

**ЦИФРОВОЕ ОБЩЕСТВО
И ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ. ИННОВАЦИИ
DIGITAL SOCIETY AND
INFORMATION TECHNOLOGIES
INNOVATIONS**

**ЦИФРОВОЕ ОБЩЕСТВО
И ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ. ИННОВАЦИИ
DIGITAL SOCIETY AND
INFORMATION TECHNOLOGIES
INNOVATIONS**

Н. Ф. Дьячкова

**ИСПОЛЬЗОВАНИЕ ОНЛАЙН-ДАНЫХ ПО
ЦЕНАМ В КАЧЕСТВЕ ОПЕРЕЖАЮЩИХ
ИНДИКАТОРОВ ДИНАМИКИ
МАКРОЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ**

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО
ХОЗЯЙСТВА И ГОСУДАРСТВЕННОЙ СЛУЖБЫ ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ
ФЕДЕРАЦИИ»
(РАНХиГС)

**ИСПОЛЬЗОВАНИЕ ОНЛАЙН-ДАННЫХ ПО ЦЕНАМ В КАЧЕСТВЕ
ОПЕРЕЖАЮЩИХ ИНДИКАТОРОВ ДИНАМИКИ МАКРОЭКОНОМИЧЕСКИХ
ПОКАЗАТЕЛЕЙ**

Дьячкова Н.Ф., Центр изучения проблем центральных банков Института
прикладных экономических исследований (ИПЭИ) Российской академии народного
хозяйства и государственной службы при Президенте РФ (РАНХиГС), с.н.с., к.э.н.,
ORCID: 0000-0001-9212-8869, dyachkova-nf@ranepa.ru

Москва

2023

FEDERAL STATE BUDGETARY EDUCATIONAL INSTITUTION OF HIGHER EDUCATION
«RUSSIAN PRESIDENTIAL ACADEMY OF NATIONAL ECONOMY AND PUBLIC
ADMINISTRATION» (RANEPA)

**USING ONLINE PRICE DATA FOR CONSTRUCTION LEADING MACROECONOMIC
INDICATORS**

N.F. Dyachkova, Center for the Study of Central Banking Problems, Institute of Applied Economic Research, Russian Presidential Academy of National Economy and Public Administration (RANEPA), senior researcher, Cand. Sci. (Econ.), ORCID ID: 0000-0001-9212-8869, dyachkova-nf@ranepa.ru

Moscow

2023

Аннотация

Технологии больших данных (big data) и онлайн-данных предоставляют макроэкономистам возможность создавать уникальные наборы данных, адаптированные к конкретным потребностям исследований и позволяющие расширить возможности как измерения, так и прогнозирования с использованием новых форм информации. Особую значимость представляют онлайн-данные по ценам, обладающие способностью к демонстрации опережающей динамики различных экономических, в том числе инфляционных процессов. **Цель** настоящего исследования – выявить возможности высокочастотных показателей и ценовых индексов показывать опережающую динамику и их возможности для краткосрочного прогнозирования. Достижению заявленной цели способствует обзор академической и эмпирической литературы для систематизации и классификации видов опережающих индикаторов и методов их построения, а также по направлениям оценки различных макроэкономических процессов и определения степени согласованности результатов, получаемых на основе низкочастотных данных, с результатами оценок на ежедневных и онлайн-данных. **Актуальность** исследования определяется способностью высокочастотных показателей, включая ценовые индексы на основе онлайн-данных, демонстрировать опережающую динамику. Работы по вопросам оценки и построения высокочастотных показателей приобретают особую значимость. Исследование проводилось с использованием дескриптивного, статистического и графического анализа, анализа больших данных, сравнительного анализа, а также регрессионного анализа. По **результатам** анализа эмпирических подходов к построению и оценке высокочастотных показателей в качестве опережающих индикаторов макроэкономических процессов можно прийти к **выводу**, что высокочастотные показатели демонстрируют лучшие возможности оценки текущих изменений в экономике и имеют не меньшую значимость, чем низкочастотные показатели при включении в VAR- & SVAR- модели. **Перспективы** дальнейшего исследования высокочастотных показателей заключаются в систематизации эмпирических подходов к оценке и прогнозу инфляционных процессов на краткосрочном промежутке времени. Результаты проведенного обзора и регрессионного анализа могут быть **использованы** в интересах статистических служб и макроэкономических ведомств РФ для мониторинга текущей ценовой статистики и получения опережающей информации для целей краткосрочного планирования.

Ключевые слова: онлайн-цены, индекс потребительских цен, ценовая статистика, большие данные

JEL: C43, E31, P42

Abstract

Today big data and online data provide macroeconomists with the advantage to create unique data sets that are conducted to specific research needs, and more than that, it enables some specific capabilities for both measurement and forecasting with the usage of new forms of data. The importance of this work is based on that the online price data has the ability to demonstrate the upcoming dynamics of various economic processes, including inflation, in short term period. The **relevance** of this work is determined by the ability of high-frequency indicators, including price indices based on online & scanner data, to evaluate the leading dynamics from microlevel and big data. This research is devoted to the issues of assessing and constructing high-frequency indicators, which acquires special significance for macroeconomic forecasting. The main **goal** of the study is to identify the capabilities of high-frequency indicators and price indices to show leading dynamics in macroeconomics and its capabilities for short-term forecasting (**subject** of the study). To achieve the stated goal, it is based on reviewing of academic and empirical literature (as the main **source** of information) to systematize and classify the types of high-frequency leading indicators and their construction methods. Also, it is conducted with the areas of measuring various macroeconomic processes and determined the degree of consistency of the results, which are obtained from low-frequency data comparatively to the results on daily and online data basis (research **objectives**). In this research were used **methods** such as descriptive, statistical and graphical analysis, the big data analysis, including the systematic approach and comparative analysis with regression modelling. Based on the empirical results of a study it is **shown** that the construction and evaluation of high-frequency indicators as leading indicators of macroeconomic processes, can be quite interesting and combat. Finally, the high-frequency indicators show better statistics dynamics for current changes in the economy and demonstrate the greater significance than low-frequency indicators in VAR- & SVAR- models (**scientific novelty** of the work). The **prospects** for further research are to show the capabilities of high-frequency indicators in forecasting inflation in the short-term period of time. In the conclusion, the results of the work **can be used** in the interests of statistical services and macroeconomic departments of the Russian Federation to monitor current price statistics and to obtain advanced information for short-term planning purposes from microlevel data.

Key words: Online price statistics, big data, consumer price index, price aggregation

JEL: C43, E31, P42

Содержание

ВВЕДЕНИЕ	7
1. Эмпирический обзор построения опережающих показателей с использованием онлайн-данных.....	8
1.1 Виды опережающих индикаторов и предикторов макроэкономических показателей	8
1.2 Сопоставление возможностей использования опережающих показателей для целей экономического анализа и инфляционного прогнозирования.....	10
2. Построение var-модели с использованием высокочастотного ценового индекса.....	19
2.1 Стандартные var-модели с включением переменных на основе онлайн-данных	190
2.2 Описание ежедневных данных и высокочастотного ценового индекса.....	201
2.3 Описание спецификаций и предварительный анализ данных.....	22
2.4 Результаты моделирования и реакции функций импульсных откликов	25
ЗАКЛЮЧЕНИЕ	31
Список использованных источников	33

Введение

Данная работа посвящена анализу применения онлайн-данных и высокочастотных показателей в макроэкономических исследованиях, а также оценке возможности построения макроэкономических прогнозов с опорой на высокочастотные данные. В работе описаны свойства и источники происхождения высокочастотных и онлайн-данных, их способность предвосхищать тренды в настоящем и будущем. Подробно рассматриваются различные эмпирические работы, которые используют высокочастотные данные для построения краткосрочных прогнозов.

В работе показывается реплицированный высокочастотный ценовой индекс для построения VAR-модели наряду с другими высокочастотными показателями с целью оценки импульсных откликов на ежедневных онлайн-данных. Авторы рассмотрели возможность оценки и прогнозирования макроэкономических процессов для различных высокочастотных показателей. Период для рассмотрения был выбран с 01.09.2020 г. по 01.06.2023 г., разделенный на два подпериода – с 01.09.2020 г. по 01.02.2022 г. и с 01.04.2022 г. по 01.06.2023 г. Данные подпериоды характеризуются различием в социально-экономических процессах из-за пандемии и выхода из этого кризисного периода, а также неопределенностью социально-экономических процессов и обострением геополитической ситуации в мире во время второго подпериода наблюдений.

1. Эмпирический обзор построения опережающих показателей с использованием онлайн-данных

1.1 Виды опережающих индикаторов и предикторов макроэкономических показателей

Внедрение новых форм электронных платежных систем и развитие Интернет-торговли привело к появлению нового цифрового типа онлайн-данных, отличного по своим характеристикам от других видов статистических данных. Сегодня онлайн-данные позволяют получить информацию оперативно в течение нескольких часов. Эти данные характеризуются высокой частотой (от англ. high frequency) и носят название высокочастотных данных, и в отличие от низкочастотных данных, таких как недельные, месячные, квартальные и годовые данные, высокочастотные данные обновляются постоянно. Высокочастотные данные позволяют следить за формированием трендов и движением цен в режиме реального времени. Это привело к появлению новой обширной области исследований [1], [2], [3], опирающейся на использование высокочастотных данных, где анализируются и используются характеристики высокочастотных данных для разработки и построения моделей прогнозирования и оценивается способность высокочастотных данных показывать опережающую динамику различных экономических процессов. Прогнозные модели с включением высокочастотных данных как опережающих показателей охватывают множество областей, связанных с оценкой поведения экономических агентов на финансовых рынках, оценку доходности и волатильности ценных бумаг, различные тенденции в ценовой динамике и в решении оптимизационных задач для целей макроэкономического прогнозирования.

Отличительной особенностью высокочастотных данных от низкочастотных является скорость обновления и получения информации (например, для высоко алгоритмической торговли ценными бумагами скорость получения и обработки поступающей информации в онлайн составляет от нескольких минут до нескольких часов), а также накопленная частота данных (см. рисунок 1). Разница между высокочастотными и низкочастотными данными, такими как месячные или квартальные показатели, связана с возможностями применения первых для построения достоверных статистических прогнозов на краткосрочных промежутках времени.



Примечание – Источник: составлено авторами

Рисунок 1 – Виды и частота обновления онлайн-данных

Существует ряд независимых суждений [4], [5], [6] о том, что высокочастотные данные, в отличие от других типов данных, способны с большей вероятностью и более высокой достоверностью предсказывать изменения в краткосрочном периоде времени и показывать статистическое превосходство в оценке трендов при прогнозировании. К примеру, для оценки факторов, влияющих на изменение стоимости ценных бумаг в течение двух следующих недель, авторы [7] указывают на то, что высокочастотные данные будут показывать тренд изменения стоимости в режиме реального времени, что позволит быстрее оценить рыночные риски и степень волатильности. Такой подход с использованием высокочастотных данных в качестве опережающих переменных носит название «бега-на-опережение» (от англ. front running) и широко используется в техническом анализе при торговле ценными бумагами на финансовых рынках.

Иначе на использование высокочастотных данных смотрят работы [8], [9], [10], в которых авторы включают высокочастотные данные в модели макроэкономического прогнозирования. Оценки из различных моделей с включением данных смешанной частоты дают неоднозначные прогнозы. Полученные из моделей оценки сложно интерпретируются с экономической точки зрения как исследователями, так и надзорными и управляющими государственными структурами [11], [12]. Вопрос эффективности и применимости высокочастотных данных по ценам в качестве опережающих показателей будет проанализирован с помощью построения прогнозной VAR-модели далее в пункте 2.

Технологии больших данных (от англ. Big data) и онлайн-данных предоставляют макроэкономистам возможность создавать наборы данных, адаптированные к конкретным потребностям измерений и исследований. Это помогает сгладить проблемы в эмпирических исследованиях, такие как ограничение выборок по времени, эндогенность данных, пропуски и ошибки измерения в данных, которые часто встречаются в наборах данных.

Теоретические работы [13], [14] по рассмотрению и внедрению ежедневных данных как переменных с особой значимостью в эконометрическом моделировании впервые были опубликованы в 1980-х годах. Эти работы были посвящены рассмотрению вопросов временных рядов и их свойств для возможностей построения краткосрочных прогнозов. Первые известные ежедневные данные по финансовым котировкам с внутридневной частотой изменения данных начали собираться Нью-Йоркской фондовой биржей (NYSE) в 1992 г. для отслеживания динамики стоимости ценных бумаг на американском фондовом рынке.

Неструктурированные внутридневные данные могут быть преобразованы в формат структурированных временных рядов путем классификации и построения их ежедневных оценок. Таким образом, получается, что из большого массива данных возможно построить опережающие высокочастотные показатели для различных целей и получения прогнозов, а также для последующей их возможной комбинации и включения в регрессионные модели в трех формах:

- сценарные опережающие показатели, использующие получаемые тренды для определения развития различных макроэкономических сценариев и шоков;
- агрегированные опережающие показатели на основе данных со смешанной частотой и структурой для целей текущего мониторинга;
- простые опережающие показатели, связанные только с одним видом макроэкономических процессов.

В эмпирических исследованиях [2], [7], [13] высказывается несколько мнений относительно того, как высокочастотные данные могут использоваться в виде макроэкономических переменных и предикторов-корректоров, т.е. переменных, способных заранее указывать максимально приближенное значение будущего прогноза к реальным данным и оперативно подстраиваться под него (т.е. само-корректироваться, от англ. self-correcting) при приближении даты прогноза. Некоторые авторы [15], занимающиеся построением финансовых моделей по оценке доходности ценных бумаг, склонны расценивать подобную способность высокочастотных данных как особую статистическую ценность, повышающую достоверность и точность прогнозов.

Другое направление работ [9], [14], [16], где авторы занимаются исследованиями в области макроэкономических процессов и прогнозированием инфляционных ожиданий с использованием высокочастотных данных в различных видах моделях векторной авторегрессии

смешанной и ежедневной частоты рассматривают этот тип данных с осторожностью и определенным скепсисом из-за природы происхождения высокочастотных и онлайн-данных. С одной стороны, подобная способность высокочастотных и онлайн-данных указывать направление движения тренда является для целей макроэкономического прогнозирования большим преимуществом, с другой стороны, большая часть моделей с использованием данных смешанной частоты не имеет однозначной интерпретации и при отсутствии стационарности временных рядов является спорной.

1.2 Сопоставление возможностей использования опережающих показателей для целей экономического анализа и инфляционного прогнозирования

При построении моделей для целей макроэкономического прогнозирования важно учесть вид и тип данных, а также методологию исследования. В таблице 1 приведены различные виды исследований, демонстрирующие возможности применения высокочастотных и онлайн-данных как опережающих показателей для целей макроэкономического прогнозирования. Примеры исследований иллюстрируют то, как использование онлайн-данных, собранных с разными целями, может помочь получить новые эмпирические результаты. Подобные работы указывают на то, что на сегодняшний день мы должны относиться к большим данным (big data) и онлайн-данным не просто как к побочным наборам данных, а как к еще одному очень ценному и обширному источнику количественной и качественной информации.

Работы, в которых онлайн-данные показали свою наибольшую эффективность, связаны с оценкой и прогнозированием уровня безработицы как для развивающихся, так и для развитых стран. Возможности использования ежедневных данных из трендов Google позволяют оценить реальный уровень безработицы в экономике и ее тенденции, а также оценить потенциал человеческого капитала и спрогнозировать уровень занятости населения в краткосрочном периоде времени [17]. Другое направление макроэкономических исследований, в которых применяются онлайн-данные и высокочастотные показатели, тесно связано с оценкой ВВП и его компонентов [1], [13]. Большая часть работ указывает на то, что оценка ВВП и его возможный рост находится в прямой зависимости от количества и объема электронных платежей, а также от уровня реального потребления. Оценка компонентов ВВП с использованием высокочастотных показателей различается для стран с разным уровнем развития платежных систем и имеет тенденцию расходиться с оценками трендов, полученных на основе низкочастотных данных [18]. Это объясняется тем, что на низкочастотных данных труднее получить текущие оценки уровня развития платежных систем, что связано с «запаздыванием» поступления информации.

Одним из современных методов построения и оценки макроэкономических трендов с включением высокочастотных показателей является метод наукастинга (от англ. nowcasting), заимствованный из геологических и погодных исследований. Суть метода наукастинга заключается в оценке текущих тенденций и их экстраполяции (переносе эффектов и тренда) на ближайшее обозримое будущее (т.е. реализации прогноза события в краткосрочном периоде времени). Метод наукастинга позволяет сделать приближенный и сильно ограниченный по времени прогноз, но с достаточно высокой вероятностью его реализации. В работе [9], опубликованной на сайте Банка России, показано применение этого метода с использованием высокочастотных данных для анализа динамики роста ВВП России с 2003 г. по 2020 г. и последующим прогнозом для различных компонентов ВВП. Результатом проведенного исследования с применением наукастинга и включением 18 высокочастотных финансовых переменных, таких как ежедневная ставка денежного рынка и ставка овернайт, индексы бирж РТС и ММВБ, было получение прогнозных оценок на основе моделей MIDAS & VARMA со скользящим средним на ближайшие три месяца для оценки роста ВВП и его компонентов. Стоит отметить, что автор также указывает на ограниченность предсказания во времени для используемого метода, а также на тот факт, что результаты моделирования долгосрочных прогнозов будут несостоятельны из-за того, что прогнозная модель в целом имеет тенденцию сходиться к выборочному среднему (см. таблицу 1).

Следующее направление исследований, где широко применяются онлайн-данные, связано с оценкой и прогнозированием инфляции. Такие работы демонстрируют противоречивые эмпирические наблюдения и дают неоднозначные оценки того, насколько точно и эффективно можно предсказать уровень инфляции с помощью наблюдений за онлайн-ценами. С одной стороны, большая часть соответствующих исследований связана с построением и репликацией высокочастотных ценовых индексов, показывающих опережающую ценовую динамику официального ИПЦ. С другой стороны, в работах часто встречаются неоднозначно интерпретируемые результаты наблюдений за ценовой динамикой. Некоторые авторы отмечают, что прогнозирование инфляции с использованием онлайн-данных довольно спорно и зависит от множества факторов, связанных как со спецификой метода оценки, так и с проблемой эндогенности данных. Стоит обратить внимание на то, что благодаря онлайн-ценам удастся получить сглаженность ценовых тенденций в долгосрочном периоде времени, когда уровни цен показывают постепенное подстраивание для достижения долгосрочного равновесия в экономике, в то время как официальная статистика на основе низкочастотных данных и офлайн цен убирает ненаблюдаемые ценовые тренды, зафиксированные в высокочастотных данных.

Таблица 1.

Список работ в области макроэкономического прогнозирования с использованием различных источников высокочастотных и онлайн-данных в качестве опережающих показателей

№	Название работы	Автор(ы) и год	Данные	Модель	Страна и период	Методы анализа и работы с онлайн-данными
Оценка и прогноз безработицы	Прогнозирование количества заявок на пособие по безработице и прогнозирование настоящего уровня безработицы с помощью трендов Google	Choi, Varian, 2009	Недельные данные занятости Бюро США, тренды Google	Линейная регрессия	США, 1987 – 2009	Наукастинг, сезонная корректировка данных
	Эконометрика на основе данных Google и прогнозирование безработицы	Askitas, Zimmermann, 2009	Ежедневные и недельные тренды Google	Модель VECM	Германия, 2004 – 2009	Наукастинг, оценка компонентов тренда при поиске работы через модель коррекции ошибок
	Предсказательная сила поисков работы через Google и прогнозирование безработицы в США	D'Amuri, Marcucci, 2012	Недельные данные занятости Бюро США, тренды Google	ARMA & ARIMA(X)	США, 1967 – 2011	Регрессионный анализ на основе временных рядов и панельных данных
	Наукастинг с помощью трендов Google: метод подбора по поиску ключевых слов	Ross, 2013	Ежедневные и недельные тренды Google	Линейная регрессия	Великобритания, 2004 – 2012	Наукастинг, метод подбора на основе количества «заявителей» (Claimant count & query data)
	Использование данных о веб-активности для повышения своевременности (оперативности) показателей официальной статистики	Reis, Ferreira, Perduca, 2014	Ежедневные тренды Google, данные Twitter	Линейная регрессия	Франция, Италия, 2005 – 2014	Регрессионный анализ на основе временных рядов с включением «больших данных» и текстовой информации (big & textual data)

Продолжение таблицы 1

Оценка и прогноз безработицы	Улучшение прогнозирования безработицы с помощью трендов Google: часть 2	Ferreira, 2014	Ежедневные тренды Google	Линейная регрессия, построение факторных моделей	Португалия, 2006 – 2014	Включение инструментальных и латентных переменных в регрессию
	В поисках информации: использование трендов Google для сокращения информационных пробелов в развивающихся странах с низкими доходами	Narita, Yin, 2018	Ежедневные тренды и данные Google, данные МВФ	Линейная регрессия	193 страны, три кластера по уровню дохода (ВВП на чел.), 2004 – 2016	Кластерный анализ, кросс-секционный регрессионный анализ и анализ временных рядов
	Прогнозирование безработицы с помощью трендов Google: оценка возраста, пола и цифрового разрыва	Garcia-Hiernaux, Mulero, 2022	Ежедневные тренды и данные Google, LinkedIn, Orange Jobs	ARIMA	Испания, 2004 – 2018	Регрессионный анализ на основе временных рядов с включением «больших данных» и текстовой информации (big & textual data), кросс-секционный анализ
ВВП и его компоненты	Прогнозирование уровня потребления: индикаторы на основе опросов против трендов Google	Schmidt, Vosen, 2009	Индексы CCI & MSCI, тренды Google	Построение факторной модели	Германия, 2004 – 2009	Построение однофакторной модели, сравнение результатов между прогнозами на основе индексов и данных трендов Google
	Платежи в системе Данкорт как своевременный индикатор розничных продаж в Дании	Carlsen, Storgaard, 2010	Электронные и Интернет-платежи, различные виды транзакций	Линейная регрессия	Дания, 1998 – 2008	Регрессионный анализ на основе временных рядов, построение опережающих оценок
	Макроэкономический наукастинг с использованием вероятностей Google	Koop, Onorante, 2013	9 групп из 121 переменных со смешанной частотой	Модели DMA & DMS, TVP-VAR	США, 1973 – 2012	Наукастинг, регрессионный анализ на основе панельных данных

Продолжение таблицы 1

ВВП и его компоненты	Прогнозирование ВВП с использованием онлайн-данных электронных платежей	Galbraith, Tkacz, 2015	Данные по платежным картам и транзакциям, данные Google	Линейная регрессия, построение факторных моделей	Канада, 2005 – 2012	Наукастинг, регрессионный анализ на основе временных рядов и панельных данных
	Использование данных платежной системы для прогнозирования экономической активности	Aprigliano, Ardizzi, Monteforte, 2019	Данные по платежным картам и транзакциям, индексы производства	LASSO & MIDAS	Италия, 1999 – 2015	Регрессионный анализ на основе временных рядов, тестирование с разбиением выборки по периодам
	Прогнозирование уровня потребления с помощью цифровых платежей: анализ на основе данных смешанной частоты	Rooj, Sengupta, 2021	Данные по электронным и Интернет-платежам	Модели ARDL, AR-MIDAS	Индия, 2005 – 2019	Регрессионный анализ на основе временных рядов
	Прогнозирование ВВП России с использованием комбинированного подхода прогнозов	Zhemkov, 2021	3 группы из 89 переменных со смешанной частотой	MIDAS & VARMA	Россия, 2003 – 2020	Наукастинг, регрессионный анализ на основе временных рядов, тестирование с разбиением выборки по периодам
	Прогнозирование темпов роста ВВП с использованием трендов Google Trends на примере США и Бразилии	Bantis, Clements, Urquhart, 2023	Для США – 115, Для Бразилии – 96, включая ежедневные данные Google	Модели DFM, LASSO	США, Бразилия, 2005 – 2019	Наукастинг, построение и оценка многофакторных моделей на основе данных смешанной частоты
	Прогнозирование ВВП Китая в среде с большим объемом данных: уроки алгоритмов машинного обучения	Zhang, Ni, Xu, 2023	89 переменных с различной частотой, включая 8 по онлайн-данным	Модели DFM, LASSO & MIDAS	Китай, 1995 – 2020	Наукастинг, регрессионный анализ на основе временных рядов с включением «больших данных» (big data)

Продолжение таблицы 1

Оценка и прогноз инфляции	Автоматический сбор данных в Интернете (парсинг веб-страниц)	Boettcher, 2015	Онлайн-цены по различным продовольственным категориям европейских мультимедийных ритейлеров	Построение ценовых индексов	Страны ЕС: Великобритания, Нидерланды, Германия, Италия, Люксембург, Норвегия, Швеция, Австрия, Бельгия, Финляндия, Словения, 2009 – 2012	Репликация и воспроизведение высокочастотных ценовых индексов и гармонизированных ценовых индексов продовольствия, агрегирование и кластеризация данных, использование «больших данных» (big data)
	Сбор данных об одежде из Интернета	Griffioen, de Haan, Willenborg, 2014	Онлайн-цены по различным видам одежды европейских торговых сетей и онлайн-ритейлеров	Построение ценовых индексов	Нидерланды, 2012 – 2014	Репликация и воспроизведение высокочастотных ценовых индексов по непродовольственным категориям, использование «больших данных» (big data)
	Проект «Миллиард цен»: использование онлайн-цен для экономических измерений и исследований	Cavallo, Rigobon, 2016	Онлайн-цены по различным категориям товаров, сбор «больших данных» (big data)	Построение индексов, модель VAR & ADL	США, Аргентина, Бразилия, Великобритания, Германия, Китай, Южная Африка, Япония, 2007 – 2015	Репликация и воспроизведение высокочастотных ценовых индексов по различным странам, сопоставление с официальными статистическими данными между странами, оценка трендов инфляции и эффектов высокочастотного ценового индекса с помощью модели векторной авторегрессии
	Прогноз инфляции цен производителей в Китае с использованием низкочастотных и смешанной частоты данных в динамических факторных моделях	Liang, Wang, Xu, 2021	Бюро статистики Китая, база данных WIND: 69 переменных, включая 11 ежедневных показателей и высокочастотный ИПЦ Китая	Модели DFM, ARMA	Китай, 2006 – 2020	Построение и оценка многофакторных моделей, регрессионный анализ временных рядов на основе ежедневных данных

Окончание таблицы 1

Оценка и прогноз инфляции	Оценка весов данных, полученных из Интернета, для индекса потребительских цен	Ayoubkhani, Thomas, 2022	Онлайн-цены по непродовольственным категориям товаров (уникальный датасет с включением качественных характеристик товаров и онлайн-данных ККТ)	Построение ценовых индексов, модель пространственной авторегрессии MAPE	Великобритания, 2011 – 2013	Построение модели для оценки весов онлайн-данных с целью корректировки построения высокочастотных индексов для сопоставления с официальными данными ИПЦ
	Потенциальные причины смещения расчетов высокочастотных ИПЦ при использовании онлайн-данных	Bialek, Roszko- Wojtowicz, 2023	Онлайн-цены по различным категориям товаров, сбор «больших данных» (big data)	Построение ценовых индексов, модели TPD & GEKS-Jevons	Польша, 2020 – 2022	Репликация и воспроизведение высокочастотных ценовых индексов, оценка возможностей и сопоставления онлайн-данных для прогнозирования инфляции
	Автоматическая классификация продуктов с использованием алгоритмов машинного обучения для ценовой статистики	Oancea, 2023	Онлайн-цены по различным категориям товаров, сбор «больших данных» (big data)	Логистическая регрессия, методы машинного обучения для обработки текстовых данных (SVM, RW)	Румыния, Болгария, страны восточной Европы, 2020 – 2022	Оценка возможностей и сопоставление онлайн-данных для прогнозирования инфляции, построение модели автоматического классификатора, упрощающего процедуру репликации официального ИПЦ на основе ежедневных данных

Примечание – Источник: составлено авторами по источникам [9], [19], [17], [20], [21], [22], [23], [18], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39].

Разница между ежедневными и месячными данными для прогнозирования поведения ключевых экономических показателей достаточно большая, особенно при включении подобных показателей в различные виды VAR-моделей. Если сфокусироваться на измерениях и последующем прогнозировании, то высокочастотные и онлайн-данные в этом случае выступают экспериментальными и новыми источниками информации, призванными улучшить расчет традиционных экономических показателей, например, в случае оценки ценовой динамики и построения ежедневного индекса потребительских цен. Большая часть исследователей при воспроизведении высокочастотных ИПЦ стремится оценить их преимущества и недостатки, понять, имеют ли онлайн-цены четкую динамику и могут ли они быть надежным источником информации в текущих экономических условиях, а не только для разового исследовательского приложения.

Одно из очевидных преимуществ онлайн-данных по ценам – низкая себестоимость сбора большого статистического массива наблюдений. Вторым важным преимуществом является ежедневная частота сбора данных. Высокая частота данных по онлайн-ценам также позволяет избежать необходимости использовать средние значения по времени, так как онлайн-данные могут указывать на изменения цен оперативнее, чем официальные данные. В-третьих, онлайн-данные по ценам позволяют включать подробную информацию обо всех продуктах, продаваемых отобранными мультиканальными и онлайн-ритейлерами. Таким образом, диапазон доступных цен внутри категорий гораздо шире, чем в данных об индексе потребительских цен. В-четвертых, в онлайн-данных нет ограниченных ценовых периодов для наблюдаемых товаров и их ценовой динамики. Цены регистрируются с первого дня, когда товар предлагается потребителям и доступен в продаже, до дня его прекращения продажи в магазине (выбытия). Традиционные методы сбора данных, напротив, обычно начинают отслеживать новые товары только тогда, когда ранее используемые для сбора цен товары больше недоступны в офлайн торговле магазинов. Наблюдение и полнота ценового ряда на отдельные товары может помочь отслеживать тенденции потребления в отношении новых товаров, вносить как неявные, так и явные корректировки качества товара-представителя, а также изучать поведение онлайн-цен на момент появления нового товара-заменителя.

Важно отметить, что онлайн-данные по ценам позволяют воспроизвести высокочастотные ценовые индексы, которые достаточно близки по сопоставлению с официальными данными потребительских цен. Ряд исследователей указывает [35], [38], что большое количество перекрывающихся рядов цен на основе онлайн-данных может упростить процедуру корректировки качественной оценки товаров, отстающей от официальных данных индексов. Существует предположение, что онлайн-данные по ценам также могут предвосхищать изменения официального уровня инфляции на несколько месяцев вперед.

Возможные причины того, почему онлайн цены могут превосходить будущие шоки инфляции и рост потребительских цен, связаны с наличием задержек во времени, заложенных в методологии, используемой для получения официальных данных. Также эти различия связаны с выборкой ритейлеров и магазинов (точек онлайн и офлайн продаж) и более быструю корректировку онлайн цен в реальном времени, чем у наблюдений в магазинах. Часть вопросов, связанных с быстротой обновления информации и более быстрой корректировкой онлайн цен, остается открытой для обсуждения. Однако предполагается, что онлайн-данные по ценам могут быть эффективными в моделях прогнозирования инфляции. В исследовании [38] показано, что прогнозирование инфляции на ограниченных выборках с использованием онлайн и офлайн данных по ценам различно. В моделях смешанной частоты, онлайн цены и их тренды могут превзойти по точности оценок модели с включением офлайн данных.

В основе многих макроэкономических исследований [7], [11] лежит фундаментальная предпосылка того, что цены в экономике в краткосрочном периоде являются жесткими. За последнее десятилетие в эмпирической литературе предпринималось множество попыток измерения жесткости цен и объяснения причин их жесткости на микроуровне. Исследования по жесткости цен и распределению их изменений стало возможным благодаря доступу к онлайн-данным по ценам на микроуровне и наборам ценовых данных, полученных с помощью парсинга и веб-скрейпинга.

Еще одна уникальная особенность онлайн-данных по ценам заключается в том, что они предоставляют информацию об относительных уровнях цен, которая недоступна при использовании официальных индексов потребительских цен. Например, эти данные показывают [35], что стоимость минимальной корзины может варьироваться сильнее, чем это фиксируется официальными статистическими службами. Наличие подобных закономерностей важно для оценки стоимости продовольствия в разные моменты времени (в кризисные и не кризисные периоды времени), особенно для развивающихся стран.

2. Построение VAR-модели с использованием высокочастотного ценового индекса

2.1 Стандартные VAR-модели с включением переменных на основе онлайн-данных

Модели векторной авторегрессии (VAR-модели) широко используются в исследованиях временных рядов для оценки и изучения взаимосвязи, существующей между переменными. Кроме того, VAR-модели также являются важными инструментами прогнозирования, которые используются в макроэкономических исследованиях [40], [11].

Разработка таких моделей является предметом многочисленных продолжающихся исследований [41], [42], [12] с 1972 г., когда была заложена основа для оценки моделирования эндогенных процессов с множеством переменных в многомерной среде вычислений. Структура VAR-моделей позволяет объяснить значения эндогенных переменных на основе их прошлых наблюдаемых значений. После разработки базового метода и его структуры был разработан ряд статистических тестов для определения характера взаимосвязей между переменными. Кроме того, более поздние разработки VAR-моделей включают в себя различные структурные декомпозиции, линейные и нелинейные ограничения, учет изменяющихся во времени параметров, структурных сдвигов, стохастической волатильности и тому подобное.

В самом простом виде модель векторной авторегрессии может быть записана следующим образом (1):

$$Y_t = c + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + u_t, \quad (1)$$

где u_t – гауссовский белый шум для размерности n при ковариационной матрице, определенной как $E[u_t] = 0$, $E[u_t u_t'] = \Sigma_u$,

c – вектор констант размерности n ,

$A_1, A_2, \dots, A_p$ – это матрицы параметров размерности $n \times n$, состоящие из элементов a_{ij}^k , где i – номер оцениваемого уравнения, j – номер переменной в модели, k – номер лага, также для всех $i, j = 1, \dots, n$, и для лага определено, что $k = 1, \dots, p$.

При объединении матриц параметров в общую матрицу Φ , определяемую как $\Phi = [A_1, \dots, A_p, c]'$, новый вектор переменных будет определяться как X_t , где $X_t = [Y_{t-1}', \dots, Y_{t-p}', 1]'$, что позволит получить более простую форму записи VAR-модели (уравнение (2)):

$$Y_t = \Phi X_t + u_t, \text{ или в виде: } Y = X\Phi + U, \quad (2)$$

где $Y = [Y_1, Y_2, \dots, Y_T]'$, $X = [X_1, X_2, \dots, X_T]'$, $U = [u_1, u_2, \dots, u_T]'$ – переменные и шоки.

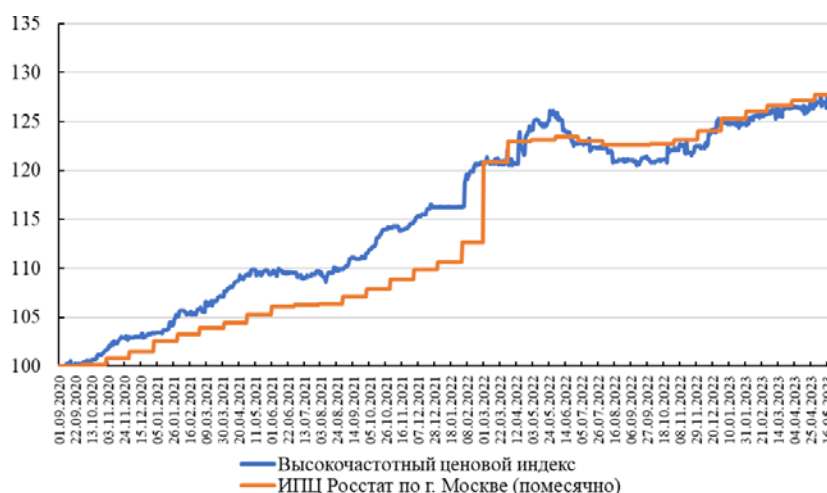
С одной стороны, при наличии положительной параметризации Y_t будет положительно зависеть от собственных прошлых значений, что приведет к регрессированию переменной самой на себя. С другой стороны, для VAR-модели естественно оценивать внутреннюю

динамику большинства переменных, которые зависят только от собственных прошлых значений.

Во многих макроэкономических исследованиях [42], [40], [43], [44] считается рациональным построение модели, зависящей не только от прошлых значений самих переменных, но и от прошлых значений других показателей. Поэтому VAR-модели с использованием высокочастотных данных будут включать в себя также лаги всех эндогенных включенных переменных на основе онлайн-данных. Одним из важных критериев состоятельности результатов в этом случае будет выступать частота наблюдений переменных и их сопоставимость во времени.

2.2 Описание ежедневных данных и высокочастотного ценового индекса

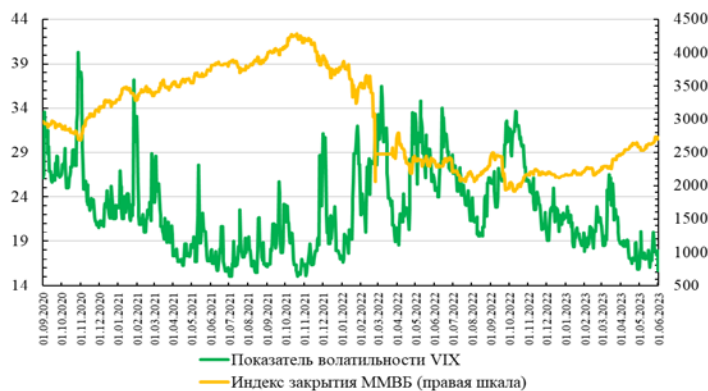
Для построения и оценки VAR-модели были собраны ежедневные данные по макрофинансовым переменным и воспроизведен авторский показатель для отслеживания ценовой динамики – это высокочастотный ценовой индекс (см. рисунок 2). Период построения составляет 1003 дня. Как видно из сопоставления с данными Росстата, высокочастотный ценовой индекс является более волатильным и динамичным по сравнению с официальными статистическими данными.



Примечание – Источник: составлено авторами

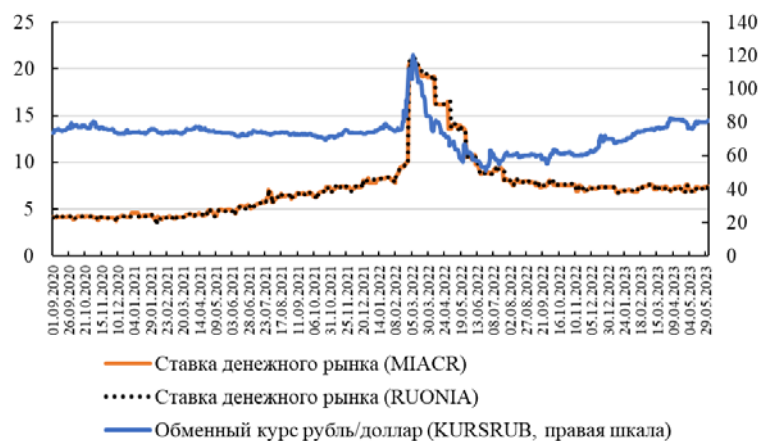
Рисунок 2 – Высокочастотный ценовой индекс и ИПЦ Росстата, ежедневные данные с 01.09.2020 г. – 01.06.2023 г.

Из макрофинансовых показателей для построения VAR-модели были собраны следующие ежедневные данные: обменный курс с сайта Банка России (KURSRUB) и дневные ставки денежного рынка (MIACR, RUONIA, см. рисунок 4), индекс закрытия ММВБ (IMOEX, см. рисунок 3), в качестве двух экзогенных переменных выступают цены на нефть двух марок Brent и Urals (см. рисунок 5), а также показатель волатильности финансовых рынков VIX (см. рисунок 3). Период сбора данных: с 01.09.2020 г. по 01.06.2023 г.



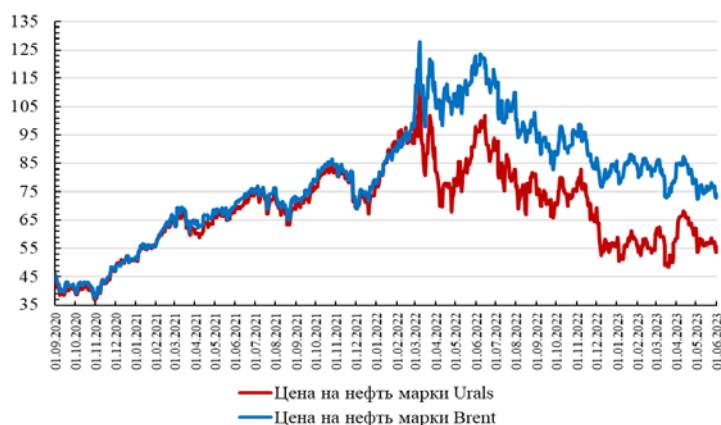
Примечание – Источник: составлено авторами

Рисунок 3 – Показатель волатильности VIX и индекс закрытия ММВБ, ежедневные данные с 01.09.2020 г. – 01.06.2023 г.



Примечание – Источник: составлено авторами

Рисунок 4 – Ставки денежного рынка и обменный курс, ежедневные данные с 01.09.2020 г. – 01.06.2023 г.



Примечание – Источник: составлено авторами

Рисунок 5 – Динамика цен на нефть марки Brent / Urals, в долларах США, ежедневные данные с 01.09.2020 г. – 01.06.2023 г.

Как видно из динамики высокочастотных макрофинансовых показателей и высокочастотного ценового индекса, в течение рассматриваемого периода имело место несколько экономических шоков и структурных сдвигов.

2.3 Описание спецификаций и предварительный анализ данных

Целью регрессионного анализа является выявление признаков того, что онлайн-данные по ценам действительно могут предвосхищать тенденции, которые находят отражение в макрофинансовых переменных российского рынка. Для построения был использован массив данных с 01.09.2020 г. по 01.06.2023 г., всего 1003 дня наблюдений, 7 переменных, наблюдения по каждой переменной есть на каждый день, включая выходные и праздничные дни. Лаги в моделях были взяты следующие: 10, 14 и 28 дней. Подобный выбор лагов в модели согласуется с эмпирическими работами [45], [41], [46] и позволяет оптимально учесть временной масштаб для исходных нестационарных временных рядов при построении VAR-модели.

Основными спецификациями для проверки прогнозов в VAR-модели были:

- Спецификация М-1: цена на нефть марки Brent/ Urals (экзогенная переменная), обменный курс, высокочастотный индекс цен, ставка денежного рынка;
- Спецификация М2-1: показатель волатильности VIX (экзогенная переменная), обменный курс, ставка денежного рынка, высокочастотный индекс цен;
- Спецификация М2-2: показатель волатильности VIX (экзогенная переменная), обменный курс, индекс ММВБ, высокочастотный индекс цен;
- Спецификация М3-1: цена на нефть марки Brent/ Urals (экзогенная переменная), индекс ММВБ, высокочастотный индекс цен, ставка денежного рынка;
- Спецификация М3-2: цена на нефть марки Brent/ Urals (экзогенная переменная), показатель волатильности VIX, высокочастотный индекс цен, ставка денежного рынка.

Стандартные VAR-модели без добавления экзогенных переменных:

- Спецификация М4-1: курс рубля, ставка денежного рынка, высокочастотный индекс;
- Спецификация М4-2: курс рубля, ставка денежного рынка, высокочастотный индекс цен, индекс ММВБ.

Для проверки стационарности (см. таблицу 2) временных рядов были взяты темпы роста для следующих переменных: цены на нефть марки Brent и Urals, обменный курс, индекс закрытия ММВБ. В первых разностях была рассчитана ставка денежного рынка, для показателя волатильности VIX был извлечен квадратный корень.

Таблица 2.

Расширенный тест Дики-Фуллера для проверки стационарности временных рядов

Расширенный тест Дики-Фуллера на «URALS» Нулевая гипотеза: Данные имеют единичный корень. Уровень значимости = 0,05 Тестовая статистика = -14,5051 Количество выбранных лагов = 22 Критическое значение 1% = -3,437 Критическое значение 5% = -2,865 Критическое значение 10% = -2,568 => Р-значение = 0,0. Нулевая гипотеза отвергается. => Ряд стационарен.	Расширенный тест Дики-Фуллера на «BRENT» Нулевая гипотеза: Данные имеют единичный корень. Уровень значимости = 0,05 Тестовая статистика = -15,4773 Количество выбранных лагов = 22 Критическое значение 1% = -3,437 Критическое значение 5% = -2,865 Критическое значение 10% = -2,568 => Р-значение = 0,0. Нулевая гипотеза отвергается. => Ряд стационарен.
Расширенный тест Дики-Фуллера на «KURSRUB» Нулевая гипотеза: Данные имеют единичный корень. Уровень значимости = 0,05 Тестовая статистика = -14,0328 Количество выбранных лагов = 21 Критическое значение 1% = -3,437 Критическое значение 5% = -2,865 Критическое значение 10% = -2,568 => Р-значение = 0,0. Нулевая гипотеза отвергается. => Ряд стационарен.	Расширенный тест Дики-Фуллера на «IMOEX» Нулевая гипотеза: Данные имеют единичный корень. Уровень значимости = 0,05 Тестовая статистика = -14,6089 Количество выбранных лагов = 22 Критическое значение 1% = -3,437 Критическое значение 5% = -2,865 Критическое значение 10% = -2,568 => Р-значение = 0,0. Нулевая гипотеза отвергается. => Ряд стационарен.
Расширенный тест Дики-Фуллера на «MIACR» Нулевая гипотеза: Данные имеют единичный корень. Уровень значимости = 0,05 Тестовая статистика = -12,3056 Количество выбранных лагов = 20 Критическое значение 1% = -3,437 Критическое значение 5% = -2,865 Критическое значение 10% = -2,568 => Р-значение = 0,0. Нулевая гипотеза отвергается. => Ряд стационарен.	Расширенный тест Дики-Фуллера на «CPI» Нулевая гипотеза: Данные имеют единичный корень. Уровень значимости = 0,05 Тестовая статистика = -14,7605 Количество выбранных лагов = 21 Критическое значение 1% = -3,437 Критическое значение 5% = -2,865 Критическое значение 10% = -2,568 => Р-значение = 0,0. Нулевая гипотеза отвергается. => Ряд стационарен.
Расширенный тест Дики-Фуллера на «VIX» Нулевая гипотеза: Данные имеют единичный корень. Уровень значимости = 0,05 Тестовая статистика = -12,3749 Количество выбранных лагов = 21 Критическое значение 1% = -3,437 Критическое значение 5% = -2,865 Критическое значение 10% = -2,568 => Р-значение = 0,0. Нулевая гипотеза отвергается. => Ряд стационарен.	

Примечание – Источник: расчеты авторов.

Формальное тестирование показало, что данные, приведенные в темпы роста, являются стационарными. Однако построение моделей на таких очищенных показателях не дало состоятельных результатов, поскольку из рядов была извлечена вся важная информация.

Поэтому для повторного построения моделей были использованы неочищенные данные, что позволило найти взаимосвязь между различными показателями. Следующим этапом проверки данных стала проверка коинтегрированности временных рядов, позволяющая подтвердить состоятельность оценок моделей при наличии нестационарности временных рядов. Для этого проводился тест коинтеграции Йохансена (см. таблицу 3).

Тест коинтеграции Йохансена применяется для оценки связанности нескольких временных рядов между собой (cointegrating within), его преимущества при тестировании нескольких временных рядов заключается в определении рангов на несколько отношений. Также тест позволяет оценить степень отношений и согласованности временных рядов.

Таблица 3.

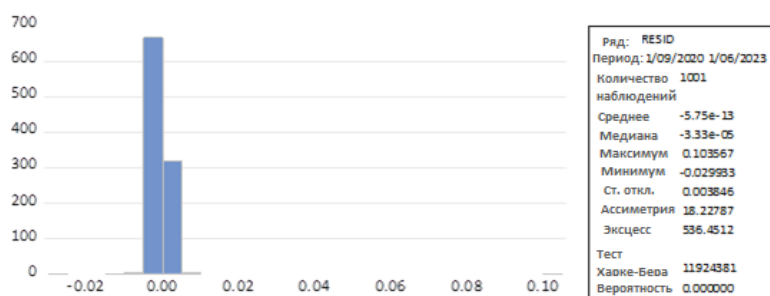
Наличие коинтеграции временных рядов на основе теста Йохансена

Имя переменной: Тестовая статистика > C(95%) => Значимость	

URALS :: 1000,22 > 111,7797 => Истина	
BRENT :: 681,07 > 83,9383 => Истина	
RUBLE :: 479,98 > 60,0627 => Истина	
IMOEX:: 293,88 > 40,1749 => Истина	
MIACR :: 137,86 > 24,2761 => Истина	
CPI :: 5,72 > 12,3212 => Ложь	
VIX : 0,31 > 4,1296 => Ложь	

Примечание – Источник: расчеты авторов.

Важно отметить, что при сопоставлении остатков из лучших спецификаций модели видно, что для неочищенных данных наблюдается некоторое левостороннее смещение наподобие «тяжелого хвоста», которое может указывать на наличие недостаточного количества объясняющих переменных (см. рисунок 6). Похожее правостороннее смещение наблюдается в ошибках моделей для переменных, приведенных к темпам роста. Мы можем связывать это с пропуском переменной, отвечающей за экономическую активность населения, однако такие данные недоступны в дневном разрезе.



Примечание – Источник: рассчитано авторами

Рисунок 6 – Распределение остатков из модели М3-1 со стационарными временными рядами

2.4 Результаты моделирования и реакции функций импульсных откликов

Разделение выборки на 2 подпериода связано с наблюдаемой неоднородностью данных во времени, что подтверждают результаты описательного анализа, проведенного в разделе выше. Также из рассмотрения был исключен временной промежуток в 59 дней (почти 2 месяца) из-за короткого периода высокой волатильности и зашумленности ежедневных данных для различных спецификаций VAR-модели, пересчитанных на двух подпериодах.

Ниже в таблице 4 приведены наблюдаемые реакции переменных на основе анализа рисунков 7, 8, 9 и 10 импульсных откликов для спецификаций, показавших наличие коинтеграции для двух и более нестационарных временных рядов и наилучшие результаты среди остальных спецификаций VAR-модели. Следует отметить, что наличие лагов в каждой из моделей различно (см. таблицу 4). Так, например, в спецификации M3-1 VAR-модели наблюдается большое количество значимых лагов, а именно: оказываются значимыми на 5% уровне лаги для 5, 11, 13 и 27 дня с включением MIACR, и лаги для 5, 7, 8, 9, 11, 27 для ставки RUONIA. Для спецификаций моделей M4-1 и M4-2 характерно наличие значимых лагов на уровне 5% для следующих дней: 5, 8, 13 и 27.

Включение показателя волатильности финансовых рынков (VIX) не дало особых результатов для большинства спецификаций VAR-модели, но показал интересную реакцию в спецификации модели M3-2. Из полученных результатов нужно отметить, что обменный курс почти всегда влияет на ставку денежного рынка, при этом в остальных моделях не всегда наблюдается обратная взаимосвязь между обменным курсом и высокочастотным ценовым индексом. Также ставка денежного рынка реагирует на показатель волатильности VIX (см. рисунок 7). Для остальных спецификаций VAR-модели не было выявлено наличия коинтеграции среди временных рядов высокочастотных показателей, а также не обнаружены значимые лаги. Дополнительно была оценена SVAR-модель для спецификации M3-1 с наибольшим количеством коинтегрированных временных рядов, показавшая сходимость временных рядов на долгосрочном промежутке времени.

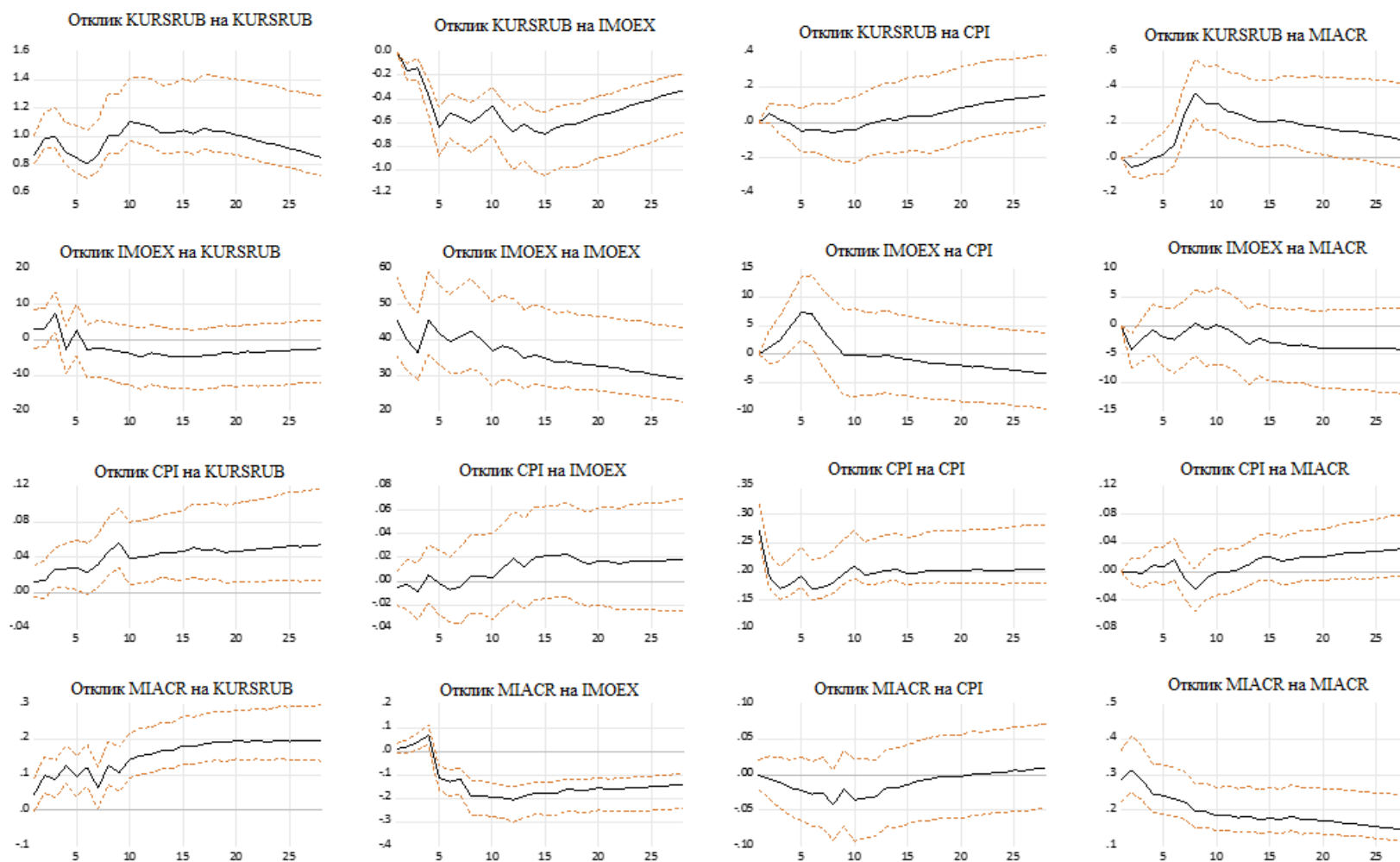
Таблица 4.

Краткое описание результатов реакций переменных для различных спецификаций VAR-модели

Номер модели и рисунок / Реакция переменных	Спецификация М3-1 (см. рисунок 7)	Спецификация М3-2 (см. рисунок 8)	Спецификация М4-1 (см. рисунок 9)	Спецификация М4-2 (см. рисунок 10)
1. Обменный курс влияет на ставку денежного рынка	Да (обесценение рубля повышает ставку)	+	+	+
2. Обменный курс влияет на ИПЦ	Да (обесценение рубля приводит к росту цен)	+	+	+
3. Ставка денежного рынка влияет на обменный курс	Да (рост ставок приводит к укреплению рубля)	+	+	+
4. Ставка денежного рынка влияет на ИПЦ	Наблюдается ложная зависимость	-	-	-
5. Рост волатильности влияет на обменный курс	Нет (-, не учтено в спецификации модели)	Да (высокая волатильность приводит к обесценению рубля)	-	-
6. Рост волатильности влияет на ставку денежного рынка	Нет (-, не учтено в спецификации модели)	Да (высокая волатильность приводит к пересмотру ставок)	-	-
7. Рост волатильности влияет на ИПЦ	Нет (-, не учтено в спецификации модели)	Да (высокая волатильность приводит к внутреннему росту цен)	-	-
8. Рост волатильности влияет на индекс ММВБ	Нет (-, не учтено в спецификации модели)	Да (высокая волатильность негативно влияет на финансовый рынок)	-	-

Примечание – Источник: составлено авторами.

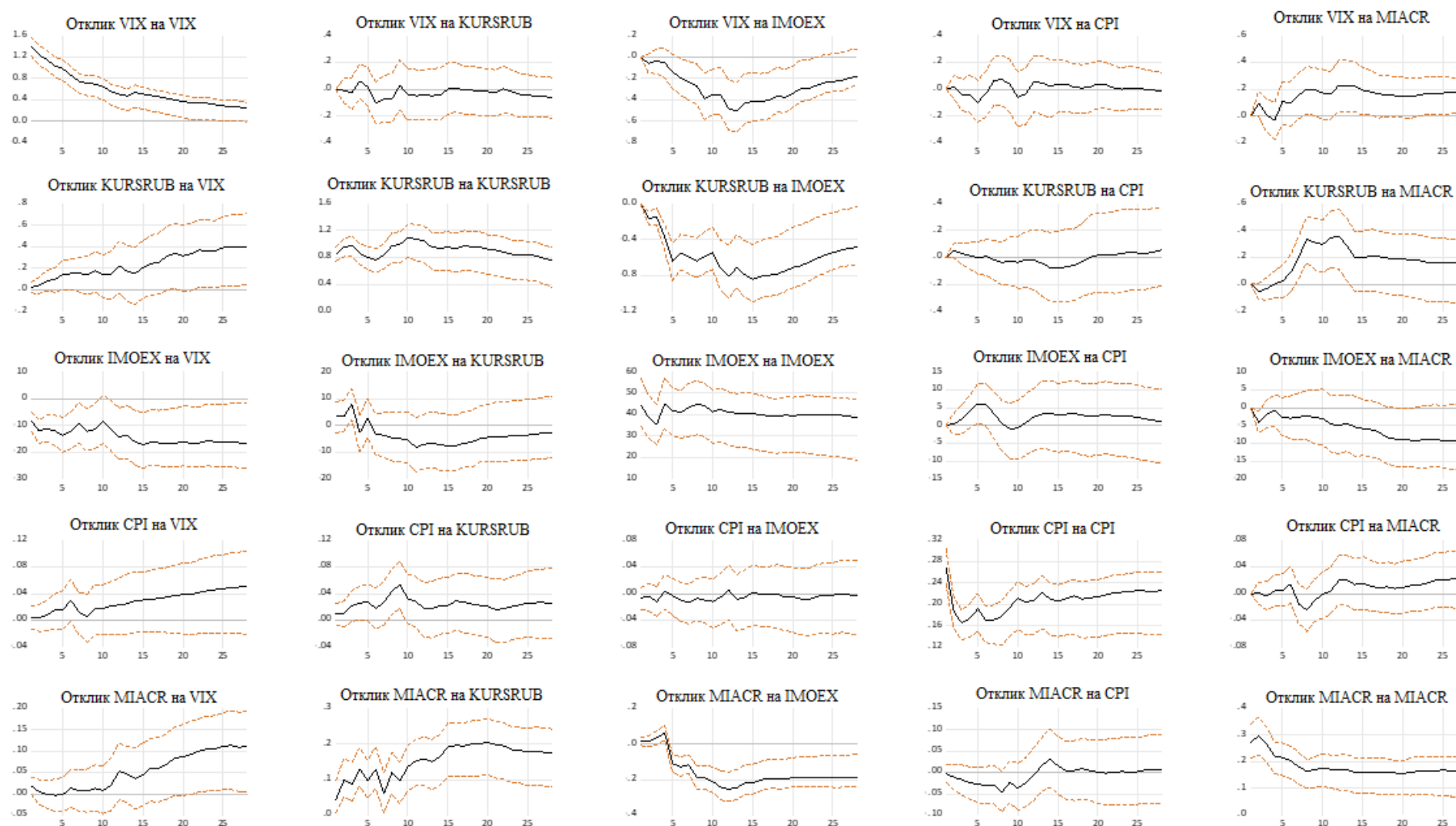
Отклик по Холецкому для единичных корней S.D. (без корректировки начальных степеней свободы)
95% ДИ с использованием процентиля Холла для 999 бутструп-повторений



Примечание – Источник: составлено авторами

Рисунок 7 – Модель МЗ-1, экзогенная переменная – цена на нефть марки Brent, лаг = 10, для периода с 01.09.2020 г. – 01.06.2023 г.

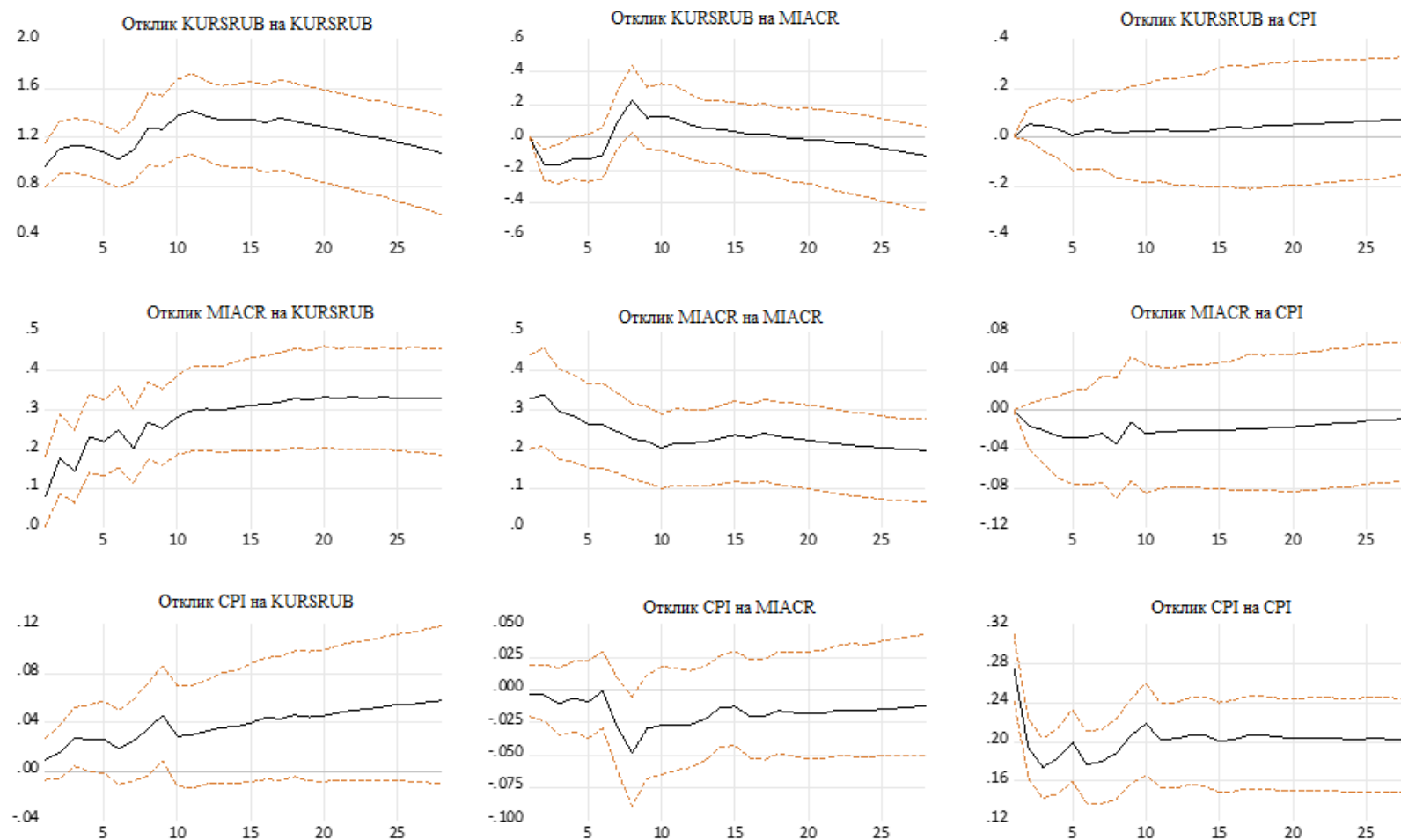
Отклик по Холецкому для единичных корней S.D. (без корректировки начальных степеней свободы)
95% ДИ с использованием процентиля Холла для 999 бутстреп-повторений



Примечание – Источник: составлено авторами

Рисунок 8 – Модель M3-2, экзогенная переменная – цена на нефть марки Brent, с включением показателя волатильности VIX, лаг= 14, для периода с 01.09.2020 г. – 01.06.2023 г.

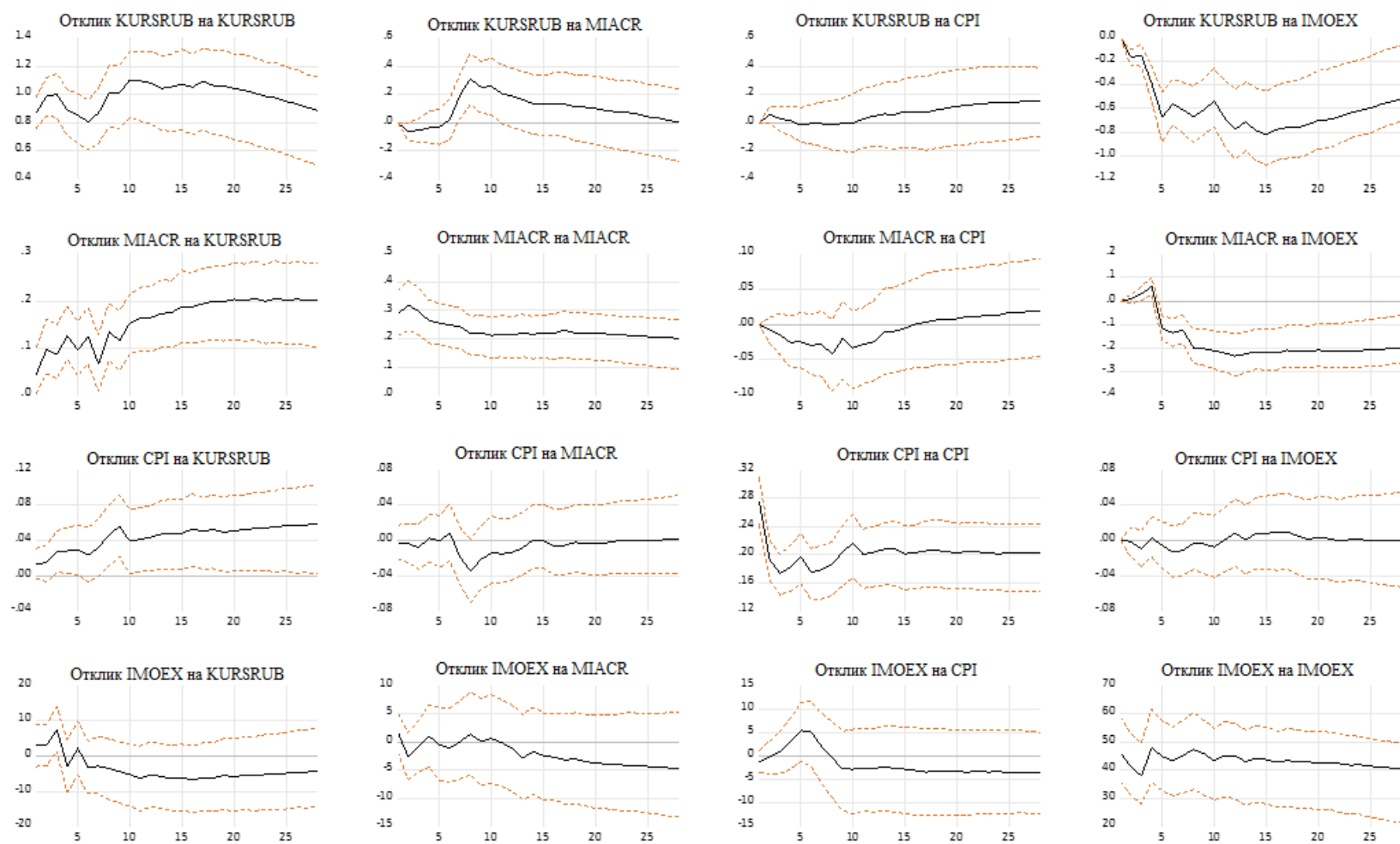
Отклик по Холецкому для единичных корней S.D. (без корректировки начальных степеней свободы)
95% ДИ с использованием процентиля Холла для 999 бутстреп-повторений



Примечание – Источник: составлено авторами

Рисунок 9 – Модель М4-1, лаг = 10, для периода с 01.09.2020 г. – 01.06.2023 г.

Отклик по Холецкому для единичных корней S.D. (без корректировки начальных степеней свободы)
95% ДИ с использованием процентиля Холла для 999 бутстреп-повторений



Примечание – Источник: составлено авторами

Рисунок 10 – Модель М4-2, лаг = 10, для периода с 01.09.2020 г. – 01.06.2023 г.

Заключение

В работе были рассмотрены различные подходы к использованию и включению высокочастотных показателей на основе онлайн-данных в качестве опережающих показателей для целей оценки и прогнозирования макроэкономических процессов. Приведенный обзор эмпирических источников иллюстрирует широкие возможности использования высокочастотных и ежедневных показателей на основе онлайн-данных по различным макроэкономическим темам – от прогнозов и оценки уровня безработицы, роста ВВП и его компонентов до оценки и прогнозирования инфляции.

Во-первых, в работе были оценены различные спецификации для VAR- & SVAR- моделей с включением высокочастотных показателей, в том числе авторского показателя высокочастотного индекса цен на основе онлайн-данных мультиканальных ретейлеров, для демонстрации способности онлайн-данных показывать опережающую динамику по сравнению с низкочастотными данными. Как видно из полученных результатов для спецификаций VAR-модели с коинтеграцией нестационарных временных рядов, что обменный курс (KURSRUB) почти всегда реагирует на ставку денежного рынка (MIACR), но подобная зависимость не наблюдается для ставки RUONIA; высокочастотный индекс цен (CPI) реагирует на обменный курс (KURSRUB), что указывает на наличие эффекта переноса курса в цены, и ставка денежного рынка (MIACR) реагирует на показатель волатильности финансовых рынков VIX.

Во-вторых, важно отметить, что в макроэкономических исследованиях с включением высокочастотных показателей существуют определенные ограничения при построении моделей, связанные с «природой» онлайн-данных. Отсутствие взаимосвязи или наличие нелинейности в моделях может объясняться тем, что у исследователей отсутствует возможность включить в модель нужные переменные с аналогичной частотой данных (например, показатель экономической активности населения). Проблема «пропуска» переменных возникает из-за сложности при сопоставимости высокочастотных онлайн-данных с другими типами низкочастотных данных.

В-третьих, проблема эндогенности и зашумленности высокочастотных данных требует более детального рассмотрения и проработки в теоретической литературе, а

также новых прикладных подходов к решению проблемы зашумленности ежедневных онлайн-данных.

Главный результат, полученный по итогам рассмотрения различных спецификаций VAR- & SVAR- моделей, заключается в проверке способности высокочастотных показателей на основе онлайн-данных показывать и воспроизводить результаты стандартных макроэкономических моделей. Корректность полученных результатов указывает на хорошие предиктивные свойства высокочастотных данных, которые можно использовать для целей оперативного мониторинга и анализа экономической ситуации.

*Материал подготовлен в рамках выполнения научно-исследовательской работы государственного задания РАНХиГС.

Список использованных источников

1. Buono D. Big data types for macroeconomic nowcasting // Eurona. 2017. pp. 1 - 53.
2. Elshendy M F Big data analysis of economic news: Hints to forecast macroeconomic indicators // International Journal of Engineering Business Management. 2017. No. 9.
3. Karsten S. Macroeconomic and macro-financial factors as leading indicators of non-performing loans: Evidence from the EU countries // Journal of Economic Studies. 2021.
4. Iyetomi H. Relationship between Macroeconomic Indicators and Economic Cycles in U.S. // Sci Rep. 2020. Vol. 10.
5. Cavallo A. The Billion Prices Project: Using Online Prices for Measurement and Research // Journal of Economic Perspectives. 2016. Vol. 2. No. 30. pp. 151-178.
6. Tingfeng Jiang T. Online prices and inflation during the nationwide COVID-19 quarantine period: Evidence from 107 Chinese websites // Finance Research Letters. 2022. Vol. 49. pp. 1544-6123.
7. Bin W. Macroeconomic indicators alone can predict the monthly closing price of major U.S. indices: Insights from artificial intelligence, time-series analysis and hybrid models // Applied Soft Computing. 2018. Vol. 71. pp. 685-697.
8. Chapman A.D. Macroeconomic Predictions using Payments Data and Machine Learning // Working paper series - Bank of Canada. 2022. pp. 1 - 46.
9. Zhemkov M. Nowcasting Russian GDP using forecast combination approach // WORKING PAPER SERIES Bank of Russia. February 2021. pp. 1 - 26.
10. Galbraith J.W Nowcasting GDP with electronic payments data // Statistics Paper Series, European Central Bank. 2015.
11. Bernanke B. Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach // NBER Working Paper Series. January 2004. pp. 1-49.
12. Bentour E.M. A ranking of VAR and structural models in forecasting // MPRA Paper 61502. 2015. pp. 1-20.

13. Clements M. Macroeconomic Forecasting with Mixed-Frequency Data: Forecasting Output Growth in the United States // *Journal of Business & Economic Statistics*. 2008. Vol. 26. No. 4. pp. 546–554.
14. Banbura M. Now-Casting and the Real-time Data Flow // *ECB Working Paper No. 1564*. 2013. pp. 1 - 55.
15. Zhang X. Novel modelling strategies for high-frequency stock trading data // *Financ Innov*. 2023. Vol. 9. No. 39.
16. Liang R. Nowcasting China's PPI inflation using low-frequency and mixed-frequency dynamic factor models // *Shanghai University of Finance and Economics - Working paper*. 2021.
17. Askitas N. Google Econometrics and Unemployment Forecasting // *German Council for Social and Economic Data (RatSWD) Research Notes No. 41*. September 2009.
18. Yin F., Narita Y. In Search of Information: Use of Google Trends' Data to Narrow Information Gaps for Low-income Developing Countries // *IMF Working Papers 2018/286*, International Monetary Fund. December 2018. pp. 1 - 51.
19. Choi H. Varian. Predicting Initial Claims for Unemployment Benefits. 2009.
20. D'Amuri F. The predictive power of Google searches in forecasting unemployment // No 891, *Temi di discussione (Economic working papers)*, Bank of Italy. 2012.
21. Ross A. Nowcasting with Google Trends : a keyword selection method // *Fraser of Allander Economic Commentary*. 2013. Vol. 37. No. 2. pp. 54-64.
22. Reis F. The use of web activity evidence to increase the timeliness of official statistics indicators // *Use of web activity evidence to increase timeliness (IAOS2014)*. 2014. pp. 1 - 35.
23. Ferreira P. Use of web activity evidence to increase timeliness (IAOS2014) // *Improving prediction of unemployment statistics with Google trends: part 2*. 2014.
24. Mulero R. Forecasting unemployment with Google Trends: age, gender and digital divide. *Empirical Economics* // *Empirical Economics*. 2023. pp. 587–605.
25. Schmidt T., Vosen S. Forecasting Private Consumption: Survey-based Indicators vs. Google Trends // *Ruhr Economic Papers*, No. 155. 2009. pp. 1 - 24.
26. Carlsen M., Storgaard P.E. Dankort payments as a timely indicator of retail sales in Denmark // *Danmarks Nationalbank Working Papers*, No. 66. 2010. pp. 1 - 28.

27. Onorante G. Macroeconomic Nowcasting Using Google Probabilities // In: *Advances in Econometrics: Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling*. Emerald Group Publishing Limited, 2019. pp. 17-40.
28. Galbraith J. Nowcasting GDP with electronic payments data // *Statistics Paper Series 10*, European Central Bank. 2015. pp. 1-20.
29. Monteforte V.A., Using Payment System Data to Forecast Economic Activity, *International Journal of Central Banking*, Vol. 15, No. 4, October 2019. pp. 55-80.
30. Rooj D., Sengupta R. Forecasting private consumption with digital payment data: A mixed frequency analysis // *ADB Working Paper Series*, No. 1249, Asian Development Bank Institute. 2021. pp. 1-19.
31. Evripidis P. Bantis M., Forecasting GDP growth rates in the United States and Brazil using Google Trends, *International Journal of Forecasting*, Vol. 39, No. 4, October–December 2023. pp. 1909-1924.
32. Zhang H.N., "Nowcasting Chinese GDP in a data-rich environment: Lessons from machine learning algorithms," *Economic Modelling*, No. 122, May 2023. pp. 106-154.
33. Boettcher I. Ottawa Group Online Pricing Project // *Automatic data collection on the Internet (web scraping)*. Ottawa. 2015. pp. 1-9.
34. Griffioen J.D. Collecting clothing data from the Internet // *Statistics Netherlands - Working Papers Series*. April 2014. pp. 1-17.
35. Cavallo F. Rigobon R., "The Billion Prices Project: Using Online Prices for Measurement and Research," *Journal of Economic Perspectives*, Vol. 2, No. 30, Spring 2016. pp. 151–178.
36. Liang W.X. Nowcasting China's PPI inflation using low-frequency and mixed-frequency dynamic factor models // *Peking University Preprints*. 2021. pp. 1-38.
37. Ayoubkhani H.T., "Estimating Weights for Web-Scraped Data in Consumer Price Indices," *Journal of Official Statistics*, Vol. 38, No. 1, 2022. pp. 5–21.
38. Białek J., Roszko- Wojtowicz E. Potential reasons for CPI chain drift bias while using electronic transaction data, *Technological and Economic Development of Economy*, Vol. 29, No. 2, 2023. pp. 564–590.
39. Oancea B., "Automatic Product Classification Using Supervised Machine Learning Algorithms in Price Statistics," *Mathematics*, No. 11, 2023. pp. 1588-1631.

40. Ankargren M. A Flexible Mixed-Frequency Vector Autoregression with a Steady-State Prior, *Journal of Time Series Econometrics*, Vol. 12, No. 2, April 2020. pp. 1-41.
41. Jonas M. B. Haslbecka L. A Tutorial on Estimating Time-Varying Vector Autoregressive Models, *MULTIVARIATE BEHAVIORAL RESEARCH*, Vol. 56, No. 1, 2021. pp. 120–149.
42. Nakajima J., "Time-Varying Parameter VAR Model with Stochastic Volatility: An Overview of Methodology and Empirical Applications," *Monetary and Economic Studies*, No. 29, 2011. pp. 107-142.
43. Roşoiu A., "Monetary Policy and Factor-Augmented VAR Model," *Procedia Economics and Finance*, Vol. 32, 2015. pp. 400-407.
44. Григорьев Р.А. "Одновременные эффекты несинхронных временных рядов" *Экономика и математические методы*, Vol. 55, No. 2, 2019. pp. 118–129.
45. Atsushi L.K., "The uniform validity of impulse response inference in autoregressions," *Journal of Econometrics*, Vol. 215, No. 2, April 2020. pp. 450-472.
46. Ghysels E., Valkanov R. The MIDAS Touch: Mixed Data Sampling Regression Models // *CIRANO Working Papers 2004s-20*, CIRANO. 2004. pp. 1-39.

**В СЕРИИ ПРЕПРИНТОВ
РАНХиГС РАССМАТРИВАЮТСЯ
ТЕОРЕТИЧЕСКИЕ
И ПРАКТИЧЕСКИЕ ПОДХОДЫ
К СОЗДАНИЮ, АКТИВНОМУ
ИСПОЛЬЗОВАНИЮ
ВОЗМОЖНОСТЕЙ
ИННОВАЦИЙ В РАЗЛИЧНЫХ
СФЕРАХ ЭКОНОМИКИ
КАК КЛЮЧЕВОГО УСЛОВИЯ
ЭФФЕКТИВНОГО УПРАВЛЕНИЯ**



РАНХиГС
РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ