the consistent application of ranked models

, and , and ,

..

9 January 2025

Online at <https://mpra.ub.uni-muenchen.de/123093/>
MPRA Paper No. 123093, posted 18 Jan 2025 21:16 UTC

Построение краткосрочных прогнозов социально-экономических показателей России на основе последовательного применения ранжированных моделей

Савинова Виктория Михайловна

**ФГБОУ ВО «РЭУ им. Г.В. Плеханова», Москва, Россия,
savinova.vm@rea.ru, <https://orcid.org/>**

Ярушев Сергей Александрович

**[ФГБОУ ВО «РЭУ им. Г.В. Плеханова», Москва, Россия,
savinova.vm@rea.ru, https://orcid.org/](https://orcid.org/)**

Иксанов Владислав Рашидович

**[ФГБОУ ВО «РЭУ им. Г.В. Плеханова», Москва, Россия,
savinova.vm@rea.ru, https://orcid.org/](https://orcid.org/)**

Аннотация. Авторы приводят методику построения моделей краткосрочного прогнозирования с использованием последовательного применения математических моделей, таких как ARIMA, множественная линейная регрессия, регрессионное дерево решений и случайный лес. Модели в рамках статьи ранжированы в порядке возрастания их вычислительной сложности и в порядке уменьшения их устойчивости и интерпретируемости. В качестве инструмента моделирования выступает разработанная сотрудниками ФГБОУ ВО «РЭУ им. Г.В. Плеханова» информационно-аналитическая система «Горизонт».

Ключевые слова: краткосрочное прогнозирование, социально-экономические показатели, эконометрические модели.

Construction of short-term forecasts of socio-economic indicators of Russia based on the consistent application of ranked models

Savinova Victoria Mikhailovna

**Plekhanov Russian University of Economics, Moscow, Russia,
savinova.vm@rea.ru, <https://orcid.org/>**

Yarushev Sergey Aleksandrovich

**Plekhanov Russian University of Economics, Moscow, Russia,
savinova.vm@rea.ru, <https://orcid.org/>**

Iksanov Vladislav Rashidovich

Plekhanov Russian University of Economics, Moscow, Russia,
savinova.vm@rea.ru, <https://orcid.org/>

Abstract. The authors present a methodology for constructing short-term forecasting models using the sequential application of mathematical models such as ARIMA, multiple linear regression, regression decision tree and random forest. The models in the article are ranked in order of increasing computational complexity and in order of decreasing stability and interpretability. The modeling tool is the information and analytical system "Horizon" developed by the staff of the Plekhanov Russian University of Economics.

Keywords: short-term forecasting, socio-economic indicators, econometric models.

Введение

В настоящее время разработано множество методов и моделей краткосрочного прогнозирования экономических показателей. Это связано с высокой степенью актуальности проблемы краткосрочного прогнозирования экономических показателей, которая обусловлена следующими причинами:

1. Быстрая реакция на изменения: Краткосрочные прогнозы позволяют руководителям быстро реагировать на изменения в рыночной среде, спросе и предложении, что может помочь избежать потерь.
2. Планирование ресурсов: краткосрочное прогнозирование помогает в эффективном распределении ресурсов, управлении запасами и оптимизации производственных процессов.
3. Финансовое планирование: в финансовом секторе краткосрочные прогнозы помогают инвесторам и трейдерам принимать обоснованные решения, основываясь на ожидаемых колебаниях цен.
4. Управление рисками: краткосрочные прогнозы позволяют лучше оценивать риски и разрабатывать стратегии их минимизации.
5. Адаптация к сезонным колебаниям: в некоторых отраслях,

таких как сельское хозяйство или розничная торговля, краткосрочное прогнозирование помогает учитывать сезонные изменения и планировать соответствующие действия.

Однако, показатели социально-экономической сферы России характеризуются высокой степенью изменчивости тенденций. Такой характер изменения динамики исследуемых показателей приводит к возможности использования малых объемов данных и наличия сложных взаимосвязей между показателями.

В качестве примера частого изменения тенденций рассмотрим график динамики темпов роста реальных располагаемых денежных доходов населения России.

Я подтверждаю, что имею право внести вклад в MPRA

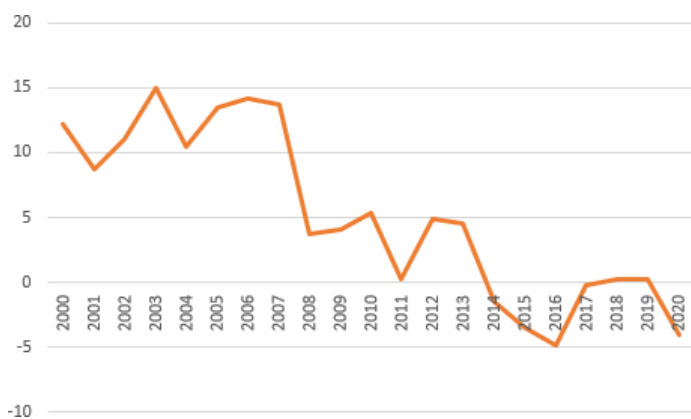


Рис. 1 Динамика темпов роста реальных располагаемых денежных доходов населения за период 2000 — 2020

На рисунке 2 видно, что в коротком временном интервале наблюдаются изменения тенденций процесса. С 2000 по 2007 год темпы роста реальных располагаемых денежных доходов населения стабилизировались благодаря экономическому росту после кризиса 1998 года, что способствовало увеличению занятости и зарплат. Однако в период экономического кризиса 2008–2011 годов темпы роста значительно замедлились.

С 2014 по 2016 год наблюдается снижение темпов роста доходов из-за ухудшения экономических условий, вызванного падением цен на нефть, ростом инфляции и ослаблением рубля, а также западными санкциями. В 2017–2019 годах доходы немного увеличились благодаря улучшению экономической ситуации, росту цен на нефть и снижению инфляции.

С 2020 года доходы снова начали снижаться из-за неблагоприятных политических и экономических условий, а также пандемии коронавируса.

Частая смена тенденций, вызванная нестабильностью социально-экономических условий, затрудняет краткосрочное прогнозирование и может негативно сказаться на точности моделей.

Описанные особенности приводит к невозможности использования единственного типа моделей прогнозирования. В связи с этим актуальным становится последовательное применение нескольких классов моделей к

исследуемым данным и выбора из них наиболее адекватной рассматриваемым процессам.

В рамках данной работы приводится краткосрочное прогнозирование на примере 5 показателей социально-экономической сферы России, включая показатели «Среднемесячная номинальная начисленная заработная плата в РФ, руб.», «Расходы населения на оплату обязательных платежей и взносов, % от ВВП», «Валовая прибыль экономики, % от ВВП», «Безработные, зарегистрированные в государственных учреждениях службы занятости, тыс. человек», «Количество занятых, тыс. человек», с использованием последовательного применения моделей временных рядов (ARIMA), множественной линейной регрессии, деревьев решений и случайного леса с целью получения приемлемых результатов. Для проведения моделирования были отобраны поквартальные данные, предоставляемые Федеральной службой государственной статистики за период с 2013 по 2021. Прогноз был построен на 2022 год с целью сравнения фактических данных и результатов краткосрочного прогнозирования.

Методика

В рамках данной работы краткосрочное прогнозирование исследуемых процессов производилось с использованием разработанной сотрудниками ФГБОУ ВО «РЭУ им. Г.В. Плеханова» информационно-аналитической системы «Горизонт». Данная система позволяет проводить краткосрочное прогнозирование временных рядов с использованием различных классов моделей.

На первом этапе необходимо провести ранжирование моделей в порядке возрастания их вычислительной сложности, уменьшения их интерпретируемости и устойчивости.

На основании данных из документации библиотеки Scikit-learn языка программирования python, которая использовалась при разработке информационно-аналитической системы «Горизонт», изменение

вычислительной сложности заявленных в данной работе методов краткосрочного прогнозирования можно представить в виде графика (рис 2).

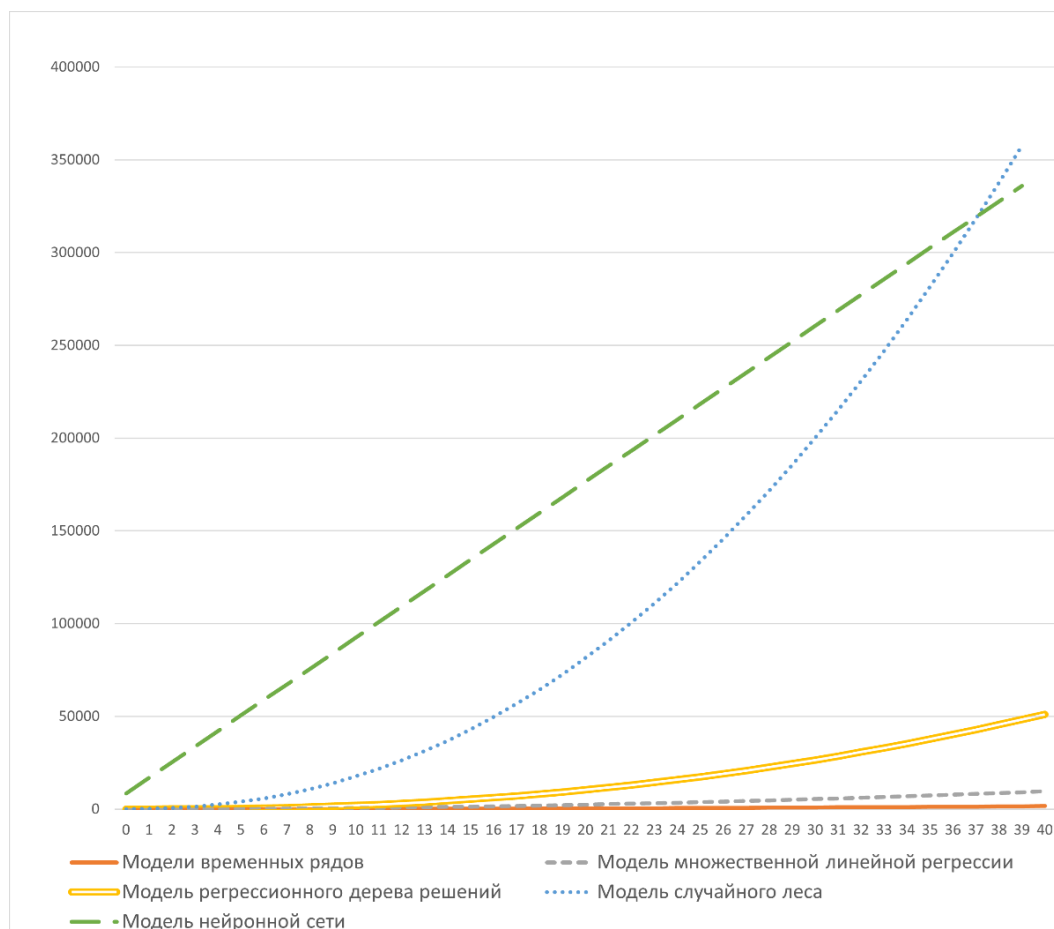


Рис.2 Изменение вычислительной сложности краткосрочного прогнозирования временных рядов

Как видно из рисунка 2, модели временных рядов имеют самую низкую вычислительную сложность, модель случайного леса — самую высокую.

Далее на основании публикаций авторов и опыта построения прогноза авторов модели были экспертно ранжированы согласно уменьшению их интерпретируемости и устойчивости.

В результате ранжирования, модели в рамках данной работы будут применяться в следующем порядке: модели временных рядов, множественная линейная регрессия, регрессионное дерево решений, случайный лес.

Рассмотрим представленные модели более подробно.

Модели временных рядов — это статистические методы, используемые для анализа данных, собранных в последовательности по времени. Они помогают

выявлять закономерности, тренды и сезонные колебания, а также делать прогнозы на основе исторических данных.

ARIMA (АвтоРегрессионная Интегрированная Модель Скользящей Средней) — это популярная статистическая модель, используемая для анализа и прогнозирования временных рядов. Она сочетает в себе три основных компонента:

1. AR (Авторегрессия): Этот компонент описывает зависимость текущего значения временного ряда от его предыдущих значений.

Например, AR(1) означает, что текущее значение зависит от одного предыдущего значения.

2. I (Интегрирование): Этот компонент используется для достижения стационарности временного ряда. Стационарный ряд имеет постоянное среднее и дисперсию во времени. Интегрирование включает в себя вычитание предыдущих значений из текущих, что помогает устранить тренды.

3. MA (Скользящая средняя): Этот компонент описывает зависимость текущего значения от случайных ошибок (остатков) предыдущих периодов. MA означает, что текущее значение зависит от одной предыдущей ошибки.

Общий вид модели временных рядов выглядит следующим образом:

$$x_t = \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}$$

где $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$ — коэффициенты уравнения, $\varepsilon_t, \dots, \varepsilon_{t-q}$ — текущее и прошедшие значения ошибок, x_t, \dots, x_{t-p} — значения временного ряда, приведенного к стационарному виду после взятия конечных разностей.

Модель ARIMA относительно проста в понимании и реализации, что делает её доступной для анализа временных рядов. Однако ARIMA не подходит для данных с нелинейными зависимостями, что ограничивает её применение в некоторых случаях.

Модели множественной линейной регрессии — это статистические методы, используемые для анализа взаимосвязи между одной зависимой

переменной и несколькими независимыми переменными. Они позволяют оценить, как изменения в независимых переменных влияют на зависимую переменную.

Основные компоненты модели множественной линейной регрессии:

1. Зависимая переменная (Y): Это переменная, которую мы пытаемся предсказать или объяснить.
2. Независимые переменные (X_1, X_2, \dots, X_n): Это переменные, которые используются для предсказания зависимой переменной.
3. Коэффициенты ($\beta_0, \beta_1, \beta_2, \dots, \beta_n$): Эти параметры модели показывают, как изменение каждой независимой переменной влияет на зависимую переменную.

Уравнение множественной линейной регрессии можно записать следующим образом:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

где:

- Y — зависимая переменная,
- β_0 — свободный член (константа),
- $\beta_1, \beta_2, \dots, \beta_n$ — коэффициенты регрессии,
- X_1, X_2, \dots, X_n — независимые переменные,
- ε — ошибка модели (остаток).

Модель легко интерпретировать, так как коэффициенты показывают, как изменение независимых переменных влияет на зависимую. Однако данная модель накладывает жесткие ограничения на остатки модели и не позволяют получить качественный результат краткосрочного прогнозирования при отсутствии линейных зависимостей в данных.

Модель регрессионного дерева решений — это метод машинного обучения, используемый для предсказания непрерывных значений. Она основана на концепции дерева решений, где данные разбиваются на подгруппы на основе значений независимых переменных. Каждая ветвь дерева представляет собой условие, а листья дерева содержат предсказания для зависимой переменной.

Преимущество данного метода заключается в высокой степени интерпретируемости. Однако данные модели склонны к переобучению, что делает процесс краткосрочного прогнозирования более трудоемким.

Модель случайного леса (Random Forest) — это метод машинного обучения, который используется как для классификации, так и для регрессии. Он основан на ансамблевом подходе, который объединяет множество деревьев решений для улучшения точности и устойчивости предсказаний.

Данная модель обладает высокой эффективностью аппроксимации данных, имеющие нелинейные зависимости. Однако у данной модели высокая вычислительная сложность и слабая степень интерпретации.

После отбора и ранжирования моделей был произведен расчет прогнозов. Для полученных результатов была проведена верификация моделей по отобранным авторами критериям. Критерии верификации для используемых моделей представлены в таблице 1.

Табл.1 Перечень критериев верификации построенных моделей

Наименование модели	Критерии верификации
Модели временных рядов	Коэффициент детерминации, средняя относительная ошибка прогноза
Множественная линейная регрессия	Коэффициент детерминации, средняя относительная ошибка прогноза, критерий Фишера, критерий Дарбина - Уотсона
Регрессионное дерево решений	Коэффициент детерминации, средняя относительная ошибка прогноза
Случайный лес	Коэффициент детерминации, средняя относительная ошибка прогноза

Представленные критерии позволяют оценить точность аппроксимации (коэффициент детерминации), качества полученных моделей (критерии Фишера и Дарбина — Уотсона), а также эмпирическую точность моделей (средняя относительная ошибка). Модель считается релевантной исследуемому процессу, если коэффициент детерминации больше, чем 0.8, средняя

относительная ошибка не превышает 10%, а для линейной регрессии коэффициент детерминации должен быть более 1,5 но менее 2,7, коэффициент Фишера больше 0,5.

Результаты

«Среднемесячная номинальная начисленная заработная плата в РФ, руб.», «Расходы населения на оплату обязательных платежей и взносов, % от ВВП», «Валовая прибыль экономики, % от ВВП», «Безработные, зарегистрированные в государственных учреждениях службы занятости, тыс. человек, «Количество занятых, тыс. человек»

В рамках данной работы были проведены экспериментальные расчеты для 5 показателей социально-экономической сферы, взятые за период с 2013 — 2022 гг. Для все 5 показателей были построены модели временных рядов. Результаты построения моделей временных рядов представлены в таблице 2.

Таблица 2. Результаты построения моделей временных рядов для исследуемых показателей

Наименование показателя	Значение коэффициента детерминации	Средняя относительная ошибка
«Среднемесячная номинальная начисленная заработная плата в РФ, руб.»	0.99	0.02
«Расходы населения на оплату обязательных платежей и взносов, % от ВВП»	0.3	0.12
«Валовая прибыль экономики, % от ВВП»	0.5	0.2
«Безработные, зарегистрированные в государственных	0.2	0.22

учреждениях службы занятости, тыс. человек		
«Количество занятых, тыс. человек»	0.7	0.1

Как видно из таблицы 2, только для одного показателя, а именно «Среднемесячная номинальная начисленная заработная плата в РФ, руб.».

Для остальных 4 показателей были построены модели множественной линейной регрессии. Факторы были отобраны из более чем 20 социально-экономических показателей посредством применения корреляционного анализа. Фактор принимался в уравнение в случае, если коэффициент корреляции по модулю больше 0.7.

Результаты построения моделей множественной линейной регрессии представлены в таблице 3.

Таблица 3. Результаты построения моделей множественной линейной регрессии для исследуемых показателей

Наименование показателя	Значение коэффициента детерминации	Средняя относительная ошибка	Критерий Дарбина Уотсона	Критерий Фишера
«Расходы населения на оплату обязательных платежей и взносов, % от ВВП»	0.8	0.07	2.1	7.3
«Валовая прибыль экономики, % от ВВП»	0.6	0.3	1.9	5
«Безработные, зарегистрированные в	0.4	0.3	2.5	7

государствен ных учреждениях службы занятости, тыс. человек				
«Количество занятых, тыс. человек»	0.6	0.1	1.4	7

Как видно из таблицы 3, только для одного показателя, а именно «Расходы населения на оплату обязательных платежей и взносов, % от ВВП».

Для оставшихся 3 показателей была построена модель регрессионного дерева решений. Результаты моделирования представлены в таблице 4.

Таблица 4. Результаты построения моделей регрессионного дерева для исследуемых показателей

Наименование показателя	Значение коэффициента детерминации	Средняя относительная ошибка
«Валовая прибыль экономики, % от ВВП»	0.9	0.1
«Безработные, зарегистрированные в государственных учреждениях службы занятости, тыс. человек»	0.7	0.15
«Количество занятых, тыс. человек»	0.8	0.13

Как видно из таблицы 4, только для одного показателя, а именно «Валовая прибыль экономики, % от ВВП».

Для оставшихся 3 показателей была построена модель случайного леса. Результаты моделирования представлены в таблице 5.

Таблица 4. Результаты построения моделей случайного леса для исследуемых показателей

Наименование показателя	Значение коэффициента детерминации	Средняя относительная ошибка
«Безработные, зарегистрированные в государственных учреждениях службы занятости, тыс. человек»	0.8	0.03
«Количество занятых, тыс. человек»	0.8	0.03

Таким образом, последовательное применение моделей к исследуемым данным позволило получить релевантные модели для всех рассматриваемых процессов.

Выводы

В рамках данной работы были отобраны модели краткосрочного прогнозирования социально-экономических показателей, включая модели временных рядов, модели множественной линейной регрессии, регрессионного дерева решений и случайного леса. Порядок моделей был обусловлен показателем вычислительной сложности, устойчивости и интерпретируемости. В результате применения рассмотренной методики удалось получить модели, которые показали адекватные результаты и могут быть использованы для краткосрочного прогнозирования исследуемых процессов.

Данное исследование выполнено в рамках государственного задания в сфере научной деятельности Министерства науки и высшего образования Российской Федерации на тему «Модели, методы и алгоритмы искусственного интеллекта в задачах экономики для анализа и стилизации многомерных данных, прогнозирования временных рядов и проектирования рекомендательных систем», номер проекта FSSW-2023-0004

This study was carried out within the framework of the state assignment in the field of scientific activity of the Ministry of Science and Higher Education of the Russian Federation on the topic "Models, methods and algorithms of artificial intelligence in economic problems for the analysis and stylization of

multidimensional data, forecasting time series and designing recommender systems",
project number FSSS-2023-0004