

**Институт экономической
политики имени Е. Т. Гайдара**

Научный труд № 186

**Макроэкономическая
политика и методы
экономического
прогнозирования**

Издательство Института Гайдара
Москва | 2026

УДК 338.22+338.27
ББК 65.050+65.054
М16

Рецензент:

А. В. Улюкаев, д.э.н., главный научный сотрудник Фонда экономической политики

М16 Макроэкономическая политика и методы экономического прогнозирования. — М. : Издательство Института Гайдара, 2026. — 192 с. — (Научные труды / Институт экономической политики имени Е.Т.Гайдара ; № 186). — ISBN 978-5-93255-743-3.

Сборник объединяет исследования, посвященные вопросам макроэкономической политики, экономико-математического моделирования и прогнозирования экономических процессов. Работы рассматривают взаимодействие инструментов денежно-кредитной и бюджетно-налоговой политики, проблемы оптимального налогообложения и функционирования пенсионных систем, а также современные методы прогнозирования экономических показателей с использованием методов машинного обучения.

Представленные исследования демонстрируют возможности современных количественных методов анализа для изучения макроэкономической динамики и разработки эффективной экономической политики.

УДК 338.22+338.27
ББК 65.050+65.054

ISBN 978-5-93255-743-3

© Институт Гайдара, 2026

Содержание

Предисловие	5
-----------------------	---

Божечкова А. В., Горюнов Е. Л., Колесник С. И., Трунин П. В.

Анализ взаимодействия денежно-кредитной и бюджетно-налоговой политики в России	7
--	---

Введение	8
--------------------	---

1. Обзор теории взаимодействия денежно-кредитной политики и бюджетно-налоговой политики	9
---	---

2. Анализ взаимосвязей бюджетно-налоговой и денежно-кредитной политики в российской экономике	16
---	----

3. Оценка последствий денежно-кредитной политики в зависимости от фискального контекста	48
---	----

Заключение	64
----------------------	----

Список использованных источников	67
--	----

Замниус А. В., Шпилевая А. Е.

Оптимальное налогообложение и пенсионные системы в моделях перекрывающихся поколений	71
--	----

Введение	72
--------------------	----

1. Оптимальное налогообложение в моделях общего равновесия	75
--	----

2. Оптимальные пенсионные системы	90
---	----

3. Ключевые аспекты моделирования бюджетно-налоговой политики России	100
--	-----

Заключение	103
----------------------	-----

Список использованных источников	105
--	-----

Каукин А. С., Косарев В. С., Хубежова Д. И., Аникутин М. Ю., Швецов О. А.

Прогнозирование индексов промышленного производства с использованием продвинутых методов нейросетевого моделирования	113
--	-----

1. Обзор современных нейросетевых архитектур для прогнозирования временных рядов	114
--	-----

2. Влияние текстовых данных на экономические переменные	125
---	-----

3. Моделирование и прогнозирование индексов промышленного производства	161
--	-----

Выводы	179
------------------	-----

Список использованных источников	183
--	-----

Предисловие

Современная макроэкономическая политика требует применения сложных аналитических инструментов, позволяющих оценивать последствия принимаемых решений и прогнозировать развитие экономических процессов. Экономико-математическое моделирование и современные методы анализа данных становятся важными инструментами исследования макроэкономической динамики.

В работах, представленных в данном сборнике, рассматриваются различные аспекты макроэкономического анализа: взаимодействие инструментов экономической политики, моделирование систем налогообложения и пенсионного обеспечения, а также методы прогнозирования экономических показателей. Использование современных методов анализа данных и машинного обучения открывает новые возможности для повышения точности экономических прогнозов и углубления анализа экономических процессов.

В совокупности исследования демонстрируют развитие методологического инструментария современной макроэкономики.

Анализ взаимодействия денежно-кредитной и бюджетно-налоговой политики в России

Божечкова А.В., кандидат экономических наук, старший научный сотрудник лаборатории денежно-кредитной политики Института Гайдара.

Горюнов Е.Л., кандидат экономических наук, заведующий лабораторией денежно-кредитной политики Института Гайдара.

Колесник С.И., научный сотрудник лаборатории денежно-кредитной политики Института Гайдара.

Трунин П.В., доктор экономических наук, руководитель научного направления «Макроэкономика и финансы» Института Гайдара.

Аннотация. Работа посвящена анализу взаимодействия фискальной и монетарной политики в России, а также роли каждой из них в ускорении инфляции в 2022–2025 гг. Исследуется, как операции Минфина России и Банка России совместно влияли на динамику денежного предложения. Проведена декомпозиция прироста денежной базы с точки зрения источников прироста. Сформулированы принципы, которым должны удовлетворять фискальная и монетарная политика, позволяющие обеспечить макроэкономическую стабильность. Обсуждаются особенности фискальной политики в 2022–2025 гг., а также вклад операций Минфина России со средствами ФНБ в динамику денежной базы. При помощи метода локальных проекций проведен сравнительный анализ откликов макроэкономических переменных на шок монетарной политики в зависимости от режима фискальной политики. Показано, что охлаждающий эффект от повышения ключевой ставки ниже в условиях, когда бюджетная политика смягчается по сравнению с ситуацией, когда бюджетная политика ужесточается.

ВВЕДЕНИЕ

За период с января 2022 по декабрь 2025 г. потребительские цены выросли почти на 40%, чему соответствует средний ежегодный темп роста, равный 8,5%. Это существенно выше, чем целевой уровень Банка России, равный 4%, а также это превышает среднегодовую инфляцию предшествующих лет (среднегодовая инфляция за период 2016–2021 гг. составила 4,7%). При этом начиная с середины 2023 г. Банк России проводит весьма жесткую политику. В период с июля 2023 г., когда ЦБ РФ начал цикл повышения ключевой ставки, по ноябрь 2025 г. ключевая ставка в среднем превышала инфляцию на 9,2 п.п.

Ускорение инфляции сопровождалось значительным ростом государственных расходов и увеличением дефицита бюджета, при этом значительная часть этого дефицита была профинансирована за счет средств ФНБ. В период с 2022 по 2024 г. номинальные расходы федерального бюджета и бюджета расширенного правительства в среднем росли на 17,5 и 16,2% в год соответственно. Номинальный ВВП за тот же период рос в среднем на 8,6% в год, то есть доля государственных расходов в ВВП существенно возросла. При этом, поскольку бюджетные доходы также выросли, дефицит бюджета, хотя и увеличился, но оставался умеренным.

В связи с этими обстоятельствами возникает следующий ряд вопросов. В какой мере рост цен был спровоцирован увеличением расходов бюджета? Как операции Минфина России со средствами ФНБ повлияли на денежное предложение и как они в конечном итоге сказались на инфляции? Как можно охарактеризовать эффективность политики Банка России? Мог ли ЦБ РФ предотвратить быстрый рост потребительских цен, и если да, то при каких условиях и какие для этого необходимы были меры? Можно ли сказать, что ускорение инфляции последних лет стало следствием провала координации политики между Минфином России и Банком России? Наконец, какие ограничения для результативности политики ЦБ РФ создает реализация бюджетно-налоговой политики и расходование средств ФНБ? В результате проведенного исследования удалось продвинуться в понимании соответствующей проблематики.

Настоящая работа состоит из трех разделов. В первом разделе дается краткий обзор экономической теории, посвященной взаимодействию фискальной и монетарной политики. Во втором разделе проводится исследование совместного влияния операций Минфина России и Банка России на динамику денежного предложения на протяжении последних 25 лет и делаются выводы о том, какие принципы согласования политики бюджетных и денежных властей способны обеспечить макроэкономическую стабильность. В третьем разделе проводится эконометрическое исследование, цель которого определить, как зависит реакция экономики на шоки монетарной политики в ситуациях, когда фискальная политика ужесточается и когда она смягчается.

1. ОБЗОР ТЕОРИИ ВЗАИМОДЕЙСТВИЯ ДЕНЕЖНО-КРЕДИТНОЙ ПОЛИТИКИ И БЮДЖЕТНО-НАЛОГОВОЙ ПОЛИТИКИ

Вопросы взаимодействия денежно-кредитной и фискальной политики широко обсуждаются в теоретических и эмпирических исследованиях с момента возникновения макроэкономики как отдельного направления экономической теории [Hicks 1937; Keynes 1936]. С точки зрения экономической теории, а также экономической политики интерес представляют следующие тесно взаимосвязанные вопросы. Во-первых, как необходимо распределить задачи между фискальными и монетарными властями, чтобы в результате добиться максимального позитивного эффекта, то есть обеспечить оптимум общественного благосостояния, обеспечив условия для устойчивого экономического роста и макроэкономической стабильности. Этот вопрос касается того, каковы области эффективного применения инструментов денежных и фискальных властей. К примеру, ответив на этот вопрос, можно установить, какое из ведомств должно нести на себе основной вес по реализации контрциклических мер, кто должен быть ответственен за макроэкономическую стабильность и т. д. Во-вторых, при каких условиях фискальная политика вступает с монетарной политикой в конфликт или создает ограничения для ее реализации, и, наоборот, при каких условиях монетарная политика препятствует эффек-

тивному проведению фискальной политики. Ответ на этот вопрос позволяет сформулировать принципы, следование которым позволит избегать конфликтов между двумя ветвями экономических властей. Правила монетарной и фискальной политики следует проверять на соответствие данным принципам. В-третьих, каковы оптимальные сочетания действий фискальных и монетарных властей в ответ на внешние шоки различной природы. Этот вопрос связан с тем, как в зависимости от характеристик и источника макроэкономического шока следует распределять роли и действия между министерством финансов и центральным банком в рамках общего контрциклического комплекса мер, чтобы в итоге удалось наиболее эффективно сгладить негативные последствия шока. В-четвертых, каким должен быть институциональный дизайн системы государственного управления экономикой, чтобы он позволил, насколько это возможно, приблизить политику к оптимальной, с учетом всех полученных из теории рекомендаций. Оптимальный дизайн должен снизить возможные проблемы, возникающие из потенциального конфликта интересов и комплекса ограничений совместимости стимулов, связанного с достижением каждой из целей. Иначе говоря, институциональная конструкция должна быть такой, чтобы каждое из ведомств имело четкую систему стимулов, следование которым позволяло бы получить близкое к оптимуму макроэкономическое регулирование. В-пятых, какие конкретные инструменты должны использовать денежные и фискальные власти.

Согласно современным теоретическим представлениям, ответы на сформулированные выше вопросы звучат так. Монетарная политика должна быть прежде всего ориентирована на обеспечение долгосрочной ценовой стабильности, а также играть роль в контрциклической политике, сдерживая темпы экономического роста, когда выпуск превышает потенциальный и, напротив, стимулируя активность, когда экономическая активность опускается ниже потенциального уровня. Самостоятельная задача обеспечения благоприятных условий для реализации задач бюджетно-налоговой политики от центрального банка не требуется, хотя в тех условиях, когда между целями фискальных и монетарных властей нет конфликта, центральный банк может содействовать

министерству финансов. Иначе говоря, нет ничего предосудительного в том, чтобы при определенных условиях, когда бюджетная экспансия не мешает центральному банку достигать своих целей, регулятор держал ставку на приемлемом для бюджета уровне, облегчающем финансирование бюджетного дефицита. Бюджетно-налоговая политика должна поддерживать государственные финансы в устойчивом состоянии и при этом обеспечивать средствами все утвержденные программы государственных расходов. Бюджетно-налоговая политика должна проводиться в соответствии с бюджетными ограничениями в предположении, что эмиссионный доход не может выступать значимым источником финансирования бюджетных расходов. Ценовая стабильность, обеспечиваемая центральным банком, формирует условия для планирования бюджета на средне- и долгосрочную перспективу, устойчивости государственного долга. В свою очередь, следование бюджетному правилу, поддержание стабильного уровня государственного долга, предсказуемость траектории бюджетной политики способствуют успешной реализации монетарной политики, направленной на поддержание низкой инфляции. Перед фискальными властями при этом не ставится задача достижения ценовой стабильности, но за ними закрепляется определенная ответственность за содействие сглаживанию экономических колебаний посредством влияния на совокупный спрос. Суммируя сказанное, можно утверждать, что министерство финансов должно заниматься финансированием государственных расходов и блюсти финансовую дисциплину, не допуская кризиса государственного долга, а центральный банк должен защищать покупательную способность национальной валюты, залогом которой является монетарная дисциплина.

Важным условием, делающим возможным достижение описанного комплекса согласованных фискально-монетарных мер, выступает независимость денежных властей от правительства, поскольку это разграничение позволяет избежать конфликта интересов между задачей поддержания монетарной дисциплины и задачей обеспечения финансирования государственных нужд.

Эти принципы стали результатом эволюции теоретических представлений о месте фискальной политики и монетарной

политики, которые менялись по мере осмысления и обобщения опыта. Среди значимых мировых событий, которые сформировали сегодняшние представления, необходимо назвать период стагфляции 70-х, Великое успокоение (Great moderation, период с середины 1980-х и до наступления мирового финансового кризиса 2008–2009 гг.), мировой финансовый кризис, стагнацию 2010-х и пандемический кризис 2020 г. с посткризисной инфляцией. Опыт как развитых, так и развивающихся стран показал, что проблема координации и согласованности денежно-кредитной и фискальной политик является существенной с точки зрения результирующего макроэкономического эффекта каждой из них. Так, например, жесткие денежно-кредитные условия повышают стоимость обслуживания государственного долга и формируют риски для устойчивости бюджетной системы, тогда как значительные бюджетные стимулы, направленные на поддержание экономического роста, приводят к ускорению инфляции и ее значительному отклонению от целевых ориентиров, что осложняет задачу обеспечения ценовой стабильности для денежных властей. В этой связи в теоретических исследованиях особое внимание уделяется механизмам координации монетарной и фискальной политик, выявлению условий, при которых достижение целей бюджетной политики осуществляется без избыточного давления на ценовую динамику.

Ранние исследования, посвященные вопросам взаимосвязи между бюджетно-налоговой и денежно-кредитной политикой, соответствуют кейнсианским и монетаристским представлениям. В рамках монетаристской концепции предполагается, что ценовая динамика полностью определяется темпами роста денежной массы. Меры фискальной политики являются нейтральными по отношению к инфляционным процессам при условии, что дефицит бюджета не финансируется за счет дополнительной денежной эмиссии [Sargent, Wallace 1981]. В кейнсианских моделях увеличение совокупного спроса за счет роста дефицита бюджета приводит к повышению выпуска и росту цен, что означает проинфляционный эффект стимулирующей бюджетной политики. Дальнейшее развитие двух представленных направлений связано с появлением фискальной теории цен и новокейнсианского

подхода, учитывающих инфляционные ожидания и предпосылку о рациональности экономических агентов.

Одной из базовых монетаристских теоретических работ, в которых рассматривается вопрос взаимодействия монетарной и фискальной политик, является исследование Сарджента и Уоллеса [Sargent, Wallace 1981]. Авторы предложили две схемы координации монетарной и фискальной политик. В первом случае денежно-кредитная политика доминирует над фискальной, что означает, что центральный банк проводит независимую денежно-кредитную политику, например устанавливая темпы роста денежной базы или процентную ставку, что позволяет ему достигать ценовой стабильности. Параметры монетарной политики оказывают влияние на бюджетное ограничение правительства, определяя доход от сеньоража и процентные платежи по государственному долгу. Во втором случае в условиях фискального доминирования политика центрального банка может оказаться значительно более экспансионистской, чем при монетарном доминировании, и подчиненной нуждам бюджетной политики в ущерб поддержанию низкой инфляции. Реализация мягкой монетарной политики в интересах фискальных властей с целью монетарного финансирования бюджетного дефицита, снижения стоимости государственных заимствований создает высокие проинфляционные риски. При фискальном доминировании монетарное правило может включать в себя переменные фискальной политики (например, отклонение государственных расходов, государственного долга от устойчивого уровня). При этом коэффициенты в монетарном правиле при инфляции оказываются крайне нестабильными, учитывая его подчиненность потребностям бюджета [Cochrane 2023].

Исследование [Kumhof et al. 2020] дало толчок развитию фискальной теории уровня цен (*fiscal theory of price level*). Данная теория предполагает, что целевой ориентир по инфляции при определенных условиях не может быть обеспечен исключительно усилиями центрального банка без корректировки фискальной политики [Aiyagari, Gertler 1985]. Направленность текущей и будущей бюджетной политики является ключевым фактором инфляции, поскольку монетарная и фискальная политики связаны общим бюджетным ограничением. Общее бюджет-

ное ограничение представляет собой равенство реальной стоимости государственного долга и реальной стоимости будущих бюджетных профицитов:

$$\frac{B_{t-1}}{P_t} = s_t + \frac{1}{1+i_t} \frac{B_t}{P_t} = s_t + E_t \sum_{j=1}^{\infty} \beta^j s_{t+j},$$

где B_{t-1} — государственный долг в период $t-1$,

P_t — уровень цен в периоде t ,

s_t — бюджетный профицит (дефицит) в периоде t ,

i_t — ставка процента, определяемая центральным банком,

β — коэффициент дисконтирования,

E_t — математическое ожидание.

Уровень цен в текущий момент времени определяется параметрами текущей и будущей бюджетной политики. Стимулирующая бюджетная политика в текущем периоде должна быть компенсирована более высокими будущими налоговыми доходами. В случае невыполнения данного условия балансирующей переменной в рамках бюджетного ограничения выступает уровень цен (P).

В исследовании [Leeper 1991] вопросы взаимодействия фискальной и монетарной политик рассматриваются в контексте рикардianского и нерикардianского режимов поведения правительства. При нерикардianском режиме фискальные власти не предполагают полного финансирования государственного долга за счет будущих налогов, полагаясь на монетарное финансирование. В этом случае цены в модели устанавливаются эндогенно и определяются степенью фискальной экспансии. В условиях рикардianского режима центральный банк определяет параметры денежно-кредитной политики (объем денежного предложения и уровень цен), с учетом которых правительство должно достичь определенного уровня первичного бюджетного профицита, чтобы гарантировать выполнение бюджетных ограничений и поддержание бюджетной устойчивости.

Вопросы взаимодействия фискальной и монетарной политик с учетом наличия совместного бюджетного ограничения детально рассмотрены в работе [Leeper 1991]. Уравнение бюджетного ограничения предполагает, что величина государственного долга равна сумме будущих бюджетных профицитов. Это означает, что не пога-

шенный в текущем периоде государственный долг будет погашен в последующие периоды за счет доходов бюджета. Классифицируя политику на активную и пассивную, автор предполагает, что активная политика направлена на достижение определенных целей (например, снижение инфляции, достижение полной занятости), тогда как пассивная политика осуществляется для поддержания устойчивости описанного выше бюджетного ограничения. Для стабильных макроэкономических условий, как правило, характерна активная монетарная политика, направленная на достижение инфляцией целевых ориентиров, и пассивная фискальная, обеспечивающая бюджетную устойчивость. Однако в нестабильной экономической ситуации для поддержания экономической активности фискальная политика может перейти в активный режим, что ведет к наращиванию величины государственного долга. В этом случае достижение макроэкономической стабильности возможно при переходе центрального банка в пассивный режим, обеспечивающий баланс между реальным государственным долгом и приведенной величиной будущих бюджетных профицитов в результате ускорения инфляции. Результаты теоретической модели показывают, что реализация фискальной и монетарной политик в активном режиме увеличивает риски стремительного роста цен и гиперинфляции.

Современный взгляд на проблемы координации монетарной и фискальной политик представлен в работе [Bianchi, Melosi 2019]. На основе теоретической модели авторы показали, что в случае реализации негативного шока спроса наилучшей стратегией для центрального банка является формирование условий для стабилизации государственного долга и отсутствие реакции на растущую инфляцию. Бюджетно-налоговая политика, напротив, должна переключиться на активный режим за счет увеличения государственного долга, положительных бюджетных стимулов. Рост инфляционных ожиданий приведет к снижению реальной ставки процента, смягчая масштабы рецессии и сдерживая темпы роста государственного долга. Скоординированные таким образом монетарная и фискальная политики обеспечат более быстрое преодоление отрицательного разрыва выпуска, чем при отсутствии координации.

В рамках современных неокейнсианских моделей с номинальной жесткостью цен фискальные стимулы в виде увеличения

государственных расходов или снижения налогов приводят к повышению совокупного спроса. При наличии ограничений со стороны совокупного предложения расширение совокупного спроса приведет к росту цен и повышению инфляционных ожиданий. Изменение процентной политики зависит от функции реакции центрального банка. Если монетарные власти стремятся не допустить отклонения выпуска от потенциального уровня, процентная ставка вырастет, сократив положительный разрыв выпуска.

Таким образом, в рамках неокейнсианских моделей взаимодействие монетарной и фискальной политики осуществляется через влияние на разрыв выпуска, тогда как в рамках фискальной теории уровня цен — через воздействие на траекторию государственного долга в реальном выражении. В рамках фискальной теории цен в стабильных условиях согласованность монетарной и фискальной политик заключается в сочетании активной монетарной политики, ориентированной на стабилизацию инфляции, и пассивной бюджетной политики, направленной на обеспечение устойчивости государственного долга. В кризисных условиях монетарная политика становится подчиненной, тогда как фискальная политика, направленная на стимулирование выпуска, становится ведущей. Проведенный обзор теоретических исследований свидетельствует о значимости учета особенностей взаимодействия денежно-кредитной и фискальной политик при проведении эмпирических оценок макроэкономического эффекта каждой из них.

2. АНАЛИЗ ВЗАИМОСВЯЗЕЙ БЮДЖЕТНО-НАЛОГОВОЙ И ДЕНЕЖНО-КРЕДИТНОЙ ПОЛИТИКИ В РОССИЙСКОЙ ЭКОНОМИКЕ

2.1. Влияние Минфина России на денежное предложение и разделение сфер монетарной и фискальной политики

Эмиссия национальной валюты является прерогативой центрального банка. Базовое представление о денежной политике связано с идеей регулирования денежного предложения, то есть управления номинальным объемом денег в обращении. На прак-

тике центральный банк создает деньги, когда он приобретает некоторый актив, а продажа актива, напротив, сокращает денежное предложение [McLeay et al. 2014]. Центральный банк может осуществлять эмиссию, приобретая ценные бумаги, включая государственные облигации или иные бумаги, иностранную валюту, или в результате предоставления кредита банкам. Получение активов может иметь как нерегулярный характер (к примеру, покупки золота или валюты в целях пополнения международных резервов), так и быть подчинены достижению каких-либо промежуточных целей, к числу которых могут относиться поддержание ставки денежного рынка на заданном уровне или обеспечение стабильности валютного курса. На стороне обязательств ЦБ при этом возникают эмитированные денежные средства в форме наличных денег в обращении или средств на корреспондентских счетах кредитных организаций, а источником формирования дополнительных объемов денежной базы становятся реализуемые регулятором операции с активами.

Хотя центральный банк имеет широкие возможности регулирования объема денежной базы, в некоторых ситуациях определенное пространство для влияния на денежную базу имеют и фискальные власти. Когда совершаются налоговые платежи или бюджет привлекает заемные средства, происходит сокращение денежной базы (и денежной массы). Обратная ситуация имеет место, когда государственный бюджет осуществляет расходы или погашает задолженность. Если потоки налоговых платежей и бюджетных расходов равномерны и согласованы, то есть saldo изъятий и вливаний денежной базы по бюджетному каналу стабильно, то операции Минфина не будут оказывать большого влияния на денежное предложение.

Возможности влияния бюджетных властей на денежную базу определяются размером остатков бюджетных средств на счетах в центральном банке. Поэтому если Минфин имеет существенные средства на счетах в ЦБ, то их расходование приведет к заметному росту денежного предложения и будет фактором динамики кредитных и денежных рынков. В подобной ситуации монетарным властям необходимо сглаживать избыточные колебания денежного предложения посредством реализации контропераций,

чтобы нейтрализовать влияние операций бюджета, особенно если они приводят к дестабилизации денежного и кредитного рынков.

В основе современного разделения фискальной и монетарной политики лежит принцип, согласно которому монетарная политика состоит в осуществлении денежной эмиссии с приоритетной целью обеспечения ценовой и макроэкономической стабильности, а фискальная политика предполагает распределение налоговых поступлений по направлениям расходования при возможных заимствованиях на рыночных условиях. Иначе говоря, в идеальной ситуации центральный банк поддерживает устойчивость национальной валюты, а Минфин следит за устойчивостью бюджета и не рассчитывает на эмиссионный доход. Главная причина институционального разделения фискальной и монетарной политики связана с необходимостью поддерживать бюджетную дисциплину. Если бюджетно-налоговая политика не отделена от денежно-кредитной, то возникает риск возникновения проблемы мягких бюджетных ограничений. Эту проблему трудно преодолеть, когда решения о размере эмиссии и о величине государственных расходов (а также источниках их финансирования) принимается единым органом. Чтобы избежать конфликта интересов, Минфин и центральный банк должны действовать как независимые ветви экономической власти, и за каждым из них закрепляется своя зона ответственности.

Однако вне зависимости от степени институционального разделения денежные и бюджетные власти оказывают влияние друг на друга, поскольку действуют в едином поле — на денежном рынке и рынке капитала. Центральный банк осуществляет операции на открытом рынке, в том числе с государственными облигациями, и управляет процентными ставками, от которых зависит стоимость заимствования для государственного бюджета, а Минфин своими действиями влияет на денежное предложение. Бюджетно-налоговая политика не только оказывает влияние на денежную массу, но и на совокупный спрос и тем самым на инфляцию, поэтому монетарные власти при принятии решения о проводимой политике должны учитывать действия Минфина. Точно так же политика ЦБ влияет не только на процентные ставки и, значит, на стоимость долга, но, помимо этого,

имеет влияние на экономическую активность, а значит, на доходы и расходы бюджета.

Когда центральный банк приобретает на вторичном рынке государственные облигации, эмитируя денежную базу, он, во-первых, выкупает государственный долг у частного сектора, а во-вторых, поддерживает цены на государственные облигации, тем самым снижая их доходность и делая условия заимствования для бюджета более выгодными. Сходная ситуация возникает, если центральный банк кредитует банки под залог государственных облигаций, поскольку в результате этой операции он получает в качестве залога госдолг и эмитирует денежную базу. Таким образом, реализация монетарной политики оказывает влияние на условия реализации фискальной политики.

Само по себе это не является проблемой до тех пор, пока сохраняется разграничение сфер монетарной и фискальной политики и поддерживается монетарная и финансовая дисциплина. Для этого достаточно, чтобы операции на открытом рынке с государственными облигациями и кредитование банков с использованием их в качестве залога инициировалось регулятором исключительно в рамках задачи достижения целей монетарной политики и не были продиктованы фискальными соображениями.

Чтобы общий результат действий фискальных и монетарных властей был благоприятен для экономики, две данные структуры, оставаясь автономными, должны находиться в постоянном диалоге. Центральный банк и минфин по возможности должны учитывать действия и задачи друг друга, минимизируя риски возникновения ситуаций, когда действия одного института будут значительно осложнять достижения целей другого института. Отделение эмиссионных функций от функций управления государственными финансами необходимо для обеспечения макроэкономической стабильности и финансовой дисциплины. Необходимым условием достижения этого результата является четкое закрепление за каждым из регулирующих органов своих задач и стратегических целей, и тогда совместные действия на едином поле не будут приводить к провалам координации. Иначе говоря, в основе разделения фискальной и монетарной политики лежат не столько инструменты и область их применения, а исходная

мотивация использования данных инструментов. Размещение облигаций государственным бюджетом и самим центральным банком будет оказывать на ликвидность банковского сектора идентичное воздействие, но эти действия будут относиться к разным политикам, поскольку размещение облигации бюджетом ставит целью профинансировать расходы, а размещение облигации центральным банком имеет целью связать избыточную ликвидность, чтобы сдерживать рост кредитования и стабилизировать инфляцию.

Современные центральные банки реализуют свою политику посредством управления ставкой денежного рынка с целью достичь таргетируемого уровня инфляции. Это означает, что вместо того, чтобы определять целевой уровень эмиссии денежной базы, позволяющий обеспечить желаемую динамику цен, центральные банки объявляют границы коридора для ставок (или просто значение ключевой ставки) по краткосрочному межбанковскому кредиту, после чего осуществляют эмиссию денежной базы в таком объеме, чтобы фактические рыночные ставки на денежном рынке оказались в границах установленного коридора [Keister et al. 2008]. Таким образом, денежные власти сперва определяют тот уровень ставок, который создаст финансовые условия, необходимые для достижения таргета по инфляции, а затем с помощью операций эмиссии или абсорбирования денежной базы приводят актуальные ставки к этому уровню. Из вышесказанного следует, что объем эмиссии диктуется выбором величины ключевой ставки, то есть можно говорить об эндогенности эмиссии, объем которой в каждый момент времени будет обусловлен конкретными условиями на денежном рынке и величиной ключевой ставки. В этом смысле если денежным властям не удастся сдерживать инфляционные процессы и привести инфляцию к целевому уровню, то это означает, что выбранный уровень ставок был недостаточно высоким и, как следствие, темп эмиссии оказался избыточным.

На первый взгляд, после того как определен и объявлен уровень ключевой ставки, конкретный выбор активов, посредством приобретения которых будет осуществлена эмиссия, не имеет самостоятельного значения. Последнее соображение также озна-

чает, что расширение денежного предложения в результате операций минфина не обязательно противоречит задачам денежных властей, поскольку эти операции могут содействовать формированию денежного предложения как раз в том объеме, который необходим для удерживания ставок денежного рынка в коридоре. Бывают, однако, и обратные ситуации, когда расходование средств бюджетом приводит к избыточному увеличению денежной базы, и тогда центральный банк должен стерилизовать избыточную эмиссию доступными инструментами.

Вместе с тем увеличение денежной базы на одну и ту же величину может иметь различные инфляционные последствия и по-разному отразиться на динамике денежной массы, в зависимости от того, какая операция стала причиной роста денежной базы. К примеру, эмиссия денежной базы в рамках операций количественного смягчения в США с наступлением мирового финансового кризиса 2008–2009 гг. не приводила к росту денежной массы, поскольку в условиях низких процентных ставок увеличение объема средств в банковской системе не оказывало выраженного стимулирующего воздействия на кредитование, процесс кредитной эмиссии не запускался, совокупный спрос оставался слабым, а инфляция — пониженной. В другом эпизоде в период 2020–2021 гг. прирост денежной базы, реализованный ФРС, привел к существенному росту денежной массы, что стало одним из важных факторов последовавшего за этим инфляционного всплеска. Такая разная реакция экономики на рост денежной базы объясняется тем, что в первом случае эмитированные резервы замещали низкодоходные и ликвидные гособлигации, то есть происходило только замещение одних активов другими активами с близкими характеристиками, что не оказывало значимого влияния на интенсивность банковского кредитования. Во втором же случае рост денежной базы был обратной стороной (с точки зрения баланса кредитных организаций) массированного распределения бюджетных трансфертов среди населения, что стало фактором быстрого разогрева спроса и роста цен после ослабления карантинных ограничений [Goryunov et al. 2023; Горюнов и др. 2021]. Приведенные выше примеры показывают, что инфляционные последствия роста денежного предложения

зависят не только от величины прироста денежных агрегатов, но и от того, что именно послужило первопричиной увеличения денежной массы. Поэтому при анализе взаимосвязи между инфляцией и ростом объема денег в обращении нельзя ограничиваться только сопоставлением количественных показателей роста и цен и темпов роста денежной массы, но обязательно обращать внимание на структуру прироста.

Главным инструментом борьбы с инфляцией является политика высоких реальных ставок. Есть множество способов, с помощью которых центральный банк может обеспечить высокие реальные ставки на денежном рынке. Во-первых, он может начать продавать активы (валюту или государственные облигации) на открытом рынке, в результате чего денежная база сократится, что повлечет за собой рост стоимости привлечения средств. Помимо этого, если ЦБ продает на открытом рынке государственные облигации, их цена упадет, а значит, вырастет их доходность и следом вырастут доходности других долговых инструментов. Во-вторых, ЦБ может увеличить стоимость фондирования банков, то есть повысить ставку рефинансирования или ставку по операциям репо. В случае если банковский сектор в предшествующие периоды преимущественно формировал запас ликвидности за счет займов у ЦБ, рост ставки повысит издержки поддержания ликвидности, что снизит спрос на резервы. При этом денежная база сократится. Если банковский сектор имеет профицит ликвидности, центральный банк может обеспечить высокие ставки на денежном рынке, либо увеличив ставку по депозитам в ЦБ, либо повысив ставку по операциям обратного репо. В этом случае высокая ставка не сократит денежную базу, но тем не менее она приведет к росту стоимости кредита на межбанковском рынке. Регулятор может добиться аналогичного эффекта, если начнет выплачивать проценты на банковские резервы, причем в этом случае денежная база также не будет сокращаться. Все описанные способы приводят к увеличению стоимости межбанковского займа. Высокие ставки денежного рынка перекладываются в высокие ставки по кредитам и депозитам, а также оказывают повышательное влияние на доходности долговых ценных бумаг широкого спектра. Все это влияет охлаждающе на совокупный спрос, а значит, и на инфляцию.

Высокие ставки по депозитам стимулируют сбережения, а высокие ставки по кредитам дестимулируют инвестиционный процесс. Помимо этого, высокие ставки и доходности повышают привлекательность отечественных активов, что сдерживает отток капитала и укрепляет национальную валюту. На укрепление национальной валюты работает также повышение склонности к сбережениям, так как при этом падает спрос на импорт и, следовательно, спрос на иностранную валюту. Более крепкая валюта способствует торможению роста цен торгуемых товаров и в конечном итоге сокращает инфляцию.

Таким образом, содержательно антиинфляционная политика, проводимая путем установления высоких ставок, заключается в оказании охлаждающего эффекта на совокупный спрос путем его межвременного перераспределения в будущие периоды или, иначе говоря, за счет стимулирования сбережений. Антиинфляционная результативность политики высоких ставок в значительной степени зависит от того, насколько чувствителен совокупный спрос к изменению процентной ставки. Также эффективность антиинфляционной политики зависит от природы инфляционного шока. Сложнее подавлять инфляционные всплески, если они вызваны ростом номинальных доходов групп населения, минимально склонных к сбережениям и поэтому нечувствительных к изменению ставки. По этой причине в целях сохранения ценовой стабильности ЦБ следует более решительно противодействовать приросту денежной базы, если он сопровождается бюджетными расходами, относящимися к категории расходов, имеющих максимальный стимулирующий эффект для совокупного и потребительского спроса.

2.2. Особенности формирования денежной базы в разные периоды

Вопросы взаимодействия денежно-кредитной и фискальной политики в российской экономике можно рассматривать, анализируя динамику и факторы денежного предложения или денежной базы. Денежная база формируется по нескольким основным каналам: операции Банка России с кредитными организациями,

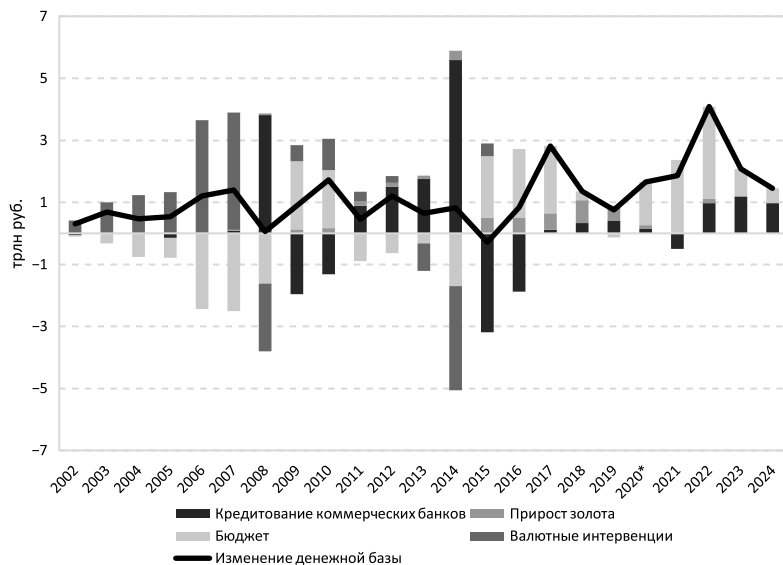


Рис. 1. Факторы формирования денежной базы в 2002–2024 гг. (трлн рублей)¹

Источник: Банк России, Минфин, расчеты авторов.

валютные интервенции Банка России и операции с золотом, операции Банка России с бюджетом и прочие операции. Важнейшим элементом бюджетно-налоговой политики в России является бюджетное правило, смысл которого состоит в сглаживании цикла, связанного с динамикой цен на нефть (обсуждение бюджетного правила см. в работах: [Гурвич, Соколов 2016; Кудрин, Соколов 2017; Аблаев, Сударев 2024; Аблаев, Магжанов 2024]). На рис. 1 представлена декомпозиция динамики денежной базы по данным факторам за период с 2002 по 2024 г. Изменение денежной

¹ Кредитование коммерческих банков — требования Банка России к кредитным организациям, покупка золота — изменение запасов монетарного золота, интервенции на валютном рынке — прирост валютных резервов с учетом валютной переоценки (за период с 2001 по 2008 г.), нетто-покупка долларов США и евро (2009–2016), операции Банка России на внутреннем валютном рынке (2017 г. — июнь 2025 г.), бюджет — операции с ФНБ, операции со счетами бюджета.

базы по бюджетному каналу с начала действия нового бюджетного правила в 2017 г. рассчитывалось как сумма операций Минфина России со средствами ФНБ и изменений остатков на счетах в Банке России и прочими операциями (например, требованиями Банка России к другим финансовым организациям).

Темпы роста денежной базы РФ существенным образом отличались на различных временных отрезках. Так, до мирового финансового кризиса 2008–2009 гг. годовые темпы прироста денежной базы составляли в среднем 35%, в посткризисный период и до кризиса конца 2014 — начала 2015 г. — 9,1%, в период активного расходования Резервного фонда в 2015–2017 гг. — 13%, в период с 2019 по 2021 г. — 7,2%, в период с начала СВО с 2022–2024 гг. — 11%. Рассмотрим динамику денежного предложения в эти периоды более подробно.

Период с 2002 по 2007 г.: профицитный бюджет и режим фиксированного курса

Данный период характеризуется сочетанием режима квазификсированного обменного курса и существенного профицита бюджета, который правительство использовало для погашения государственного долга и формирования накоплений в стабилизационном фонде. В 2000-е гг. благоприятная конъюнктура мировых рынков сырья, устойчивый рост цен на энергоносители обеспечивали существенный приток иностранной валюты в российскую экономику. Рост цен на нефть с 25 долларов за баррель Brent в 2002 г. до 73 долларов за баррель в 2007 г. привел к значительному увеличению доходов бюджета. Сальдо счета текущих операций платежного баланса увеличилось с 28 млрд долларов США в 2002 г. до 74 млрд долларов в 2007 г., а профицит бюджета расширенного правительства за тот же период вырос с 1,5% ВВП до 6,1% ВВП. Положительное сальдо счета текущих операций в условиях режима квазификсированного валютного курса монетизировалось Банком России через проведение валютных интервенций вплоть до мирового финансового кризиса.

В 2006–2007 гг. приток иностранной валюты в страну по текущему счету (92,3 млрд долларов США и 72,2 млрд долларов США)

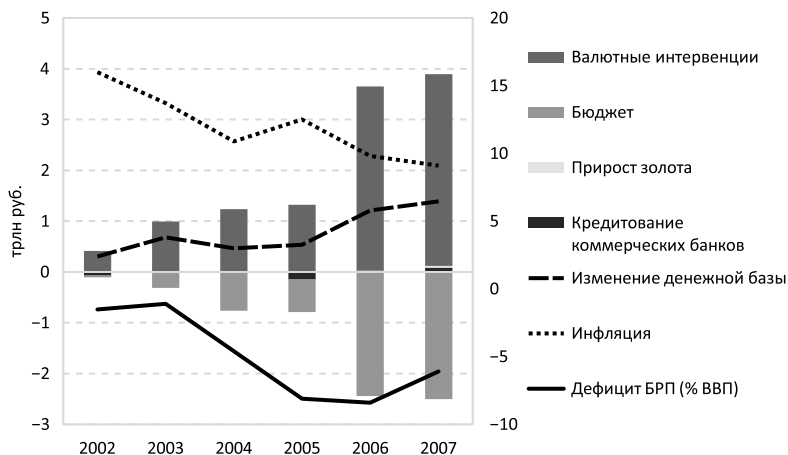


Рис. 2. Источники формирования денежной базы (трлн рублей), инфляция (%) и дефицит бюджета расширенного правительства (% ВВП) в 2002–2007 гг.

Источник: составлено авторами.

сопровождался также притоком капитала в частный сектор российской экономики (43,7 млрд долларов США в 2006 г., 87,8 млрд долларов США в 2007 г.).

На протяжении большей части периода с 2002 по 2007 г. увеличение международных резервов происходило одновременно с сопоставимым по величине ростом денежной базы (рис. 2). Необходимость устранения монетарных предпосылок для инфляции, отсутствие эффективных инструментов для операций на открытом рынке стали стимулом для использования альтернативных инструментов стерилизации избыточного предложения денег. Основным из них стало накопление профицита государственного бюджета на счетах Минфина России в Федеральном казначействе (средства Правительства РФ на счетах в банке России). Экспортная выручка выкупалась Банком России за счет эмиссии денежной базы, значительная часть которой затем изымалась в форме налогов и пошлин у экспортеров и перечислялась на счета Минфина России в ЦБ РФ. Общий чистый прирост денежной базы

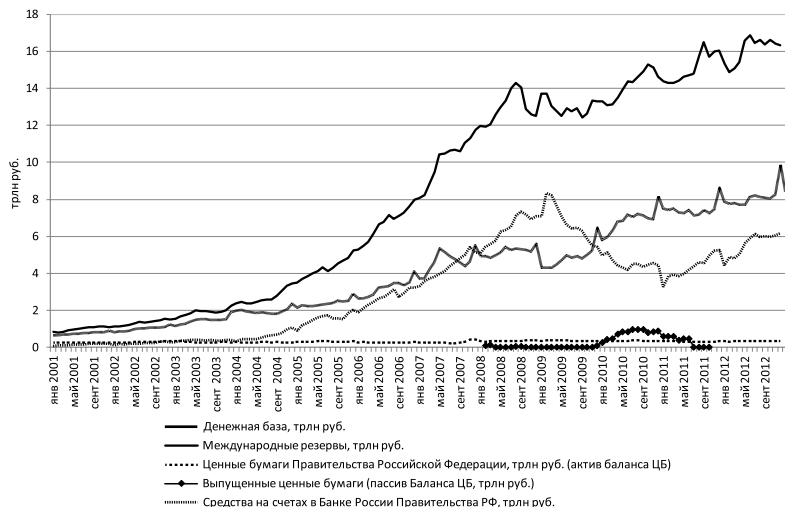


Рис. 3. Динамика международных резервов и отдельных составляющих баланса Банка России (млн рублей) в январе 2001 — декабре 2012 г.

Источник: ЦБ РФ.

за 2002–2007 гг. составил 4,6 трлн рублей, что стало результатом роста денежной базы на 11,4 трлн в результате валютных интервенций ЦБ РФ и стерилизации денежной базы в размере 6,7 трлн рублей в результате политики Минфина РФ.

При этом государство активно гасило взятые ранее зарубежные валютные кредиты: с 2002 по 2007 г. внешний государственный долг сократился с 103 млрд долларов до 44,7 млрд долларов. Эти операции приводили к одновременному сокращению резервов и снижению остатков на бюджетных счетах и поэтому не оказывали влияния на денежную базу. Таким образом, в указанном периоде расширение денежного предложения осуществлялось денежными властями путем покупки валюты на валютном рынке, а бюджетная политика, напротив, оказывала сдерживающее влияние на инфляцию.

В среднем за каждый год рассматриваемого периода денежная база росла на 35%, и денежная масса росла близкими темпами

(36,7% по агрегату M2X). Среднегодовые темпы роста номинального ВВП были равны 24,6%. Таким образом, рост денежного предложения опережал рост номинального ВВП, что привело к увеличению монетизации экономики. Инфляция в 2002–2007 гг. следовала нисходящему тренду, медленно снижаясь с 15,8% в 2002 г. до 9,0% в 2007 г. Среднегодовые темпы инфляции составили 12%.

2008–2010 годы: мировой финансовый кризис — острая фаза и восстановление

В кризисный и посткризисный периоды валютные интервенции ЦБ перестали быть ключевым фактором роста денежной базы. В конце 2008 — начале 2009 г. Банк России проводил масштабные валютные интервенции, пытаясь предотвратить обесценение рубля в условиях быстрого снижения цен на нефть. В 2008 г. началась острая фаза глобального финансового кризиса, для которой было характерно падение цен на мировых рынках сырьевых товаров, включая нефть, и отток капитала с рынков развивающихся стран, что приводило к сильнейшему понижательному давлению на обменные курсы их валют. В России ЦБ РФ, стремясь противодействовать ослаблению рубля, осуществлял валютные интервенции, продавая валюту и изымая рублевую денежную базу. В результате за полгода объем международных резервов РФ сократился более чем на треть с 598 млрд долларов США в августе 2008 г. до 380 млрд долларов США в феврале 2009 г. Вклад операций по продаже валюты для поддержания курса в сокращение денежной базы в 2008 г. составил 2,2 трлн рублей, притом что на начало 2008 г. денежная база в широком определении составляла 5,5 трлн рублей. Поскольку продажа валюты приводила к снижению ликвидности банковского сектора, ЦБ РФ одновременно экстренно и в больших объемах предоставлял средства банковскому сектору. На пике кризиса в начале 2009 г. суммарная задолженность кредитных организаций перед Банком России составила около 3 трлн рублей, тогда как до кризиса подобная задолженность практически отсутствовала.

Проводимая бюджетная политика вела к сокращению денежной базы, поскольку в 2008 г. бюджет расширенного правитель-

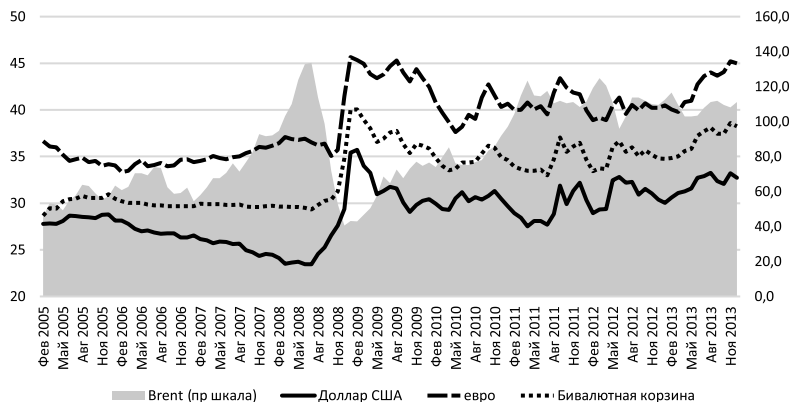


Рис. 4. Курсы доллара, евро, бивалютной корзины (рубли) и цена нефти Brent (доллар/баррель)

Источник: Банк России, IMF.

ства был сведен с профицитом около 2 трлн рублей (4,9% ВВП), что привело к увеличению остатков на счетах правительства в ЦБ и сокращению денежной базы на 1,6 трлн рублей. Таким образом, по итогам 2008 г. денежная база осталась практически неизменной, так как продажи валюты центральным банком и накопление правительством средств на счетах снизило денежную базу на 3,8 трлн рублей, но это сокращение было компенсировано расширением кредитования банков со стороны Банка России.

В последующие два года произошел разворот данных тенденций: покупка валюты ЦБ РФ совместно с расходованием средств Минфином России при дефицитном бюджете приводили к росту денежной базы, в то время как кредитные организации гасили кредиты, взятые у ЦБ РФ в период острой фазы кризиса, что снижало денежную базу. После мирового финансового кризиса Банк России продолжил проводить валютную политику в рамках режима управляемого плавления валютного курса. Гибкость механизмов курсообразования существенно возросла. Теперь ЦБ РФ не фиксировал курс, а устанавливал коридор, в рамках которого курс мог свободно плавать. При достижении курсом границы коридора ЦБ РФ осуществлял интервенции для удержания курса

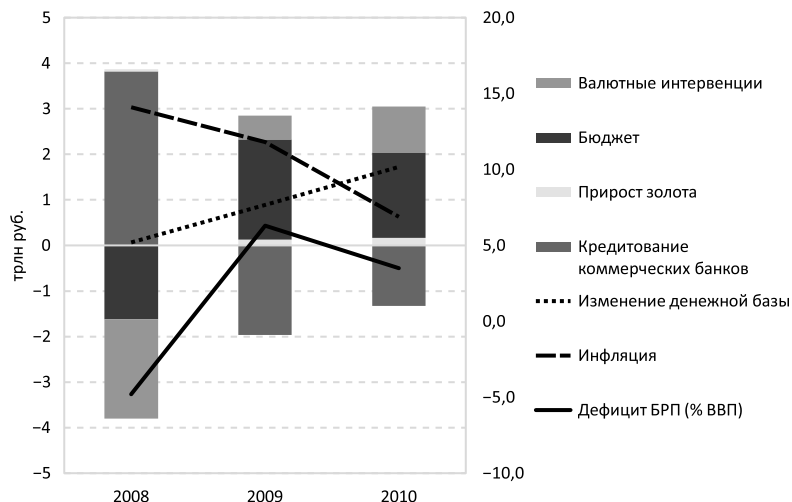


Рис. 5. Источники формирования денежной базы (трлн рублей), инфляция (%) и дефицит бюджета расширенного правительства (% ВВП) в 2008–2010 гг.

Источник: составлено авторами.

в пределах диапазона, но после исчерпания некоторого лимита на поддержание курса происходил сдвиг коридора. С февраля 2010 г. Банк России приступил к расширению операционного интервала допустимых значений рублевой стоимости бивалютной корзины, а также стал сокращать величину накопленных валютных интервенций, приводящих к сдвигу границ операционного интервала. Интервенционная политика в данный период проводилась несимметрично: в моменты наиболее серьезного укрепления номинального курса рубля к бивалютной корзине ЦБ РФ активно прибегал к интервенциям, практически не вмешиваясь в функционирование валютного рынка в периоды ослабления рубля. Высокие цены на нефть в посткризисные годы и растущее сальдо платежного баланса (50 млрд долларов в 2009 г. и затем 70 млрд и 100 млрд долларов в 2010 и 2011 г. соответственно) стали важным фактором роста денежной базы и накопления резервов центральным банком.

В результате падения нефтегазовых доходов и роста расходов, связанных с антикризисными программами, бюджет расширенного правительства в 2009 г. был сведен с дефицитом, равным 2,5 трлн рублей. В 2010 г. дефицит сократился и составил 1,6 трлн рублей. Так как источником финансирования дефицита были преимущественно остатки средств на счетах, это привело к увеличению денежной базы в 2009 и 2010 г. на 2,2 и 1,9 трлн рублей соответственно. Сворачивание антикризисной политики в 2011 г. привело к возвращению бюджетного профицита.

Всего за период 2008–2010 гг. денежная база выросла на 2,7 трлн рублей. Среднегодовой прирост денежной базы составил 14,2% и в основном этот прирост реализовался в 2009 и 2010 г. Основной вклад в данный прирост внесли операции Минфина России по финансированию дефицита бюджета (+2,5 трлн рублей), в то время как операции с валютой сократили денежную базу на 0,3 трлн рублей. Изменение денежной базы, обусловленное рефинансированием банковского сектора, составило всего около 0,5 трлн. Таким образом, в 2008–2010 гг. денежная база росла преимущественно из-за того, что Минфин России, осуществляя расходы на антикризисные программы, возвращал в экономику средства из Резервного фонда, изъятые из нее в предшествующий период, когда бюджет был профицитным. Валютная политика, напротив, вела к изъятию ликвидности ввиду исключительных размеров продажи валюты в 2008 г. При этом впервые значимым каналом формирования денежной базы стали операции рефинансирования банков. Темпы роста широкой денежной массы за рассматриваемый трехлетний период были равны 20,3%, что выше темпов роста денежной базы. Это объясняется эффектом переоценки валютных депозитов, поскольку они включаются в широкую денежную массу, а за указанный период рубль ослаб против ведущих валют на 22%. Среднегодовая инфляция составила 10,9%, а всего цены выросли за 2008–2010 гг. на 36%.

2011–2013 годы: посткризисный период

Данный период выделяется, во-первых, тем, что ЦБ РФ повысил гибкость курсообразования, установив достаточно широкий коридор для возможных колебаний курса рубля и сократил

интервенции на валютном рынке, расширив при этом операции рефинансирования банков, а во-вторых, нормализацией бюджетных показателей после повышенных дефицитов посткризисных лет. Это привело к тому, что расширение денежной базы происходило преимущественно за счет кредитов, предоставленных банкам от ЦБ РФ, а роль валютных интервенций снизилась. Иначе говоря, практика эмиссии за счет покупок валюты была замещена практикой эмиссии под обязательства банков, что стало результатом сознательной политики российских денежных властей, направленной на модернизацию монетарной политики в России, частью которой является использование ставки денежного рынка в качестве инструмента политики и регулирование ликвидности посредством операций рефинансирования.

На протяжении 2011–2012 гг. курс бивалютной корзины колебался в пределах 33–37 рублей, а ЦБ РФ осуществлял как покупки, так и продажи валюты (см. рис. 6). С весны 2013 г. на валютном рынке из-за ожиданий сворачивания программы количественного смягчения ФРС США и снижения сальдо текущего счета РФ стала усиливаться тенденция к ослаблению рубля и ЦБ РФ стал расходовать валютные резервы, сдерживая это ослабление и расширяя коридор допустимых колебаний курса. За 2013 г. курс доллара вырос с 34,8 до 38,4 рублей, при этом ЦБ потратил 17 млрд долларов резервов, сократив денежную базу на 0,9 трлн рублей.

С 1 октября 2013 г. Банком России осуществлена корректировка механизма курсовой политики, в соответствии с которой параметры операций Банка России по покупке и продаже иностранной валюты на внутреннем валютном рынке определялись с учетом операций Федерального казначейства по пополнению или расходованию средств Резервного фонда и Фонда национального благосостояния. В частности, объемы операций Банка России по покупке или продаже иностранной валюты на внутреннем валютном рынке, устанавливаемые исходя из цели сглаживания волатильности обменного курса рубля, изменялись на величину, эквивалентную объему валютных операций казначейства. Введенная корректировка позволяла сглаживать резкие колебания курса российского рубля и поддерживать стабильность валютного и денежного рынков в периоды осуществления Федеральным

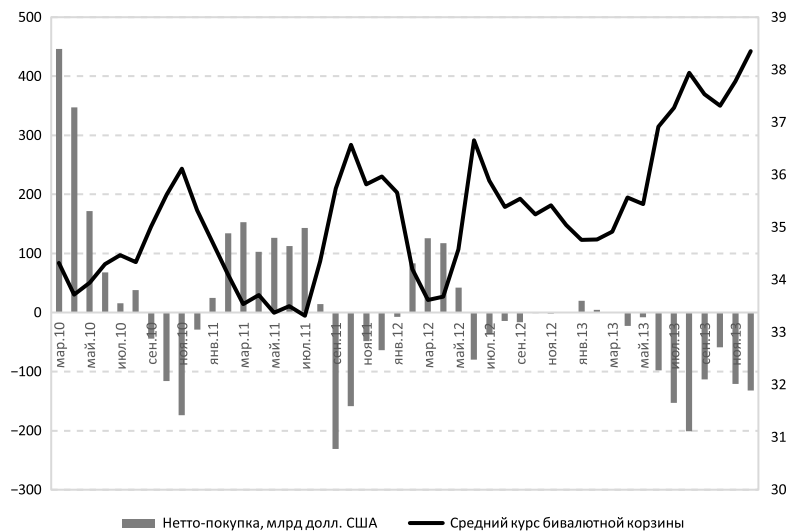


Рис. 6. Валютные интервенции ЦБ и динамика рублевой стоимости бивалютной корзины в 2010–2013 гг.

Источник: ЦБ РФ.

казначейством валютных операций. Тем не менее реализация бюджетного правила в таком виде оставалась не нейтральной по отношению к денежному предложению (сокращение денежной базы при накоплении средств в Резервном фонде и денежная эмиссия при их расходовании) и компенсировалась операциями по кредитованию банковского сектора регулятором.

Начиная с III квартала 2011 г. задолженность банков перед ЦБ РФ вновь начала увеличиваться, и к началу 2013 г. она составляла уже более 2,5 трлн рублей. Всего за 2011 г. прирост денежной базы, обусловленный расширением операций рефинансирования, составил 0,9 трлн рублей, а за 2012 и 2013 гг. соответственно 1,5 трлн рублей и 1,7 трлн рублей.

В 2011–2013 гг. Банк России реализовывал реформы, целью которых была подготовка банковской системы к переходу на режим инфляционного таргетирования и использования в качестве главного инструмента денежной политики процентной

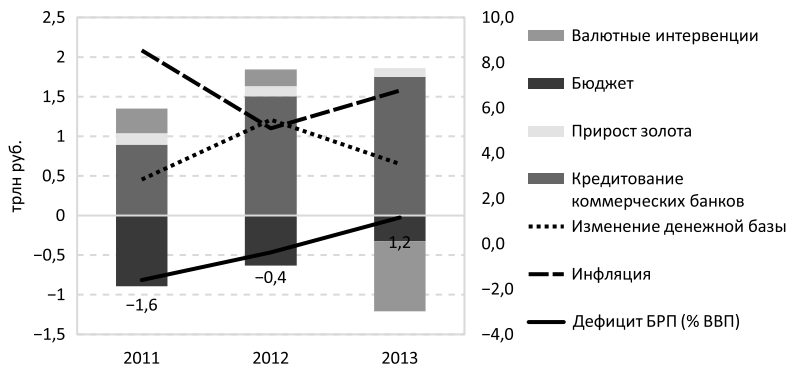


Рис. 7. Источники формирования денежной базы (трлн рублей), инфляция (%) и дефицит бюджета расширенного правительства (% ВВП) в 2011–2013 гг.

Источник: составлено авторами.

ставки по краткосрочному кредитованию. Это сопровождалось сужением процентного коридора и повышением роли операций репо как инструмента регулирования ликвидности банковского сектора. Увеличение объема операций рефинансирования было естественным следствием данного процесса.

Влияние Минфина России на динамику денежного предложения в 2011–2013 гг. было сдерживающим. Бюджет расширенного правительства был профицитным в 2011 г. (860 млрд рублей, или 1,4% ВВП) и 2012 г. (260 млрд рублей, или 0,3% ВВП), чему способствовали высокие цены на нефть, которые в эти годы находились на исторических максимумах. В связи с замедлением роста ВВП в 2013 г. бюджет расширенного правительства стал дефицитным (848 млрд рублей, или 1,3% ВВП). В соответствии с бюджетным правилом отчисления дополнительных нефтегазовых доходов в резервные фонды привязаны к мировым ценам на нефть и не связаны с текущим бюджетным балансом. В 2013 г. возникла ситуация, когда при дефицитном бюджете правительство сберегало часть доходов в фондах. Следствием превышения фактических цен на нефть в 2013 г. уровня базовой цены, заложенной в бюджетное правило, стало зачисление в Резервный

фонд дополнительных нефтегазовых доходов, что привело к сокращению денежной базы на 0,3 трлн рублей.

Темпы роста денежной базы и денежной массы в 2011–2013 г. были сравнительно низкими и близкими по величине. Среднегодовые темпы прироста денежной базы составили 9,4%, а общий ее объем за три года увеличился на 2,3 трлн рублей. Рост денежной базы происходил из-за расширения кредитования банковского сектора (+4,1 трлн рублей), при этом операции Минфина России сдерживали этот рост (–1,9 трлн рублей). Широкая денежная масса росла среднегодовыми темпами, равными 11%, при среднегодовой инфляции равной 6,8%.

2014–2017 годы: переход к инфляционному таргетированию и дезинфляционный период

Характерные черты процесса формирования денежной базы в 2014–2017 гг. были сходными с аналогичными процессами в 2008–2010 гг. Причиной обоих кризисных эпизодов стало падение цен на нефть, негативный эффект которого в 2014 г. был усилен наложением на Россию санкций и оттоком иностранных инвестиций. Хотя гибкость курсообразования в 2014 г. была выше, чем в 2008 г., но к свободному плаванию рубля Банк России перешел только осенью 2014 г. До этого момента ЦБ РФ старался поддержать обменный курс за счет продаж валюты, в результате чего за январь — ноябрь 2014 г. 140 млрд долларов международных резервов были израсходованы, что привело к потере около 16% резервов. Сокращение денежной базы по причине продаж валюты, составившее 3,4 трлн рублей, денежные власти компенсировали массированным увеличением кредитования банковского сектора, что объяснялось стремлением восполнить отток ликвидности. За 2014 г. задолженность банковского сектора перед ЦБ РФ выросла с 3,7 до 9,3 трлн рублей. Таким образом, данные операции суммарно увеличили денежную базу на 5,6 трлн рублей. Так же, как и в 2008 г., результатом операций Минфина России стало сокращение денежной базы — по итогам 2014 г. эти операции сократили денежную базу на 1,7 трлн рублей.

В течение 2015–2016 гг. банки активно гасили займы, взятые у Банка России, что сопровождалось общим сокращением денеж-

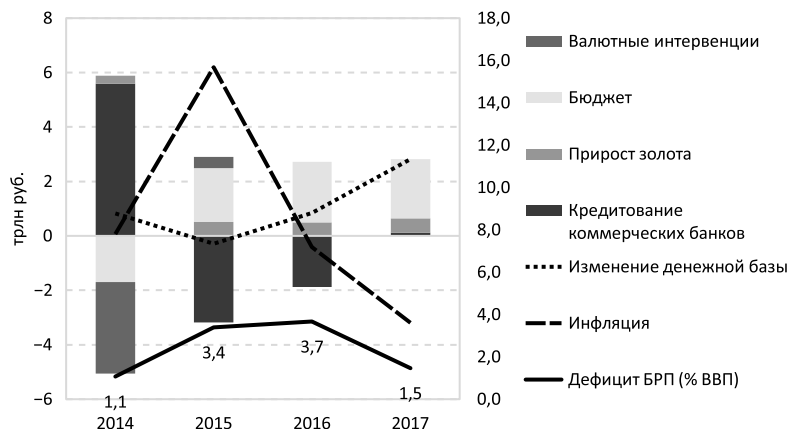


Рис. 8. Источники формирования денежной базы (трлн рублей), инфляция (%) и дефицит бюджета расширенного правительства (% ВВП) в 2014–2017 гг.

Источник: составлено авторами.

ной базы на 5,1 трлн рублей за два года. К началу 2017 г. задолженность кредитных организаций перед ЦБ РФ снизилась до уровня менее 500 млрд рублей. Поскольку в условиях экономического спада бюджет расширенного правительства стал дефицитным, правительство обратилось к средствам Резервного фонда для финансирования дефицита. В 2015 г. дефицит бюджета расширенного правительства был равен 2,8 трлн рублей (3,4% ВВП), 3,1 трлн рублей в 2016 г. (3,7% ВВП) и 1,4 трлн рублей в 2017 г. (1,9% ВВП). В связи с изменениями остатков на счетах Правительства РФ в Банке России денежная база в 2015 г. увеличилась на 2,0 трлн рублей, а в 2016 и 2017 она выросла на 2,2 трлн рублей. Соответствующий приток ликвидности в банковский сектор по бюджетному каналу стал главной причиной сокращения задолженности кредитных организаций по операциям рефинансирования. Так как в конце 2014 г. ЦБ РФ перешел к режиму свободного плавления рубля, он прекратил регулярные значимые покупки иностранной валюты с целью оказывать влияние на обменный курс. Вместе с тем Банк России с 2015 по 2017 г. ежегодно приобретал моне-

тарное золото, что увеличивало денежную базу на сумму около 0,5 трлн рублей ежегодно.

В целом период 2014–2017 гг. характеризуется высокой ролью бюджетного канала увеличения денежной базы. В условиях кризиса и увеличившихся дефицитов Правительство РФ обратилось к собственным суверенным фондам. Их расходование добавило к объему денежной массы 4,7 трлн рублей. Вклад рефинансирования в формирование денежной базы был выраженным в отдельные годы, но в целом за период по этому каналу денежная база выросла всего на 0,6 трлн рублей. Это объясняется тем, что рост кредитования банков со стороны ЦБ РФ объяснялся выросшим спросом на ликвидность со стороны банковского сектора в ситуации высокой неопределенности, который затем к началу 2017 г. вернулся к прежним уровням по мере нормализации ситуации. Всего за период 2014–2017 гг. продажа валюты для противодействия ослаблению рубля в 2014 г. и последующие покупки монетарного золота привели к сокращению денежной базы на 1,1 трлн рублей.

В период 2014–2017 гг. темп прироста денежной базы снизился до 8,1% в среднем ежегодно при ежегодном среднегодовом приросте широкой денежной массы на 11,2%. Инфляция, которая в 2015 г. поднялась до уровня 15,7% после 7,8% в 2014 г., затем достаточно быстро перешла к снижению и уже в 2017 г. оказалась на исторически минимальном уровне в 3,6%. Первоначальный инфляционный всплеск был прежде всего связан с обесценением рубля, который потерял 60% своей стоимости против ведущих валют.

2018–2019 годы: макроэкономическая стабильность

Двухлетний период 2018–2019 гг. может рассматриваться как спокойный в том смысле, что в данный период экономика России не сталкивалась со значимыми шоками, сохраняя рост ВВП на уровне 2–3% при низкой инфляции, относительно стабильном курсе рубля и профицитном бюджете. Всего за 2018–2019 гг. денежная база выросла на 2,1 трлн рублей при достаточно высоком среднегодовом темпе прироста, равном 14,7%. На протяжении 2018–2019 гг. денежная масса росла медленнее (по 8,8% в среднем

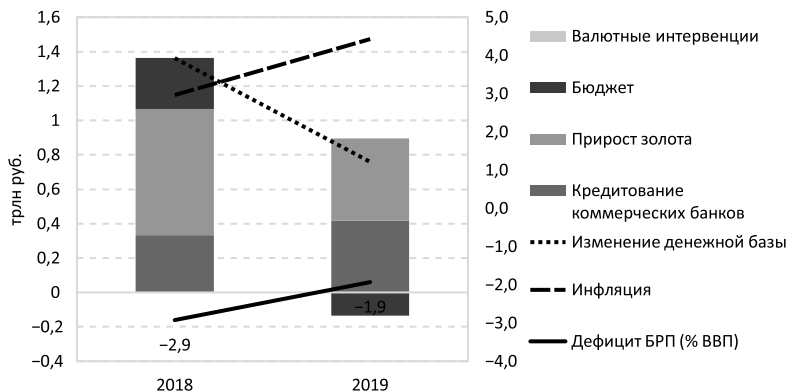


Рис. 9. Источники формирования денежной базы (трлн рублей), инфляция (%) и дефицит бюджета расширенного правительства (% ВВП) в 2018–2019 гг.

Источник: составлено авторами

за год), чем денежная база, а инфляция по итогам 2018 и 2019 гг. составила соответственно 3,0% и 4,4%.

Основной вклад в прирост денежной базы внесли операции с золотом и операции рефинансирования банков. Покупка монетарного золота Банком России повысила денежную базу на 737 млрд рублей в 2018 г. и на 477 млрд рублей в 2019 г. (+1,2 трлн рублей за два года). Аналогичные цифры, соответствующие каналу формирования денежной базы через кредитование банков, равны 332 млрд рублей и 416 млрд рублей (+0,7 трлн рублей за два года). Темпы роста широкой денежной массы были ниже, чем темп роста денежной базы.

В 2017 г. с целью обеспечения нейтральности операций со средствами правительственных фондов по отношению к денежному предложению, а также ориентации бюджетной политики на стабилизацию динамики дефицита государственного бюджета и государственного долга было введено новое бюджетное правило. Банк России приступил к покупке (продаже) иностранной валюты на внутреннем валютном рынке ЦБ РФ по поручению Минфина России в зависимости от соотношения фактиче-

ской и базовой цены на нефть². Дополнительные нефтегазовые доходы конвертировались в иностранную валюту и направлялись на пополнение ФНБ. Дополнение бюджетного правила операциями с валютой позволяет избежать колебаний денежной базы, вызванных действиями Минфина России, поскольку при покупке иностранной валюты на внутреннем рынке с целью пополнения ФНБ Банк России увеличивает денежное предложение, которое сократилось ранее на этапе сбора налоговых поступлений. Введение описанной процедуры привело к тому, что отчисления бюджетных средств в Фонд национального благосостояния не приводили к изменению объема денежной базы. Всего за 2018 г. по бюджетному каналу было добавлено 0,3 трлн рублей денежной базы, а на следующий год операции Минфина России привели к сокращению денежной базы на 0,2 трлн рублей. При этом по итогам 2018 и 2019 г. в ФНБ было перечислено 0,9 и 4,1 трлн рублей дополнительных нефтегазовых доходов соответственно, а в 2018 г. 1,1 трлн рублей был изъят из фонда для финансирования дефицита пенсионной системы. Данные операции не оказали какого-либо заметного влияния на денежную базу, что говорит об эффективной работе механизма «зеркалирования» операций Минфина России Центральным банком в этом периоде.

2020–2021 годы: кризис COVID-19 и восстановление

На 2020 и 2021 г. пришелся кризис, связанный с пандемией и локдауном, а также последующее восстановление экономики. Спад ВВП по итогам 2020 г. составил 2,7%, а в 2021 г. экономика выросла на 5,9%. В ответ на пандемию бюджетные власти России существенно увеличили расходы, направляя их на антикризисные цели. Расходы бюджета расширенного правительства в 2020 г. составили 39,6% ВВП при соответствующем значении предшествующего 2019 г. на уровне 34,1% ВВП. Падение доходов привело к тому, что в 2020 г. сформировался значительный дефицит в 4% ВВП, который покрывался в основном за счет заимствований. Норматив

² 40 долларов за баррель (нефть марки Urals) в 2017 г. и начиная с 2018 г. эта цена подлежала ежегодной 2-процентной индексации.

лизация экономической ситуации в 2021 г. привела к увеличению доходов бюджета (включая восстановление нефтегазовых поступлений, сократившихся в 2020 г. на 40%), в результате чего с учетом падения расходов в долях ВВП баланс бюджета расширенного правительства в 2021 г. оказался положительным и составил 0,8% ВВП.

В период 2020–2021 гг. денежная база увеличилась на 3,5 трлн рублей. Операции со средствами ФНБ не оказывали значительного влияния на динамику денежной базы, поскольку ЦБ РФ проводил своевременную стерилизацию. Так, за период 2020–2021 гг. из ФНБ на финансирование бюджетного дефицита было направлено 0,3 трлн рублей.

В соответствии с бюджетным правилом за этот период чистые покупки валюты Банком России по поручению Минфина России должны были составить 2,3 трлн рублей, а фактические покупки составили 2,0 трлн рублей. Следовательно, за счет отклонения объемов продажи валюты от механизма, установленного бюджетным правилом, денежная база сократилась на 0,3 трлн рублей. Одно из объяснений такого неполного зеркалирования состоит в том, что в 2020–2021 гг. бюджет реализовывал экстренные антикризисные меры, поэтому правительству были разрешены отдельные точечные траты средств ФНБ, которые не согласуются с бюджетным правилом. По-видимому, Банк России принял решение не зеркалировать в полной мере эти операции, так как они реализовывались вне регулярного бюджетного правила.

Помимо финансирования дефицита бюджета, средства ФНБ инвестировались в разрешенные активы. На эти цели в 2020–2021 гг. было направлено 3,5 трлн рублей. Согласно механизму бюджетных правил (о нейтральном влиянии операций со средствами ФНБ на денежную базу) Банк России должен был продать валюту на эту сумму, но этого не происходило, что также привело к росту денежной базы на сумму инвестирования. Итоговое влияние операций с ФНБ на денежную базу в 2020–2021 гг. составило 3,0 трлн рублей. В ситуации устойчивого профицита ликвидности, сформировавшегося еще в 2017 г. в результате эмиссии за счет расходования средств суверенных фондов, рост денежной базы в результате предоставления кредитов банковскому сектору был отрицательным и составил лишь -0,3 трлн рублей. Другими

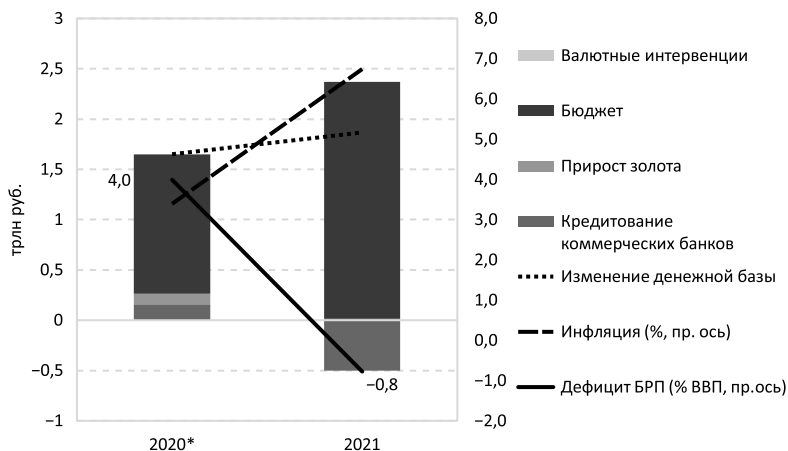


Рис. 10. Источники формирования денежной базы (трлн рублей), инфляция (%) и дефицит бюджета расширенного правительства (% ВВП) в 2020–2021 гг.

Источник: составлено авторами.

словами, кредитные организации, получая приток ликвидности по бюджетному каналу, предпочитали гасить имеющиеся обязательства. Прирост денежной базы, обусловленный покупкой монетарного золота, составил всего 0,1 трлн рублей.

По итогам двухлетнего периода можно заметить, что основным источником роста денежной базы стали неотзеркалированные операции Минфина России, в то время как операции с банками приводили к сокращению денежного предложения. Денежная масса за два года увеличилась почти на 30%. Ключевая ставка была снижена с приходом пандемии, и к ее увеличению Банк России перешел только весной 2021 г. по мере ускорения инфляции, которая после умеренно превышающего цель значения в 4,9% в 2020 г. ускорила в 2021 г. до 8,4%. Хотя ограничения на стороне совокупного предложения, обусловленные карантинными мерами, послужили дополнительным источником роста цен в 2021 г., есть основания утверждать, что к числу важных факторов ускорения инфляции в период восстановления экономики можно отнести рост денежного предложения, связанный с бюджетными операциями.

2022–2024/2025 годы: обострение геополитической напряженности, бюджетная экспансия и дезинфляционная политика

С началом специальной военной операции в 2022 г. бюджетные расходы значительно выросли и в целом роль бюджета как фактора совокупного спроса значительно усилилась. В рассматриваемый трехлетний период бюджет расширенного правительства был умеренно дефицитным: в 2022 г. дефицит был равен 1,3% ВВП, в 2023 г. — 2,2%, в 2024 г. — 1,6%. При этом для финансирования дефицита использовались не только заемные средства, но также и средства ФНБ. Большая часть изъятий средств из ФНБ на финансирование дефицита и инвестиционных проектов при- шлась на 2022–2023 гг. (8,8 трлн рублей).

Это привело к тому, что в период расходования средств ФНБ без соответствующего зеркалирования бюджетные расходы стали одной из причин ускорения роста денежной базы. Так, в период с 2022 по 2024 г. основной вклад в увеличение денежной базы внесли операции Минфина России со средствами ФНБ, которые не были стерилизованы (6,1 трлн рублей, или 80% от общего при- роста денежной базы).

Отсутствие зеркалирования со стороны Банка России связано с приостановкой действия бюджетного правила в марте 2022 г., что также предполагало отсутствие регулярных продаж валюты со стороны ЦБ РФ. Помимо изъятий средств из ФНБ, обусловлен- ных бюджетным правилом, в особых «разовых» случаях Правитель- ство также использовало средства ФНБ для антикризисных мер. Подобные «разовые» изъятия Банк России не зеркалировал вплоть до 2024 г., когда он перешел к полному зеркалированию всех опера- ций Минфина России с ФНБ, включая разовые изъятия и вложения средств ФНБ в российские активы. Сам ЦБ РФ не давал официаль- ных пояснений о причинах, по которым он не продавал валюту, компенсируя эффект от расходования средств ФНБ на денежное предложение. Возможное объяснение состоит в том, что на тот момент денежные власти считали, что подобные вмешательства в валютный рынок в условиях санкций и чрезмерной неопреде- ленности создаст слишком высокие риски быстрого расхода- ния международных резервов. Заметим, что во втором полугодии

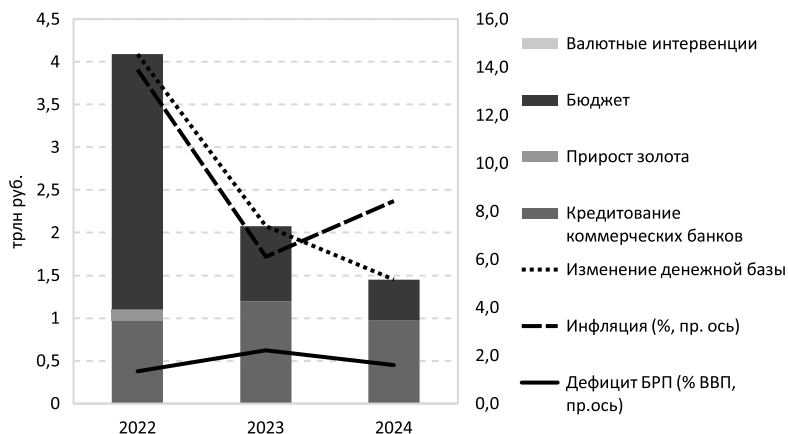


Рис. 11. Источники формирования денежной базы (трлн рублей), инфляция (%) и дефицит бюджета расширенного правительства (% ВВП) в 2022–2024 гг.

Источник: составлено авторами.

2023 г. ЦБ РФ отказался от регулярных покупок валюты в рамках обновленного бюджетного правила, ссылаясь на то, что такие покупки дестабилизируют валютный рынок. К осени 2023 г. рубль значительно ослаб. Курс доллара вырос к августу до 95 рублей, при том что в начале года курс доллара был ниже 70 рублей.

Положительный вклад в динамику денежной базы внесло кредитование коммерческих банков Банком России: в 2022–2024 гг. оно достигло 3,1 трлн рублей. В период с 2022 по 2024 г. прирост денежной базы за счет покупки золота Банком России составил 0,1 трлн рублей. Отрицательный вклад (–1,7 трлн рублей) в изменение денежной базы в период 2022–2024 гг. связан с изменением остатков бюджета на счетах в Банке России и с прочими операциями, влияющими на денежную базу (например, с требованиями Банка России к другим финансовым организациям). Всего за три года денежная база выросла на 38%.

После первого шока весной 2022 г., когда выпуск сократился на 1,4%, в последующие два года темпы экономического роста были высокими и составили 4,1% и 4,3%. Данный рост сопро-

вождался быстрым приращением кредитования и увеличением денежной массы, объемы которых за три года увеличились соответственно на 61% и на 57%. Интенсификация экономической активности привела к ускорению инфляционных процессов: по итогам 2022 г. цены выросли на 13,9%, а по итогам 2023 и 2024 г. — на 6,1% и 8,4% соответственно. Реакция ЦБ РФ на ускорение инфляции последовала только в августе 2023 г., когда он начал цикл ужесточения политики. Рост ключевой ставки был значительным — если в июне 2023 г. она составляла 7,5%, то к концу того года выросла до 16%. Однако запоздалый переход к ужесточению стал дополнительным фактором увеличения денежного предложения и роста цен.

Важным отличием последнего периода от остальных периодов, рассмотренных выше, является сонаправленный вклад увеличения кредитования банков и операций с бюджетом в рост денежной базы. До 2022 г. вклады денежной и бюджетной политики в денежную базу противодействовали друг другу. Как правило, противоположная направленность воздействия объяснялась совместным стремлением Банка России компенсировать возникающий в результате операций Минфина России отток и приток ликвидности. Банковская система прошла в 2008–2010 гг. и 2014–2017 гг. два цикла оттока-притока средств по бюджетному каналу с одновременным накоплением-погашением задолженности банков перед ЦБ РФ. Однако в 2022–2024 гг. рост денежной базы в результате операций с ФНБ шел одновременно с увеличением кредитования Банком России кредитных организаций. Также в период 2022–2024 гг. банки увеличили кредитование реального сектора и населения, что во многом стало результатом программ льготного кредитования.

Заметную роль в инфляционных процессах последних лет сыграл фактор фискальной политики, которая оказала сильный стимулирующий эффект на экономическую активность. В целом бюджетно-налоговую политику после 2022 г. можно охарактеризовать как экспансионистскую, причем значимый сдвиг в политике произошел в 2023 г. В 2022 г. государственные расходы росли как в номинальном выражении, так и в долях ВВП. Следствием опережающего роста расходов по отношению к доходам стал от-

рицательный баланс как федерального бюджета, так и бюджета расширенного правительства, притом что в 2021 г. оба бюджета были профицитными.

Начиная с 2023 г. проявляется новое качество бюджетно-налоговой политики — она перестает быть контрциклической, как в предшествующие годы. Для российской бюджетно-налоговой политики до 2023 г. было характерно сонаправленное изменение бюджетного баланса и темпов роста выпуска. В периоды, когда темпы роста ВВП России увеличивались, наблюдалось повышение бюджетного баланса (за счет роста доходов и снижения расходов в долях ВВП), а в период торможения экономической активности, напротив, по причине падения бюджетных доходов и роста расходов на поддержку экономики рос бюджетный дефицит. В этом отношении период, начавшийся в 2023 г., существенно отличается, поскольку в эти годы экономика вышла на высокие темпы роста экономики (среднегодовой рост ВВП в 2023–2024 г. составил 4,2%) после спада на 1,4% в 2022 г., и при этом происходило заметное расширение дефицита бюджета расширенного правительства до 2,2 и 1,6% ВВП в 2023 и 2024 г. соответственно, после дефицита в 1,3% ВВП в кризисный 2022 г. Такие величины дефицитов выглядят умеренно в сравнении с дефицитами, характерными для развитых стран, однако для России, где фискальная политика в последние десятилетия была достаточно консервативной, такое сочетание высокой экономической активности, рекордно низкой безработицы и сравнительно высоких бюджетных дефицитов является необычным. Таким образом, можно говорить о наличии признаков заметной бюджетной экспансии после 2023 г., которая привела к росту совокупного спроса. Получение достоверных количественных оценок вклада бюджетной экспансии в рост потребительских цен затруднено, но тем не менее есть достаточные основания полагать, что экспансия стала одним из важных факторов ускорения инфляции последних лет.

Таким образом, с 2022 г. российская экономика испытала действие различных сильных проинфляционных шоков, к числу которых относятся отрицательные шоки предложения (рост издержек производителей по причине санкций, ухудшение условий

**Таблица 1. Избранные показатели денежного обращения
в соответствии с выбранной периодизацией**

		2002– 2007	2008– 2010	2011– 2013	2014– 2017	2018– 2019	2020– 2021	2022– 2024
Изменение денежной базы	(среднегодовые темпы роста, %)	35,0	14,2	9,4	8,1	14,7	9,9	11,6
	(прирост за период, %)	493,9	48,6	28,2	40,1	14,4	20,9	37,5
	(трлн рублей)	4,6	2,7	2,3	4,2	2,1	3,5	7,6
в том числе из-за:								
операций с ФНБ и остатками бюджета на счетах в ЦБ РФ*	(трлн рублей)	–6,7	2,5	–1,9	4,7	0,2	3,7	4,3
Кредитование Банком России коммерческих банков	(трлн рублей)	–0,2	0,5	4,1	0,6	0,7	–0,3	3,1
интервенции ЦБ РФ, покупки золота и др.	(трлн рублей)	11,4	–0,3	0,0	–1,1	1,2	0,1	0,1
Изменение М2Х	(среднегодовые темпы роста, %)	36,7	20,3	17,9	11,2	8,8	12,2	15,5
	(прирост за период, %)	568,4	66,7	56,9	76,0	18,1	29,8	56,6
	(трлн рублей)	12,2	9,5	13,5	17,3	9,9	19,2	47,4
Мультипликатор		2,4	3,1	3,6	4,4	3,7	3,9	4,3
Прирост кредита:	%	48,7	13,1	80,7	36,7	23,2	35,9	61,1
предприятиям	%	44,3	13,8	63,0	42,8	15,3	33,2	67,6
населению	%	82,8	10,6	143,8	22,3	45,0	42,0	47,5
М2Х/ВВП	%	31,2	43,5	48,6	58,8	59,0	66,1	63,1
Изменение номинального ВВП (за скользящий год)	%	24,6	12,5	15,3	5,9	9,8	6,4	16,3
Изменение реального ВВП (за скользящий год)***	%	7,1	1,0	н/д	0,2	2,3	0,8	2,6
ИПЦ	%	12,0	10,9	6,8	8,6	3,7	5,1	9,5
Валютный курс	%	–20,0	24,5	10,3	76,0	–7,0	6,3	36,9
Реальный эффективный валютный курс	%	47,1	7,0	7,2	–14,1	–4,8	–9,5	4,3

Источник: расчеты авторов по данным Росстата, Банка России.

торговли, рост неопределенности и инфляционных ожиданий) и положительные шоки спроса (бюджетная экспансия, рост инфляционных ожиданий). Российский опыт последних лет показал, что возможности регулятора по охлаждению совокупного спроса в описанных условиях существенно ограничены.

При этом Банк России не имел возможности предотвратить инфляционный шок или существенно его смягчить по двум причинам. Во-первых, для этого необходимо было начать цикл ужесточения политики значительно раньше июля 2023 г. Однако макроэкономические данные, доступные в первом полугодии 2023 г., включая данные о динамике инфляции, давали противоречивую картину и не указывали явно на наличие признаков надвигающейся новой инфляционной волны. Поскольку решения о ставке оказывают влияние на инфляцию с лагом в 3–6 кварталов, возможности своевременно и эффективно охладить спрос у денежных властей практически не было.

Во-вторых, уровень ставок, необходимый для стабилизации ценовой динамики, был настолько высок, что с заметной вероятностью мог спровоцировать разворачивание масштабного кризиса ликвидности с угрозой потери финансовой стабильности и устойчивости банковского сектора, а также вызвать каскад банкротств предприятий реального сектора с разрывом производственных цепочек и риском быстрого сползания экономики в рецессию. В данных обстоятельствах руководство ЦБ РФ приняло решение реализовать сценарий мягкой посадки ценой временного ускорения потребительской инфляции.

2.3. Выводы

1) Политика фиксированного обменного курса неоптимальна, поскольку не обеспечивает ценовой стабильности и при этом делает экономику уязвимой к внешним шокам. Такая политика может проводиться в условиях недостаточного развития финансовой системы и институтов госрегулирования, поскольку имплементация режима фиксированного курса не требует сложных институциональных структур. Вместе с тем режим фиксированного

курса делает экономики более уязвимыми ко внешним шокам и к тому же обладает рядом черт, делающих его неустойчивым.

2) С точки зрения обеспечения ценовой стабильности оптимальным представляется сочетание в меру консервативной фискальной политики с бюджетным правилом при полном зеркалировании операций бюджета для нейтрализации эффекта расходования/накопления средств ФНБ. Обменный курс при этом должен быть плавающим, а денежные власти должны таргетировать инфляцию. Примером такого периода является период 2018–2019 г., когда экономический рост на уровне 2–3% сочетался с инфляцией, близкой к целевому уровню.

3) Эффективность доступных центральному банку инструментов антиинфляционной политики существенно снижается в условиях, когда бюджетная политика значительно отклоняется от контрциклического режима и приводит к значительному стимулированию совокупного спроса. Для стабилизации соответствующего положительного шока спроса со стороны бюджетной политики денежным властям необходимо значительно ужесточать монетарную политику, что создает предпосылки для кризиса в финансовом секторе и обостряет диспропорции между сектором экономики.

3. Оценка последствий денежно-кредитной политики в зависимости от фискального контекста

3.1. Эконометрические исследования влияния фискальной политики на эффективность монетарных инструментов и элементы теории

В последние десятилетия денежно-кредитная политика (ДКП) играла центральную роль в стабилизации экономической активности и инфляции. После глобального финансового кризиса 2008 г. центральные банки как развитых, так и развивающихся экономик расширили набор инструментов, сочетая традиционные изменения ключевых ставок с нетрадиционными мерами (QE, forward guidance). Однако практика показала, что эффектив-

ность этих мер неоднородна: при прочих равных однотипные решения центрального банка приводили к разным макроэкономическим последствиям. Этот факт переместил фокус исследований со «среднего» эффекта ДКП на контекст ее проведения, в том числе и на взаимодействие ДКП с фискальной политикой [Горюнов и др. 2021].

Фискальная политика (через государственные расходы, налоги и управление долгом) напрямую влияет на совокупный спрос, относительные цены и распределение доходов, а также формирует ожидания домохозяйств, фирм и финансовых рынков. В условиях высокой долговой нагрузки или активного бюджетного стимулирования монетарные импульсы могут быть усилены либо ослаблены. Например, при фискальной консолидации ужесточение ДКП, как правило, усиливает дезинфляционные эффекты за счет снижения спроса. Напротив, при фискальной экспансии те же монетарные меры могут оказаться менее действенными из-за компенсирующего действия государственных расходов или налоговых послаблений, которые поддерживают совокупный спрос.

Теоретически вопрос взаимодействия фискальной и монетарной политики занимает центральное место в работе [Sargent, Wallace 1981], где показано, что при фискальном доминировании возможности центрального банка по контролю инфляции ограничены. Впоследствии в статье [Leeper 1991] были формализованы режимы «активной/пассивной» политики, М. Вудфорд [Woodford 1995] развил фискальную теорию уровня цен, а Симс [Sims 2011] акцентировал роль смены режимов во времени: фискальные дисбалансы и особенности бюджетного финансирования меняют эффективность монетарной политики через совокупный спрос, процентные ставки и инфляционные ожидания. Анализ взаимосвязи монетарной и фискальной политики с опорой на методологию Липера проделан в ряде работ [Белова, Перевышин 2017ab; Аблаев 2025].

Практическое измерение этой взаимосвязи хорошо иллюстрирует российский опыт: как отмечают Дробышевский и соавторы в статье [Дробышевский и др. 2025], нестерилизованные бюджетные операции, прежде всего использование средств ФНБ, расширяли денежное предложение и ослабляли дезинфляционный

эффект жесткой ДКП. Этот пример подчеркивает необходимость системно учитывать фискальный режим при оценке силы трансмиссии монетарных шоков.

Восприятие бюджетных операций влияет на инфляционные ожидания: домохозяйства и бизнес видят, что, несмотря на высокую ставку, государство продолжает активно тратить, стимулируя доходы и кредитную активность. Дополнительно фискальная политика может менять эффективность монетарной трансмиссии через поведение государственных компаний, которые часто ориентируются не только на рыночную стоимость заимствований, но и на бюджетную поддержку, а также через систему льготных кредитов, субсидированных ставок и госгарантий. Если значимая часть инвестиционной и потребительской активности завязана на такие программы, то чувствительность экономики к изменениям ключевой ставки снижается: для получателей льготного финансирования фактическая цена кредита определяется не решением центрального банка, а параметрами бюджетных программ. Тогда ужесточение ДКП оказывает более сильное влияние на те сегменты экономики, которые полностью зависят от рыночных ставок, и заметно меньше на сектора, интегрированные в систему прямой или косвенной фискальной поддержки. Отсюда возможна ситуация, когда при формально высокой ставке совокупные финансовые условия для части экономики остаются мягкими, что приводит к более слабой реакции инфляции и выпуска на монетарные шоки.

Таким образом, оценка эффективности ДКП требует учета не только текущих макроэкономических индикаторов, но и фискального контекста: бюджетного режима (консолидация или экспансия), параметров долговой устойчивости, масштабов и формы участия государства в кредитовании экономики. Поэтому важен исследовательский вопрос о том, изменяется ли трансмиссия монетарных шоков в зависимости от режима фискальной политики и при каких условиях бюджетная политика усиливает или смягчает воздействие ДКП на совокупный спрос и инфляцию.

Эмпирические исследования последних лет все чаще фиксируют, что эффект денежно-кредитной политики может зависеть от фискального контекста — уровня государственного долга,

бюджетного баланса и координации решений правительства и центрального банка. Если в ранних работах эффективность монетарных шоков оценивалась как усредненная величина, то современная литература стремится выявить режимную и структурную зависимость трансмиссии, а также различия между фазами фискальной политики.

В недавнем исследовании [Kloosterman et al. 2024], проведенном на данных 10 стран еврозоны за 1999–2019 гг., авторы используют метод локальных проекций с плавным переходом между режимами фискальной политики (panel smooth transition LP), где границей служит изменение циклически скорректированного первичного сальдо бюджета. Показано, что один и тот же монетарный шок дает противоположные результаты в зависимости от фискального режима. Смягчение монетарной политики повышает выпуск и инфляцию лишь при мягкой фискальной политике, тогда как в периоды консолидации бюджета реакции на аналогичный монетарный шок статистически незначимы или даже отрицательны. Аналогично, ужесточение ДКП снижает выпуск и цены только тогда, когда бюджетная политика ограничивает расходы. Интерпретация авторов базируется на так называемом эффекте богатства — фискальные решения меняют богатство частного сектора и, следовательно, модифицируют поведение домохозяйств и фирм в ответ на изменение процентных ставок. В совокупности результаты [Kloosterman et al. 2024] подтверждают, что эффективность ДКП зависит от согласованности с бюджетным курсом.

Схожий вывод получают авторы работы [Caramp, Feilich 2024]. Используя новокейнсианскую модель с нерикардиянской фискальной политикой и безопасным уровнем государственного долга, они показывают, что рост долговой нагрузки снижает эластичность выпуска и занятости по процентной ставке. При высоком долге повышение ставки приводит к положительному эффекту богатства для держателей государственных облигаций, то есть часть населения получает больший доход от процентных выплат, что компенсирует сокращение спроса. Авторы оценивают реакцию промышленного производства и безработицы на монетарные шоки и обнаруживают, что при увеличении долга на одно стандартное отклонение реакция промышленного производства на монетарное

ужесточение снижается примерно на 0,75 п.п., а реакция безработицы — на 0.1 п.п. Тем самым подтверждается, что высокий уровень долга ослабляет трансмиссию монетарной политики, делая экономику менее чувствительной к изменениям ставки.

Эмпирический анализ в международном разрезе представлен в исследовании [Deb et al. 2023]. Используя данные 33 стран за 1991–2023 гг., авторы показывают, что монетарная политика наиболее действенна, когда ее направление совпадает с направлением фискальной политики, то есть при координации действий (например, одновременное ужесточение для борьбы с перегревом или совместное смягчение в кризис). Когда же политики движутся в противоположных направлениях, эффективность каждой из них снижается.

Отдельный пласт литературы посвящен влиянию структуры долга и финансовых условий на передачу ДКП. Так, Ким и Лим [Kim, Lim 2020] исследуют 23 страны за 1984–2015 гг. и показывают, что при высоком уровне частного долга и распространенности кредитов с плавающей ставкой реакция потребления и инвестиций на монетарные шоки усиливается. Напротив, при высоком уровне государственного долга чувствительность совокупного спроса к ставке снижается, что согласуется с результатами [Cagarmp, Feilich 2024]. Эти два результата подчеркивают двойственность долгового фактора: частный долг усиливает, а государственный ослабляет действие монетарных мер.

Роль временной изменчивости и институциональных условий подробно анализируется в работе [Frigerio, Magnifico 2024], где применяется модель TVP-VAR для анализа данных по странам Европейского союза. Авторы показывают, что после 2010-х гг. зависимость реакции экономики на монетарные шоки от бюджетного режима возросла: в периоды консолидации инфляция реагирует слабее, а в периоды бюджетного стимулирования выпуска — сильнее. В интерпретации авторов это отражает «новую нормальность» — высокий долг и мягкая фискальная политика делают традиционные инструменты центрального банка менее эффективными.

Авторы исследования [Cloyne et al. 2020] обнаружили, что величина фискального мультипликатора существенно зависит

от поведения монетарных властей. Используя разложение Kitagawa—Blinder—Oaxaca в рамках метода локальных проекций, авторы количественно оценивают взаимодействие фискальной и монетарной политики и демонстрируют, что единого «фискального мультипликатора» не существует. Его значение эндогенно определяется степенью монетарного отклика: при активной (жесткой) реакции денежно-кредитной политики фискальный мультипликатор близок к нулю, тогда как в условиях пассивного или более мягкого режима он может достигать двух. Этот результат подтверждает, что эффективность фискальных мер неотделима от монетарного режима, в рамках которого они реализуются. Аналогичная зависимость, вероятно, характерна и для монетарного мультипликатора, что еще раз подчеркивает необходимость оценивать последствия денежно-кредитной политики с учетом фискального контекста и поведения бюджетных властей.

В совокупности эмпирическая литература последних лет демонстрирует устойчивый вывод: фискальная среда существенно меняет силу и направление действия денежно-кредитной политики. В странах и периодах с высокой долговой нагрузкой и бюджетной экспансией монетарные меры оказываются менее эффективными в подавлении инфляции и регулировании спроса. При этом эффект неоднороден не только между странами, но и во времени: одна и та же экономика может демонстрировать разные реакции на монетарные шоки в зависимости от текущего фискального режима. Это делает учет взаимодействия двух политик ключевым элементом при оценке эффективности трансмиссионного механизма денежно-кредитной политики.

3.2. Эмпирические стратегии: как измерять зависимость трансмиссии ДКП от фискального режима

Традиционно для оценки эффекта денежно-кредитной политики на макроэкономические переменные используется векторная авторегрессия (VAR). В такой модели каждая переменная представляется как линейная функция лагов переменных системы и теку-

щих значений структурных шоков. Модель VAR(p) записывается в общем виде как:

$$Y_t = A_1 Y_{t-1} + \dots + A_p Y_{t-p} + U_t$$

$$U_t = B_0 \varepsilon_t,$$

где Y_t — вектор эндогенных переменных, $A_1, \dots, A_p, B_0$ — матрицы коэффициентов, U_t — вектор ошибок, представляющий собой линейную комбинацию структурных шоков (ε_t).

Для выделения структурного монетарного шока — отклонения политики центрального банка от ее систематической реакции на экономические условия (правило Тейлора) — необходимо задать идентификационные ограничения. Чаще всего применяются рекурсивные схемы (разложение Холецкого), ограничения на знаки (sign-restrictions), высокочастотные идентификации (по движению ставок или цен фьючерсов до и после решений ЦБ) или внешние инструменты (проху-SVAR). После выделения монетарного шока ($\widehat{m}_t$) можно построить импульсные отклики (IRF), отражающие динамический эффект от неожиданного единичного изменения процентной ставки на выпуск, инфляцию и другие переменные.

Однако линейная VAR имеет ограничения. Она предполагает, что эффект монетарной политики постоянен во времени и одинаков во всех макроэкономических состояниях. В реальности это допущение может быть слишком строгим. Кроме того, при ограниченной выборке оценка большого числа параметров ведет к росту дисперсии оценок и снижает надежность выводов. Поэтому в последние годы активно применяются более гибкие эмпирические стратегии, позволяющие ослабить эти ограничения, в частности подход локальных проекций [Jordà 2005].

Вместо того чтобы восстанавливать всю динамическую систему, метод локальных проекций оценивает отклик переменной на шок напрямую для каждого горизонта времени. Для каждого горизонта строится отдельная регрессия вида

$$y_{t+h} - y_{t-1} = \alpha_h + \beta_h \widehat{m}_t + \sum_{i=1}^p \Gamma_{h,i} X_{t-i} + v_{t,h},$$

X_{t-i} — вектор лагов контрольных переменных. Набор коэффициентов $\{\beta_h\}_{h=0}^H$ образует импульсную функцию отклика переменной

на монетарный шок. Как показали авторы работы [Plagborg-Møller, Wolf 2021], при корректной спецификации результаты локальных проекций и VAR совпадают.

Для анализа того, как сила трансмиссии монетарной политики меняется в зависимости от фискального режима, стандартная спецификация модифицируется. В модель вводится количественный индикатор фискального состояния, например уровень долга, его прирост или изменение циклически скорректированного первичного бюджетного баланса ($\Delta CAPB_t$). Чтобы учесть зависимость эффекта монетарного шока от этого показателя, оценивается регрессия с взаимодействием шока и индикатора:

$$y_{t+h} - y_{t-1} = \alpha_h + \beta_h \widehat{m}_t + \delta_h S_{t-1} + \gamma_h (\widehat{m}_t \times S_{t-1}) + \sum_{i=1}^p \Gamma_{h,i} X_{t-i} + v_{t,h},$$

где S_{t-1} отражает фискальное состояние.

В этом случае отклик переменной (y) на монетарный шок через (h) периодов равен:

$$IRF_h(S_{t-1}) = \beta_h + \gamma_h S_{t-1}.$$

Если коэффициент γ_h статистически значим, это означает, что фискальная политика действительно изменяет силу воздействия ДКП. Подобный подход применен, например, в работе [Sargam, Feilich 2024], где авторы показали, что при высокой долговой нагрузке реакция выпуска на ужесточение монетарной политики в США становится заметно слабее.

Другой вариант основан на использовании фиктивного индикатора фискального режима. Так, [Deb et al. 2023] определяют бинарную переменную D_t , равную единице, если фискальная политика носит сдерживающий характер ($\Delta CAPB_t > 0$), и нулю, если бюджетная политика, стимулирующая ($\Delta CAPB_t < 0$). Тогда оценивается модель с двумя различными коэффициентами отклика:

$$y_{t+h} - y_{t-1} = \alpha_h + \beta_{h1} D_t \widehat{m}_t + \beta_{h2} (1 - D_t) \widehat{m}_t + \sum_{i=1}^p \Gamma_{h,i} X_{t-i} + v_{t,h}.$$

Сравнение β_{h1} и β_{h2} показывает, отличается ли реакция экономики на монетарный шок в зависимости от фискального режима.

Вариацией этого подхода является вероятностная спецификация, предложенная в [Kloosterman et al. 2024]. Авторы вводят функцию принадлежности к режиму $F(\Delta CAPB_t) \in [0,1]$, интерпретируемую как вероятность нахождения экономики в режиме фискальной экспансии или сдерживания. Коэффициенты модели взвешиваются этой вероятностью, что позволяет плавно переходить между режимами и избегать резких границ между ними. Основная спецификация модели принимает следующий вид:

$$y_{t+h} - y_{t-1} = \sum_r F^r(\Delta CAPB_t) [\alpha_{hr} + \beta_{hr} \widehat{m}_t^{\{ДКП\}} + \sum_{i=1}^p \Gamma_{ir} X_{t-i}] + v_t.$$

Однако, как показывают недавние работы, использование нелинейных моделей требует особой осторожности. В статье [Gonçaves et al. 2024] отмечается, что в моделях state-dependent local projections корректность оценок зависит от природы режима: если фискальное состояние эндогенно реагирует на макроэкономические шоки, оценки отражают лишь предельный отклик на бесконечно малый шок, а при крупных шоках могут быть сильно смещены. В то же время [Kolesár, Plagborg-Møller 2024] показывают, что линейные VAR и локальные проекции на наблюдаемые шоки часто дают более надежные оценки средних предельных эффектов, чем неправильно специфицированные нелинейные модели. Поэтому при переходе к моделям с взаимодействием шоков важно сохранять баланс между гибкостью спецификации и интерпретируемостью результатов.

С учетом этих методологических выводов в настоящем исследовании используется более интерпретируемый подход — векторная авторегрессия с элементом нелинейности в виде взаимодействия монетарного шока и фискального индикатора. Такой формат позволяет сохранить преимущества VAR-моделей (прозрачность структуры и сопоставимость импульсных откликов) и одновременно учесть возможную зависимость силы трансмиссии от состояния бюджета. В отличие от полностью нелинейных или пороговых моделей этот метод допускает прямую интерпретацию коэффициентов взаимодействия как меры изменения эффекта монетарной политики в различных фискальных режимах.

Формально уравнения системы в матричном виде могут быть записаны как

$$Y_t = A_1 Y_{t-1} + \dots + A_p Y_{t-p} + B_0 \widehat{m}_t + B_1 (S_t \times \widehat{m}_t) + \sum_{i=0}^p \Gamma_i X_{t-i} + \eta_t,$$

где $\widehat{m}_t$ — идентифицированный монетарный шок, S_t — индикатор фискальной позиции, X_t — вектор контрольных переменных. Матрица коэффициентов B_1 отражает, как фискальный режим модифицирует эффект монетарного шока. В отличие от линейного VAR здесь импульсные отклики рассчитываются условно по состоянию фискального индикатора. Например, мгновенный импульсный отклик переменных Y_t на шок $\widehat{m}_t$ будет равен

$$IRF_0(S_t = \bar{S}) = \left. \frac{\partial Y_t}{\partial \widehat{m}_t} \right|_{S_t = \bar{S}} = B_0 + B_1 \bar{S}.$$

Интерпретация таких IRF предполагает, что реакция экономики на монетарные импульсы различается в зависимости от того, находится ли бюджет в фазе консолидации ($\bar{S}_1$) или стимулирования ($\bar{S}_2$).

Таким образом, современные эмпирические стратегии позволяют гибко учитывать влияние фискальной политики на трансмиссию монетарной. Используемый индикатор может быть количественным (например, уровень долга), бинарным (на основе изменений первичного бюджетного баланса) или вероятностным (оценка принадлежности к фискальному режиму). Включение коэффициента взаимодействия в VAR-модель представляется наиболее естественным способом эмпирически проверить гипотезу о режимной зависимости трансмиссии денежно-кредитной политики.

3.3. Описание данных

В качестве переменной состояния используется фискальный импульс, характеризующий направленность и интенсивность фискальной политики. Базовый показатель определяется как отрицательное изменение сальдо бюджета расширенного правительства, то есть сальдо бюджета консолидированного бюд-

жета РФ и бюджетов государственных внебюджетных фондов, ($-\Delta \text{Сальдо}_t$): положительные значения отражают фискальное смягчение (рост дефицита), а отрицательные — ужесточение. Такой индикатор фиксирует совокупный эффект бюджетных решений, включая действие автоматических стабилизаторов, и позволяет напрямую связать изменение сальдо с динамикой совокупного спроса.

В дополнение используется фискальный импульс, скорректированный на фазу экономического цикла и мировые цены на нефть, построенный по методике [Perevyshin, Kolyadenko 2024]. Корректировка необходима, поскольку значительная часть колебаний бюджетного баланса в России обусловлена не дискреционными решениями, а внешними факторами, прежде всего динамикой нефтегазовых доходов. Очистка сальдо от этих влияний позволяет выделить именно активную, управляемую компоненту фискальной политики и оценить ее взаимодействие с денежно-кредитными мерами.

Для устранения сезонности применяются два подхода. Во-первых, метод X-13ARIMA, во-вторых, метод скользящего годового окна, который определяет фискальный импульс как изменение годовой суммы бюджетного сальдо, что устраняет сезонность по построению и сглаживает шум, хотя делает индикатор менее чувствительным к кратковременным колебаниям.

Оба подхода дают сопоставимые результаты и обеспечивают проверку устойчивости оценок. Как видно на рис. 12–13, оба варианта индикатора фиксируют схожие фазы фискальной политики: умеренное стимулирование в 2011–2014 гг., ужесточение в 2015–2018 гг. и заметное смягчение начиная с 2019 г., особенно в период пандемии 2020–2021 гг. и последующий период 2022–2024 гг. Различия между рядами касаются главным образом амплитуды краткосрочных колебаний, но не общей последовательности фискальных режимов.

В модели используются квартальные данные за период I квартал 2012 — I квартал 2025 г. Вектор эндогенных переменных включает реальный ВВП, индекс потребительских цен (ИПЦ), ключевую ставку Банка России и номинальный обменный курс рубля к доллару США. Идентификация монетарного шока осуществляется

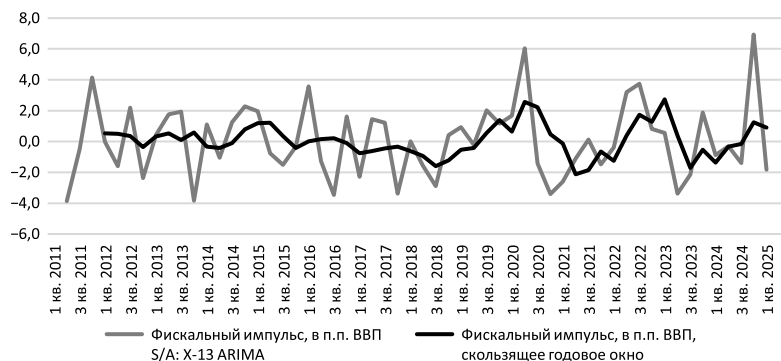


Рис. 12. Фискальный импульс, оцененный двумя способами на основе изменений дефицита бюджета, в % к ВВП.

Источник: Минфин России, расчеты авторов.

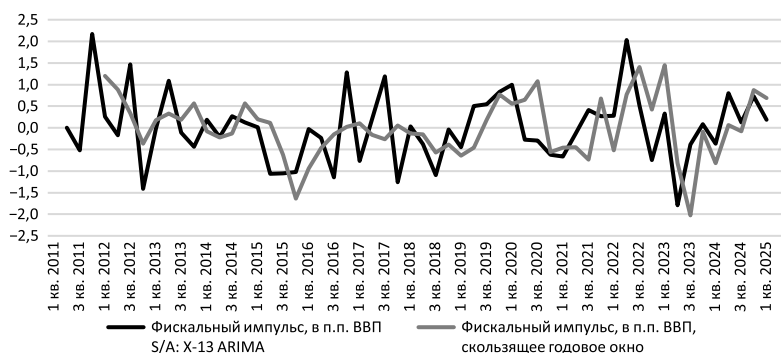


Рис. 13. Фискальный импульс, оцененный двумя способами на основе изменений дефицита бюджета в процентах к ВВП, циклически скорректированный.

Источник: Минфин России, расчеты авторов.

с помощью разложения Холецкого при указанном порядке переменных. Реальный ВВП берется в постоянных ценах; ИПЦ агрегируется из месячных данных и используется для расчета квартальной инфляции; ключевая ставка представлена как среднее значение за квартал; обменный курс учитывается в логарифмах. Все ряды очищены от сезонности. Так как в данных отсутствует

коинтеграция, то все переменные, кроме ставки, взяты как лог-разности. Источником данных являются Росстат и Банк России.

В качестве экзогенной переменной используется Bloomberg Commodity Index (BCOM), отражающий динамику мировых цен на 23 сырьевых товара, сгруппированных по четырем категориям: энергоресурсы, металлы, сельхозтовары и животноводство. Индекс представляет собой взвешенную корзину фьючерсных контрактов с ограничением на долю одного товара (не более 33%) и одной группы (не более 15%). Использование BCOM позволяет учесть влияние внешней ценовой конъюнктуры, в частности колебаний нефтяных и продовольственных цен, на инфляцию, курс рубля и динамику внутреннего выпуска, что особенно важно для российской экономики с высокой зависимостью от мировых товарных рынков. Кроме того, в систему добавлены дамми-переменные для кризисных эпизодов, чтобы исключить влияние экстремальных наблюдений и структурных сдвигов на оценку параметров VAR.

Фискальная переменная рассматривается экзогенно, поскольку ограниченное число наблюдений (53 точки) не позволяет достоверно оценить ее динамику внутри системы. Такое сочетание переменных обеспечивает баланс между компактностью модели и экономической интерпретируемостью, позволяя проследить взаимодействие денежно-кредитных, фискальных факторов в российской экономике за исследуемый период. Число лагов для всех переменных равно 2.

3.4. Результаты эконометрического моделирования

Импульсные отклики показывают, как изменяются ключевые макроэкономические показатели — выпуск, уровень цен, процентная ставка и валютный курс — в ответ на монетарный шок, по сравнению с гипотетической ситуацией, в которой этого шока не произошло. Иными словами, это разница между траекторией экономики «с шоком» и «без шока», при условии, что никаких побочных шоков или потрясений не происходит. В данной работе рассматривается положительный монетарный шок, то есть неожиданное повышение ключевой ставки на 1 процентный пункт

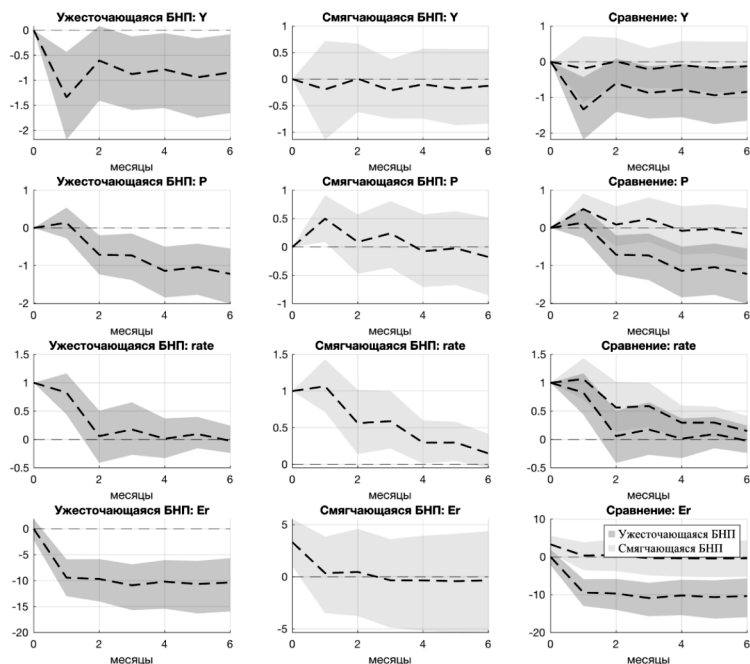


Рис. 14. Условные импульсные отклики на положительный монетарный шок при различных фискальных режимах. По строкам представлены реакции: выпуска (ВВП), инфляции (ИПЦ), ключевой ставки и номинального обменного курса рубль/доллар. Доверительные интервалы в 68%.

Источник: расчеты авторов.

при прочих равных условиях (например, с 20% до 21% годовых). Такой шок интерпретируется как ужесточение денежно-кредитной политики, направленное на снижение инфляции.

Ниже представлены условные импульсные отклики на положительный монетарный шок (неожиданное ужесточение политики), оцененные на подвыборке за I квартал 2012 — II квартал 2023 г. при использовании в качестве переменной состояния текущего фискального импульса, скорректированного на циклические колебания и рассчитанного по скользящему годовому окну (рис. 14).

На графиках показаны траектории изменения эндогенных переменных для двух режимов — ужесточающейся (первый столбец) и смягчающейся (второй столбец) фискальной политики, а также их наложение (третий столбец), что позволяет визуально оценить различия в реакции экономики на монетарные шоки при различных фискальных условиях.

Кумулятивные отклики для выпуска и цен построены по переменным в логарифмах, поэтому их значения можно интерпретировать как процентное отклонение уровня показателя от сценария без шока. В режиме ужесточающейся фискальной политики монетарный шок приводит к заметному снижению выпуска: уже в первые два квартала (полгода) уровень ВВП становится примерно на 1% ниже, чем в базовом сценарии, и это отклонение остается статистически значимым. В режиме смягчающейся фискальной политики кумулятивный отклик выпуска неотличим от нуля: первоначальная реакция мала, иногда даже слегка положительна, и быстро исчезает.

Для уровня цен в режиме ужесточающейся фискальной политики наблюдается последовательное снижение: через год кумулятивный отклик соответствует снижению уровня цен примерно до 1% относительно сценария без шока. При смягчающейся фискальной политике реакция цен более слабая: на краткосрочном горизонте возможен небольшой положительный сдвиг (так называемый *price puzzle*, или загадка цен), затем траектория переходит в отрицательную область, но остается ближе к нулю, чем в режиме ужесточающейся БНП. Это означает, что один и тот же рост ставки дает более выраженный дезинфляционный эффект, если бюджетная политика не поддерживает совокупный спрос.

Динамика ключевой ставки отражает постепенное ее возвращение к исходному уровню, то есть тому уровню, который соответствует состоянию экономики без монетарного шока. Значимых различий в динамике ставки между двумя режимами фискальной политики нет. Затухание происходит не сразу, поскольку ЦБ РФ действует по правилу Тейлора с некоторой инерционной компонентой (зависимости ставки от предыдущих значений). Реакция обменного курса более чувствительна к фискальному контексту: при ужесточающейся фискальной полити-

ке рубль заметно укрепляется и сохраняет более низкий уровень курса по сравнению с базовым сценарием, тогда как при смягчающейся политике укрепление слабее и менее устойчиво.

На первый взгляд, результаты указывают на режимную зависимость трансмиссии ДКП: один и тот же рост ключевой ставки дает более сильные реальные и ценовые эффекты при фискальном ужесточении. Однако интерпретацию необходимо сопровождать оговорками, связанными прежде всего с доступностью и частотой данных. Фискальный индикатор состояния берется в том же квартале, что и эндогенные переменные, пусть и сглаженный годовым окном; значит, мы не можем полностью исключить эндогенность регрессора и потенциальное смещение из-за двусторонней причинно-следственной связи. Строгое решение потребовало бы лагирования переменной состояния, инструментирования или постановки с экзогенно заданной переходной вероятностью режима.

Кроме того, короткая выборка и наличие крупных структурных сдвигов (2014–2015, 2020, 2022) ограничивают статистическую мощность и делают доверительные интервалы чувствительными к спецификации. Сам показатель фискального импульса подвержен ошибкам измерения: разные способы сезонной и циклической корректировки могут влиять на амплитуды оценок. Наконец, в текущей постановке фискальная переменная рассматривается экзогенно, а модель не допускает, что монетарный шок сам по себе может способствовать переключению фискального режима. По этим причинам приведенные отклики следует интерпретировать с осторожностью: «как изменилась бы экономика при данном, фиксированном состоянии бюджетной политики». Эти ограничения не отменяют основного вывода, но задают рамки его применимости.

3.5. Обсуждение результатов

Полученные результаты показывают, что сила и знак реакции экономики на ужесточение денежно-кредитной политики потенциально могут зависеть от фискального режима. В режимах уже сточающейся БНП один и тот же шок ставки дает более глубокое

и устойчивое снижение выпуска и инфляции и сопровождается более заметным укреплением рубля. При смягчении бюджетной политики реальные и ценовые эффекты оказываются слабее и кратковременнее, что согласуется с идеей частичной компенсации со стороны бюджета. Условные прогнозы подтверждают этот вывод: при заданных внешних условиях и одинаковых монетарных шоках переход к мягкой бюджетной политике поддерживает выпуск, но приводит к более высокой траектории цен; жесткая бюджетная политика быстрее помогает добиться дезинфляции ценой более слабой динамики ВВП.

Интерпретировать итог следует с учетом ограничений: короткая выборка, возможная эндогенность фискального индикатора и ошибки измерения означают, что оцененные эффекты — условные по заданному состоянию фискальной политики, а доверительные интервалы достаточно широкие. Тем не менее результаты указывают на практический вывод для макроэкономической политики: координация фискальных и монетарных решений повышает предсказуемость и эффективность стабилизации экономики после шоков, тогда как разнонаправленные импульсы ослабляют трансмиссию монетарной политики.

ЗАКЛЮЧЕНИЕ

Российская экономика столкнулась в 2022–2025 гг. со значительным ускорением инфляции. Темп роста потребительских цен устойчиво и значительно отклонился от цели. Банк России проводил жесткую монетарную политику начиная с середины 2023 г., причем уровень реальных ставок, который пришлось устанавливать регулятору для приведения инфляции к цели, оказался высоким. Несмотря на жесткую политику, привести инфляцию к цели получится только в 2026–2027 гг.

Одной из причин роста инфляции стал стимулирующий эффект, источником которого послужила фискальная политика. В последние десятилетия бюджетно-налоговую политику можно было характеризовать как контрциклическую, поскольку при росте выпуска и увеличении доходов бюджета фискальные власти

сдерживали рост расходов, а в периоды кризиса и падения выпуска бюджет, напротив, тратил существенные средства на меры поддержки экономики. В результате такой политики при повышении темпов роста ВВП дефицит сокращался. Однако начиная с 2023 г. Минфин России значимо отклонился от этого курса. Рост государственных расходов в 2022–2025 гг. способствовал росту совокупного спроса, в результате чего в российской экономике сложилось сочетание высоких темпов экономического роста и стимулирующей бюджетной политики.

При этом важная особенность периода с 2022 по 2025 г. состоит в том, что в этом периоде Банк России не осуществлял полное зеркалирование операций Минфина России с Фондом национального благосостояния. Отсутствие стерилизации привело к дополнительному приросту денежной базы на 4,3 трлн рублей. В связи с этим возникли сомнения в том, что Банк России может полностью контролировать денежное предложение. Мы приходим к выводу, что такое заключение неверно, поскольку Банк России имел все инструменты сдерживания денежного предложения, а также фактический объем денежного предложения соответствовал устанавливаемому регулятором уровню ключевой ставки.

Если бы ЦБ РФ начал цикл ужесточения раньше июля 2023 г., то, вероятно, удалось бы избежать такого существенного роста инфляции и, как следствие, такого роста ключевой ставки, какой наблюдался в реальности. Следует, однако, также иметь в виду, что макроэкономические данные, доступные в первом полугодии 2023 г., включая данные о динамике инфляции, давали противоречивую картину и не указывали явно на наличие признаков надвигающейся новой инфляционной волны. Поскольку решения о ставке оказывают влияние на инфляцию с лагом в 3–6 кварталов, возможности своевременно и эффективно охладить спрос у денежных властей практически не было.

Кроме того, с 2022 г. российская экономика испытала действие сильных проинфляционных шоков, к числу которых относятся отрицательные шоки предложения (рост издержек производителей в связи с санкциями, ухудшение условий торговли, рост неопределенности и инфляционных ожиданий) и положительные шоки спроса (бюджетная экспансия, рост инфляционных

ожиданий). Российский опыт последних лет показал, что возможности ЦБ РФ по охлаждению совокупного спроса в описанных условиях существенно ограничены. Уровень ставок, необходимый для стабилизации ценовой динамики, был настолько высок, что с заметной вероятностью мог спровоцировать масштабный кризис ликвидности с угрозой потери финансовой стабильности и устойчивости банковского сектора, а также вызвать банкротства предприятий реального сектора с разрывом производственных цепочек и риском рецессии. Кроме того, еще более высокие ставки проценты вызвали бы еще большее укрепление рубля, дополнительно осложняя финансовое положение российских экспортеров. В данных обстоятельствах руководство Банка России приняло решение реализовать сценарий мягкой посадки, ценой временного ускорения потребительской инфляции.

Мы также приходим к выводу, что смягчение бюджетно-налоговой политики затрудняет достижение цели по инфляции. Проведенное эконометрическое исследование показало, что охлаждающий эффект от ужесточения монетарной политики оказывается более сильным в условиях, когда бюджетно-налоговая политика находится в цикле ужесточения. Таким образом, существует синергетический эффект от одновременного действия жесткой монетарной политики и ужесточающейся фискальной. Это указывает на то, что сдерживание инфляционного давления при смягчении бюджетно-налоговой политики потребует от Центрального банка проведения более жесткой политики.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Aiyagari S., Gertler M. The backing of government bonds and monetarism//Journal of Monetary Economics. 1985. Vol. 16. No. 1. P. 19–44.
2. Bianchi F., Melosi L. The dire effects of the lack of monetary and fiscal coordination//Journal of Monetary Economics. 2019. June. Vol. 104. P. 1–22.
3. Caramp N., Feilich E. Monetary policy and government debt//Journal of Money, Credit and Banking. 2024. DOI:10.1111/jmcb.13215.
4. Cloyne J. S. et al. Decomposing the fiscal multiplier//National Bureau of Economic Research Working Paper Series. 2020. No. w26939.
5. Cochrane J. The fiscal theory of the price level. Princeton, NJ.: Princeton University Press, 2023. P. 584.
6. Deb M. P. et al. Monetary policy transmission heterogeneity: cross-country evidence//International Monetary Fund. Washington, DC., 2023.
7. Frigerio L., Magnifico L. Household indebtedness and the effectiveness of monetary policy: evidence from the Eurozone. 2024.
8. Gonçalves S. et al. State-dependent local projections//Journal of Econometrics. 2024. Vol. 244. No. 2. P. 105702.
9. Goryunov S. M. et al. Factors of global inflation in 2021–2022//Russian Journal of Economics. 2023. Vol. 9. No. 3. P. 219–244. DOI 10.32609/j.ruje.9.111967.
10. Hicks J. R. Mr. Keynes and the "Classics"; A Suggested Interpretation//Econometrica. 1937. Vol. 5. No. 2. P. 147–159.
11. Jordà Ò. Estimation and inference of impulse responses by local projections//American economic review. 2005. Vol. 95. No. 1. P. 161–182.
12. Keister T. et al. Divorcing money from monetary policy//Economic Policy Review. 2008. Vol. 14. No. 2.
13. Keynes J. M. The General Theory of Employment, Interest and Money. London: Macmillan, 1936.
14. Kim Y., Lim H. Transmission of monetary policy in times of high household debt//Journal of Macroeconomics. 2020. Vol. 63. C. 103168.
15. Kloosterman R., Bonam D., van der Veer K. The effects of monetary policy across fiscal regimes//Journal of Macroeconomics. 2024. Vol. 81. P. 103616.
16. Kolesár M., Plagborg-Møller M. Dynamic causal effects in a nonlinear world: the good, the bad, and the ugly//arXiv preprint arXiv:2411.10415. 2024.

17. Kumhof, M. et al. Simple monetary rules under fiscal dominance // *Journal of Money, Credit and Banking*. 2010. Vol. 42. No. 1. P. 63–92.
18. Leeper E. M. Equilibria under ‘active’ and ‘passive’ monetary and fiscal policies // *Journal of Monetary Economics*. 1991. Vol. 27. No. 1. P. 129–147.
19. McLeay M., Radia A., Thomas R. Money creation in the modern economy // *Bank of England quarterly bulletin*. Q1. 2014.
20. Perevyshin Y., Kolyadenko P. Contribution of Fiscal Policy to Inflation in the Russian Economy // Available at SSRN 5038767. 2024.
21. Plagborg-Møller M., Wolf C. K. Local projections and VARs estimate the same impulse responses // *Econometrica*. 2021. Vol. 89. No. 2. P. 955–980.
22. Sargent T., Wallace N. Some unpleasant monetarist arithmetic // *Federal reserve bank of minneapolis quarterly review*. 1981. Vol. 5. No. 3. P. 1–17.
23. Sims C. A. Stepping on a rake: The role of fiscal policy in the inflation of the 1970s // *European Economic Review*. 2011. Vol. 55. No. 1. P. 48–56.
24. Woodford M. Price-level determinacy without control of a monetary aggregate // *Carnegie-Rochester conference series on public policy*. North-Holland. 1995. Vol. 43. P. 1–46.
25. Аблаев Э. Ю. Подходы к определению режимов взаимодействия бюджетно-налоговой и денежно-кредитной политики // *Проблемы прогнозирования*. (2025) № 3 (210). С. 159–172. DOI 10.47711/0868-6351-210-159-172.
26. Аблаев Э. Ю., Сударев Д. А. Бюджетные правила в системе приоритетов бюджетной политики России // *ЭКО*. 2024. № 5 (599). С. 202–223. DOI 10.30680/ECO0131-7652-2024-5-202-223.
27. Аблаев Э. Ю., Магжанов Т. Р. Бюджетные правила и контрциклическая бюджетная политика в России // *Проблемы прогнозирования*. 2024. № 3 (204). С. 30–42. DOI 10.47711/0868-6351-204-30-42.
28. Белова Е., Перевышин Ю. (2017а) Проблемы взаимодействия фискальной и монетарной политики в странах, использующих режим инфляционного таргетирования // *Российское предпринимательство*. Т. 18. № 2. С. 133–146.
29. Белова Е., Перевышин Ю. (2017b) Устойчивость государственного долга и инфляции в рамках взаимодействия фискальной и монетарной политики // *Аудит и финансовый анализ*. № 2. С. 171–179.
30. Горюнов Е. Л., Дробышевский С. М., Мау В. А., Трунин П. В. Что мы (не) знаем об эффективности инструментов ДКП в современном мире? // *Вопросы экономики*. 2021. № 2. С. 5–34. DOI 10.32609/0042-8736-2021-2-5-34.

31. Гурвич Е., Соколов И. Бюджетные правила: избыточное ограничение или неотъемлемый инструмент бюджетной устойчивости? // Вопросы экономики. 2016. № 4. С. 5–29.
32. Дробышевский С. М., Перевышин Ю. Н., Синельников-Мурылев С. Г., Трунин П. В. Причины повышенной инфляции в российской экономике // Вопросы экономики. 2025. № 1. С. 5–31.
33. Кудрин А., Соколов И. Бюджетные правила как инструмент сбалансированной бюджетной политики // Вопросы экономики. 2017. № 11. С. 5–32.

Оптимальное налогообложение и пенсионные системы в моделях перекрывающихся поколений

Замниус А. В., аспирант кафедры микро- и макроэкономического анализа экономического факультета МГУ имени М. В. Ломоносова, научный сотрудник лаборатории математического моделирования экономических процессов Института Гайдара, zamnius@ier.ru

Шпилевая А. Е., научный сотрудник лаборатории математического моделирования экономических процессов Института Гайдара, старший преподаватель кафедры моделирования и системного анализа Финансового университета при Правительстве Российской Федерации, angelina.shpilevaya@gmail.com

Аннотация. В работе систематизируется опыт построения моделей общего равновесия в контексте оптимального налогообложения и пенсионных систем. Актуальность исследования обусловливается, с одной стороны, современными демографическими тенденциями и уже проводящимися реформами в России, а с другой — недостаточной разработанностью этих тем в отечественной литературе. Показано, что, во-первых, величина эластичности предложения труда может в значительной степени повлиять не только на размер оптимального подоходного налога, но и на сам тип шкалы налогообложения. Во-вторых, в условиях растущей с возрастом эластичности предложения труда обложение налогом доходов от капитала вместо труда может приводить к увеличению общественного благосостояния. В-третьих, если не склонные к риску индивиды существуют в условиях неопределенности, система социального обеспечения может приносить пользу индивидам, частично заменяя отсутствующие страховые рынки. Наконец, если экономика динамически неэффективна, распределительная пенсионная система позволяет справиться с проблемой накопления. На основе проделанного обзора приводится ряд специфических элементов, которые модель общего равновесия для России должна учитывать.

ВВЕДЕНИЕ

В ближайшем будущем российская экономика столкнется с рядом вызовов, обусловленных изменениями в демографии. Так, в течение следующих 20–30 лет в России ожидается снижение общей численности населения и изменение демографической структуры, выражающееся в увеличении доли пожилого населения¹. Рост расходов на пенсионное обеспечение и здравоохранение на фоне снижающейся численности рабочей силы ставит вопрос о способах финансирования растущих государственных расходов. Некоторые шаги в направлении смягчения последствий для бюджета были предприняты в 2016 г. и 2018 г. В 2016 г. была приостановлена индексация пенсий работающим пенсионерам, в 2018 г. была начата реформа повышения пенсионного возраста. С 1 января 2025 г. индексация пенсий работающим пенсионерам возобновлена, а также осуществлен переход от двухступенчатой шкалы подоходного налогообложения граждан (13% для доходов ниже 5 млн рублей; 15% для доходов выше 5 млн рублей) к значительно более прогрессивной пятиступенчатой (13% для доходов ниже 2,4 млн рублей; 15% для доходов ниже 5 млн рублей; 18% для доходов ниже 20 млн рублей; 20% для доходов ниже 50 млн рублей; 22% для доходов выше 50 млн рублей). Совокупность демографических тенденций и проводимых реформ ставит вопрос о том, как происходящие изменения в пенсионной и налоговой сфере повлияют на российскую экономику: как изменится уровень благосостояния населения, основные макроэкономические показатели. Ответ на данный вопрос может быть получен из экономической литературы, посвященной вопросам налогообложения и пенсионным реформам.

Ключевым инструментом изучения последствий налоговой и пенсионной политики является модель общего экономического равновесия с перекрывающимися поколениями (OLG), предложенная Даймондом и Самуэльсоном [Samuelson 1958; Diamond 1965]. Преимущество этого класса моделей перед другими состоит в двух аспектах. С одной стороны, они позволяют рассматривать эко-

¹ United Nations: World Population Prospects. <https://population.un.org/wpp/> (дата обращения 01.04.2025).

номику как систему, в которой экономические агенты — домохозяйства, фирмы и государство — взаимодействуют друг с другом на различных рынках, стремясь достичь одновременного равновесия на каждом из них. С другой стороны, OLG-модели позволяют реалистично описать поведение множества одновременно живущих поколений индивидов, для каждого из которых модель отражает изменения в потребительском поведении, предложении труда и сбережениях на протяжении всего жизненного цикла.

Для изучения эффектов налоговой и пенсионной политики используются различные модификации OLG-моделей, включающие предпосылки о наличии или отсутствии идиосинкратических и/или систематических рисков дохода. Идиосинкратические риски возникают, когда доход индивидов в процессе экономической деятельности подвержен случайным индивидуальным шокам, называемым в литературе идиосинкратическими. Такие шоки уникальны, зависят только от особенностей индивида и не связаны с общей макроэкономической конъюнктурой (утрата человеком места работы из-за закрытия компании). Идиосинкратические шоки следует отличать от систематических шоков дохода, которые затрагивают всю экономику в целом в результате кризисов, политических потрясений и т.д. В литературе по оптимальному налогообложению представлены модели с идиосинкратическими рисками [Conesa, Krueger 2006; Conesa et al. 2009], эндогенным накоплением человеческого капитала [Da Costa, Santos 2018; Blandin, Peterman 2019], учетом состава домохозяйств и решением о занятости его членов [Karabarbounis 2016; Guner et al. 2012] и т.д. В контексте изучения пенсионных систем широко используются модификации, учитывающие несовершенство финансовых рынков [İmrohoroglu et al. 1995], неопределенность продолжительности жизни [Bruce, Turnovsky 2013; Yaari 1965; Blanchard 1985], наличие систематических и идиосинкратических рисков [İmrohoroglu et al. 1995; Krueger, Kubler 2006; Harenberg, Ludwig 2017; Conesa, Krueger 1998], наличие альтруистических связей между поколениями [Caballé, Fuster 2003] и др.

В настоящее время количество работ, посвященных оценке макроэкономических эффектов налогообложения и пенсий для России в контексте OLG-моделей, крайне мало. Так, работа

[Nesterova 2024] изучает эффекты перехода к прогрессивной системе налогообложения на основе глобальной вычислимой модели перекрывающихся поколений (CGE-OLG), а работа [Зубарев, Нестерова 2019] анализирует последствия повышения пенсионного возраста так же в рамках CGE-OLG-модели.

Потребность в систематизации мирового опыта построения моделей оптимального налогообложения и пенсионных систем обусловлена недостаточной разработанностью этих тем в отечественной литературе и необходимостью моделирования последствий происходящих в России реформ. В связи с этим в статье приводится систематизация литературы по теме оптимального налогообложения и пенсионных реформ. Совместное изучение двух предметов обусловлено тесной связью этих сфер бюджетно-налоговой политики. Пенсии выплачиваются из государственного бюджета, который формируется за счет налоговых поступлений, в числе которых налог на доходы физических лиц (НДФЛ), налог на прибыль организаций, НДС и другие. Снижение численности трудоспособного населения и рост числа пенсионеров создают дополнительную нагрузку на бюджет. Государству приходится искать новые источники финансирования, повышать налоги или сокращать расходы на другие сферы. Более прогрессивная налоговая система может частично компенсировать эти демографические вызовы, увеличивая поступления от высокодоходных групп населения, но в то же время создавать искажения на рынке труда, влияя на решение домохозяйств о количестве отработанных часов. Увеличение ставки налога на прибыль может увеличить налоговые поступления в бюджет, но вместе с тем снизить равновесные ставки заработной платы ввиду растущих расходов фирм и уровень инвестиций в стране. Ввиду сложности изучаемого предмета наиболее адекватным инструментом для моделирования последствий реформ представляются модели перекрывающихся поколений. По этой причине в приведенном обзоре мы сосредоточиваем внимание именно на этих моделях. Обзор литературы охватывает широкий спектр моделей, которые различаются набором базовых предпосылок, что позволяет соотносить получаемые результаты с постановками моделей и выбирать из них ту, которая сможет наиболее адекватно отразить особенности российской экономики.

Статья структурирована следующим образом. В первой части статьи внимание уделяется ключевым принципам оптимального налогообложения в моделях перекрывающихся поколений, в частности эластичности предложения труда, как отправной точке при разработке оптимального налогообложения. Затем приводится обзор моделей оптимального налогообложения капитала и дохода и основные результаты, полученные исследователями на основе моделей с различными предпосылками. Во второй части статьи рассматриваются модели перекрывающихся поколений, в которых изучаются особенности поведения домохозяйств, накопления капитала и других показателей долгосрочного равновесия в условиях распределительной и накопительной пенсионных систем. Наряду со сравнительным анализом состояний долгосрочного равновесия приводятся модели, в которых исследуются особенности переходной траектории от одного типа пенсионной системы к другой в рамках детерминированных моделей перекрывающихся поколений. Далее рассматриваются расширения этих моделей с учетом волатильности доходов, позволяющие изучать дополнительные эффекты от налогообложения и пенсионной системы. В заключении приводятся выводы о том, как результаты приведенных исследований могут быть применены при разработке модели для России, на основе которой можно получить наиболее полное представление о влиянии текущих налоговых и пенсионных реформ на ключевые макроэкономические показатели в долгосрочном периоде.

1. ОПТИМАЛЬНОЕ НАЛОГООБЛОЖЕНИЕ В МОДЕЛЯХ ОБЩЕГО РАВНОВЕСИЯ

1.1. Эластичности предложения труда и оптимальное налогообложение

Подходные налоги являются одной из ключевых составляющих политики любого государства, так как они представляют собой один из основных источников государственного дохода. В то же время подходный налог — это важный инструмент для перерас-

пределения дохода между гражданами, который позволяет регулировать уровень неравенства в обществе для достижения наиболее справедливого распределения.

Строго говоря, «справедливость» можно определять по-разному. Традиционный взгляд экономистов на проблему справедливости сводится к решению задачи центрального планировщика, который стремится распределить доход таким образом, чтобы максимизировать благосостояние всего общества. Конкретное решение такой задачи во многом зависит от вида функции благосостояния, однако ключевой вывод остается одним — перераспределение дохода от богатых к бедным оптимально, а следовательно, справедливо. Так, согласно классическому утилитаристскому подходу, сформулированному еще в XVIII в. английским философом Дж. Бентамом [Bentham 1996], общественная функция благосостояния представляет собой сумму индивидуальных функций полезностей налогоплательщиков. Поэтому если предельная полезность от потребления убывает по мере роста дохода, более равномерное распределение дохода будет приводить к увеличению общественного благосостояния [Piketty, Saez 2013].

Вместе с тем любое перераспределение будет осуществляться за счет потерь в эффективности. Так, любое увеличение ставок подоходного налога будет снижать стимулы к предложению труда, так как на одну единицу предложенного труда индивид будет получать меньший доход — один из основополагающих выводов теории оптимального налогообложения, полученный в работе [Mirrlees 1971]. Отсюда вытекает противоречие принципов эффективности и справедливости, исходя из которого в работе [Mirrlees 1971] было введено понятие «оптимальной системы налогообложения». Такая система призвана создавать определенный баланс между конкурирующими принципами.

Ключевой принцип оптимального налогообложения заключается в том, чтобы в большей степени облагать факторы производства, предложение которых неэластично, то есть снижение величины предложения данного фактора в ответ на рост ставок налога на доход от него будет слабым [Mirrlees 1971; Ramsey 1927]. Труд является одним из ключевых факторов производства, поэтому высокоэластичный труд (например, замужних женщин [Замниус,

Полбин 2021; Замниус и др. 2022] или людей предпенсионного возраста) следует облагать по меньшей ставке, чем низкоэластичный (например, молодых людей или женатых мужчин среднего возраста).

Так как далее речь пойдет о реакции предложения труда на те или иные изменения в заработных платах, целесообразно кратко обсудить виды эластичностей предложения труда. Так, в литературе принято выделять три основных вида эластичностей: по Маршаллу, Хиксу и Фришу. Эластичность предложения труда по Маршаллу (или некомпенсированная эластичность) показывает изменение отработанных часов в ответ на изменение реальной ставки заработной платы, которое учитывает в себе как эффект дохода, так и эффект замещения. Эластичность предложения труда по Хиксу (или компенсированная эластичность) учитывает в себе лишь эффект замещения. Для примера рассмотрим ситуацию повышения ставки заработной платы. С одной стороны, будет наблюдаться рост альтернативных издержек досуга, которые способствуют увеличению отработанных часов (эффект замещения). С другой же стороны, индивид будет получать больше за то же самое количество отработанных часов, что оказывает негативное воздействие на отработанные часы (эффект дохода). Эластичность предложения труда по Фришу измеряет реакцию отработанных часов в ответ на изменение заработной платы при постоянстве потребления. Такая реакция возможна в двух ситуациях. Либо когда речь идет об эффекте межвременного замещения, то есть о перераспределении отработанных часов между двумя периодами в зависимости от изменения соотношения заработных плат в эти периоды, из-за чего эластичность по Фришу часто называют межвременной эластичностью замещения труда. Либо когда речь идет о реакции отработанных часов в ответ на краткосрочные шоки заработных плат, которые не влияют на потребительское решение индивида. Подробный обзор представлен в [Замниус, Полбин 2021; Замниус и др. 2022].

Для лучшего понимания связи эластичностей и оптимальных ставок налогообложения имеет смысл обсудить результаты двух фундаментальных теоретических работ по оптимальному налогообложению: [Mirrlees 1971] и [Summers 1981]. В модели

Миррлиса выпуск зависит от высокоэластичного агрегированного предложения труда, из-за чего повышение подоходного налога создавало значительные потери в интенсивности труда и, следовательно, выпуске. Исходя из этого автор приходит к выводу, что оптимальной будет, скорее, пропорциональная система с относительно низкой ставкой для индивидов с наибольшим трудовым доходом и относительно малым уровнем трансфертов для наименее богатых индивидов, из-за чего налоговая система в рамках описанных предпосылок оказывается неэффективным инструментом регулирования уровня неравенства в обществе.

Здесь необходимо оговориться о двух ключевых моментах в модели Миррлиса. Во-первых, в модели [Stern 1976] было показано, что оптимальные ставки налогообложения могут быть значительно выше, если допустить, что предложение труда менее эластично. Во-вторых, Миррлис предполагал, что функция полезности индивида зависит только от потребления, что исключает наличие эффекта дохода и, следовательно, делает компенсированную и некомпенсированную эластичности предложения труда эквивалентными. Однако в работах [Blomquist 1983; Aaron, Reisman 1981] было показано, что даже при низкой некомпенсированной эластичности прогрессивное налогообложение может в значительной степени воздействовать на предложение труда, если компенсированная эластичность достаточно велика.

В работе [Summers 1981] рассматривалась модель с перекрывающимися поколениями с неэластичным предложением труда, где существует три вида налогов: на трудовой доход, на капитал и на продажи. В отличие от модели Миррлиса, повышение подоходного налога незначительно влияет на выпуск. В то же время рост налога на капитал уменьшает ставку процента, из-за чего инвестиции в капитал становятся менее выгодными, а в долгосрочном периоде падает производительность труда и, следовательно, заработная плата. Исходя из этого, государству выгоднее сосредоточить свое внимание на подоходном налоге и на налоге на продажи, не облагая капитал в принципе. Именно поэтому от значений эластичности предложения труда зависят не только оптимальные ставки подоходного налога, но и ставки по другим видам налогов.

Отсюда возникает вопрос: насколько эластично предложение труда на самом деле? Согласно классическим работам по оценке эластичности предложения труда по Фришу на микроданных США за 1968–1980 гг., ее значение лежит в промежутке от 0 до 0,5 [Altonji 1986; MaCurdy 1981]. Однако в более поздних работах было показано, что эти оценки могут быть занижены в силу разных причин: использование слабых инструментальных переменных при оценке функции предложения труда [Pistaferri 2003], наличие ограничений на займы [Domeij, Floden 2006], накопление человеческого капитала [Imai, Keane 2004] и т. п. (более подробно см.: [Keane 2011]). Более того, в большинстве микроэконометрических работ рассматривается лишь реакция отработанных часов на шоки заработных плат, в то время как волатильность агрегированного предложения труда в большей степени объясняется колебаниями занятости, а не отработанных часов [Chetty et al. 2011; Hansen 1985; Kydland 1995; Ohanian, Raffo 2012].

Эндогенное накопление человеческого капитала играет особую роль в оценке эластичности предложения труда, потому что за счет него эластичность меняется с течением жизненного цикла [Imai, Keane 2004]. Рассмотрим классическую постановку модели жизненного цикла индивида:

$$\begin{cases} \mathbb{U} = \sum_{t=1}^T \beta^{t-1} U(C_t; N_t) \rightarrow \max_{\mathbb{C}; \mathbb{N}}, \\ A_{t+1} = (1+r)A_t + W_t N_t - C_t, \end{cases} \quad (1)$$

где полезность $U(\bullet)$ зависит от объема потребления C_t и отработанных часов N_t в периоде t , A_t — запас активов, W_t — ставка заработной платы, β — коэффициент дисконтирования.

В условиях, когда W_t задана экзогенно, предельная норма замещения труда потреблением равна ставке заработной платы:

$$MRS_{N,C} = \frac{\partial U / \partial N_t}{\partial U / \partial C_t} = -W_t. \quad (2)$$

При накоплении человеческого капитала условие (2) будет выглядеть иначе. Предположим, что накопление человеческого капитала осуществляется по механизму *learning-by-doing* (LBD):

$$\begin{aligned} W_{t+1} &= WH_{t+1} \\ H_{t+1} &= H(H_t; N_t), \end{aligned} \quad (3)$$

где W — рыночная ставка заработной платы на единицу эффективного труда, H_t — запас человеческого капитала, $H(\bullet)$ — функция накопления человеческого капитала. Мы будем предполагать, что $H(\bullet)$ является выпуклой, дважды дифференцируемой и возрастает по обоим аргументам.

При добавлении в модель (1) механизма эндогенного накопления человеческого капитала (3) индивиды могут влиять на свою будущую заработную плату за счет интенсивности труда в настоящем, приобретая новые знания и опыт. Таким образом, увеличение будущей производительности выступает в роли побочного продукта текущего труда. Это, безусловно, является некоторым упрощением, так как в реальности индивиды могут получать знания и за пределами рынка труда, однако оно позволяет продемонстрировать эффекты накопления человеческого капитала на принятие решений о предложении труда.

Выпишем модель жизненного цикла с эндогенным накоплением человеческого капитала в виде уравнения Беллмана:

$$\begin{aligned} V_t(A_t; H_t) &= \max_{C_t, N_t} \{U[C_t; N_t] \\ &+ \beta V_{t+1}[(1+r)A_t + WH_t N_t - C_t; H(H_t; N_t)]\} \end{aligned} \quad (4)$$

и выведем для нее предельную норму замещения труда потреблением:

$$MRS_{N,C}^H = \frac{\partial U / \partial N_t}{\partial U / \partial C_t} = -WH_t - \frac{\partial H_{t+1}}{\partial N_t} \frac{\partial V_{t+1} / \partial H_{t+1}}{\partial V_{t+1} / \partial A_{t+1}}. \quad (5)$$

Видно, что теперь, в отличие от (2), эффективная ставка заработной платы представляет собой сумму рыночной ставки WH_t и отдачи от накопленного человеческого капитала $\frac{\partial H_{t+1}}{\partial N_t} \frac{\partial V_{t+1} / \partial H_{t+1}}{\partial V_{t+1} / \partial A_{t+1}}$ [Shaw 1989]. Его наличие обуславливает непостоянство реакции отработанных часов на шоки заработных плат на протяжении жизненного цикла. Так, в начале трудовой карьеры индивид стремится получить опыт и навыки, необходимые для повышения своей будущей производительности. В условиях высокой

отдачи от накопления человеческого капитала и низкой фактической заработной платы реакция отработанных часов на шок будет слабой. Предпенсионный возраст характеризуется низкой отдачей от накопления человеческого капитала и высокой фактической заработной платой, из-за чего реакция отработанных часов на шок будет значительно сильнее. Другими словами, эластичность отработанных часов индивида по заработной плате с возрастом будет увеличиваться. Исходя из этого, одинаковые колебания в заработной плате в молодом и предпенсионном возрастах будут по-разному отражаться на величине эффективной заработной платы, из-за чего реакция предложения труда на шок заработной платы будет с возрастом увеличиваться. Эмпирические подтверждения росту эластичности предложения труда по мере старения были представлены в работах [French, Stafford 2017; Imai, Keane 2004; Iskhakov, Keane 2021; Wallenius 2011] и др.

Логичным решением в такой ситуации представляется применение снижающегося с возрастом подоходного налога. Так, в работах [Erosa, Gervais 2002] было показано, что в рамках модели с перекрывающимися поколениями оптимальным было бы облагать трудовой доход по разным ставкам в зависимости от возраста. Однако на практике этот сценарий нереалистичен в силу этических причин [Mirrlees et al. 2011], поэтому [Erosa, Gervais 2002] показали, что налог на капитал позволяет аппроксимировать снижающийся с возрастом подоходный налог. Продемонстрировать эту логику можно на примере двухпериодной версии модели (1) (допустим, что период 1 — молодость, а период 2 — старость):

$$\begin{cases} \mathbb{U} = U(C_1; N_1) + \beta U(C_2; N_2) \rightarrow \max_{C_1; C_2; N_1; N_2} \\ C_1 + A_2 = (1 - \tau_1^{(W)})W_1N_1 + [1 + (1 - \tau^{(K)})r]A_1, \\ C_2 = (1 - \tau_2^{(W)})W_2N_2 + [1 + (1 - \tau^{(K)})r]A_2, \end{cases} \quad (6)$$

где $\tau_t^{(W)}$ — ставка подоходного налога, $\tau^{(K)}$ — ставка налога на капитал.

Уравнение Эйлера и предельная норма замещения труда потреблением для такой задачи выглядят следующим образом:

$$\frac{\partial U}{\partial C_1} = [1 + (1 - \tau^{(K)})r]\beta \frac{\partial U}{\partial C_2} \quad (7)$$

$$MRS_{N,c} = \frac{\partial U / \partial N_t}{\partial U / \partial C_t} = (1 - \tau_t^{(w)}) w_t. \quad (8)$$

С их помощью может быть получена межвременная относительная цена труда [Туманова, Шагас 2004]:

$$\beta \frac{\partial U / \partial N_2}{\partial U / \partial N_1} = \frac{(1 - \tau_2^{(w)}) w_2}{[1 + (1 - \tau^{(K)}) r] (1 - \tau_1^{(w)}) w_1}. \quad (9)$$

Условие (9) демонстрирует эффект межвременного замещения труда: чем выше реальная заработная плата в настоящем по сравнению с будущим, тем более привлекателен сегодняшний труд; чем выше ставка процента, тем больший доход будет получен от текущей заработной платы, если индивид приобретет активы. Таким образом, падающий с возрастом подоходный налог $\tau_2^{(w)} < \tau_1^{(w)}$ делает будущий труд более привлекательным. Однако аналогичный эффект может быть получен при постоянном во времени подоходном налоге $\tau_2^{(w)} = \tau_1^{(w)}$ и положительном налоге на капитал $\tau^{(K)} > 0$.

Стоит отметить, что положительный налог на капитал, полученный в модели перекрывающихся поколений [Erosa, Gervais 2002], противоречит результатам классических работ, где изучались оптимальные размеры налога на капитал в рамках моделей с бесконечно живущими индивидами [Chamley 1986; Judd 1985]. В работе [Judd, 1999] было показано, что налоги искажают оптимальное решение индивида, так как создают разрывы между предельной нормой замещения часов досуга потреблением в начальном периоде и предельной нормой трансформации часов досуга в потребление в начальном периоде, однако если при наличии подоходного налога этот разрыв постоянен во времени, то в условиях положительного налога на капитал он экспоненциально возрастает во времени. Однако [Aiyagari 1995] показал, что описанный результат в отношении налога на капитал для данного класса моделей оказывается неустойчивым, если присутствуют ограничения на займы и случайные индивидуальные шоки дохода, наличие которых способствует возникновению сбережений по мотиву предосторожности, поскольку застраховаться от таких шоков невозможно.

Таким образом, размер эластичности предложения труда может в значительной степени повлиять не только на размер оптимального подоходного налога, но и на сам тип шкалы налогообложения. Более того, задача оптимизации системы подоходного налогообложения в значительной степени усложняется за счет изменения эластичности предложения труда с возрастом. Наиболее общепринятым способом аппроксимации изменяющегося с возрастом подоходного налога в моделях перекрывающихся поколений является использование налога на капитал. Отсюда возникает вопрос, какие факторы влияют на формирование оптимального налога на капитал?

1.2. Оптимальный налог на капитал в моделях перекрывающихся поколений

Одной из ключевых работ по оптимизации системы налогообложения, которую многие современные исследователи рассматривают в качестве бенчмарка, считается [Conesa et al. 2009], где авторами была использована DSGE-OLG модель². Цель этой работы заключается в том, чтобы показать, при каких условиях оптимальная ставка налогообложения на капитал будет положительной в долгосрочном периоде, и в объяснении механизмов взаимодействия ставок налога на трудовой и капитальный доходы. Ключевые предпосылки модели состоят в наличии индивидуальных шоков дохода, гетерогенности индивидов по способностям и ограничений на займы. Предложение труда описывается в модели достаточно просто: ставки заработных плат формируются экзогенно, то есть не учитывается накопление человеческого капитала, в модели отсутствует механизм принятия решения о занятости, все индивиды выходят на пенсию в определенном возрасте.

² DSGE-OLG (Dynamic Stochastic General Equilibrium Overlapping Generations) — это динамическая стохастическая модель общего равновесия с перекрывающимися поколениями. Она сочетает в себе элементы динамических и стохастических моделей общего равновесия с учетом возрастных групп населения, что позволяет анализировать долгосрочные экономические процессы, включая взаимодействие различных поколений, их потребление, сбережения и трудовую деятельность.

Согласно результатам [Conesa et al. 2009], оптимальный налог на капитал для США составил 36%, в то время как подоходная шкала оказалась плоской со ставкой 23% и налоговым вычетом в 7200 долларов при среднем доходе в 42000 долларов. Ключевым преимуществом такой системы является то, что она стимулирует предложение труда людей старших возрастов, что особенно важно, учитывая тот факт, что пик профиля эффективности достигается примерно в 50 лет.

Согласно анализу чувствительности результатов [Conesa et al. 2009], первым ключевым элементом модели, который способствовал получению такого результата, является относительно высокая эластичность предложения труда по Фришу, которая отрицательно зависит от количества отработанных часов. Учитывая тот факт, что в моделях жизненного цикла профиль отработанных часов представляет собой некоторую убывающую кривую, эластичность предложения труда будет увеличиваться с возрастом.

Полученный [Conesa et al. 2009] положительный налог на капитал представлял собой несколько необычный для того времени результат, поэтому работа [Peterman 2013] была посвящена более детальному изучению его причин в рамках модели с перекрывающимися поколениями для США. Так, существуют два фактора, каждый из которых снижает оптимальный налог на капитал примерно на треть: эластичность предложения труда не меняется с возрастом; случайное наследство, возникающее вследствие неопределенной длительности жизни (*англ.* accidental bequests) и капитал могут облагаться по разным ставкам. Первое объясняется тем, что у государства больше нет стимулов аппроксимировать изменяющийся с возрастом подоходный налог. Второе объясняется тем, что случайное наследство является неэластичным источником дохода, в отличие от остальных видов капитального дохода, поэтому государству выгоднее облагать наследство в первую очередь, тем самым минимизируя искажения от налогообложения.

В качестве менее важных факторов, положительно влияющих на размер налога на капитал, [Peterman 2013] выделяет наличие государственного долга и ограничений на заимствования. Наличие государственного долга способствует росту индивидуальных сбережений, что увеличивает оптимальный налог на капитал.

Логика для второго заключается в следующем: так как индивиды предпочитают сглаживать свое потребление во времени, то, сталкиваясь с горбообразным профилем доходов, на ранних этапах жизненного цикла они стремятся заимствовать средства за счет своих будущих доходов. Ограничения на займы препятствуют сглаживанию потребления, однако решению этой проблемы может поспособствовать налоговая политика: поскольку индивид, как правило, накапливает больше активов в более позднем возрасте, увеличение налога на доход от капитала и снижение налога на трудовой доход приведет к тому, что большая часть налогового бремени распределится на поздние этапы жизненного цикла, что делает профиль потребления более сглаженным во времени. Следовательно, ограничения на заимствования могут способствовать увеличению налога на капитал.

Важной работой, где изучается оптимальный размер налога на капитал, является [Freestone 2020]. Автор предложил добавить в модель [Conesa et al. 2009] разные типы домохозяйств, которые могут состоять либо из двух супругов, либо из одного. Более того, каждый период с некоторой вероятностью возможна смена супружеского статуса. Ключевой результат заключается в том, что в оптимуме налог на капитал для США становится практически нулевым, а шкала подоходного налога становится плоской со ставкой в 30%. Описанная модификация порождает два эффекта: для женатых вероятность развода в следующем периоде способствует формированию избыточных сбережений; для одиноких домохозяйств возможность вступить в брак в следующем периоде снижает стимулы к сбережению из-за своего рода проблемы безбилетника. На основе численных симуляций автором было показано, что чистый эффект будет отрицательным, то есть объем сбережений в экономике упадет. Исходя из этого, низкий налог на капитал может способствовать решению этой проблемы.

Учет состава домохозяйства создает для его членов дополнительные механизмы страхования от возможных шоков доходов с помощью совместного предложения труда [Blundell et al. 2016]. Так, [Freestone 2020] показал, что по мере снижения корреляции шоков заработной платы супругов, у них пропадают стимулы к осуществлению сбережений по мотиву предосторожности,

что равносильно росту эластичности сбережений по ставке процента. Это означает, что оптимальный налог на капитал должен быть понижен, и служит подтверждением наличия связи между эластичностью сбережений по ставке процента и оптимальным налогом на капитал [Summers 1981].

Таким образом, ключевым драйвером положительного налога на капитал является растущая с возрастом эластичность предложения труда и невозможность государством разделять доход от капитала и случайного наследства. В качестве побочных факторов, положительно влияющих на размер оптимальной ставки налога на капитал, можно выделить наличие государственного долга и ограничений на заимствования. Однако учет состава домохозяйств значительно снижает размер ставки налога на капитал в силу возникновения рисков, связанных со сменой супружеского статуса.

1.3. Прогрессивность оптимальной шкалы подоходного налога в моделях перекрывающихся поколений

Возвращаясь к работе [Conesa et al. 2009], вторым важным элементом, обуславливающим его результат, являются нестрахуемые индивидуальные шоки производительности труда, которые оказывают значительное влияние на шкалу подоходного налога. Так, при добавлении в модель перманентных различий по способностям, оптимальная система подоходных налогов становится прогрессивной, так как у государства появляются стимулы к перераспределению богатства от способных в пользу менее способных. Добавление же ограничений на займы не оказало значимого влияния на систему налогообложения, так как прогрессивная схема налогообложения позволяет смягчить эти ограничения, а наличие индивидуальных шоков доходов способствует появлению сбережений по мотиву предосторожности на ранних этапах жизненного цикла.

Однако не только наличие неопределенности в доходах обуславливает прогрессивность шкалы подоходного налога. В работе [Gervais 2012], где автором использовалась модель с пере-

крывающимися поколениями без учета эндогенного накопления человеческого капитала и индивидуальных шоков для экономики США, было показано, что в оптимальной системе налогообложения, где размер ставки подоходного налога не может зависеть от возраста, шкала подоходного налога должна быть прогрессивной в силу того, что доходы индивидов растут с течением возраста. Таким образом, несмотря на растущую во времени эластичность предложения труда, умеренно прогрессивная шкала подоходного налога с высокой ставкой налога на капитал позволяют сгладить профили потребления и доходов во времени, как это делает изменяющийся с возрастом подоходный налог.

Другим важным фактором, обуславливающим прогрессивность оптимальной шкалы подоходного налога, является накопление человеческого капитала. В модели [Conesa et al. 2009] возрастание эластичности предложения труда по Фришу обеспечивается лишь за счет спецификации функции полезности, поэтому часть последующих исследований была посвящена эндогенизации процесса накопления человеческого капитала. Так, в работах [Da Costa, Santos 2018; Peterman 2016] была предложена модификация модели [Conesa et al. 2009], где автором было учтено эндогенное накопление человеческого капитала с помощью механизма LBD, который обеспечивает рост эластичности предложения труда по Фришу с возрастом. Введение эндогенного накопления человеческого капитала делает прогрессивный подоходный налог еще более искажающим в силу того, что появляется связь между текущим предложением труда и будущим объемом человеческого капитала, из-за чего оптимальная шкала становится более плоской. Оптимальный налог на капитал же по-прежнему остается положительным, чтобы аппроксимировать эффект от убывающего с возрастом подоходного налога.

В силу того что эндогенное накопление человеческого капитала оказывает значимое влияние на поведение индивидов, в работе [Blandin, Peterman 2019] сопоставляется влияние различных механизмов его эндогенизации на оптимальные ставки подоходного и капитального налогов в рамках модели с перекрывающимися поколениями по мотивам [Conesa et al. 2009]. Так, помимо LBD, авторы рассматривают механизм *learning-or-doing* (LOD) [Bek-

ker 2009; Ben-Porath 1967]. Суть механизма LOD состоит в том, что общий фонд времени распределяется не только между досугом и работой, но и образованием — именно оно способствует совершенствованию человеческого капитала. В силу того что инвестиции в человеческий капитал, которые в условиях LOD выполняют роль субститута рабочего времени, осуществляются в основном в молодом возрасте, предложение труда молодых работников теперь становится более эластичным, чем у людей старшего возраста, из-за чего представляется разумным ввести подоходный налог, который будет увеличиваться с возрастом. Однако рост подоходного налога делает инвестиции в человеческий капитал менее привлекательными, из-за чего направление изменения подоходного налога с возрастом перестает быть однозначным.

Согласно результатам [Blandin, Peterman 2019], при LOD в экономике присутствует большее количество молодых людей с низким доходом, для которых остро стоит проблема ограничений на займы, так как теперь им приходится посвящать часть своего времени накоплению человеческого капитала. Из-за этого шкала подоходного налога оказывается более прогрессивной, чем при LBD. Более того, прогрессивная система способствует перераспределению доходов между поколениями, то есть от более старых, которые в среднем более богатые [Conesa, Krueger 2006; Erosa, Gervais 2002; Gervais 2012], к молодым, из-за чего прогрессивный налог аппроксимирует растущий с возрастом подоходный налог.

Другим фактором, который может определять оптимальную степень прогрессивности налоговой системы, является совместное решение о предложении труда членами домохозяйства. Так, учет эффекта дополнительного работника, суть которого заключается в том, что в ответ на отрицательный шок дохода одного супруга второй супруг будет увеличивать интенсивность своего труда с целью сгладить траекторию потребления всего домохозяйства во времени [Lundberg 1985], может в значительной степени снижать потребность домохозяйства в страховке, обеспечиваемой государством за счет прогрессивного налогообложения. Именно к этому результату пришли авторы работы [Wu, Krueger 2021].

Подавляющее большинство работ по оптимизации системы налогообложения в рамках моделей общего равновесия игнори-

руют решение о занятости, сосредоточивая свое внимание лишь на реакции отработанных часов [Keane 2022]. Игнорирование реакции занятости является серьезным упрощением в силу того, что соответствующие эластичности оказываются значительно выше, чем для отработанных часов [Keane 2016]. Более того, мужчины и женщины по-разному реагируют на шоки трудовых доходов, как с точки зрения отработанных часов, так и вероятности работать в принципе. Так, в силу того что эластичности предложения труда у женщин в среднем выше, чем у мужчин [Keane 2011], налоговые реформы могут создавать для них негативные стимулы к предложению труда. В работах [Guner et al. 2012; Holter et al. 2019; Karabarounis 2016], где изучается экономика США, были представлены модификации модели с перекрывающимися поколениями, учитывающие решение о занятости каждого из членов домохозяйства. Авторами было показано, что значительная часть реакции предложения труда на изменение налоговой системы приходится именно на вероятность работать. Более того, в работе [Holter et al. 2019] было продемонстрировано, что рост прогрессивности шкалы подоходного налога особенно сильно дестимулирует труд замужних женщин. Однако отметим, что такой эффект во многом может быть связан с тем, что в некоторых странах (Германия, Дания, США, Бельгия, Франция) налоги платятся со всего дохода домохозяйства, поэтому при сильном разрыве в производительности супругов доход менее производительного будет облагаться по избыточной высокой ставке и сокращать его предложение труда [Holter et al., 2019], поэтому переход к более плоской системе налогообложения был бы более разумным. В частности, [Karabarounis 2016] показал, что изменения уровня занятости и отработанных часов с возрастом следуют перевернутой U-образной траектории, из чего следует, что оптимальный подоходный налог должен следовать аналогичной горбообразной траектории.

Таким образом, на степень прогрессивности шкалы налогообложения влияют различные факторы: с одной стороны, возрастная динамика трудовых доходов, гетерогенность по способностям и наличие индивидуальных шоков производительности труда, с другой стороны, тот или иной механизм накопления че-

ловческого капитала. Более того, при учете состава домохозяйства и гетерогенности с точки зрения реакции на шоки трудовых доходов, высокая степень прогрессивности, с одной стороны, может быть избыточна в силу того, что домохозяйства, помимо собственного труда и накоплений, могут использовать труд супругов в качестве недостающего механизма страховки от идиосинкратической неопределенности, а с другой стороны, может оказывать значительное негативное воздействие на труд менее продуктивных членов домохозяйства, если налогом облагается совокупный доход домохозяйства, а не отдельных его представителей.

2. ОПТИМАЛЬНЫЕ ПЕНСИОННЫЕ СИСТЕМЫ

2.1. Виды пенсионных систем и пенсионные сбережения

Существует несколько видов пенсионных систем. В рамках накопительной системы каждый человек индивидуально откладывает часть дохода на свою пенсию. Распределительная система предполагает перераспределение части дохода молодого населения в пользу пожилого. Как правило, пенсионные системы разных стран не бывают исключительно распределительными или накопительными, а сочетают в себе элементы обеих систем в различных пропорциях. В частности, в России действует смешанная пенсионная система, сочетающая в себе накопительную и распределительную части. Накопительная часть была заморожена в 2014 г., что де-факто привело к полностью распределительной пенсионной системе.

Устройство пенсионной системы оказывает влияние на ключевые макроэкономические показатели, такие как потребление, сбережения индивидов и предложение труда. Значительное количество дискуссий происходит вокруг вопроса о том, какое влияние пенсионная система оказывает на сбережения индивидов.

Эконометрические оценки [Feldstein 1974] указывают, что в условиях отсутствия пенсионной системы в США, частные сбережения были бы на 30–50% выше. Как отмечал Фридман [Friedman 1957], система социального обеспечения будет снижать потреб-

ность в сбережениях, поскольку индивидам нет необходимости оставлять резервы для обеспечения потребления в старости. Аналогичный результат получает [Munnell 1974], однако оценки оказались ниже.

[Leimer, Lesnoy 1982] обнаруживают ошибку в исследовании [Feldstein 1974]. Переменная, связанная с социальным обеспечением, которую использовал Фелдстейн, была серьезно искажена в процессе работы с данными. Исправление этой ошибки существенно меняет предполагаемое влияние социального обеспечения на сбережения. После ее исправления и оценки с использованием той же спецификации, что у [Feldstein 1974], авторы приходят к диаметрально противоположным выводам о том, что система социального обеспечения США не оказала значимого отрицательного влияния на сбережения, более того, данные послевоенного периода свидетельствуют о том, что социальное обеспечение, напротив, увеличило сбережения.

Из исследования [Katona 1964], выполненного на данных по США за 1957–1962 гг., следует, что люди, участвующие в частных пенсионных планах, могут сберегать не меньше, а в некоторых случаях и больше. Такое поведение автор объясняет с помощью психологической гипотезы градиента цели. Согласно этой гипотезе, чем ближе человек к цели, тем больше его усилия, а конкретные и достижимые вознаграждения оказывают более стимулирующее воздействие, чем наказание. Индивид, понимающий, что после выхода на пенсию он не будет иметь достаточно средств, чтобы обеспечить желаемый уровень потребления, будет стремиться работать больше, чтобы достигнуть своей цели. К тем же выводам приходит [Cagan 1965]. В своем исследовании он сравнивал сбережения индивидов, участвующих в частных пенсионных планах, со сбережениями индивидов, не участвующих в них, используя данные выборочного обследования американских домохозяйств 1963 г. (пространственные данные). В среднем сбережения в форме прироста наличных, сберегательных счетов и ценных бумаг составили 2,8% дохода у индивидов с частными пенсионными планами против 2,1% в группе без таких планов. Сбережения в форме платежей по страхованию жизни, аннуитетов и долевого участия недвижимом имуществе составили 5,9% против 5,7%. Такое поведение

индивидов он объясняет особым психологическим эффектом: когда человек участвует в индивидуальном пенсионном плане, он острее осознает важность сбережений к старости.

Исследования [Boyle, Murray 1979; Boadway, Clark 1986], проведенные на канадских данных, показывают, что государственные пенсионные планы Канады не оказали значимого влияния на сбережения домохозяйств.

Вопрос избыточных или недостаточных сбережений рассматривается также в рамках моделей перекрывающихся поколений с позиции динамической эффективности. Так, избыточные сбережения могут обуславливать целесообразность распределительной пенсионной системы [İmrohoroglu et al. 1995; Spremann 1984; Aaron 1966], а недостаточные — вести к низкой капиталовооруженности и выпуску. Эти и другие вопросы, связанные с пенсионными системами, будут рассмотрены далее на примере моделей перекрывающихся поколений.

2.2. Анализ перехода от распределительной пенсионной системы к накопительной в моделях перекрывающихся поколений

Исследования, связанные с оптимальным налогообложением и способами обеспечения пожилых индивидов доходами, приобретают все большую популярность в свете дискуссий о необходимости реформирования распределительной пенсионной системы ввиду низкой рождаемости и растущей продолжительности жизни. В рамках этих исследований изучаются статические и динамические аспекты перехода от распределительной системы к накопительной. В первом случае исследователи прибегают к сравнительному анализу устойчивых долгосрочных равновесий, во втором — изучают траектории перехода от одного равновесия к другому, на которых с помощью различных фискальных инструментов стремятся минимизировать потери переходного поколения от смены пенсионных систем.

В простой модели, где индивиды рациональны, рынки совершенны, отсутствуют искажающие налоги и имеет место экзогенное предложение труда, внедрение накопительной пенсионной

системы будет эквивалентно ситуации, в которой нет системы пенсионного страхования, и индивиды самостоятельно сберегают средства, чтобы обеспечить потребление в старости в случае, если государство ссужает налоговые поступления фирмам для оплаты производительного капитала, по аналогии с тем, как это происходит с индивидуальными сбережениями. В тех же условиях распределительная система не будет нейтральной, в ней всегда имеет место эффект вытеснения капитала, что ведет к более высоким процентным ставкам, а более низкая капиталовооруженность сокращает предельный продукт труда, уменьшая заработную плату. Однако введение распределительной системы способно увеличить благосостояние индивидов в случае, если темп роста населения превышает норму отдачи от капитала. Эта особенность распределительной системы доказана в теоретической работе [Aaron 1966]. Если темп роста населения слишком высок, происходит перенакопление капитала, что снижает процентную ставку. В этом случае экономика динамически неэффективна и внедрение распределительной пенсионной системы ведет к парето-улучшению [Spremann 1984] в отношении всех поколений, в том числе и переходного. В условиях быстрорастущего населения каждое поколение может быть уверено в том, что оно будет иметь достаточно высокую пенсию ввиду высокой численности следующих поколений.

Исследование [Abel et al. 1989] указывает на то, что большая часть развитых экономик относится к разряду динамически эффективных. Если экономика динамически эффективна (и процентная ставка выше темпов роста населения) [Samuelson 1975], снятие распределительной пенсионной системы привело бы к увеличению благосостояния будущих поколений. Однако в этой ситуации возникает вопрос о том, что произойдет с благосостоянием переходного поколения и будет ли оно согласно на такой переход. В случае перехода от распределительной системы к накопительной, молодое поколение будет вынуждено обеспечивать живущих пенсионеров и в то же время откладывать на собственную пенсию. Это приводит нас к вопросу о том, существует ли такая траектория перехода, которая будет парето-улучшением для переходного поколения.

Теоретические исследования [Breyer 1989; Homburg 1990], не привязанные к конкретной стране, показывают, что если предложение труда является экзогенным, то не существует траектории перехода от распределительной пенсионной системы к накопительной, которая являлась бы парето-улучшением. Это означает, что невозможно найти такую траекторию, которая улучшила бы благосостояние переходного поколения, не ухудшив положения хотя бы одного будущего поколения. Этот результат в рамках стандартной односекторной модели перекрывающихся поколений получен как для малой открытой, так и для закрытой экономики.

В моделях, в которых используется эндогенное предложение труда и искажающие налоги, распределительная пенсионная система также создает эффект вытеснения капиталовооруженности, снижающий благосостояние индивидов, если экономика является динамически эффективной. Однако в этих условиях возможен парето-эффективный переход к накопительной системе, что доказано в теоретическом исследовании [Breyer, Straub 1993]. Когда взносы в пенсионную систему формируются за счет пропорциональных налогов, в условиях эндогенного труда в экономике формируются искажения. Если в процессе перехода от распределительной системы к накопительной используются паушальные налоги, то за счет уменьшения потерь мертвого груза возможно сформировать фонд, обеспечивающий парето-эффективный переход. Это становится возможным, поскольку при переходе от искажающей системы к неискажающей системе высвобождаются средства из-за снижения потерь мертвого груза, которые можно использовать для компенсации переходному поколению.

[Huang et al. 1997] рассматривают два сценария перехода от распределительной к накопительной пенсионной системе, которые могут быть политически осуществимы, калибруя OLG модель по данным экономики США. В первом случае правительство внезапно прекращает выплаты в рамках текущей пенсионной системы, компенсируя пенсионерам доходы путем выпуска государственного долга, который выплачивается в течение последующих 40 лет. Компенсация подбирается для каждой когорты таким образом, чтобы полезность среднего индивида в когорте

оставалась неизменной. В результате такой реформы в долгосрочном равновесии совокупный физический капитал увеличивается на 40%, отдача от капитала снижается с 6,8% до 4,4%, а ставка налога на трудовой доход снижается с 33,9% до 14%. На переходной траектории выплата государственного долга осуществляется за счет увеличения на 40 лет ставки налога на трудовой доход до 38%. Второй сценарий предполагает другой способ финансирования пенсий с использованием государственного долга. В этом случае государство приобретает права на физический капитал, чтобы пособия по социальному обеспечению можно было финансировать за счет доходов от частного капитала, находящегося в государственной собственности. В этом случае государство стремится сосредоточить достаточно большой запас частного физического капитала, чтобы дохода от этого капитала было достаточно для выплаты пенсионных пособий по социальному обеспечению в переходный период. Эта схема оказывает аналогичное воздействие на живущие поколения, но приносит больше выгод для следующих поколений. Независимо от сценария в долгосрочном периоде ставка налога на трудовой доход установится на более низком уровне, а благосостояние индивидов вырастет, однако благосостояние поколений, родившихся сразу после реформы, снижается, поскольку в переходный период государство временно вынуждено устанавливать более высокие ставки налога на труд.

[Conesa, Krueger 1998] акцентируют внимание на политической осуществимости перехода от распределительной пенсионной системы к накопительной. Почему если в долгосрочном периоде благосостояние в случае переключения на накопительную систему растет, то не становится больше стран с такой системой и мы не наблюдаем страны, находящиеся на этапе перехода? Авторы показывают, что в процессе перехода большинство избирателей понесут потери благосостояния, поэтому они будут придерживаться статус-кво. Анализ проводится на основе модели перекрывающихся поколений, откалиброванной по данным экономики США, и авторы обращают внимание на важность учета эффектов общего равновесия, заключающихся в изменении цен на факторы производства. Так, например, снижение процентных

ставок делает невыгодным переход для молодых, богатых и высокопроизводительных индивидов в результате чего они будут блокировать этот переход. Если абстрагироваться от эффектов общего равновесия, часть из них поддержала бы решение о переходе.

Таким образом, в условиях стареющего населения остро стоит вопрос целесообразности использования распределительной пенсионной системы и ее экономической эффективности, а также возможности перехода от распределительной системы к накопительной. Если труд является экзогенным, этот переход невозможен без потери благосостояния для переходного поколения, поскольку оно, осознавая эти потери, будет стремиться к поддержанию статус-кво. Если труд эндогенен, а налоги являются искажающими, то переход возможен с помощью замены искажающих налогов паушальными, поскольку потери переходного поколения компенсируются снижением потерь мертвого груза. Описанные выше модели не учитывали фактор неопределенности, который имеет особое значение в контексте пенсионных систем. Если рынки страхования отсутствуют или несовершенны, пенсионная система становится механизмом, обеспечивающим частичную страховку от ряда рисков. Когда население не склонно к риску, наличие такой страховки может существенно увеличить благосостояние, даже несмотря на присутствие в экономике искажающих налогов. Это подтверждается рядом исследований в области пенсионных систем, проведенных с использованием моделей перекрывающихся поколений.

2.3. Пенсионные системы в моделях перекрывающихся поколений с неопределенной продолжительностью жизни, систематическими и идиосинкратическими рисками

Одной из первых моделей, в которую была включена неопределенность в контексте исследований пенсионной системы, стала модель перекрывающихся поколений Яари [Yaari 1965]. В этой модели индивиды имеют неопределенную продолжительность жизни, сталкиваются с некоторой вероятностью умереть, которая зависит от возраста. На базе такой модели Яари показывает,

что в условиях риска смерти индивидуальная ставка дисконтирования индивида увеличивается пропорционально вероятности умереть. Из этого следует, что индивид в большей степени будет дисконтировать полезность, получаемую от будущего потребления, а значит, будет больше потреблять сегодня и меньше сберегать. Яри отмечает, что риски, связанные с продолжительностью жизни, можно застраховать путем покупки аннуитета у страховой компании. Индивид ежегодно будет получать фиксированную выплату до тех пор, пока не умрет. Если индивид не оставляет наследства, оптимальным для него будет держать все свои активы в аннуитетах, то есть полностью «застраховаться» на случай смерти.

Данная модель получила развитие в модели Бланшара [Blanchard 1985], известной как «модель вечной молодости». Если в модели Яри вероятность умереть зависела от возраста, то в модели Бланшара, она постоянна. Хотя предпосылка Яри более реалистична, упрощение Бланшара позволяет аналитически выводить зависимости на агрегированном уровне, является довольно гибким и удобным для анализа. Динамическое поведение индивида в отношении потребления и сбережений в этой модели может быть описано с помощью пары дифференциальных уравнений, а получаемое в результате решения равновесие будет единственным и устойчиво. Основным недостатком этой модели является демографическая структура, предполагающая существование слишком большого количества пожилых индивидов в сравнении с тем, которое есть в реальности. Чтобы получить более точные оценки влияния пенсионной системы на экономику в модели перекрывающихся поколений, необходимо использовать более соответствующую действительности демографическую структуру. Анализ влияния распределительной пенсионной системы на рост и благосостояние при реалистичных предпосылках о смертности и демографической структуре населения, то есть тех, которые используют фактические данные, представлен в модели [Bruce, Turnovsky 2013]. Авторы обнаруживают, что в долгосрочном равновесии производительность труда в экономике без распределительной системы растет на 0,71 п.п. быстрее, чем в экономике с распределительной пенсионной системой, что объясняется

более низкими сбережениями в рамках этой пенсионной системы. Ожидаемая полезность индивидов также оказывается выше в экономике без социального обеспечения.

Следующая категория рисков, рассматриваемая в рамках моделей пересекающихся поколений при моделировании пенсионных реформ, связана с идиосинкратическими рисками. [İmrohoroğlu et al. 1995], используя модель перекрывающихся поколений, откалиброванную к американским данным, анализируют эффекты от распределительной пенсионной системы в условиях, где индивиды сталкиваются с риском смерти и индивидуальными шоками доходов, а рынки займов и аннуитетов не доступны. Чтобы застраховаться от будущих колебаний доходов и обеспечить потребление в пожилом возрасте, индивиды сберегают. В старости их потребление обеспечивается частично за счет сбережений и частично за счет пенсионной системы. Пенсионные выплаты формируются за счет пропорционального налога на труд молодого населения. Авторы указывают, что, если пенсионное обеспечение будет составлять 60% дохода в старости, благосостояние индивидов будет существенно выше, чем в экономике без пенсионной системы. Прирост благосостояния происходит частично за счет устранения динамической неэффективности экономики, вызванной перенакоплением капитала, частично за счет того, что пенсионная система заменяет отсутствующие рынки аннуитетов, позволяя сглаживать потребление в пожилом возрасте. Как мы отмечали выше, степень динамической неэффективности характеризуется разницей между темпами роста населения и доходностью капитала. Если темпы роста населения высоки, то для устранения динамической неэффективности потребуется более масштабная пенсионная система и соотношение между пенсионными выплатами и собственными сбережениями на пенсии должно будет расти.

В упомянутой выше модели перекрывающихся поколений [Conesa, Krueger 1998] также анализируется роль идиосинкратических рисков. Авторы подчеркивают, что пенсионная система может повышать благосостояние индивидов через два канала. Первый связан с тем, что пенсионная система предоставляет частичную страховку от индивидуальных шоков дохода. Второй

связан с тем, что пенсионная система заменяет отсутствующие рынки аннуитетов.

[Krueger, Kubler 2006] разрабатывают теоретическую модель пересекающихся поколений со стохастическими шоками производства и несовершенными рынками, на которых доходы от капитала и заработные платы несовершененно коррелированы. Модель не калибруется к конкретным данным какой-либо страны, используется для обобщенного анализа. В отличие от моделей с индивидуальными шоками, в данном случае шок является систематическим и ему подвержены абсолютно все индивиды в экономике. Внедрение распределительной пенсионной системы в этих условиях, если отвлечься от эффекта вытеснения капитала в общем равновесии, представляет собой парето-улучшение, если индивиды не склонны к риску. Однако эффекты общего равновесия, заключающиеся в изменении относительных цен в экономике при переходе к накопительной системе, перевешивают положительный эффект от страховки либо просто сводят его к нулю, если для калибровки модели используются значения несклонности к риску и межвременной эластичности замещения, обычно используемые в литературе.

[Harenberg, Ludwig 2017] изучают эффекты распределительной пенсионной системы в модели перекрывающихся поколений с обоими типами рисков: с систематическими и идиосинкратическими. Авторы разрабатывают OLG модель, калибруя ее к американским данным за период 1990–2000 гг. Именно такой подход, по мнению авторов исследования, более уместен, рассмотрение этих типов риска по отдельности является неполным и ведет к некорректным результатам. Во-первых, это связано с тем, что в течение жизненного цикла индивиды сталкиваются с обоими типами риска. Во-вторых, правильно разработанная пенсионная система может способствовать частичному страхованию от обоих типов риска. Аналитически авторы показывают, что совокупная выгода от совместного страхования обоих рисков больше, чем сумма выгод от страхования от отдельных компонентов.

Таким образом, распределительная пенсионная система в условиях неопределенности оказывает не только негативное влияние на экономику, снижая капиталовооруженность и произ-

водительность труда. Если не склонные к риску индивиды существуют в условиях неопределенности, которая может принимать форму систематических и идиосинкратических рисков, система социального обеспечения может приносить полезность индивидам, частично заменяя отсутствующие страховые рынки. Кроме того, если экономика динамически неэффективна, распределительная пенсионная система позволяет справиться с проблемой перенакопления.

3. КЛЮЧЕВЫЕ АСПЕКТЫ МОДЕЛИРОВАНИЯ БЮДЖЕТНО-НАЛОГОВОЙ ПОЛИТИКИ РОССИИ

Численное моделирование бюджетно-налоговой политики в России на основе моделей перекрывающихся поколений требует учета целого ряда специфических факторов. Результаты последних исследований дают важные ориентиры для построения таких моделей.

Во-первых, модель должна быть откалибрована на основе оценок эластичностей предложения труда именно по России. До недавнего времени в макроэкономической литературе по моделям общего равновесия использовались оценки эластичностей предложения труда по другим странам в силу того, что таких оценок для России не было. Недавние исследования же показали, что предложение труда в России значительно менее эластично, чем во многих других странах [Замниус, Полбин 2021; Замниус и др. 2022], поэтому реакция отработанных часов на изменение заработных плат не может быть описана коэффициентами эластичности для других стран.

Во-вторых, модель должна реалистично описывать изменение заработных плат с возрастом для разных социально-демографических групп населения России. В российской литературе достаточно мало работ посвящено оценке внутрикогортных профилей заработных плат, а представленные в них модели эволюции заработной платы в большинстве случаев не предназначены для построения прогнозов, что ограничивает их применение в рамках разработки моделей общего равновесия. Например, в работах

[Benzell et al. 2015; Nesterova 2024; Зубарев, Нестерова 2019, 2022], посвященных изучению тех или иных аспектов российской экономики на основе OLG-моделей, использовались оценки зарплатных профилей из [Auerbach, Kotlikoff 1987] для США. Однако в работе [Замниус и др. 2023] была предложена модель эволюции заработной платы, которая была специально разработана для калибровки зарплатных траекторий. Ее результаты уже были применены в работах [Шпилева и др. 2023; Шпилева и др. 2024].

В-третьих, модель должна учитывать неопределенность в доходах. В работе [Мартьянова, Полбин 2022] были построены оценки ненаблюдаемой части эволюции заработной платы, которая аппроксимируется AR(1)-процессом. Это стандартная спецификация для стохастического процесса производительности в рамках моделей общего равновесия. В работе [Zamnius 2024] был рассмотрен более сложный процесс с временной и постоянной компонентами, где первая представляет собой белый шум, а вторая — случайное блуждание. Так, результаты показали, что волатильность постоянной составляющей ненаблюдаемой компоненты заработной платы оказалась в несколько раз выше в России, чем в США, что свидетельствует о существенном вкладе неопределенности в формирование заработной платы в России и напрямую влияет на результаты моделирования налоговой и пенсионной политики. В этих условиях повышение налогов, особенно если оно осуществляется для обеспечения более высокого коэффициента замещения пенсий, может иметь положительные эффекты на благосостояние индивидов как в молодом, так и в пожилом возрасте. Введение прогрессивной шкалы подоходного налога в условиях низкой эластичности предложения труда и более волатильных доходов будет создавать более существенные положительные эффекты, если сравнивать, например, с США, и в ряде случаев могут нивелировать отрицательные эффекты от искажающего налогообложения. Произведя расчеты на базе модели OLG, исследователи могут оценить границы, при которых повышение налоговых ставок может приводить к приросту благосостояния населения в России.

В-четвертых, модель должна описывать поведение домохозяйств, состоящих из двух супругов, которые принимают

совместное решение об индивидуальном предложении труда, чтобы учесть механизм страхования за счет совместного предложения труда. В работе [Zamnius 2024] было показано, что для российских домохозяйств именно этот механизм выступает в роли ведущего инструмента для сглаживания траектории потребления во времени, а не собственное предложение труда.

В-пятых, в контексте влияния реформ на пенсионную систему и благосостояние различных когорт населения необходимо не ограничиваться сравнительным анализом долгосрочных равновесий, а исследовать возникающие положительные и отрицательные эффекты в динамике. Приведенные исследования свидетельствуют о том, что пенсионная реформа, направленная на формирование накопительной системы, приводит к улучшению благосостояния всех когорт населения в долгосрочном периоде, но отрицательно сказывается на благосостоянии ныне живущих. По этой причине необходимо рассмотрение механизмов компенсации потерь в динамике, с учетом страховой роли налогообложения и пенсионной системы и эластичности предложения труда людей разных возрастов.

В-шестых, модель должна учитывать эндогенное накопление человеческого капитала, особенно в контексте российского рынка труда. Как показано в работах [Becker 2009; Ben-Porath 1967], инвестиции в образование и профессиональные навыки существенно влияют на динамику заработков в течение жизни. Налоговая политика должна минимизировать искажающее воздействие на стимулы к обучению. Ввиду протекающих демографических процессов средний уровень производительности труда в России будет падать из-за снижения доли людей молодого возраста (с высокой производительностью) и увеличения доли людей пожилого населения (с низкой производительностью). Смягчение отрицательных последствий старения населения может быть достигнуто переквалификацией людей пожилого возраста, созданием для них стимулов к инвестициям в человеческий капитал со стороны налоговой системы.

В-седьмых, при моделировании пенсионной системы важно учитывать неопределенность продолжительности жизни [Yaari 1965; Harenberg, Ludwig 2019]. В России, где ожидаемая продол-

жительность жизни ниже, чем в развитых странах, и рынки страхования развиты слабее, распределительная пенсионная система служит важным инструментом страхования рисков. Как отмечают [Breyer, Straub 1993; Krueger, Kubler 2006], даже экономически неэффективная пенсионная система может улучшать благосостояние за счет снижения неопределенности, особенно для малообеспеченных групп. Кроме того, при оценке последствий пенсионных реформ важно принимать во внимание межпоколенческие трансферты [Caballé, Fuster 2003]. В России, где семьи традиционно играют важную роль в поддержке пожилых, формальные пенсии могут частично вытеснять неформальную помощь. Это требует аккуратного учета альтруистических мотивов в моделях OLG.

ЗАКЛЮЧЕНИЕ

В данной работе были рассмотрены модели перекрывающихся поколений, используемые для анализа оптимальных систем налогообложения и пенсионных систем. В обзоре рассмотрены работы с систематическими и идиосинкратическими рисками, эндогенным и экзогенным предложением труда, накоплением человеческого капитала, гетерогенными домохозяйствами. При разработке оптимальной налоговой системы необходимо учитывать меняющиеся в течение жизненного цикла эластичности предложения труда, различную реакцию членов домохозяйств на шоки доходов, наличие специфических рисков. Учитывая динамику эластичности предложения труда во времени, государство может прибегать к положительному налогу на капитал, который частично может имитировать снижающийся с возрастом подоходный налог.

В условиях стареющего населения большое количество работ посвящено анализу распределительной пенсионной системы и возможности перехода к накопительной. Как правило, такой переход сопровождается потерями благосостояния для того поколения, на жизнь которого приходится пенсионная реформа. Это приводит к тому, что индивиды не поддержат решение о пере-

ходе к другой пенсионной системе, даже если в долгосрочном периоде она увеличит благосостояние всех поколений. Экономика с распределительной пенсионной системой зачастую характеризуется более низким уровнем капиталовооруженности, производительности труда и выпуска, чем экономика без пенсионной системы. Этот вывод может быть опровергнут, если в модели учитываются систематические и идиосинкратические риски, а также неопределенная продолжительность жизни. В условиях неопределенности пенсионная система позволяет частично заменить отсутствующие страховые рынки и, таким образом, повысить благосостояние индивидов.

Количественная оценка последствий реформ в области налогообложения и пенсионной системы для России может быть получена путем численного имитационного анализа модели перекрывающихся поколений. На основе проделанного обзора мы приводим ряд специфических для России элементов, которые такая модель должна учитывать. Это позволит повысить качество количественного анализа налоговых и пенсионных реформ, сформировать более обоснованные рекомендации для экономической политики России.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- Замниус А. В., Полбин А. В. Оценка межвременной эластичности замещения предложения труда для замужних женщин в России // Прикладная эконометрика. 2021. Т. 64, № 4. С. 23–48. <https://doi.org/10.22394/1993-7601-2021-64-23-48>
- Замниус А. В., Полбин А. В., Синельников-Мурылев С. Г. Заработная плата, возраст и экономический рост: оценки для России // Вопросы экономики. 2023. № 6. С. 94–116. <https://doi.org/10.32609/0042-8736-2023-6-94-116>
- Замниус А. В., Полбин А. В., Синельников-Мурылев С. Г. Эластичность предложения труда по заработной плате у женатых мужчин в России // Экономический журнал Высшей школы экономики. 2022. Т. 26. № 2. С. 177–212. <https://doi.org/10.17323/1813-8691-2022-26-2-177-212>
- Зубарев А. В., Нестерова К. В. Оценка последствий пенсионной реформы в России в глобальной CGE-OLG модели // Экономический журнал Высшей школы экономики. 2019. Т. 23. № 3. С. 384–417. <https://doi.org/10.17323/1813-8691-2019-23-3-384-417>
- Зубарев А. В., Нестерова К. В. Фискальная консолидация в условиях пандемии // Вопросы экономики. 2022. № 7. С. 5–26. <https://doi.org/10.32609/0042-8736-2022-7-5-26>
- Мартыанова Е. В., Полбин А. В. Анализ динамики доходов домохозяйств России на основе базы данных РМЭЗ // Финансы: теория и практика. 2022. Т. 26. № 6. С. 271–287. <https://doi.org/10.26794/2587-5671-2022-26-6-271-287>
- Туманова Е. А., Шагас Н. Л. Макроэкономика. Элементы продвинутого подхода. М.: ИНФРА-М, 2004. 400 с.
- Шпилевая А. Е., Полбин А. В., Синельников-Мурылев С. Г. Имитационный анализ OLG-модели с гетерогенными предпочтениями и способностями к обучению // Экономический журнал Высшей школы экономики. 2024. Т. 28. № 1. С. 44–80. <https://doi.org/10.17323/1813-8691-2024-28-1-44-80>
- Шпилевая А. Е., Полбин А. В., Синельников-Мурылев С. Г. Разработка OLG-модели с гетерогенными предпочтениями и способностями к обучению для анализа политики в сфере высшего образования // Экономический журнал Высшей школы экономики. 2023. Т. 27. № 3. С. 449–469. <https://doi.org/10.17323/1813-8691-2023-27-3-449-469>

- Aaron H. The social insurance paradox // *Canadian Journal of Economics and Political Science/Revue canadienne de economiques et science politique*. 1966. Vol. 32. No. 3. P. 371–374. <https://doi.org/10.2307/139995>
- Aaron H. J., Pechman J. A. How taxes affect economic behavior. Washington, DC: Brookings Institution, 1981. No. 14.
- Abel A. B., Mankiw N. G., Summers L. H., Zeckhauser R. J. Assessing dynamic efficiency: Theory and evidence // *The Review of Economic Studies*. 1989. Vol. 56. No. 1. P. 1–19.
- Aiyagari S. R. Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting // *Journal of Political Economy*. 1995. Vol. 103. No. 6. P. 1158–1175. <https://doi.org/10.1086/601445>
- Altonji J. G. Intertemporal substitution in labor supply: Evidence from micro data // *Journal of Political Economy*. 1986. Vol. 94. No. 3. P. 176–215.
- Becker G. S. Human capital: A theoretical and empirical analysis, with special reference to education. Chicago: University of Chicago Press, 2009.
- Ben-Porath Y. The production of human capital and the life cycle of earnings // *Journal of Political Economy*. 1967. Vol. 75. No. 4. P. 352–365.
- Bentham J. The collected works of Jeremy Bentham: An introduction to the principles of morals and legislation. Oxford: Clarendon Press, 1996.
- Benzell S. G. et al. Simulating Russia's and other large economies' challenging and interconnected transitions. National Bureau of Economic Research, 2015. No. 21269.
- Blanchard O. Debts, deficits and finite horizons // *Journal of Political Economy*. 1985. Vol. 93. P. 223–247.
- Blandin A., Peterman W. B. Taxing capital? The importance of how human capital is accumulated // *European Economic Review*. 2019. Vol. 119. P. 482–508. <https://doi.org/10.1016/j.euroecorev.2019.08.007>
- Blomquist N. S. The effect of income taxation on the labor supply of married men in Sweden // *Journal of Public Economics*. 1983. Vol. 22. No. 2. P. 169–197. [https://doi.org/10.1016/0047-2727\(83\)90064-6](https://doi.org/10.1016/0047-2727(83)90064-6)
- Blundell R. et al. Consumption inequality and family labor supply // *American Economic Review*. 2016. Vol. 106. No. 2. P. 387–435. <https://doi.org/10.1257/aer.20121549>
- Boadway R., Clark W. S. The government budget, the accumulation of capital, and long-run welfare // *Fiscal and Monetary Policy*. 1986. Vol. 21.
- Boyle P., Murray J. Social security wealth and private saving in Canada // *The Canadian Journal of Economics/Revue canadienne d'Economie*. 1979. Vol. 12. No. 3. P. 456–468.

- Breyer F., Straub M. Welfare effects of unfunded pension systems when labor supply is endogenous// *Journal of Public Economics*. 1993. Vol. 50. No. 1. P. 77–91. [https://doi.org/10.1016/0047-2727\(93\)90061-W](https://doi.org/10.1016/0047-2727(93)90061-W)
- Bruce N., Turnovsky S. J. Social security, growth, and welfare in overlapping generations economies with or without annuities// *Journal of Public Economics*. 2013. Vol. 101. P. 12–24. <https://doi.org/10.1016/j.jpubeco.2013.02.007>
- Caballé J., Fuster L. Pay-as-you-go Social Security and the Distribution of Altruistic Transfers// *The Review of Economic Studies*. 2003. Vol. 70. No. 3. P. 541–567.
- Cagan P. The Effect of Pension Plans on Aggregate Savings: Evidence from a Sample Survey. National Bureau of Economic Research, 1965.
- Chamley C. Optimal taxation of capital income in general equilibrium with infinite lives// *Econometrica*. 1986. Vol. 54. No. 3. P. 607–622.
- Chetty R. et al. Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins// *American Economic Review*. 2011. Vol. 101. No. 3. P. 471–475. <https://doi.org/10.1257/aer.101.3.471>
- Conesa J. C. et al. Taxing capital? Not a bad idea after all!// *American Economic Review*. 2009. Vol. 99. No. 1. P. 25–48. <https://doi.org/10.1257/aer.99.1.25>
- Conesa J. C., Krueger D. Social Security Reform with Heterogeneous Agents// *Review of Economic Dynamics*. 1999. Vol. 2. No. 4. P. 757–795. <https://doi.org/10.1006/redy.1998.0039>
- Conesa J. C., Krueger D. On the optimal progressivity of the income tax code // *Journal of Monetary Economics*. 2006. Vol. 53. No. 7. P. 1425–1450. <https://doi.org/10.1016/j.jmoneco.2005.03.016>
- Da Costa C. E., Santos M. R. Age-dependent taxes with endogenous human capital formation// *International Economic Review*. 2018. Vol. 59. No. 2. P. 785–823. <https://doi.org/10.1111/iere.12288>
- Diamond P. A. National debt in a neoclassical growth model// *American Economic Review*. 1965. Vol. 55. No. 5. P. 1126–1150.
- Domeij D., Floden M. The labor-supply elasticity and borrowing constraints: Why estimates are biased// *Review of Economic Dynamics*. 2006. Vol. 9. No. 2. P. 242–262. <https://doi.org/10.1016/j.red.2005.11.001>
- Erosa A., Gervais M. Optimal taxation in life-cycle economies// *Journal of Economic Theory*. 2002. Vol. 105. No. 2. P. 338–369. <https://doi.org/10.1006/jeth.2001.2877>
38. Feldstein M. Social security, induced retirement, and aggregate capital accumulation// *Journal of Political Economy*. 1974. Vol. 82. No. 5. P. 905–926.

- Freestone O. Economic inequality over the life cycle in Australia. Ph.D. thesis, Australian National University (ANU), 2020.
- French S., Stafford T. Returns to Experience and the Elasticity of Labor Supply. UNSW Business School Research Paper, 2017.
- Friedman M. A Theory of the Consumption Function. Princeton, N.J.: Princeton University Press, 1957.
- Gervais M. On the optimality of age-dependent taxes and the progressive US tax system // Journal of Economic Dynamics and Control. 2012. Vol. 36. No. 4. P. 682–691. <https://doi.org/10.1016/j.jedc.2011.11.011>
- Guner N. et al. Taxation and household labour supply // The Review of Economic Studies. 2012. Vol. 79. No. 3. P. 1113–1149.
- Hansen G. D. Indivisible labor and the business cycle // Journal of Monetary Economics. 1985. Vol. 16. No. 3. P. 309–327. [https://doi.org/10.1016/0304-3932\(85\)90039-X](https://doi.org/10.1016/0304-3932(85)90039-X)
- Harenberg D., Ludwig A. Idiosyncratic risk, aggregate risk, and the welfare effects of social security // International Economic Review. 2019. Vol. 60. No. 2. P. 661–692. <https://doi.org/10.1111/iere.12365>
- Holter H. A. et al. How do tax progressivity and household heterogeneity affect Laffer curves? // Quantitative Economics. 2019. Vol. 10. No. 4. P. 1317–1356. <https://doi.org/10.3982/QE653>
- Homburg S. The efficiency of unfunded pension schemes // Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die Gesamte Staatswissenschaft. 1990. Vol. 146. No. 4. P. 640–647.
- Huang H. et al. Two computations to fund social security // Macroeconomic Dynamics. 1997. Vol. 1. No. 1. P. 7–44.
- Imai S., Keane M. P. Intertemporal labor supply and human capital accumulation // International Economic Review. 2004. Vol. 45. No. 2. P. 601–641. <https://doi.org/10.1111/j.1468-2354.2004.00138.x>
- İmrohoroglu A., İmrohoroglu S., Joines D. H. A life cycle analysis of social security // Economic Theory. 1995. Vol. 6. P. 83–114.
- Iskhakov F., Keane M. Effects of taxes and safety net pensions on life-cycle labor supply, savings and human capital: The case of Australia // Journal of Econometrics. 2021. Vol. 223. No. 2. P. 401–432. <https://doi.org/10.1016/j.jeconom.2020.01.023>
- Judd K. L. Redistributive taxation in a simple perfect foresight model // Journal of Public Economics. 1985. Vol. 28. No. 1. P. 59–83. [https://doi.org/10.1016/0047-2727\(85\)90020-9](https://doi.org/10.1016/0047-2727(85)90020-9)
- Judd K. L. Optimal taxation and spending in general competitive growth models // Journal of Public Economics. 1999. Vol. 71. No. 1. P. 1–26. [https://doi.org/10.1016/S0047-2727\(98\)00054-1](https://doi.org/10.1016/S0047-2727(98)00054-1)

- Karabarbounis M. A road map for efficiently taxing heterogeneous agents // American Economic Journal: Macroeconomics. 2016. Vol. 8. No. 2. P. 182–214. <https://doi.org/10.1257/mac.20140274>
- Katona G. Private Pensions and Individual Savings. Ann Arbor: Survey Research Center, Institute for Social Research, University of Michigan, 1964.
- Keane M.P. Labor supply and taxes: A survey // Journal of Economic Literature. 2011. Vol. 49. No. 4. P. 961–1075. <https://doi.org/10.1257/jel.49.4.961>
- Keane M.P. Life-cycle labour supply with human capital: econometric and behavioural implications // The Economic Journal. 2016. Vol. 126. No. 592. P. 546–577. <https://doi.org/10.1111/ecoj.12363>
- Keane M.P. Recent research on labor supply: Implications for tax and transfer policy // Labour Economics. 2022. Vol. 77. No. 102026. <https://doi.org/10.1016/j.labeco.2021.102026>
- Krueger D., Kubler F. Pareto-improving social security reform when financial markets are incomplete! // American Economic Review. 2006. Vol. 96. No. 3. P. 737–755. <https://doi.org/10.1257/aer.96.3.737>
- Kydland F.E. Business Cycles and Aggregate Labor Market Fluctuations. Princeton: Princeton University Press, 1995.
- Leimer D.R., Lesnoy S.D. Social security and private saving: New time-series evidence // Journal of Political Economy. 1982. Vol. 90. No. 3. P. 606–629. <https://doi.org/10.1086/261077>
- Lundberg S. The added worker effect // Journal of Labor Economics. 1985. Vol. 3. No. 1, Part 1. P. 11–37. <https://doi.org/10.1086/298069>
- MaCurdy T.E. An empirical model of labor supply in a life-cycle setting // Journal of Political Economy. 1981. Vol. 89. No. 6. P. 1059–1085. <https://doi.org/10.1086/261023>
- Mirrlees J. An exploration in the theory of optimum income taxation // The Review of Economic Studies. 1971. Vol. 38. No. 2. P. 175–208. <https://doi.org/10.2307/2296779>
- Mirrlees J. et al. The Mirrlees Review: conclusions and recommendations for reform // Fiscal Studies. 2011. Vol. 32. No. 3. P. 331–359.
- Modigliani F. The life cycle hypothesis of saving, the demand for wealth and the supply of capital // Social Research. 1966. Vol. 33. No. 2. P. 160–217.
- Munnell A.H. The Effect of Social Security on Personal Saving. Cambridge, Mass.: Ballinger, 1974.
- Nesterova K.V. Estimating the effect of Progressive Personal Income Tax Schedule for Russia in a global CGE-OLG model // Journal of Tax

- Reform. 2024. Vol. 10. No. 2. P. 334–354. <https://doi.org/10.15826/jtr.2024.10.2.172>
- Ohanian L. E., Raffo A. Aggregate hours worked in OECD countries: New measurement and implications for business cycles // *Journal of Monetary Economics*. 2012. Vol. 59. No. 1. P. 40–56. <https://doi.org/10.1016/j.jmoneco.2011.11.005>
- Palmer E. Public and private pensions and saving in Sweden // *Conjugating Public and Private: The Case of Pensions*. Geneva: International Social Security Association, 1987.
- Peterman W. B. Determining the motives for a positive optimal tax on capital // *Journal of Economic Dynamics and Control*. 2013. Vol. 37. No. 1. P. 265–295.
- Peterman W. B. The effect of endogenous human capital accumulation on optimal taxation // *Review of Economic Dynamics*. 2016. Vol. 21. P. 46–71. <https://doi.org/10.2139/ssrn.1991840>
- Piketty T., Saez E. Optimal labor income taxation // *Handbook of public economics*. Vol. 5. Amsterdam: Elsevier, 2013. P. 391–474.
- Pistaferri L. Anticipated and unanticipated wage changes, wage risk, and intertemporal labor supply // *Journal of Labor Economics*. 2003. Vol. 21. No. 3. P. 729–754. <https://doi.org/10.1086/374965>
- Ramsey F. P. A Contribution to the Theory of Taxation // *The Economic Journal*. 1927. Vol. 37. No. 145. P. 47–61. <https://doi.org/10.2307/2222721>
- Samuelson P. A. An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money // *Journal of Political Economy*. 1958. Vol. 66. No. 6. P. 467–482. <https://doi.org/10.1086/258100>
- Samuelson P. A. Optimum social security in a life-cycle growth model // *International Economic Review*. 1975. Vol. 16. No. 3. P. 539–544. <https://doi.org/10.2307/2525994>
- Shaw K. L. Life-cycle labor supply with human capital accumulation // *International Economic Review*. 1989. Vol. 30. No. 2. P. 431–456. <https://doi.org/10.2307/2526656>
- Spremann K. Intergenerational contracts and their decomposition // *Zeitschrift für Nationalökonomie*. 1984. Vol. 44. No. 3. P. 237–253.
- Stern N. H. On the specification of models of optimum income taxation // *Journal of Public Economics*. 1976. Vol. 6. No. 1–2. P. 123–162. [https://doi.org/10.1016/0047-2727\(76\)90044-X](https://doi.org/10.1016/0047-2727(76)90044-X)
- Summers L. H. Taxation and capital accumulation in a life cycle growth model // *American Economic Review*. 1981. Vol. 71. No. 4. P. 547–560.

- Yaari M.E. Uncertain lifetime, life insurance, and the theory of the consumer//The Review of Economic Studies. 1965. Vol. 32. No. 2. P. 137–150. <https://doi.org/10.2307/2296058>
- Wallenius J. Human capital accumulation and the intertemporal elasticity of substitution of labor: How large is the bias?// Review of Economic Dynamics. 2011. Vol. 14. No. 4. P. 577–591. <https://doi.org/10.1016/j.red.2011.06.002>
- Wu C., Krueger D. Consumption insurance against wage risk: Family labor supply and optimal progressive income taxation//American Economic Journal: Macroeconomics. 2021. Vol. 13. No. 1. P. 79–113. <https://doi.org/10.1257/mac.20180125>
- Zamnius A. Estimating Labor Supply Elasticities in the Russian Federation Using a Two Earners Life Cycle Model. 2024. Available at SSRN: <https://ssrn.com/abstract=4822593>, <http://dx.doi.org/10.2139/ssrn.4822593>

Прогнозирование индексов промышленного производства с использованием продвинутых методов нейросетевого моделирования

Каукин А. С., кандидат экономических наук, заведующий лаборатории отраслевых рынков и инфраструктуры Института Гайдара.

Косарев В. С., научный сотрудник лаборатории отраслевых рынков и инфраструктуры Института Гайдара.

Хубежова Д. И., младший научный сотрудник лаборатории отраслевых рынков и инфраструктуры Института Гайдара.

Аникутин М. Ю., младший научный сотрудник лаборатории отраслевых рынков и инфраструктуры Института Гайдара.

Швецов О. А., младший научный сотрудник лаборатории отраслевых рынков и инфраструктуры Института Гайдара.

Аннотация. В работе исследуются возможности повышения качества краткосрочного прогнозирования отраслевых индексов промышленного производства России с использованием современных нейросетевых моделей и новостных данных. Представлен обзор методов прогнозирования временных рядов и подходов к извлечению экономически значимой информации из текстов. Разработаны новостные индексы на основе RuBERT, больших языковых моделей LLaMA и DeepSeek, а также технологии RAG, обеспечивающей дополнение анализируемых новостей релевантным контекстом. Рассмотрены методы оценки отраслевой релевантности, переранжирования и суммаризации новостей. Проведено сравнение одномерных и многомерных моделей N-HITS с эконометрическими моделями ARIMA и VAR при прогнозировании 13 отраслевых показателей на горизонте до трех месяцев. Качество прогнозов оценено посредством кросс-валидации временных рядов с учетом величины ошибки и точности предсказания направления изменений. Эмпирические результаты свидетельствуют об улучшении прогнозов при включении новостных индексов и преимуществах нейросетевого подхода для большинства рассматриваемых показателей. Наиболее выраженный эффект достигается при использовании больших языковых моделей с расширенным контекстом и учетом отраслевой релевантности новостей. Полученные результаты показывают перспективность применения больших языковых моделей для обработки текстовых данных и построения опережающих индикаторов промышленной динамики.

1. ОБЗОР СОВРЕМЕННЫХ НЕЙРОСЕТЕВЫХ АРХИТЕКТУР ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

1.1. Рекуррентная нейронная сеть с долгосрочной памятью (LSTM)

Подход был разработан [1] в 1997 г. с целью решения проблем стандартных рекуррентных нейронных сетей (RNN — Recurrent neural network), неспособных эффективно обучаться на задачах с длинными временными зависимостями. LSTM (Long Short-Term Memory) стала первой практически работоспособной RNN-архитектурой, сохраняющей информацию на длительных интервалах времени, и именно это сделало ее прорывной технологией, используемой и в настоящее время, в таких задачах, как обработка текста, речи, временных рядов.

LSTM — это рекуррентная ячейка, некоторая совокупность таких ячеек, объединенных последовательно и/или параллельно, является нейронной сетью. Архитектура такой ячейки представляет собой функцию, которую условно можно разделить на несколько составляющих, которые визуализированы на рис. 1 [2].

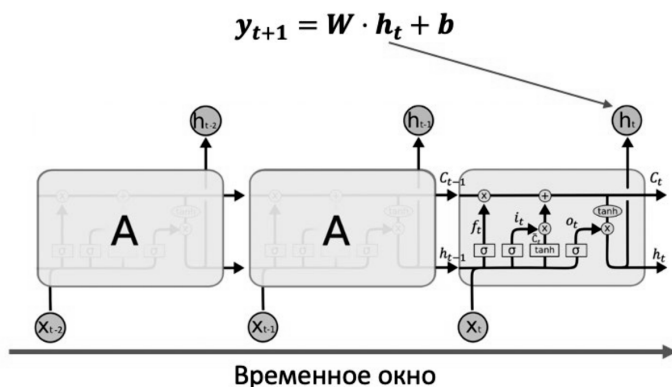


Рис. 1. Устройство LSTM ячейки

Вектор верхнего уровня (C_{t-1}), передающий информацию в реплику ячейки, хранит информацию о предыдущих состояниях и отвечает за долгосрочную память модели. С течением времени

не предполагается его существенной мутации. Вектор нижнего уровня (h_{t-1}) зависит от прошлых значений сети, но может значительно меняться от текущего входа и отвечает за краткосрочную память модели.

Ниже в виде последовательных уравнений представлена формула, описывающая процесс получения модельного вывода, соответствующая графику выше.

$$i_{tn} = \sigma(W_{in} \cdot x_t + U_{in} \cdot h_{\{t-1\}} + b_{in}), \quad (1)$$

$$f_t = \sigma(W_f \cdot x_t + U_f \cdot h_{\{t-1\}} + b_f), \quad (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot x_t + U_c \cdot h_{\{t-1\}} + b_c), \quad (3)$$

$$C_t = f_t \cdot C_{\{t-1\}} + i_t \cdot \tilde{C}_t, \quad (4)$$

$$o_t = \sigma(W_o \cdot x_t + U_o \cdot h_{\{t-1\}} + b_o), \quad (5)$$

$$h_t = o_t \cdot \tanh(C_t), \quad (6)$$

$$y_{t+1} = W \cdot h_t + b, \quad (7)$$

где,

f_t — доля информации, которую нужно «забыть»;

W — матрица весов;

h_{t-1} — краткосрочная память ячейки;

x_t — новые данные;

C_t — вектор долгосрочной памяти;

i_t — доля долгосрочной памяти, передающаяся в следующую реплику ячейки;

o_t — доля краткосрочной памяти, передающаяся в следующую реплику ячейки,

y_{t+1} — прогноз модели.

Обычно нейронная сеть включает несколько рекуррентных ячеек. Добавление рекуррентных нейронов может осуществляться параллельно на одном слое или в несколько слоев, что приводит к глубокой нейронной сети. Это позволяет находить в дан-

ных более сложные закономерности. Однако увеличение числа нейронов пропорционально увеличивает количество параметров, подлежащих оценке. Так, при использовании двух параллельных рекуррентных ячеек требуется 75 параметров. При наличии 75 параметров и относительно небольшого числа наблюдений во времени, которые имеются в случае месячных данных, возникает риск переобучения, что ограничивает применение данной архитектуры.

Описание основных гиперпараметров LSTM модели и принцип их выбора

Кроме выбора числа нейронов на одном слое, существует ряд гиперпараметров в нейросетевой модели, выбор которых имеет определяющее значение. Основной, но далеко не полный список, включает:

- число рекуррентных нейронов на одном слое;
- глубина слоев;
- функции активации внутренних слоев;
- параметры регуляризации;
- инициализация весов модели;
- временное окно, обрабатываемое рекуррентной ячейкой;
- параметры оптимизатора;
- число итераций градиентного спуска (эпохи);
- размер подвыборки для стохастического градиентного спуска (размер мини-батча).

Глубина нейронной сети, выраженная в количестве слоев, влияет на ее способность моделировать многоуровневые иерархические представления данных. Глубокие сети могут способны захватывать более сложные зависимости. Однако увеличение глубины слоев увеличивает риск градиентного затухания или взрывного роста градиентов, что затрудняет обучение. Подбор глубины требует балансировки между способностью сети к обобщению и риском переобучения. На практике глубина сети подбирается эмпирически с использованием методов кросс-валидации.

Функции активации играют ключевую роль в добавлении нелинейности в модель, что позволяет сети обучаться сложным зави-

симостям в данных. Наиболее популярные функции активации для рекуррентных сетей включают гиперболический тангенс ($\tanh$) и сигмоиду. Выбор функции активации зависит от специфики задачи: сигмоида и гиперболический тангенс хорошо работают в задачах с ограниченным диапазоном значений, а более простые функции, такие как ReLU и ее модификации, могут быть полезны в глубоких сетях для предотвращения проблем затухания градиента. В процессе подбора оптимальной архитектуры мы используем тестирование нескольких функций активаций, что является стандартной практикой при настройке гиперпараметров.

Регуляризация используется для предотвращения переобучения, особенно в сложных моделях с большим числом параметров. В рекуррентных сетях могут применяться методы регуляризации, такие как L1 и L2-регуляризация, которые ограничивают значения весов и помогают избежать их чрезмерного увеличения. Также часто используется метод отсечения (dropout), который временно отключает случайные нейроны во время обучения, снижая вероятность переобучения. Правильный выбор коэффициентов регуляризации и методов регуляризации существенно влияет на стабильность и обобщающую способность модели.

Инициализация весов в рекуррентных сетях имеет большое значение, поскольку неправильная инициализация может привести к медленному обучению или даже к застреванию в локальных минимумах. Стандартные методы инициализации, такие как метод инициализации Ксавье или инициализация Хе, позволяют распределить начальные веса так, чтобы избежать взрывного роста или затухания градиентов. Хорошая инициализация весов помогает градиентному спуску начать оптимизацию с точки, которая потенциально ближе к глобальному минимуму, что улучшает сходимость и позволяет добиться лучшего качества обучения.

Длина временного окна, обрабатываемого рекуррентной ячейкой, определяет, насколько далеко модель учитывает прошлые значения временного ряда при построении прогнозов. Слишком короткое окно может привести к потере важной информации, тогда как слишком длинное — к избыточной сложности и увеличению риска переобучения. Последнее возникает из-за

уменьшения количества образцов для обучения и тестирования модели, что может быть критически важно в случае небольшого количества точек во времени. Оптимальная длина временного окна определяется экспериментально в зависимости от свойств временного ряда. Обычно используется перебор некоторого набора значений длины окна. Критерием выбора, является значение ошибки на тестовых данных.

Выбор оптимизатора и его параметров, таких как скорость обучения, имеет критическое значение для сходимости модели. Популярные оптимизаторы, такие как Adam, RMSprop и SGD, показывают различные результаты в зависимости от специфики задачи и размера данных. Скорость обучения определяет шаг обновления весов в процессе градиентного спуска. От выбора шага градиентного спуска в том числе зависит процессорное время, затрачиваемое на обучение модели, но главным образом параметр влияет на конечный оптимум, к которому приходит модель и, соответственно, на обобщающую способность и непосредственно качество прогноза. Часто на практике используется оптимизатор Adam. Оптимальный выбор его параметров проводится экспериментально с применением поиска по сетке.

Количество эпох определяет, сколько раз модель пройдет по всему набору обучающих данных. Если количество эпох недостаточно, модель может недообучиться, в то время как чрезмерное количество эпох может привести к переобучению. Оптимальное число эпох подбирается на основе валидационной выборки с применением ранней остановки, которая прерывает обучение, когда ошибка на валидационной выборке перестает уменьшаться.

Размер подвыборки (мини-батча) для одной итерации стохастического градиентного спуска в процессе обратного распространения ошибки влияет на качество обучения и скорость сходимости. Малый размер мини-батча позволяет учитывать большее количество случайных изменений, что может улучшить обобщающую способность модели, но также приводит к более шумным обновлениям весов. Слишком большой размер мини-батча может уменьшить обобщающую способность. Оптимальный размер подбирается экспериментально в зависимости от объема данных.

1.2. Градиентный бустинг (Cat Boost)

CatBoost — это алгоритм градиентного бустинга¹ на решающих деревьях, разработанный компанией «Яндекс»². Он предназначен для эффективной работы с табличными данными и особенно хорошо справляется с задачами, где присутствует большое количество категориальных признаков. В отличие от других реализаций градиентного бустинга CatBoost использует собственные подходы к обработке таких признаков и подходы борьбы с переобучением.

Принцип работы CatBoost основан на последовательном построении ансамбля деревьев решений, каждый из которых аппроксимирует антиградиент функции потерь. На каждом шаге новое дерево добавляется к существующей модели, уменьшая ошибку предсказания. Такая стратегия позволяет последовательно улучшать качество модели. В отличие от классических методов, CatBoost применяет упорядоченный бустинг — особый способ построения деревьев, при котором для каждой строки обучающей выборки используются только предыдущие по порядку примеры. Это существенно снижает вероятность утечки информации и переобучения, особенно при работе с категориальными признаками.

Еще одной ключевой особенностью является использование симметричных деревьев (oblivious trees), где на каждом уровне применяется одно и то же условие разделения ко всем ветвям. Такая структура упрощает и ускоряет обучение, а также делает модель более устойчивой. Для категориальных признаков CatBoost применяет специальную технику кодирования, при которой категориальные значения преобразуются в числовые, основываясь на статистике по целевой переменной. При этом используется случайное сглаживание и упорядоченная схема кодирования, чтобы избежать утечки целевой информации.

Для задач регрессии в CatBoost можно использовать различные функции потерь, такие как среднеквадратичная ошибка (RMSE), средняя абсолютная ошибка (MAE), квантильная ошибка

¹ <https://education.yandex.ru/handbook/ml/article/gradientnyj-busting>

² <https://catboost.ai/docs/en/concepts/tutorials>

и другие. Эти функции определяют, насколько хорошо модель соответствует фактическим значениям, и используются в процессе оптимизации.

Модель CatBoostRegressor имеет множество гиперпараметров, которые можно настраивать в зависимости от характера задачи. Основные из них включают количество деревьев, скорость обучения, глубину дерева, параметры регуляризации, стратегию семплирования объектов и особенности обработки категориальных признаков. Кроме того, можно указать тип вычислений — на CPU или GPU, выбрать метод бутстрепа, настроить раннюю остановку обучения при отсутствии улучшений на валидационном наборе и многое другое.

Описание основных гиперпараметров CatBoostRegressor модели и принцип их выбора.

- *Скорость обучения (learning_rate).* Чем меньше значение, тем более «плавно» происходит обучение: модель делает небольшие шаги в направлении уменьшения ошибки. При малом learning_rate требуется больше итераций, чтобы достичь хороших результатов, но итоговая модель обычно оказывается более точной. Типичные значения находятся в диапазоне от 0,01 до 0,1. При использовании малого learning_rate необходимо увеличить количество итераций (параметр iterations) — обычно от 500 до 3000 и выше. Если learning_rate выше 0,2, модель может быстро переобучиться, особенно на небольших датасетах.
- *Глубина деревьев (depth)* определяет сложность базовых моделей и, соответственно, способность улавливать сложные зависимости. На практике чаще всего используют значения от 6 до 10. Глубокие деревья (например, с depth = 12 или больше) могут дать прирост точности на больших наборах данных, но также повышают риск переобучения и увеличивают время обучения. На небольших выборках разумно ограничить глубину до 6–8 уровней.
- *Регуляризация.* Регуляризация в CatBoost осуществляется с помощью параметра l2_leaf_reg, который добавляет штраф за большие веса в листьях дерева. Этот параметр особенно важен при наличии шума в данных и помогает стабилизиро-

вать модель. Обычно он варьируется в пределах от 1 до 10. Чем выше значение, тем сильнее регуляризация и тем более сглаженные (но, возможно, менее точные) будут предсказания.

- *Бутстреп*. Параметры, связанные с бутстрепом, управляют тем, как данные семплируются для построения деревьев. При использовании `bootstrap_type='Bayesian'` можно дополнительно контролировать семплирование через параметр `bagging_temperature`: низкое значение приводит к более равномерному отбору объектов, тогда как высокое усиливает фокус на «информативных» примерах. Если вы хотите получить более стабильную модель, лучше установить `bootstrap_type='Bayesian'` с `bagging_temperature` около 1.
- *Категориальные признаки*. При наличии категориальных признаков необходимо указать их явно. CatBoost автоматически применяет упорядоченное кодирование категорий, и от правильной передачи этого параметра зависит корректность обработки таких признаков.
- *Ранняя остановка* (`early_stopping_rounds`). Использование ранней остановки, особенно при наличии валидационного набора. Этот параметр позволяет прервать обучение, если качество на валидации не улучшается в течение заданного количества итераций. Это эффективный способ избежать переобучения без необходимости подбирать точное количество итераций вручную.

В целом подход к выбору гиперпараметров обычно следующий: сначала фиксируется небольшое значение `learning_rate` (например, 0,05), затем подбирается достаточное количество итераций (150–300), глубина дерева (6–10) и далее проводится перебор `l2_leaf_reg`, `bootstrap`-параметров и глубины для нахождения оптимального баланса между качеством и устойчивостью модели.

1.3. Нейронная иерархическая интерполяция для прогнозирования временных рядов (N-HITS)

Модель NHITS (Neural Hierarchical Interpolation for Time Series Forecasting) — это архитектурный подход для прогнозирования временных рядов, предложенный в 2022 г. исследователями

из Meta AI [3]. На платформе используется версия, реализованная в Python-пакете NeuralForecast.

Основной принцип работы NHITS заключается в последовательном построении прогноза с помощью нескольких блоков, каждый из которых отвечает за аппроксимацию структуры временного ряда на определенном уровне детализации. Вместо авторегрессионного подхода, когда предсказание делается шаг за шагом и может накапливать ошибки, NHITS использует схему, в которой весь горизонт прогноза строится одновременно. Каждый блок вносит свой вклад, уточняя прогноз по сравнению с предыдущими итерациями. Итоговое предсказание формируется как сумма всех вкладов, накопленных от начала до конца модели.

Каждый блок NHITS работает с остатком прогноза. То есть он получает на вход ошибку между текущим приближением и истинным значением и обучается приближать именно эту ошибку, чтобы корректировать общее предсказание. Это реализует идею остаточного обучения, аналогичную используемой в N-BEATS, но с более сложной структурой обработки и более эффективными представлениями.

Архитектура модели организована иерархически. Блоки обучаются в определенном порядке, от грубых приближений к детализированным. Такой подход позволяет эффективно учитывать структуру данных на разных временных масштабах, например сначала улавливать долгосрочные тренды, затем сезонные паттерны и, наконец, краткосрочную динамику. Это особенно важно для сложных временных рядов с многослойной внутренней структурой.

Входными данными для NHITS являются исторические значения временного ряда, временные признаки, такие как день недели или месяц, а также, при необходимости, экзогенные переменные, отражающие внешние влияния. Модель может обрабатывать как одномерные ряды, так и панели временных рядов с множественными объектами. Выходом модели является вектор прогнозируемых значений на заданный горизонт, полученный как сумма выходов всех блоков.

Важнейшими параметрами являются количество блоков, структура каждого блока, длина входного окна и длина прогноза.

*Описание основных гиперпараметров LSTM модели
и принцип их выбора*

- *Размер входного окна* (`input_size`), то есть количество предыдущих наблюдений, на которых базируется прогноз. Его выбор должен учитывать сезонность, тренды и тип прогнозируемого ряда. Если входное окно слишком короткое, модель может не уловить закономерности. Если слишком длинное — могут появиться шум и переобучение. В большинстве случаев разумно выбирать окно в два-три раза больше, чем горизонт прогноза. Например, если нужно предсказать 12 шагов вперед, можно использовать 36–48 входных точек.
- *Горизонт прогноза* — это длина выходного вектора, то есть на сколько шагов вперед модель должна предсказать. NHITS подходит как для коротких, так и для длинных горизонтов, но при большом горизонте стоит уделить больше внимания структуре блоков.
- *Количество блоков* (`n_blocks`) задает глубину архитектуры. Каждый блок приближает определенный уровень временной структуры, от грубого к детальному. Чем больше блоков, тем больше возможностей у модели улавливать сложные зависимости. Однако увеличение количества блоков может приводить к проблеме переобучения и требует больше данных и времени на обучение. В типичных случаях достаточно 2–4 блоков. Для очень сложных или сезонных данных может понадобиться 6 и более.
- В каждом блоке модель использует *многослойный перцептрон* (MLP), структура которого определяется числом слоев (`n_layers`) и размером скрытых представлений (`mlp_units`). Увеличение числа слоев и нейронов позволяет повысить выразительность модели, но также увеличивает риск переобучения. При небольшом объеме данных рекомендуется использовать простую структуру, например два слоя по 512–1024 нейрона. Если данных много, можно экспериментировать с более сложной архитектурой.
- *Регуляризация*. NHITS применяет dropout внутри блоков, чтобы повысить устойчивость к переобучению. Параметр `dropout_prob` определяет вероятность зануления нейронов в процессе

- обучения. Типичные значения варьируются от 0,1 до 0,3. Если модель переобучается, стоит увеличить dropout; если недообучается — уменьшить.
- *Скорость обучения* (learning_rate). Малое значение learning_rate замедляет обучение, но повышает устойчивость к шуму и переобучению. Обычно используется значение от $1e-4$ до $1e-3$. Более высокие значения возможны при наличии большого объема данных и хорошем регуляризационном контроле.
 - *Количество итераций оптимизации* (max_steps). Параметр зависит от сложности задачи и объема данных. При этом важно использовать метод ранней остановки на валидационной выборке, чтобы избежать переобучения модели.
 - *Подбор гиперпараметров NHITS* — это баланс между длиной входного окна, глубиной и сложностью блоков, степенью регуляризации и горизонтом прогноза. Оптимальный выбор зависит от структуры временного ряда, цели анализа и наличия вычислительных ресурсов. Лучше всего гиперпараметры подбирать с помощью кросс-валидации или на отложенной выборке, ориентируясь на метрики ошибки прогноза.

1.4. Глубокая нейронная архитектура с обратными и прямыми остаточными связями (N-BEATS)

N-BEATS — более ранняя модель, которая послужила основой модели N-HITS. Модель также состоит из нескольких блоков, каждый последующий блок получает на вход остатки прогнозирования предыдущего блока. Ее основное отличие от модели N-HITS состоит в том, что в каждом блоке модель совершает прямой прогноз вместо иерархической интерполяции.

Блоки модели N-BEATS объединены в стеки для улавливания определенной компоненты временного ряда: тренда или сезонности. Каждый блок состоит из полносвязных слоев. Выходы полносвязных слоев проецируются в некоторый векторный базис: для выявления тренда используются полиномиальные базисные функции, для выявления сезонных закономерностей — гармони-

ческие базисные функции. Итоговый прогноз модели представляет собой сумму прогнозов всех блоков.

2. ВЛИЯНИЕ ТЕКСТОВЫХ ДАННЫХ НА ЭКОНОМИЧЕСКИЕ ПЕРЕМЕННЫЕ

2.1. Использование моделей с архитектурой трансформера (BERT) для анализа текстовых данных

Одним из наиболее распространенных способов анализа тональности текста является использование моделей-трансформеров. Отличительной чертой моделей BERT (Bidirectional encoder representations from transformers) является наличие слоя внимания, который позволяет улавливать контекстуальные зависимости между словами в тексте, поэтому данный класс моделей может быть использован для анализа влияния текстовых данных на экономические показатели, а также для построения индикаторов (новостных индексов), улучшающих прогнозную силу моделей.

К примеру, в работах Costola M. [4], Fazlija B., Harder P. [5], Kim J., Kim H. S., Choi S. Y. [6] использовали модель BERT для анализа сентимента финансовых новостей с целью прогнозирования цен на акции. В работе Costola было продемонстрировано, что сентимент BERT дает независимый вклад в объяснение колебаний доходности и волатильности, увеличивая объясняющую силу модели ($\text{adjusted } R^2 \sim 0,25$ для доходности S&P 500 против $\sim 0,15-0,2$ без сентимента). При этом было отмечено, что сентимент лучше объясняет волатильность, чем классические переменные, такие как VIX (индекс ожиданий волатильности рынка) и Financial Stress Index (индекс финансового стресса).

В работе Fazlija B., Harder P. [5] прогнозная модель, использующая сентимент-анализ финансовых новостей на основе BERT, превзошла по точности модель случайного блуждания: при применении простой стратегии на основе среднего значения сентимента новостного контента точность достигла 51,8% против 49,9% у случайной модели (авторы прогнозировали направление изме-

нения цены — рост или падение, преобразовав задачу в бинарную классификацию). Наибольшего улучшения удалось достичь при использовании классификатора Random Forest с входными признаками на основе сентимента текста новостей — точность выросла до 56,1%, а годовая доходность составила 16,26% по сравнению с отрицательной доходностью –9,75% у модели случайного блуждания.

В работе Kim J., Kim H. S., Choi S. Y. [6] прогнозная модель, использующая сентимент-анализ на основе BERT, сравнивалась с базовой моделью LSTM, основанной только на рыночных данных без учета новостного сентимента. Результаты показали, что модель с индексом BERT значительно превзошла по точности модель без сентимента: коэффициент детерминации (R^2) вырос с 0,83 до 0,90.

Используя аннотированный корпус твитов, исследователи в работе Li X., Tang Z. [7] проверяли различные подходы к определению сентимента сообщений, связанных с инфляцией. Среди протестированных моделей — логистическая регрессия, наивный байес, SVM, случайный лес, градиентный бустинг, а также модель FinBERT. В среднем традиционные модели показывали точность в диапазоне от 28% до 61% на тестовой выборке, в то время как модель BERT достигла 79,99%. Полученные с помощью BERT оценки сентимента использовались далее для построения временного индекса сентимента и его включения в модели прогнозирования инфляции. Однако авторы не измеряли прирост точности прогноза при добавлении сентимента в модель инфляции, но отметили, что при его включении коэффициенты лагов инфляции становились незначимыми, а индекс сентимента — оказывался статистически значимым.

Для более глубокого анализа текстовой информации используются техники дообучения и фاین-тюнинга моделей BERT. Такой подход делает модель более специализированной для конкретного языка или задачи. Одной из таких является RuBERT — модель, дообученная на датасетах ParaPhraser, RuSentiment и SDSJ Task В исследователями Kuratov Y., Arkhipov M [8], что позволило достичь F1-меры для точности определения тональности коротких русскоязычных текстов в 72,64%, что на 1 п.п. относительно

базовой модели. Для улучшения результата решения определенной задачи на специфическом датасете используется тонкая настройка (fine-tuning), которая в отличие от дообучения предполагает изменение не всех параметров модели, а только некоторой части — весов внимания. Smetanin, Komarov [9] проанализировали, как тонкая настройка BERT и ruBERT влияет на качество классификации сентимента. На датасете Kaggle Russian News Dataset, состоящим из 8200 русскоязычных новостей, размеченных на позитивный, нейтральный и негативный класс, при тонкой настройке ruBERT было получено улучшение метрики F1-score с 70,0% до 73,63%, на мультиязыковом BERT F1-score был повышен 66,29% до 71,36%.

2.1.1. Реконструкция биологических путей путем применения выборочного поэтапного обучения к (очень) маленьким языковым моделям

В работе Saha P., Reimer J. и соавторов [10] проводилось сравнение большой языковой модели LLaMA и классической модели BERT. Обе модели имеют архитектуру трансформера, LLaMA использует только слой декодера и имеет 1 млрд параметров, модель BERT имеет 110 млн параметров.

Цель работы состояла в установлении еще не задокументированных молекулярных взаимодействий, которые позволяют выявить реакции иммунных клеток на бактерии туберкулеза.

Датасет был собран авторами самостоятельно. Из рецензируемых статей по теме были извлечены названия белков двух видов, вид белка определялся путем решения задачи NER с применением языковых моделей, предобученных в биомедицинской области. Полученный датасет взаимосвязей между белками был проверен с помощью базы данных Pathway Common и корпуса BEL Selventa Large Corpus. Эталонный набор взаимосвязей затем преобразовывался в INDRA-утверждения и после удаления противоречивых и избыточных утверждений был получен расширенный датасет взаимосвязей.

В полученном датасете объясняемыми переменными выступили "источник", "цель", целевой переменной — "тип взаимо-

действия". "Тип взаимодействия" являлся бинарной переменной (положительное и отрицательное взаимодействие).

Для LLaMA запрос был сформулирован как «Белок "источник" регулирует белок "цель" положительно или отрицательно?». Авторы использовали модель BERT как Masked Language Model (MLM), поэтому для первоначального запуска ей необходимо подать набор примеров, чтобы модель понимала требуемый формат входных и выходных данных. В связи с этим запрос для BERT формулировался как «Белок "источник" регулирует белок "цель" положительно или отрицательно? Он регулирует "тип взаимодействия"», то есть выходом модели будет одно слово — "тип взаимодействия".

Полученный датасет был разбит в соотношении 80% для обучающей выборки (в нее вошло по 93 примера каждого класса) и 20% для валидационной выборки (в нее вошли 23 случайно выбранные пары примеров). Авторами был проведен единственный раунд файн-тюнинга (в статье этот этап назван сравнением в рамках zero-shot подхода; использование термина спорно, так как подход zero-shot не подразумевает применение тонкой настройки или других методов дообучения) обеих моделей на всей обучающей выборке. LLaMA показала сильное смещение к одному классу при общей точности 50% и высокой степени уверенности в предсказаниях (энтропия 0,43). BERT показала более высокую точность равную 54%, но с значительно большей неопределенностью (энтропия 0,95).

Для получения более качественного результата к модели BERT применялось поэтапное обучение, всего было проведено два эксперимента, которые различались в зависимости от того, по какому принципу в обучающую выборку добавлялись примеры на новой итерации. Такой процесс поэтапного обучения применялся только к модели BERT, сравнения с моделью LLaMA при прочих равных не проводились.

В первом эксперименте процесс файн-тюнинга описывался следующим образом: на каждой итерации из набора данных случайным образом выбиралось 5 наборов по 10 запросов (по 5 на каждый класс), которые затем по 10 раз оценивались моделью. На основе этих предсказаний было отобрано одно вы-

сокоэнтропийное, но верное предсказание (отражающее неопределенность модели), и одно низкоэнтропийное, но ошибочное предсказание (отражающее избыточную уверенность модели), для каждого класса, что в сумме давало до 4 новых образцов за итерацию. Пример с высокой энтропией и правильным ответом модели позволил уточнить границы классификации, усиливая «положительное подкрепление». Пример с низкой энтропией и ошибочным ответом модели позволил скорректировать переобучение, добавляя «отрицательное подкрепление». Отобранные образцы добавлялись в обучающий набор на следующей итерации и помечались как использованные, чтобы избежать повторов.

Процесс фэйн-тюнинга повторялся в 25 циклах, после каждого была посчитана информационная энтропия, при этом качество полученной модели варьировалось в зависимости от первоначального набора примеров для тонкой настройки. Таким образом, общее качество модели составило от 69% до 83% в зависимости от случайного выбора начального набора примеров, при этом в процессе тонкой настройки было задействовано около 21% имеющейся обучающей выборки.

Во втором эксперименте число примеров для фэйн-тюнинга было увеличено до четырех: теперь выбиралось по два неправильных примера каждого класса с наименьшей и второй наименьшей энтропией. Таким образом, точность модели повысилась до 90% при использовании 23% имеющейся обучающей выборки.

Авторами были проанализированы молекулярные взаимосвязи, используемые в качестве примеров для тонкой настройки. Общая прогностическая точность была значительно выше для стратегии, использующей только ошибочные предсказания с низкой энтропией (93,0% против 75,7%, $p=0,03$). При этом точность отнесения к положительному классу оказалась схожей для обеих стратегий (89,3% против 82,7%, $p=0,424$), разница в отнесении к отрицательному классу оказалась более заметной (96,7% против 68,7%, $p=0,112$), хотя и не достигла статистической значимости. Разница между количеством примеров в обучающих наборах оказалась статистически незначимой (в первом эксперименте — 93 примера (116 уникальных молекулярных видов), во втором — 84 (111 уникальных молекулярных видов)).

Второй подход с отбором неправильно классифицированных примеров с наименьшей и второй наименьшей энтропией авторы выделяют как более предпочтительный, так как, помимо более высокой точности, модели удастся лучше обучаться на критически важных взаимодействиях.

Авторами показано, что в специфических задачах можно достичь высокого качества без использования тяжелых моделей, требующих большого количества ресурсов и обучающих данных.

Хотя в выводах статьи утверждается преимущество модели BERT, сравнения моделей LLaMA и BERT с использованием однотипного подхода дообучения few-shot не проводилось, поэтому корректным выводом в данном случае является: дообученная модель BERT на специфических данных может демонстрировать результаты лучше, чем базовая LLaMA.

2.1.2. Интерпретируемое машинное обучение для Macro Alpha: Тематическое исследование новостных настроений

В статье Zhang Y. [11] описывается процесс построения индекса настроения на основе ежедневных новостей, находящихся в открытом доступе, который далее используется в качестве экзогенной переменной в прогнозных моделях (логистической регрессии и XGBoost). Устойчивость полученных результатов была подтверждена на кросс-валидации, значимый вклад признаков, отражающих новостные настроения, был подтвержден после проведения процедуры SHAP³ (SHapley Additive exPlanations). Исследование проводилось на данных спотовых валютных курсов EUR/USD и USD/JPY, а также фьючерсов на 10-летние казначейские облигации США (ZN), имеющих суточную частоту наблюдений.

Автором были собраны новости за период с 1 января 2015 г. по 30 апреля 2025 г. из базы данных GDELT⁴ (Global Database

³ Метод интерпретации машинного обучения, основанный на концепции значений Шепли (метод из кооперативной теории игр, распределяющий выигрыш между всеми участниками игры; в случае SHAP — между признаками модели).

⁴ Крупнейшая база данных, хранящая информацию о социальных, политических и экономических событиях. Данные собираются из репортажей по всему миру, и с помощью анализа настроений создается каталог человеческого поведения и убеждений.

of Events, Language, and Tone). Каждая новость содержала в себе такие метаданные, как дата, действующие лица, код события, шкала Гольдштейна (показатель влияния события) и URL-адрес источника. Затем по коду события GDELT были отобраны экономические и политические события (код GDELT в диапазоне от 100 до 199 обычно относится к консультациям, заявлениям и дипломатическим или экономическим мероприятиям). Внутри каждого дня отобранные по кодам события ранжировались по частоте упоминания в текстах новостей, затем были выбраны до 100 наиболее популярных заголовков новостей, имеющих метку Гольдштейна.

С помощью модели FinBERT (модель BERT, тонко настроенная на финансовые новости) для каждого заголовка была получена вероятность принадлежности к позитивному, нейтральному и негативному классу, затем была посчитана непрерывная оценка тональности $s_{i,t}$:

$$s_{i,t} = P_{pos} - P_{neg} \in [-1, +1], \quad (8)$$

где P_{pos} — вероятность отнесения заголовка к позитивному классу; P_{neg} — вероятность отнесения заголовка к негативному классу.

Соответственно, значения, близкие к -1 , свидетельствовали о негативном настроении новости, близкие к 1 — о позитивном настроении, близкие к 0 — о нейтральном или неопределенном настроении.

Полученные оценки тональности новостей были агрегированы в ежедневные признаки:

Среднее настроение:

$$S_t = \frac{1}{N_t} \sum_{i=1}^{N_t} s_{i,t}, \quad (9)$$

Дисперсия настроения:

$$\sigma_t = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} (s_{i,t} - S_t)^2}, \quad (10)$$

Количество новостей:

$$V_t = N_t, \quad (11)$$

Логарифмированное количество новостей:

$$\log 1 + N_t, \quad (12)$$

Влияние новости:

$$AI_t = S_t \times \log 1 + N_t, \quad (13)$$

Средняя оценка Гольдштейна:

$$\overline{G}_t = \frac{1}{N_t} \sum_{i=1}^{N_t} g_{i,t}, \quad (14)$$

Дисперсия оценки Гольдштейна:

$$\sigma_t^G = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} (g_{i,t} - \overline{G}_t)^2}, \quad (15)$$

где

N_t — число новостей (заголовков) за день;

$g_{i,t}$ — оценка Гольдштейна для i -го заголовка.

Для описания динамических состояний были рассчитаны дополнительные признаки:

Лаговые признаки для S_t , σ , V_t , $\overline{G}_t$;

5- и 20-дневные скользящие средние:

$$MA_5(S_t) = \frac{1}{5} \sum_{k=0}^4 S_{t-k}, \quad (16)$$

Динамика настроения:

$$\Delta S_t = MA_5(S_t) - MA_{20}(S_t), \quad (17)$$

А также скользящее стандартное отклонение и скользящие суммы.

Полученный набор экзогенных переменных характеризовал как текущее состояние, так и динамику новостных настроений и использовался в качестве входных данных для прогнозных моделей.

Данные спотовых курсов валютных пар EUR/USD, USD/JPY и непрерывного ряда фьючерсов на 10-летние казначейские об-

лигации (ZN) взяты из Yahoo Finance. Логарифмическая доходность на следующий день определялась как:

$$r_{t+1} = \log P_{t+1} - \log P_t, \quad (18)$$

где

P_t — цена закрытия дня t .

Целевая переменная являлась бинарной: кодировалась 1, если доходность следующего дня была положительной (происходило движение вверх), и кодировалась 0, если доходность следующего дня была отрицательной (происходило движение вниз). В прогнозных моделях использовались экзогенные переменные, рассчитанные с использованием информации, доступной до момента закрытия рынка дня t . Эти признаки были использованы для прогнозирования направления рынка от закрытия дня t до закрытия дня $t + 1$ для избежания проблемы утечки данных. В качестве дополнительных рыночных признаков были использованы лаговая доходность (доходность от закрытия дня $t - 1$ до закрытия дня t) и историческая волатильность (20-дневное годовое стандартное отклонение дневных логарифмических доходностей). Для логистической регрессии все непрерывные признаки были стандартизированы для численной стабильности модели. Для XGBoost стандартизация признаков не проводилась, так как, по утверждению автора, модель инвариантна к монотонным преобразованиям отдельных признаков.

Для всех целевых переменных были построены:

- логистическая регрессия с регуляризацией L2 (сила регуляризации подбиралась с помощью кросс-валидации по временным рядам на каждом тренировочном наборе)

- ансамбль деревьев с градиентным бустингом (XGBoost), гиперпараметры подбирались методом grid search в сочетании с 5-кратной кросс-валидацией по временным рядам на тренировочных данных.

Качество моделей измерялось на 5 окнах кросс-валидации с расширяющимся окном, на первом этапе тренировочная выборка включала в себя данные за первые два года. После получения предсказаний торговых сигналов (меток 0 и 1) на тесто-

вой выборке была посчитана ежедневная доходность стратегии. К тому же из полученных доходностей вычитались транзакционные издержки: 0,02% для валютных пар FX, 0,05% для фьючерсов на казначейские облигации (ZN). Метриками качества выступили AUC (площадь под ROC-кривой), точность, годовой коэффициент Шарпа и CAGR (годовая совокупная доходность).

Модели XGBoost продемонстрировали высокую и стабильную эффективность по всем целевым активам, значительно превосходя базовую логистическую регрессию. Коэффициенты Шарпа, превышающие 4,6, свидетельствуют о высокой способности генерировать альфа-результат, особенно примечательный с учетом включения реальных транзакционных издержек и строгой методологии OOS-оценки. Показатели CAGR для моделей XGBoost также существенны, что указывает на значительный потенциал прироста капитала.

В то же время модели логистической регрессии демонстрируют слабую способность к устойчивой генерации альфы, особенно по казначейским облигациям, что подчеркивает важность учета нелинейных взаимосвязей и сложных взаимодействий признаков, которые XGBoost успешно выявляет.

Для исследования важности признаков в сложной нелинейной модели XGBoost применяется метод SHAP. SHAP предоставляет единую структуру для интерпретации модели, присваивая каждому признаку «значение важности» для каждого отдельного прогноза, основанного на принципах теории кооперативных игр. Эти значения представляют собой предельный вклад каждого признака в прогноз по сравнению с исходным уровнем (среднее значение прогноза по данным обучения).

Из анализа диаграммы SHAP следует, что решения модели XGBoost в основном определяются сигналами, выведенными из новостного индекса. Наибольший вклад в прогноз внесли дисперсия настроения, вклад новости, среднее настроение, лаги настроения и скользящие средние, величина Гольдштейна, волатильность рынка.

В работе показан единый подход построения прогнозных моделей для нескольких классов активов. Использование нелинейной прогностической модели XGBoost превосходит простую линейную регрессию по всем перечисленным метрикам, а использование

новостных признаков вносит значительный вклад в улучшение результатов работы модели. Использование открытой базы данных и моделей с открытым исходным кодом делает полученные результаты прозрачными и исследование воспроизводимым.

2.1.3. Передовые методы глубокого обучения для анализа расшифровок телефонных разговоров о доходах: методологии и приложения

В статье Zakir U., Daykin E. и соавторов [12] приводится сравнение моделей, основанных на архитектуре трансформера BERT, в задаче анализа настроений в расшифровках звонков о доходах. Авторы действуют из предположения о том, что любая общедоступная информация может влиять на цену активов на открытом рынке, поэтому имеет смысл воспользоваться неэффективностью рынка, выраженной в информации, которая хранится в расшифровках звонков, для получения конкурентного преимущества перед другими инвесторами на рынке и увеличения чистой прибыли.

В качестве расшифровок с источников Bloomberg и Refinitiv были взяты отчеты о доходах за последние четыре квартала для компаний, входящих в индекс S&P 500.

Далее для каждой даты из расшифровки были получены ближайшие котировки акций и значения индекса S&P 500 за 90 дней до отчета, 2 дня до отчета, 2 дня после отчета и 90 дней после отчета. Собранные тексты были предобработаны, пропущенные данные были удалены.

Полученные тексты авторы размечали на классы в соответствии с правилом:

Для каждой новости была посчитана динамика цены:

$$price_{movement} = \frac{sp_{t+2} - sp_{t-2}}{sp_{t-2}} \times 100, \quad (19)$$

где

sp_{t+2} — цена акции на момент закрытия через 2 дня после публикации отчета;

sp_{t-2} — цена акции на момент закрытия за 2 дня до публикации отчета.

Если полученная динамика цены оказывалась меньше -3 , то новость считалась отрицательной, соответственно, если динамика цены превышала значение 3 , то новость считалась положительной (При установлении таких порогов для негативного и позитивного класса было получено хорошее распределение классов. — *Прим. авторов*). Остальным новостям была присвоена нейтральная метка класса. В рамках предложенного авторами подхода не учитывалась семантическая окраска отчета, разметка устанавливалась только на основе динамики цены, поэтому установленные классы могли не всегда справедливо отражать действительность. Датасет был разбит на тренировочную, валидационную и тестовую выборку, для всех моделей на тренировочной выборке подбирались оптимальные параметры.

Точность базовой модели FinBERT составила 18,58%, F1-score составил 12,22%, при этом модель не могла уловить негативный оттенок текстов и все тексты классифицировала как положительные или негативные. После тонкой настройки точность возросла до 52,21%, F1-score также возрос до 49%, проблема с распознаванием негативных текстов была устранена.

Поверх базовой модели BERT авторами был создан классификатор настроений. Полученная модель хотя и показала приемлемое распределение классов, но корреляция настроений с движением цен оказалась невысокой. Точность модели составила 41%, F1-score составил 42%.

Структура модели ULMFit состоит из архитектуры SequentialRNN, в частности рекуррентной нейронной сети на основе LSTM. Структура модели включает в себя несколько уровней LSTM, за которыми следуют уровни отсева и пакетная нормализация. Выходной слой состоит из линейного слоя с функцией активации ReLU, за которым следует еще один слой для классификации. Общее количество обучаемых параметров при использовании всех текстовых данных для каждой модели составляет 62 650. На выход модель подает вероятности классов, полученные на уровне softmax. ULMFit была предварительно обучена на большом массиве общих текстовых данных по типу Wikipedia и Yelp. Авторами модель была дообучена анализу настроений на финансовых отчетах. Точность модели составила 37%, F1-score — 39%.

Перечисленные модели имели общий недостаток — контекст был ограничен 512 токенами, в то время как максимальная длина контекста в полученном датасете могла достигать 16 000 токенов. Для решения проблемы авторы использовали модель Longformer — модель-трансформер для обработки длинных последовательностей. Длина контекста модели составляет 4096 токенов. Итоговая точность модели составила 45%, F1-score — 41%. Несмотря на увеличение длины входной последовательности модель не показала глобального улучшения результатов, при этом затраты на время и вычислительные ресурсы значительно возросли.

Невозможность достижения приемлемых результатов классификации авторы связывают с наличием большого количества приукрашенной риторики в текстах, из-за которой модели не могут распознать негативный оттенок. Таким образом, у модели с архитектурой BERT лучше результаты в данной задаче, а модель FinBERT, учитывающая знания в специфической области, смогла достичь лучшего качества классификации телефонных звонков о доходах.

2.1.4. Сентимент трейдинг с использованием больших языковых моделей

В первой части исследования Kirtac K., Germano G. [13] текстовые данные преобразовываются в численный индекс настроения одним из четырех способов, во второй — полученный индекс используется в эконометрической модели для прогнозирования доходности акций.

Исследование проводилось для компаний, состоящих в AMEX, NASDAQ и NYSE. Из базы данных CRSP⁵ (Center for Research in Security Prices) собраны ежедневные данные о доходах компаний (цены на акции, объемы торгов, рыночную капитализацию), торгующих на основных фондовых биржах США. Тексты новостей взяты из Refinitiv. Из всего пула статей были отобраны новости, относящиеся к отдельным акциям с доступной доходностью в течение трех дней. Также для избежания избыточности новостей

⁵ Центр исследований, располагающий полной коллекцией данных о ценах на ценные бумаги, доходности и об объемах торгов на фондовых рынках NYSE, AMEX и NASDAQ.

была использована оценка новизны, основанная на сходстве между статьями: если новая статья слишком похожа на более старую, опубликованную в течение последних 20 дней, согласно показателю косинусного сходства, равному 0,8 или более, то она исключалась. После применения указанной фильтрации из 2 732 845 новостей о 6214 уникальных компаниях было отобрано 965 375 статей за период с 1 января 2010 г. по 30 июня 2023 г.

Для каждой новости была посчитана трехдневная избыточная доходность. Соответственно, новости с положительной избыточной доходностью помечались как имеющие положительную оценку и новости с неположительной доходностью — как имеющие негативную оценку. По мнению авторов, этот метод полезен для понимания взаимосвязи между настроениями в финансовых новостях и соответствующими изменениями цен на акции в последующие два дня, хотя опять же не учитывает семантику новости и опирается исключительно на величину избыточной доходности, на которую могли повлиять другие факторы. Из полученного датасета было случайно выбрано по 20% новостей для тестовой и валидационной выборки, тренировочный набор включил в себя около 193 000 новостей.

Модели OPT⁶ (Open Pre-trained Transformers) и BERT были тонко настроены на имеющихся данных, FinBERT и словарь Лоуграна—Макдональда⁷ в дополнительной настройке не нуждались, так как изначально специализируются на обработке текстов финансовой тематики. На тестовой выборке наиболее точной моделью оказалась OPT (точность составила 74,4%), затем BERT (72,5%) и FinBERT (72,2%). Модель на словарях показала худший результат — точность составила 50,1%. Эта точность показывает, насколько хорошо модель связывает настроения в финансовых новостях с доходностью акций за трехдневный период.

⁶ Набор моделей, основанных моделях класса GPT. В отличие от GPT они находятся в открытом доступе и обладают значительно меньшим размером (в статье используется OPT 2.7B вместо GPT-3 175B).

⁷ Англоязычный словарь для анализа финансовых текстов. Слова в словаре помечены одним из шести классов настроения.

Полученные с помощью различных моделей метки настроений далее использовались в панельной регрессионной линейной модели, где целевой переменной являлась доходность акции компании на следующий день, а в качестве экзогенных переменных выступили вектор значений настроений, фиксированные индивидуальные эффекты и фиксированные эффекты времени. Линейная модель выбрана для явного выявления влияния текстовых представлений в финансовом анализе.

Модель с оценками ОРТ продемонстрировала сильную корреляцию с доходностью акций на следующий день, о чем свидетельствовали значимые коэффициенты в различных спецификациях модели. Модели с оценками FinBERT и BERT оказались несколько хуже, хотя полученные оценки также демонстрируют значимую связь с доходностью акций. К тому же прогностическая сила моделей возрастала, когда два показателя, построенные языковыми моделями, используются в качестве экзогенных переменных в одной и той же регрессии. Словарная модель Лоуграна—Макдональда обладала наименьшей прогностической способностью среди рассмотренных моделей.

Для каждой языковой модели было создано три типа портфелей: длинные, короткие позиции и «длинные-короткие» позиции. Состав этих портфелей зависел от оценки настроений, которая присваивалась отдельным акциям каждый день. Длинные портфели включали в себя акции с самыми высокими 20% оценками настроений, короткие портфели — с самыми низкими 20% оценками настроений. Длинные-короткие портфели — это стратегии самофинансирования, которые одновременно предполагают открытие длинных позиций по акциям с самыми высокими 20% оценками настроений и коротких позиций по акциям с самыми низкими 20% оценками настроений. Транзакционные издержки также учитываются.

Оценка стратегий проводится с использованием ключевых финансовых показателей: коэффициента Шарпа, средней дневной доходности, стандартного отклонения дневной доходности и максимальной просадки.

Наиболее высокую эффективность с поправкой на риск показала длинная-короткая стратегия с использованием оценок ОРТ,

коэффициент Шарпа этой стратегии составил 3,05, в то время как аналогичная стратегия с использованием оценок на словарях имела коэффициент Шарпа равный 1,23 (такая стратегия оказывается хуже рыночного портфеля, взвешенного по стоимости). Стратегия long-short OPT показала впечатляющий выигрыш в 355%. Это подчеркивает мощные прогностические возможности передовых языковых моделей в прогнозировании динамики рынка. Стратегии long-short BERT и long-short FinBERT, также демонстрируют значительную прибыль в размере 235% и 165%, что резко контрастирует с традиционными рыночными портфелями, которые с трудом превышают 1%. В то же время словарная модель Лоуграна—Макдональда, широко используемая в финансовых исследованиях, показала доходность всего в 0,91%. Это явное несоответствие свидетельствует о том, что модели, основанные на словарях, не позволяют эффективно интерпретировать нюансы современных финансовых новостей так же эффективно, как более продвинутые языковые модели.

В результате проведенного исследования была показана важность использования сложных языковых моделей при разработке инвестиционных стратегий, основанных на новостных настроениях. Традиционные методы, основанные на словарях не позволяют в достаточной мере проанализировать тексты финансовых новостей и вывести индикаторы, несущие в себе значимую информацию о динамике показателей.

2.1.5. FinAI-BERT: Модель на основе архитектуры трансформера для поиска упоминаний ИИ в финансовых отчетах на уровне предложений

В работе Zafar M.B. [14] была произведена разработка модели FinAI-BERT (доработка существующей модели BERT), предназначенной для детекции финансовых текстов, в которых упоминается использование технологий искусственного интеллекта. Ранее подобная задача решалась с помощью специализированных словарей или подхода TF-IDF, однако оба метода не способны учитывать контекстные особенности финансовых текстов. Цель работы заключалась в том, чтобы отличить тексты со значимой

информацией (например, внедрение технологий искусственного интеллекта) от текстов с расплывчатыми упоминаниями ИИ. По мнению авторов, раскрытие существенной информации об ИИ коррелировало с увеличением расходов на НИОКР и подачей патентных заявок и таким образом могло помочь в оценке стоимости компании в долгосрочной перспективе.

В основу модели FinAI-BERT была взята базовая модель BERT, которая дообучалась на размеченном вручную датасете. В датасет вошли тексты 669 годовых отчетов 85 американских компаний за период с 2015 по 2023 г. Корпус был поделен на предложения, и после предобработки каждое предложение помечалось как связанное или не связанное с искусственным интеллектом. Первоначально разметка проводилась согласно тематическому словарю, в который входили выражения „машинное обучение“, „глубокое обучение“, „генеративный ИИ“ и прочие, затем авторами вручную проверялось соответствие полученной разметки смыслу предложения. Полученный датасет был сбалансирован по классам, далее был поделен на обучающую и тестовую выборку (80% и 20% соответственно).

Базовая модель BERT была тонко настроена на всей обучающей выборке и достигла точности классификации — 99,37% и значения 0,993 метрики F1-score. Также для прогнозов модели была посчитана метрика Brier Score, измеряющая точность вероятностных прогнозов. Ее значение оказалось 0,00, что говорит о полном соответствии вероятностей принадлежности к классам и фактических классов.

Качество прогнозов FinAI-BERT было сравнено с основными моделями, используемыми для задач классификации, среди них: логистическая регрессия, Naive Bayes, Random Forest, XGBoost. Дообученная модель превосходит классические модели по метрикам Accuracy, Precision (для ИИ), Recall (для ИИ), F1-Score (для ИИ).

Несмотря на то что поставленная задача не являлась сложной и для более простых методов, предложенный авторами способ помог автоматизировать процесс и решить некоторые проблемы предшествующих методов (например, неполноту существующего словаря и необходимость его постоянного пополнения). К тому же тонкая настройка моделей, основанных на архитектуре

трансформеров, не является прорывным подходом, однако полученный результат очередной раз подтвердил способность таких моделей приспосабливаться к специфическим задачам, которые требуют понимания профессиональной лексики.

2.1.6. Temporal Fusion Transformers — модель для интерпретируемого многогоризонтного прогнозирования временных рядов

TFT — это глубокая нейросетевая модель для прогнозирования временных рядов. Для анализа временных зависимостей разного масштаба Lim B., Arik S. O. и соавторы [15] предлагают использовать рекуррентные слои для обработки временных рядов и интерпретируемые слои самовнимания для долгосрочных зависимостей.

Модель состояла из GRN (Gated Residual Network) блоков, содержащих в себе функцию активации ELU (Exponential Linear Unit), с помощью которой регулировалось добавление нелинейности к данным, и механизм гейтинга (Gated Linear Unit), который подавлял наименее значимую информацию. Такое устройство блока, по мнению авторов, позволяло регулировать сложность модели в зависимости от входных данных, сохраняя при этом способность адаптации модели к сложным закономерностям.

Первый этап работы модели состоял в отборе наиболее релевантных признаков, этот процесс реализовался в блоках Variable Selection, которые также состояли из блоков GRN. В модель могли быть поданы статические признаки, исторические признаки и известные будущие признаки. Категориальные признаки преобразовывались в эмбединги с помощью Entity Embeddings, непрерывные признаки обрабатывались через линейные комбинации. Обработанные признаки подавались на вход GRN слоев (каждый признак обрабатывался отдельным блоком GRN, в котором каждой временной точке присваивался вес важности), затем между признаками, в зависимости от их релевантности, распределялись веса, таким образом происходил отбор наиболее важной информации, а данные, вносящие неопределенность и шум в работу модели, подавлялись.

В блоках GRN при обработке динамических признаков статические признаки могли использоваться как контекстные (вспо-

могательные). Для этого сначала статические признаки отдельно обрабатывались GRN и преобразовывались в несколько контекстных векторов, которые были использованы в:

- блоках Variable Selection при отборе динамических признаков;
- LSTM-энкодере модели для улавливания временных зависимостей;
- блоках GRN для дополнения динамических признаков.

Чтобы уловить последовательные зависимости во временных рядах, авторы использовали слои LSTM энкодера-декодера, причем начальные скрытые состояния энкодера и ячейки инициализировались контекстными (статическими) векторами. Благодаря такой системе LSTM могла сразу учитывать особенности входных временных рядов и более точно улавливать закономерности разного масштаба и специфики.

Одной из ключевых особенностей модели является наличие интерпретируемого слоя внимания. Интерпретируемость достигается за счет использования единой матрицы значений (V). Результаты всех голов внимания суммируются, поэтому вклад каждой в итоговый прогноз становится более явным. Благодаря предложенному механизму можно оценить вклад каждой временной точки в итоговый прогноз и выделить те, которые имеют наибольшее влияние при прогнозировании.

Модель обучалась с использованием квантильной регрессии для получения в качестве предсказаний квантилей предсказаний целевой переменной, однако модель легко адаптируется для построения детерминированного прогноза.

Качество модели было проверено на датасетах по расходу электроэнергии (UCI Electricity Load Diagrams Dataset), загруженности дорог (UCI PEM-SF Traffic Dataset), ритейла (Favorita Grocery Sales Dataset), волатильности ценных бумаг (библиотека OMI). Прогностическая способность нескольких временных точек была сравнена с другими популярными моделями для одновременного прогноза (Seq2Seq, MQRNN) и с моделями, которые строят прогноз итеративно (DeepAR, DSSM, ConvTrans). Также было произведено сравнение с моделями ARIMA, ETS, TRMF, DeepAR, DSSM and ConvTrans без использования экзо-

генных переменных. На всех датасетах TFT показала лучшее качество среди ранее использованных моделей. По сравнению с TFT на датасете по расходу электроэнергии значение функции потерь у ARIMA больше на 180% в случае 0,9 квантильного прогноза и больше на 278% в случае медианного прогноза (0,5 квантиль); на датасете по загруженности дорог значение функции потерь ARIMA превышает TFT на 135% для 0,9 квантиля и на 94% для 0,5 квантиля.

После проведения процедуры удаления каждого компонента модели для исследования их вклада в работу авторы пришли к выводу, что наибольшее влияние на производительность оказывают компоненты энкодера-декодера для обработки временных закономерностей и модифицированные слои внимания. Удаление этих компонентов приводит в среднем к увеличению потерь на 6%, на некоторых датасетах потери могут возрасти на 20%. Вторыми по важности для работы модели оказались использование энкодера статистических признаков, порождающего контекстные векторы (его отсутствие увеличивает потери на 2,6%), и механизм отбора релевантных признаков (его отсутствие увеличивает потери на 4,1% в среднем). Удаление механизмов гейтинга приводит к увеличению потерь на 1,9%.

По задумке авторов используемые функции активации предоставили модели возможность одинаково хорошо анализировать как простые, так и сложные нелинейные зависимости в данных. Архитектура модели позволила учесть статические и динамические признаки разных временных периодов, извлечь из них наиболее важные компоненты и на их основе получить распределение вероятных исходов или детерминированное предсказание. Модифицированный слой внимания внес наибольший вклад в прогностическую силу модели, а также сделал модель интерпретируемой, подсветив важнейшие временные точки. Проведенные на датасетах эксперименты подтвердили способность модели качественно обрабатывать временные последовательности различной специфики. К тому же стоит учитывать, что сложная архитектура модели требует обучения большого числа параметров, поэтому проблема переобучения на малых средних выборках остается актуальной.

2.1.7. Анализ финансовых настроений для предварительно подготовленных языковых моделей, включающих знание словаря и нейтральные признаки

Исследование Sun Y., Yuan H., Xu F. [16] посвящено разработке механизма для совершенствования повышения точности обработки финансовых текстов. Авторами была предложена модель EnhancedFinSentiBERT, которая способна глубоко анализировать финансовые тексты благодаря трем механизмам: предварительному обучению в финансовой области, интеграции словарных знаний и извлечения нейтральных признаков.

Этап предварительного обучения модели на финансовых текстах авторы мотивировали необходимостью лучшего понимания профессиональной лексики и выражений. Этап включения словарных знаний сочетал в себе как неявные представления, полученные из слоев сети, так и явные знания из специализированных словарей. Дополнительная информация из словарей об интенсивности тональности финансовых терминов служила внешним знанием, дополняющим скрытые представления моделей. Способность модели хорошо распознавать нейтральный оттенок в финансовых текстах была необходима для снижения уровня неопределенности модели при классификации позитивного и нейтрального класса. Выходы трех этапов агрегировались и подавались на классификационный слой.

Для предобучения модели, как и в случае оригинального BERT, авторами использовались задачи MLM (Masked Language Modeling) и NSP (Next Sentence Prediction). Предварительное обучение проводилось на корпусе финансовых новостей и аналитических отчетов за период с 2010 г. по сентябрь 2024 г., объемом около 1 Гб текста и более 9 млн предложений, тексты были собраны авторами самостоятельно. Модель прошла процесс быстрого обучения, после анализа графиков динамики функции потерь при обучении авторы сделали вывод о достижении сходимости обучения и, следовательно, о хорошей адаптации полученной модели к финансовой специфике текста.

Модель прошла процесс быстрого обучения, после анализа графиков динамики функции потерь при обучении авторы сделали

вывод о достижении сходимости обучения, что говорит о хорошей адаптации модели к финансовой специфике текста.

Для интеграции словарных знаний в модель был использован «ML-лексикон»⁸, который в отличие от словаря Лоуграна—Макдональда способен улавливать контекстуальные зависимости. Информация из лексикона о частоте слов была преобразована через логарифмическую шкалу для расчета весов слов и нормализована для балансировки влияния слов с высокой и низкой частотой. Вероятности положительной, отрицательной и нейтральной тональности из исходного лексикона далее использовались для расчета полярности и интенсивности тональности.

Для каждого слова входной последовательности выбирался вектор признаков из предобработанного словаря, после чего попадал в модуль обработки словарных эмбедингов. Внутри блока вектор преобразовывался в BERT-эмбединг, обрабатывался с помощью MLP и проходил через слой многоголового внимания. Одновременно с этим с помощью предобученной на финансовых текстах модели обрабатывалась начальная входная последовательность. Выходные данные соединялись с контекстным словарным представлением с помощью линейного слоя. Блок для обработки нейтральных текстов состоял из 2 линейных слоев, 2 функций активации ReLU, механизма многоголового внимания и слоя усредняющего пулинга. Перед подачей данных в этот блок данные преобразовывались в BERT-эмбединг. Механизм внимания позволял захватить нейтральные признаки на нескольких уровнях, которые далее усреднялись и превращались в единый вектор нейтральных представлений. Такая архитектура улучшала способность модели отличать нейтральное настроение текста от позитивного и негативного.

Для обучения блока были собраны тексты, состоящие из объективных утверждений, фактических отчетов и нейтральных выражений в финансовых текстах из датасета SEntFiN⁹. Также из источ-

⁸ Предложен Garcia et al. (2023). Применяя robust multinomial inverse regression (robust MNIR) модель к 85 530 расшифровкам отчетов о доходах, Гарсия и др. связали появление n-грамм с одновременными реакциями цен на акции, установив таким образом эмоциональную полярность слов.

⁹ Датасет с заголовками финансовых новостей, в котором указан положительный, нейтральный или негативный сентимент финансовых организаций, упомянутых в заголовках.

ников Reuters, CNBC, Investing.com и Seeking Alpha было отобрано большое количество нейтрально-объективных статей, из которых дополнительно извлекались нейтральные предложения.

Прогнозная сила модели была проверена на 3 датасетах. На датасете англоязычных финансовых новостей (Financial Phrase-Bank) с 50% согласием экспертов модель показала F1-Score 87,0%, на данных с 100% согласием экспертов значение метрики оказалось выше и достигло 98,0%. Предложенная авторами модель превзошла все базовые модели, среди них: BERT-base, XLNet, GPT-4, Llama 2, FinBERT, BloombergGPT. На наборе данных FiQA Task 1 (неформальные финансовые тексты из социальных сетей) EnhancedFinSentiBERT опережает все базовые модели с F1-score 88,0%. В задаче классификации заголовков финансовых новостей (датасет Headline) модель также показала лучший результат с F1-score 97,6%. Llama 2 достигла F1-меры 94,2%, XLNet — 91,3%, FinBERT — 90,8%, BERT-base — 89,5%, GPT-4 — 86,0%, BloombergGPT — 82,2%.

Предложенная авторами архитектура модели учитывает и корректирует недостатки предшествующих исследований в области анализа настроений финансовых текстов. Согласно выводам исследователей, благодаря ансамблевому подходу к построению прогноза, а также взаимодействию словарных знаний с контекстными представлениями BERT через механизм многоголового внимания, модель показала прогнозную силу в классификации финансовых текстов, которая превзошла наиболее популярные модели.

2.2. Использование больших языковых моделей (LLM) для анализа текстовых данных

Следующим этапом развития подходов для определения сентимента стало появление больших языковых моделей (LLM). Эти модели обучаются на больших объемах текстовых данных, что позволяет им улавливать сложные семантические и синтаксические зависимости. Одним из значительных преимуществ LLM является гибкость при правильном подходе к промпт инженерии (prompt engineering). Этот процесс включает формулировку вопросов и инструкций, которые направляют модель на реше-

ние конкретной задачи. Благодаря своей архитектуре LLM могут быстро адаптироваться к изменениям в промте и решать широкий спектр задач, без необходимости повторного обучения модели, что полезно в случае построения новостных индексов, так как позволяет менять только промпт, не требуя отдельной модели для каждого макроэкономического индекса или новостного индекса, при этом обеспечивая высокую точность в разнообразных контекстах.

В исследовании Бондаренко, Елисеева, Ермоленко [17] были дообучены RuGPT-3 и RuBERT на датасетах Russian-language reviews и News Category Dataset. Результаты показали, что RuGPT-3 превосходит RuBERT, обеспечивая средний прирост точности классификации на 8,45% и увеличение усредненной F1-меры на 10,95%.

В англоязычной научной среде имеется большее количество исследований в области семантического анализа текстов с использованием LLM. Например, работы Sun X. [18], Thomas G. J. [19], Zhang B. [20] посвящены анализу настроений новостей на финансовых рынках и тональности постов в социальных сетях.

Так, например, в упомянутом исследовании Zhang B. [20] было показано, что тонкая настройка Llama-3 на корпусах данных Twitter financial news sentiment, FiQA dataset и FPB Dataset, значительно повышает качество анализа настроений новостей финансовых рынков относительно базовой модели: качество модели (F1-score) растет с 0,397 до 0,739 на датасете FPB и с 0,363 до 0,811 на датасете Twitter Test. При этом тонко настроенная Llama-3 превзошла в точности ранее разработанные модели для определения семантики финансовых новостей: FinBERT и BloombergGPT.

Также в работе Zhang B. [20] был предложен способ улучшить качество определения семантики с помощью системы расширенной поисковой генерации (от *англ.* RAG system). Предложенная технология позволила индивидуально обогатить контекст каждой новости с помощью других текстов, имеющих связанную тематику. Для этого в запрос модели авторы добавляли иные новостные публикации и пресс-релизы. Благодаря предложенному подходу авторам удалось улучшить качество классификации текстов: метрика F1-score модели GPT-4.0 возросла с 0,652 до 0,708, а Llama-3 с 0,811 до 0,842.

В работе Thomas G. J. [19] исследовалась возможность улучшения производительности компактных языковых моделей за счет LLM. Авторами было продемонстрировано, что более простые языковые модели могут быть дообучены до сопоставимого результата, но требуют большого количества качественно размеченных данных. Для устранения недостатка данных авторы предлагают использовать генеративные возможности крупных языковых моделей, таких как GPT-4 Omni. В частности, была разработана модель TinyFinBERT, которая, будучи значительно меньшей по размеру, достигает производительности, сопоставимой с FinBERT, при этом ее точность в анализе тональности финансовых новостей составляет от 97,5% до 107,5% от показателей FinBERT.

В исследовании Sun X. [18] поднималась проблема правильного подхода к сентимент анализу. Классический подход подразумевает то, что одна модель принимает решение о тональности текста единожды. Авторами была предложена стратегия переговоров множества языковых моделей. Основой предлагаемой стратегии является система «генератор-дискриминатор», в которой одна языковая модель предлагает решение по настроению текста, в то время как другая оценивает это решение. В результате используя обсуждение между GPT-3.5 и GPT-4 авторам удалось повысить точность определения сентимента относительно базовых моделей.

2.2.1. Оценка эффективности LLM для анализа настроений заголовков финансовых новостей

Работа Muhammad I., Rospocher M. [21] посвящена сравнению эффективности больших языковых моделей и традиционных инструментов в анализе тональности финансовых новостных заголовков по отношению к экономической активности конкретных компаний. Традиционными инструментами авторы выбрали языковые модели семейства BERT и NLP инструмент VADER. Исследование отвечает на два вопроса: насколько эффективны LLM в zero-shot сценариях для анализа тональности текста по отношению к компаниям по сравнению с традиционными методами и как различаются результаты моделей OpenAI (ChatGPT-3.5,

ChatGPT-4, ChatGPT-4o) и Google (Gemini 1 Pro, Gemini 1.5 Flash, Gemini 1.5 Pro) в таких задачах.

Для анализа был собран датасет¹⁰ из 1476 заголовков Bloomberg о компаниях Amazon, Netflix, Nvidia и Alphabet, отобранных по ключевым словам, связанным с названиями компаний. Три эксперта в области финансов и экономики разметили заголовки, оценивая тональность (положительную, отрицательную или нейтральную) с точки зрения инвестора, следуя единой инструкции. Надежность разметки подтверждена коэффициентом Krippendorff's alpha¹¹ (0,82), что указывает на высокую согласованность. Датасет слегка несбалансирован: 49% заголовков — положительные, 28% — отрицательные, 23% — нейтральные.

В исследовании использовались три традиционные модели: VADER — инструмент на основе правил для анализа тональности на уровне всего документа или одного предложения, широко применяемый в социальных медиа, но также используемый в финансовой сфере; DistilFinRoBERTa — облегченная версия FinBERT, предобученная на финансовых текстах, таких как Financial PhraseBank, и адаптированная для классификации тональности в финансовом домене; DeBERTa-v3-base-ABSA-v1.1 — модель для аспектно ориентированного анализа тональности (ABSA¹²), обученная на 30 тыс. примерах и донастроенная на 180 тыс. данных из различных датасетов, что позволяет ей фокусироваться на тональности по отношению к конкретным объектам. DeBERTa-v3 тестировалась в двух вариантах: предобученная версия и версия с дообучением на финансовых данных (5-кратная кросс-валидация, 10 эпох, обученная на NVIDIA RTX 4090).

Также тестировались большие языковые модели от OpenAI (ChatGPT-3.5, ChatGPT-4, ChatGPT-4o) и Google (Gemini 1 Pro, Gemini 1.5 Flash, Gemini 1.5 Pro) в zero-shot режиме, то есть без дополнительной настройки. Все модели получали один и тот же

¹⁰ https://github.com/iftikharm95/Target-Level_Financial_Sentiment_Analysis

¹¹ Krippendorff's alpha — статистическая мера согласованности между несколькими аннотаторами, учитывающая случайное совпадение при разметке данных.

¹² Аспектно ориентированный анализ тональности (Aspect Based Sentiment Analysis) — это подвид анализа тональности, который определяет отношение к конкретному аспекту основного предмета обсуждения.

промт: «Evaluate the sentiment conveyed by the headline with respect to the {target} from an investment perspective. Employ a three-tier scale for this assessment: allocate a score of (-1) for a negative sentiment, (0) for a neutral sentiment, and (+1) for a positive sentiment. Assign a neutral score of 0 if the headline is vague or does not clearly express positive or negative implications about the {target}».

Target в тексте промта — это целевая компания, относительно которой производится анализ тональности. Настройки температуры ставились по умолчанию — 0,7 для ChatGPT, 0,9 для Gemini 1 Pro и 1.0 для Gemini 1.5 Flash и Gemini 1.5 Pro. Эффективность оценивалась по метрикам точности, полноты, F1 и общей точности с учетом дисбаланса классов, что важно, так как в датасете 49% заголовков были положительными, 28% — отрицательными и 23% — нейтральными.

Дообученная DeBERTa-v3 показала лучшие результаты: точность 0,87, макро F1 0,84, значительно превосходя предобученную версию (точность 0,52, макро F1 0,54). VADER (точность 0,49, макро F1 0,48) и DistilFinRoBERTa (точность 0,51, макро F1 0,50) показали низкую эффективность из-за ориентации на анализ уровня предложения или документа, а не на уровне объектов. Среди LLM лучшие результаты продемонстрировали ChatGPT-4 (точность 0,86, макро F1 0,83) и Gemini 1,5 Pro (точность 0,83, макро F1 0,81), что близко к дообученной DeBERTa, несмотря на отсутствие дообучения. Преимущество DistilFinRoBERTa над VADER обусловлено ее предобучением на финансовых текстах, что улучшает понимание рабочей области, в то время как успех DeBERTa связан с ее механизмом разделенного внимания, который повышает качество анализа контекста, и специализацией на ABSA, что делает ее более подходящей для задач анализа тональности на уровне объектов.

Набор данных, ограниченный 1476 заголовками из одного источника, достаточен для сравнения моделей, но его разнообразие могло бы быть шире. Согласованность экспертов требует улучшения, так как разметка существенно влияет на метрики. Тем не менее результаты подчеркивают потенциал LLM как практических решений без дообучения для решения задач разметки новостей по тональности.

2.2.2. Использование больших языковых моделей для анализа настроений и разработки инвестиционной стратегии на финансовых рынках

В исследовании Mun Y., Kim N. [22] рассматривается применение больших языковых моделей для анализа тональности финансовых новостей и их использование для разработки инвестиционных стратегий. Исследование отвечает на два вопроса: могут ли открытые LLM обеспечить стабильность и прибыльность портфеля, используя анализ тональности новостей и могут ли такие методы повышения эффективности ответов, как CoT¹³, SuperICL¹⁴ и LLM Bootstrapping¹⁵, увеличить доходность портфеля открытых LLM при использовании анализа тональности новостей.

Исследование опиралось на набор данных, включающий новостные статьи и данные о ценах акций для 30 крупнейших компаний по рыночной капитализации, зарегистрированных на Nasdaq, за период с 2021 по 2023 г. Новостные статьи были собраны с использованием API GNews, который предоставлял ежедневные материалы, относящиеся к этим компаниям. Изначально было получено 21 599 статей, но после фильтрации, исключающей статьи, за которыми не следовали торговые дни, для анализа осталось 13 696 статей. Данные о ценах акций, в частности ежедневные цены закрытия, были получены из библиотеки Yahoo Finance для оценки волатильности цен.

В исследовании для анализа тональности финансовых новостей и разработки инвестиционных стратегий использовались как дискриминативные, так и генеративные LLM. Среди дискриминативных моделей были задействованы BERT, FinBERT и RoBERTa. Эти модели изначально были предварительно об-

¹³ CoT (Chain of Thought) — методика промптинга, которая побуждает модель к пошаговому рассуждению при выполнении задачи, улучшая понимание контекста и точность анализа.

¹⁴ SuperICL (Super-In-Context Learning) — методика промптинга, расширяющая возможности In-Context Learning через структурированные примеры в промпте, что повышает эффективность.

¹⁵ Bootstrapping — метод улучшения производительности модели путем итеративного использования собственных предсказаний для дообучения или уточнения результатов.

учены на финансовых данных, а затем дообучены для задачи классификации тональности. Генеративные модели, примененные в исследовании, включали Llama 3.1, Mistral и Gemma 2. Эти модели были выбраны за их способность не только генерировать текст, но и выполнять задачи анализа тональности с использованием продвинутых техник промптинга, таких как CoT, SuperICL и Bootstrapping. Модели тестировались в различных конфигурациях: от базовых версий до вариантов с базовым и продвинутым промптингом, чтобы определить наиболее эффективные методы для построения инвестиционных портфелей.

Пример промпта методом SuperICL: System Message: You are a stock analyst specializing in assessing sentiment in financial news articles. Prompt: Based on the article titled {Title}, determine whether the tone is positive, negative, or neutral towards {Symbol}. The sentiment prediction from BERT is {BERT Answer} Please respond in the following format, and omit any reasoning If your sentiment differs from the reference prediction, provide a brief reason why Sentiment: [positive/negative/neutral] Reason: [Provide reason only if your sentiment differs from the reference]

Инвестиционная стратегия формировалась на основе тональности, которую выдала модель. Решения о покупке или продаже принимались на основе сентимента, рассчитанного за предыдущий день. Если сентимент для компании был положительным, акции этой компании покупались. Если сентимент был отрицательным, акции компании продавались или открывалась короткая позиция, если акций в портфеле не было. При нейтральной тональности никаких операций не производилось.

Результаты оценивались с помощью ключевых финансовых метрик, таких как среднедневная доходность, стандартное отклонение, коэффициент Шарпа, максимальная просадка и итоговая доходность.

Исследователи получили значимые результаты, демонстрирующие эффективность LLM для анализа тональности и разработки инвестиционных стратегий. Лучшие показатели достигнуты с использованием генеративной модели Llama 3.2 в комбинации с техникой SuperCL и BERT-ICL в долгосрочной стратегии: коэффициент Шарпа составил 0,81, итоговая доходность — 54,78%. Это

существенно превосходит рыночный индекс Nasdaq 30 с коэффициентом Шарпа 0,55 и доходностью 30,10%. Генеративные модели, такие как Llama 3.1 и Llama 3.2, в целом показали более высокую эффективность по сравнению с BERT и RoBERTa, что подчеркивает их преимущество в анализе тональности.

Применение продвинутых техник промптинга, особенно SuperCL, значительно улучшило результаты генеративных моделей, обеспечивая более высокую доходность. Долгосрочные стратегии оказались успешнее краткосрочных и комбинированных, что может быть связано с рыночными тенденциями в период исследования. Стоит отметить, что авторы никак не проверили количественно, насколько разные модели качественно справляются с определением тональности. Они лишь предоставили пример, где показали, что BERT может справляться с этой задачей хуже Llama 3.2. Можно судить о том, что лучше для разметки только по итоговым метрикам портфелей, где генеративные LLM впереди.

2.2.3. Анализ и классификация финансовых настроений: сравнительное исследование тонко настроенных моделей глубокого обучения

В исследовании Nasiopoulos D. K., Roumeliotis K. I. и соавторов [23] применение LLM для анализа финансовой тональности текстов с целью повышения точности прогнозирования рыночных трендов и поддержки инвестиционных решений. Исследователи стремились ответить на три ключевых вопроса:

- 1) какова точность классификации тональности современными генеративными LLM по сравнению с традиционными методами машинного обучения;
- 2) насколько эффективны и масштабируемы облегченные модели, такие как GPT-4o-mini, в сравнении с более крупными аналогами;
- 3) как zero-shot и fine-tuned подходы влияют на эффективность моделей в задаче анализа тональности.

Для исследования был использован объединенный датасет Financial Sentiment Analysis, включающий данные из FiQA и Financial PhraseBank. Датасет содержит 5842 финансовых текстов, аннотиро-

ванных по тональности 16 экспертами с уровнем согласованности 74,9%. Распределение меток неравномерное: 3130 нейтральных, 1852 положительных и 860 отрицательных образцов. Для устранения дисбаланса применялась стратегия стратифицированной выборки, в результате чего итоговый датасет разделен на обучающую (1548 наблюдений), валидационную (516) и тестовую (516) выборки с равным распределением меток.

Для получения тональности были использованы как традиционные модели машинного обучения, так и современные трансформерные архитектуры. Среди традиционных моделей применялись SVM, случайный лес и логистическая регрессия, которые служили в качестве базовых моделей для сравнения. Гиперпараметры подбирались через байесовскую оптимизацию через Optuna. Для SVM оптимизировались такие параметры, как тип ядра, параметр регуляризации C и параметры векторизации текста TF-IDF. Random Forest настраивался по количеству деревьев, максимальной глубине и другим параметрам, в то время как Logistic Regression оптимизировалась по параметру регуляризации C и комбинации штрафа и решателя. Эти модели, несмотря на их вычислительную эффективность, показали более низкую точность (0,64–0,65) по сравнению с LLM, что подчеркивает их ограничения в обработке сложных языковых структур финансовых текстов.

Среди трансформерных моделей исследовались BERT, FinBERT, GPT-4o и GPT-4o-mini. BERT и FinBERT, модель, специально предобученная на финансовых текстах, были донастроены на объединенном датасете FiQA и Financial PhraseBank с использованием байесовской оптимизации гиперпараметров, таких как размер пакета, скорость обучения, количество эпох и коэффициент выбывания. BERT достиг точности 0,81, слегка опередив FinBERT с 0,8, что может быть связано с более широким охватом семантики в предобученной модели BERT. GPT-4o и GPT-4o-mini, закрытые модели от OpenAI, тестировались в двух режимах: zero-shot и после дообучения через API OpenAI с использованием JSONL-файлов для парного формата "запрос-ответ". В zero-shot режиме GPT-4o показал точность 0,79, а GPT-4o-mini — 0,78, но после дообучения обе модели достигли одинаковой точности 0,87, де-

монстрируя значительное улучшение за счет адаптации к домену. GPT-4o-mini выделяется своей вычислительной эффективностью, требуя меньше времени (2054 с) и затрат (1,54 доллара) на дообучение по сравнению с GPT-4o (3621 с, 12,83 доллара).

Дообучение GPT-моделей происходило на тех же обучающих и валидационных наборах данных, что использовались для других моделей. Процесс дообучения проводился через API OpenAI с использованием JSONL-файлов, содержащих пары "запрос-ответ". Файлы загружались через интерфейс OpenAI, а дообучение выполнялось с стандартными гиперпараметрами (3 эпохи, размер пакета 3, множитель скорости обучения 1,8), что позволило моделям глубже понимать контекст финансовых новостей.

Для структурированной передачи данных авторами использовался формат JSONL-файлов: каждая строка содержала объект с массивом `"messages"`, включающим три элемента: системное сообщение (`"role": "system"`, контекст задачи), пользовательский запрос (`"role": "user"`, с текстом предложения и инструкцией) и ответ модели (`"role": "assistant"`, JSON с меткой тональности). Например: `{"messages": [{"role": "system", "content": "You are an AI assistant specializing in financial sentiment classification."}, {"role": "user", "content": "You are an AI assistant specializing in financial sentiment classification. Your task is to analyze each financial sentence and classify it as negative, positive, or neutral. Provide your final classification in the following JSON format without explanations: {\"Sentiment\": \"sentiment_tag\"}. \nFinancial sentence: ..."}, {"role": "assistant", "content": \"{\"Sentiment\": \"neutral\"}"}]}`.

По итогам исследования авторы сделали следующие выводы:

1. BERT, FinBERT, GPT-4o и GPT-4o-mini, демонстрируют значительное превосходство над традиционными методами машинного обучения в задаче анализа тональности.
2. Облегченная модель GPT-4o-mini выделяется как эффективная альтернатива более крупным моделям, таким как GPT-4o, являясь балансом точности, цены и скорости.
3. Подходы без предварительного обучения и с дообучением существенно различаются по эффективности.

2.2.4. Трансформация анализа настроений в финансовой сфере с помощью ChatGPT

В исследовании Fatouros G., Soldatos J. и соавторов [24] рассматривается применение LLM, таких как ChatGPT, для анализа тональности в финансовой сфере с целью повышения точности прогнозирования изменения курсов валют и поддержки стратегических торговых решений. Исследователи стремились ответить на следующие вопросы: могли ли генеративные LLM эффективно выполнять анализ тональности финансовых текстов в zero-shot сценарии и как их эффективность отличается от дискриминативных моделей, таких как FinBERT, в задачах классификации тональности.

Авторы статьи использовали датасет, собранный с платформ Forex Live1 and FXstreet в период с января по март 2023 г. Данные содержат 2291 заголовок. Разметку заголовков проводили исследователи, основываясь на потенциальном краткосрочном влиянии заголовка на соответствующую валютную пару. Распределение меток получилось относительно равномерным: 815 нейтральных, 767 позитивных и 709 негативных.

Для анализа тональности финансовых текстов использовались две ключевые модели: FinBERT и ChatGPT. FinBERT использовалась через библиотеку Transformers от Hugging Face, что позволило загрузить предобученные веса модели и донастроить ее на размеченном датасете. FinBERT классифицировала тональность заголовков, предоставляя вероятностные оценки уверенности модели в каждом классе. Модель была протестирована на публичном наборе данных, чтобы определить ее способность предсказывать тональность с использованием метрик, таких как точность, полнота, F1-score и Sentiment Mean Absolute Error (S-MAE).

Второй моделью, исследуемой в работе, была ChatGPT 3.5-turbo, примененная в zero-shot режиме, то есть без дополнительного обучения. В экспериментах исследовались различные варианты промптов (табл. 1), которые включали информацию о финансовых инструментах, таких как валютные пары (AUDUSD, EURCHF, EURUSD, GBPUSD, USDJPY).

Таблица 1. Варианты промптов, которые использовались в статье

Название	Промпт
GPT-P1	Act as a financial expert holding {ticker}. How do you feel about the headline {headline}? Answer in one token: positive, negative, or neutral.
GPT-P2	Act as a financial expert. Classify the sentiment for {ticker} based only on the headline {headline}. Answer in one token: positive, negative, or neutral.
GPT-P3	Act as a sentiment analysis model trained on financial news headlines. Classify the sentiment of the headline {headline}. Answer in one token: positive, negative, or neutral.
GPT-P4	Act as an expert at forex trading holding {ticker}. Based only on the headline {headline}, will you buy, sell or hold {ticker} in the short term? Answer in one token: positive for buy, negative for sell, or neutral for hold position.
GPT-P5	Act as an expert at forex trading holding {ticker}. Based only on the following list of headlines {ticker_daily_headlines}, will you buy, sell or hold {ticker} in the short term? Answer in one token: positive for buy, negative for sell, or neutral for hold position
GPT-P6	Act as a sentiment analysis service of a financial platform. Based only on the following list of headlines {all_daily_headlines}, provide a summary of the daily sentiment for the forex pairs: {tickers}. Provide only the sentiment per forex pair in JSON format eg {'USDJPY': 'positive', 'EURUSD': 'neutral'}. The sentiment can be positive for buy, negative for sell, or neutral for hold position.

Основные выводы исследователей показали, что ChatGPT с оптимизированным промптом GPT-P4 продемонстрировала лучшие результаты в задаче анализа тональности финансовых текстов, достигнув точности 0,71–0,73, макро F1-оценки 0,77 и S-MAE 0,26–0,30, значительно превосходя FinBERT (точность 0,56, макро F1 0,61, S-MAE 0,54). Стоит отметить, что, рассмотрев по отдельности категории тональности, FinBERT не обошел ChatGPT ни по одной из метрик ни в одной категории. FinBERT, несмотря на дообучение, оказалась менее эффективной из-за ограниченной способности улавливать сложные контекстные нюансы финансовых новостей. ChatGPT, напротив, показала высокую производительность в zero-shot режиме, особенно с промптами, включающими информацию о валютных парах, что подчеркивает ее способность к интерпретации сложных текстов. Успех ChatGPT связан с ее генеративной природой и способностью учитывать контекст, что делает ее более подходящей для анализа тональ-

ности в финансовой сфере по сравнению с дискриминативной моделью FinBERT.

Для анализа влияния сентиментов на курсы валют авторы использовали корреляцию между предсказанными сентиментами и доходностью валютных пар. GPT-P4 показала более высокую корреляцию (0,55–0,64), тогда как для FinBERT эти значения варьировались от 0,16 до 0,29. Это указывает на то, что сентименты, полученные с помощью ChatGPT, имели более сильную связь с реальными движениями рынка, чем сентименты от FinBERT. Кроме того, авторы проводили оценку направленной точности, которая измеряет, насколько точно модель предсказывает направление движения валютной пары на основе сентимента. Там GPT имеет разброс метрики от 0,61 до 0,67 в зависимости от промпта, что превосходит модель FinBERT, которая имеет 0,52–0,59.

2.2.5. Контекстуализированный анализ настроений с использованием больших языковых моделей

В исследовании Breitung C., Kruthof G., Müller S. [25] изучаются возможности LLM в применении экономического анализа для прогнозирования отраслевых последствий макроэкономических новостей. Исследователи ставили целью ответить на четыре ключевых вопроса:

- 1) способны ли LLM эффективно контекстуализировать тональность новостных заголовков;
- 2) варьируется ли способность языковых моделей точно контекстуализировать новостные заголовки в зависимости от различных тематических областей и доменов, таких как разные виды товаров;
- 3) является ли эффективность языковых моделей устойчивой к изменениям в дизайне промптов;
- 4) повышает ли увеличение детализации информации об отрасли точность определения сентимента.

Авторы статьи создали датасет из новостных заголовков о ценах на 40 видов товаров из набора Ravenpack за период с 2000 г. по май 2023 г., полученного через платформу Wharton Research Data Services (WRDS). Заголовки с оценкой релевантности ниже 95

из 100 были исключены. Также новости отбирались по содержанию, исключая новости о конкретных фирмах и несущественные сообщения. Для каждого заголовка с использованием модели Falcon-T GenAI были сгенерированы два дополнительных заголовка, отражающих рост или снижение цен, с последующей ручной проверкой соответствия. Отраслевые настроения (позитивные, нейтральные, негативные) для каждого товара определялись вручную на основе экономических принципов, учитывающих прямое влияние цен на отрасли производства и потребления. Чтобы сделать распределение товаров равномерным авторы отобрали по 40 заголовков для каждого из 40 видов, получив 1600 наблюдений в выборке.

В исследовании для анализа контекстуализированной тональности¹⁶ новостных заголовков применялись пять языковых моделей: Falcon-7B, LLaMA-2-7B, LLaMA-2-13B, GPT-3 и GPT-4. Эти модели также использовались без дообучения. Для обработки ответов моделей применялась постобработка, включающая удаление повторяющихся заголовков и упоминаний изменения цен, чтобы извлечь чистые метки настроений, такие как позитивные, нейтральные или негативные. В исследовании применялось 7 стратегий промптинга для оценки влияния структуры запросов на генерацию отраслевых настроений, они представлены в табл. 2.

Исследование показало, что GPT-4 и GPT-3, демонстрируют высокую способность к контекстуализированному анализу тональности новостных заголовков о ценах на товары с точностью до 73,3% и 67,6% соответственно, в то время как открытые модели, такие как Falcon-7B и LLaMA-2, достигают точности 38,9–54,9%. Производительность моделей варьируется в зависимости от тематики, с высокой точностью для таких товаров, как пшеница (91,7%), и меньшей для электроэнергии (63,3%). Дизайн промптов существенно влияет на результаты: прямой и косвенный промптинг оказались наиболее эффективными, тогда как нарративный и встроенный нарративный подходы снижали точность.

¹⁶ Контекстуализированная тональность — подход к анализу тональности текстов, при котором сентимент определяется с учетом контекста, такого как отрасль или компания, позволяя выявлять разнонаправленные эффекты одной новости на различные субъекты.

Таблица 2. Стратегии промптинга для оценки влияния структуры запросов на генерацию отраслевых настроений

Категория	Промпт
Прямой	How does the news headline affect the profit of firms within the {industry} industry? Answer with 'increase' if profits increase, 'decrease' if profits decrease or 'unaffected' if profits are unaffected. News headline: '{head}'
Косвенный	Consider a world where the news headline is '{head}'. In this world, what would likely happen to the profits of companies in the '{industry}' industry? Answer with 'increase' if profits increase, 'decrease' if profits decrease or 'unaffected' if profits are unaffected.
Гипотетический	Imagine you are a financial analyst. How does the news headline affect the profit of firms within the {industry} industry? Answer with 'increase' if profits increase, 'decrease' if profits decrease or 'unaffected' if profits are unaffected. News headline: '{head}'
Встроенный нарративный	A news anchor is reporting on the headline '{head}'. As an expert in the '{industry}' industry, they turn to you for your take. Would you say this would likely increase, decrease, or have no effect on profits?
Множественный выбор	After the news headline '{head}' breaks out, companies in the '{industry}' industry might experience: A) Increased profits B) Decreased profits C) Unaffected profits. Which one do you think is most likely?
Аналоговый	Think of the '{industry}' industry as a ship sailing in the sea. Given the news headline '{head}', is it sailing towards a treasure island (profits increase), a storm (profits decrease), or is it just calm waters ahead (profits unaffected)?
Нарративный	Imagine a chapter in a book titled '{head}', where the main characters are firms in the '{industry}' industry. What happens to these characters' fortunes: do they rise (profits increase), fall (profits decrease), or remain the same (profits unaffected)?

Остальные методы показали среднюю эффективность. Увеличение детализации контекста не всегда повышало производительность, особенно при использовании неоднозначных терминов в отраслевых описаниях, что подчеркивает важность сбалансированного подхода к категоризации отраслей.

3. МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ ИНДЕКСОВ ПРОМЫШЛЕННОГО ПРОИЗВОДСТВА

Основываясь на экономических предпосылках и предшествующих исследованиях, для текущей работы было выделено 3 отрасли промышленного производства: производство кокса, нефте-

продуктов и ядерных материалов, химическое производство, производство резиновых и пластмассовых изделий.

В настоящей работе предлагается проверить следующие гипотезы:

- 1) продвинутые нейросетевые архитектуры демонстрируют более высокую устойчивость и точность прогнозирования по сравнению с традиционными моделями;
- 2) использование новостных индексов улучшает прогнозные характеристики моделей;
- 3) совместное прогнозирование индексов промышленного производства снижает ошибку по сравнению с независимыми моделями.

3.1. Оценка тональности новостей и построение новостных индексов

Предполагается, что новостные индексы передают некоторую информацию об экономических показателях и тем самым улучшают прогнозную силу моделей. Для построения новостного индекса по ключевым словам из единственного источника были отобраны тексты ежедневных новостей по тематикам промышленного производства, безработицы и инфляции. Далее каждая новость классифицировалась одним из способов как положительная, нейтральная или негативная. Позитивным новостям присваивалась метка 1, нейтральным — 0 и негативным — -1. Метка класса умножалась на вероятность принадлежности к классу, и таким образом была получена эмоциональная оценка новости. Полученные значения усреднялись для каждого месяца, затем сглаживались скользящим средним за 3 месяца. Таким образом в исследовании было построено несколько новостных индексов, один из них продемонстрирован на рис. 2.

Основной проблемой при построении новостных индексов является обеспечение релевантности и качества обработки неструктурированного текстового потока. Классические методы, основанные на ключевых словах или базовых моделях для расчета сентимента (например, FinBERT), не учитывают контекст



Рис. 2. Динамика индекса промышленного производства и новостного индекса, взятого с лагом в 5 месяцев

новостей и не строят их причинно-следственные связи для анализа, а продвинутые методы, такие как RAG системы, изначально имеют примитивную методологию обогащения контекста новостей, что ограничивает качество разметки их сентимента.

В рамках собственного исследования предлагается комплексный подход к совершенствованию новостных индексов за счет модернизации RAG-системы. В отличие от классической реализации, мы модернизировали систему по трем ключевым направлениям: отбор релевантных новостей, представление информационного контекста и классификация тональности. В результате была разработана система, включающая пять ключевых модификаций:

- Рассуждение — внедрение механизма, при котором модель анализирует контекст новости перед оценкой сентимента, что повышает обоснованность классификации.
- Обобщение контекста — сжатие большого контекста новостей до двух-трех предложений для концентрации релевантной информации и снижения шума.
- Переранжирование — отбор наиболее релевантных новостей из пула с помощью отдельной ML-модели.
- Оценка релевантности — введение метрики, количественно оценивающей степень связи новости с конкретным таргетом (например, ИПП).
- Тонкая настройка (Fine-tuning) — адаптация базовых языковых моделей к предметной области с использованием эф-

фективных методов дообучения, таких как LoRA (Low-Rank Adaptation), в рамках архитектуры Mixture of Experts (MoE).

Прямая классификация тональности новостей без построения причинно-следственных связей может приводить к систематическим ошибкам, обусловленным наличием скрытых экономических механизмов. Поверхностный анализ часто игнорирует опосредованные эффекты, существенно влияющие на итоговую оценку.

Внедрение шага «рассуждения», в рамках которого модель формулирует логическую цепочку от события, описанного в новости, к его возможным последствиям для промышленности, повысит обоснованность и точность итоговой оценки сентимента.

Для реализации этапа каузального анализа была применена большая языковая модель DeepSeek-R1, специализирующаяся на задачах логического вывода. Модель часто в рассуждениях строила детализированные причинно-следственные цепочки от исходного события к его потенциальному влиянию на промышленное производство.

В результате тестирования, полученный новостной индекс коррелировал с индексом промышленного производства в среднем на 0,03 п.п. выше, чем изначальный индекс. Результаты демонстрируют статистически значимое улучшение предсказательной силы предложенного подхода по сравнению с изначальным методом.

Одной из ключевых проблем при обработке новостного потока для построения индекса является избыточность информации. Объемные тексты могут содержать значительное количество шума и второстепенных деталей, которые смещают эмбединги сообщения, затрудняя последующую классификацию и снижая релевантность индекса. Целью данного подхода была разработка и валидация метода суммаризации, направленного на сокращение исходного новостного контекста до ключевых утверждений без потери смысловой нагрузки, значимой для прогнозирования ИПП.

Для суммаризации использовались продвинутое языковые модели (Large Language Models, LLMs), оптимизированные для работы с большими контекстами. На основе сравнительного тестирования нескольких моделей (включая ruT5_sum, GPT-3_sum и ru-mbart-large-summ) была выбрана модель ruT5_sum, которая

визуальным анализом показала наилучшее сочетание смысловой точности и лаконичности в генерации на русском языке.

Процесс суммаризации был организован следующим образом:

- Исходные новостные тексты, собранные из целевых источников (ТАСС, РБК, Интерфакс и др.), подвергались очистке от нетекстовых элементов и форматов.
- Формирование промпта: для обеспечения стандартизированного подхода был разработан специализированный промпт-шаблон: «Обобщи следующую новость до двух-трех предложений, выделив основное событие и его потенциальное влияние на промышленное производство в России: [Текст новости]».
- Данный промпт направлял модель не на простое сокращение текста, а на тематическую фокусировку.

Также нами была изучена возможность повышения релевантности контента через переранжирование новостей. Исходной точкой послужил традиционный подход к формированию новостной выборки на основе ключевых слов, который, хотя и обеспечивает общее тематическое соответствие, обладает существенными ограничениями. С другой стороны, методы кластеризации и тематического анализа демонстрируют сложную настройку и менее стабильный результат при отборе новостей по отраслям промышленного производства. Алгоритмы опирались на лексическую и семантическую близость, но не на экономическую.

Гипотеза данного направления исследования заключалась в том, что разметка с помощью большой языковой модели новостей по степени их релевантности целевым показателям позволит существенно повысить концентрацию полезного сигнала в финальном индексе и улучшить его прогнозную силу.

Для реализации переранжирования использовалась модель DeepSeek-R1. В отличие от базового подхода, где новости агрегировались по принципу наличия ключевых слов, подход переранжирования работал по следующему алгоритму:

Первичный семантический отбор: для каждой целевой новости отбиралось 25 семантически близких новостей среди всей базы новостей за последние четыре месяца.

Переранжирование: на втором этапе применялась большая языковая модель с промптом: «Выбери из предложенного списка ново-

стей ту, которая наиболее релевантна для [конкретная новость]. Новости: [список из 25 предварительно отобранных новостей].»

Модель анализировала смысловые нюансы и контекстную релевантность, выбирая топ-3 наиболее релевантных новостей. Расчет сентимента: итоговые новости помещались в контекст к конкретной новости и рассчитывался сентимент. Внедрение двухэтапной системы переранжирования показало значимое улучшение прогнозных характеристик индекса. Улучшение прогнозной силы: композитная метрика качества прогноза для индексов, построенных с применением двухэтапного переранжирования, показала прирост в среднем на 5,8 процентных пункта. Наиболее значимое улучшение Directional Accuracy (на 7–10 п.п.) было зафиксировано для технологически сложных секторов, таких как "Машиностроение" и "Производство электрооборудования".

Полученные тестовые результаты продемонстрировали, что переранжирование новостей позволяет выделить информацию с наиболее сильным прогнозным сигналом, особенно в секторах со сложными производственными цепочками, где контекст новостных событий может быть не связан с лексической близостью новостей.

Подход оценки релевантности заключался в введении метрики, количественно оценивающей степень релевантности новости к конкретному таргету (например, ИПП). Данный подход позволил оценить уровень релевантности новости к целевому показателю по непрерывной шкале.

Для оценки релевантности большая языковая модель (LLM) получала задание присвоить каждой новости числовую оценку от 0 до 1, где 0 означает крайне слабую или косвенную связь с промышленным производством, а 1 — прямое и значительное воздействие.

Промпт для оценки формулировался следующим образом: «Оцени по шкале от 0 до 1, насколько данная новость релевантна к индексу промышленного производства России». Итоговый новостной индекс сентимент взвешивался по релевантности, что позволило учесть с разной силой как общие экономические шоки, так и локальные релевантные и сконцентрироваться на событиях, несущих наиболее значимый прогнозный сигнал.

3.2. Прогнозные модели: спецификации, описание принципа построения одномерных и многомерных подходов; модели с новостным индексом и без

Для проверки гипотез настоящего исследования были построены модели для одномерного и многомерного прогнозирования (в нашем случае — одновременного прогнозирования нескольких индексов промышленного производства. В рамках нейросетевого подхода в обоих случаях используется модель N-HITS, в рамках эконометрического подхода одномерный прогноз строится с помощью ARIMA, многомерный — с помощью VAR. Причем внутри каждого подхода сравнивалось качество моделей с новостным индексом и без него. N-HITS — нейросетевая модель с полносвязным слоем на выходном уровне, который позволяет получать в качестве прогноза как единственную точку, так и вектор прогнозов, поэтому было решено использовать ее для проверки гипотез исследования. На первом этапе были построены модели для прогнозирования индексов производства кокса, нефтепродуктов и ядерных материалов; резиновых и пластмассовых изделий; химического производства. Подобный набор индексов был выбран, так как рассматриваемые отрасли связаны в рамках производственной цепочки: продукция одной отрасли может быть использована в качестве сырья в других отраслях.

Для корректной работы моделей и получения более точных прогнозов была поставлена задача отбора экзогенных переменных для группы индексов производства. Задача была решена в два этапа: на первом этапе были отобраны экзогенные переменные, имеющие корреляцию больше 0,4 по абсолютной величине с целевой переменной. На втором этапе для избежания мультиколлинеарности сильно коррелированные признаки удалялись: если между двумя признаками корреляция по абсолютной величине превышала 0,6, то один из признаков удалялся. Для проверки одной из гипотез исследования во все модели в качестве экзогенной переменной добавлялся новостной индекс, разработанный с применением моделей BERT, LLaMA, DeepSeek.

Корреляции целевых переменных с отобранными экзогенными переменными представлены в табл. 3–5.

Таблица 3. Матрица корреляций производства кокса, нефтепродуктов и ядерных материалов с экзогенными переменными

	Производство кокса, нефтепродуктов и ядерных материалов	Индекс промышленного производства	Импорт
Производство кокса, нефтепродуктов и ядерных материалов	1		
Индекс промышленного производства	0,58	1	
Импорт	0,44	0,51	1

Таблица 4. Матрица корреляций химического производства с экзогенными переменными

	Химическое производство	Динамика оборота оптовой торговли	Стоимость нефти
Химическое производство	1		
Динамика оборота оптовой торговли	0,51	1	
Стоимость нефти	-0,44	-0,16	1

Таблица 5. Матрица корреляций резиновых и пластмассовых изделий производства с экзогенными переменными

	Производство резиновых и пластмассовых изделий	Динамика оборота розничной торговли	Импорт	Процентные ставки по кредитам финансовым организациям
Производство резиновых и пластмассовых изделий	1			
Динамика оборота розничной торговли	0,65	1		
Импорт	0,62	0,53	1	
Процентные ставки по кредитам финансовым организациям	-0,41	-0,15	-0,49	1

Таблица 6. Характеристика спроса и предложения для целевых переменных

Производство кокса, нефтепродуктов и ядерных материалов		
Индекс промышленного производства	Предложение	В смежных отраслях промышленности могут быть задействованы продукты кокса, нефтепродукты. Увеличение общего выпуска будет требовать увеличения выпуска отдельных компонентов
Импорт	Предложение	Необходимое сырье и оборудование может быть импортным, поэтому при росте общего импорта возрастет объем производства
Химическое производство		
Динамика оборота оптовой торговли	Спрос	Рост спроса на товары, в производстве которых использованы продукты из резины или пластмассы
Стоимость нефти	Предложение	Нефть и нефтепродукты могут быть задействованы в химическом производстве, поэтому при увеличении цены на нефть издержки повышаются и производство сокращается
Производство резиновых и пластмассовых изделий		
Динамика оборота розничной торговли	Спрос	Рост спроса на товары, в производстве которых использованы продукты из резины или пластмассы
Импорт	Предложение	Необходимое сырье и оборудование может быть импортным, поэтому при росте общего импорта возрастет объем производства
Средневзвешенные процентные ставки по кредитам, предоставленным кредитными организациями нефинансовым организациям в рублях свыше 1 года	Предложение	При росте ставок организации берут меньше кредитов, объем оборотного капитала снижается и выпуск падает

Полученные наборы экзогенных переменных можно охарактеризовать со стороны спроса и предложения в отрасли, пояснения представлены в табл. 6.

В настоящем исследовании для разметки тональности новостей были использованы малая языковая модель BERT и несколько больших языковых моделей (LLaMA и в рамках нового подхода с помощью ризонинговой модели DeepSeek R1). Динамика полученных индексов сравнивается на рис. 3.



Рис. 3. Динамика индекса промышленного производства и трех новостных индексов

Для сравнения моделей при прочих равных экзогенные переменные, отобранные для каждой целевой переменной, были объединены и использовались как в одномерных, так и в многомерных моделях. Все модели были обучены прогнозировать следующие три месяца.

В эконометрических моделях VAR и ARIMA прогноз на несколько точек строится рекурсивно, подразумевается переиспользование прогнозов модели для предсказания следующей точки, ошибка модели накапливается с каждой итерацией. Для получения прогнозов трех последующих значений были использованы одни и те же экзогенные переменные за последний доступный период, а также лаги целевых переменных. Число лагов было выбрано перебором на основе критерия BIC.

В случае N-HITS прогноз трех точек производится моментно, без переиспользования предыдущих предсказаний модели. Для получения прогнозов трех последующих значений были использованы 12 лагов целевых переменных и 12 последних значений экзогенных переменных. Схема получения одномерных и многомерных прогнозов представлена на рис. 4.

В исследовании предполагается, что используемая нейросетевая модель N-HITS способна улавливать нелинейные взаи-

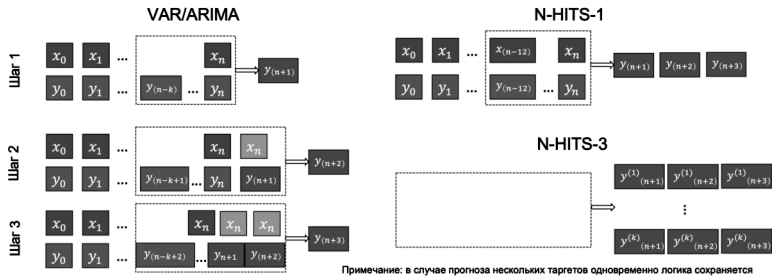


Рис. 4. Схемы получения прогнозов

Таблица 7. Группы индексов промышленного производства

Группа индексов	Индексы промышленного производства	Связь между индексами в группе
Промышленные материалы	- Производство кокса, нефтепродуктов и ядерных материалов - Производство резиновых и пластмассовых изделий - Химическое производство - Производство прочих неметаллических минеральных продуктов	Нефтепродукты являются сырьем для производства продукции химического производства, изоляционных материалов. Нефтепродукты и химия используются при производстве неметаллических продуктов (обжиг стекла, керамики, добавки для бетона)
Машиностроение и техника	- Производство машин и оборудования - Производство транспортных средств и оборудования - Производство электрооборудования, оптического оборудования - Металлургическое производство, производство готовых металлических изделий	Электро- и оптическое оборудование, металлические заготовки и готовые детали используются в производстве машин, транспорта и прочей спецтехники и оборудования
Пищевое производство	- Производство пищевых продуктов, напитков, табака - Обработка древесины и производство изделий из дерева - Целлюлозно-бумажное производство	Переработка органических природных ресурсов. Производство потребительской продукции и упаковки для нее. Изготовление изделий для транспортировки и хранения продукции (деревянные поддоны и паллеты)
Одежда и обувь	- Текстильное и швейное производство - Производство кожи, изделий из кожи, обуви	Схожие циклы производства, связанные со сменой модных трендов, сезонов. Одни и те же каналы сбыта продукции

мосвязи между несколькими целевыми показателями, поэтому на втором этапе 13 индексов промышленного производства были объединены авторами в группы, индексы внутри группы связаны по технологической цепочке или иным признакам. Полученные группы индексов промышленного производства и предполагаемая связь между индексами представлены в табл. 7. Аналогично первому этапу для каждой группы были отобраны экзогенные переменные и построены модели.

3.3. Принцип сравнения подходов прогнозирования

Вывод о качестве моделей был сделан по средним значениям метрик на 10 окнах кросс-валидации. Кросс-валидация — это процедура для оценки качества модели. При проведении кросс-валидации для временных рядов наблюдения в выборке не смешиваются, и обучающая выборка идет строго перед тестовой (рис. 5).

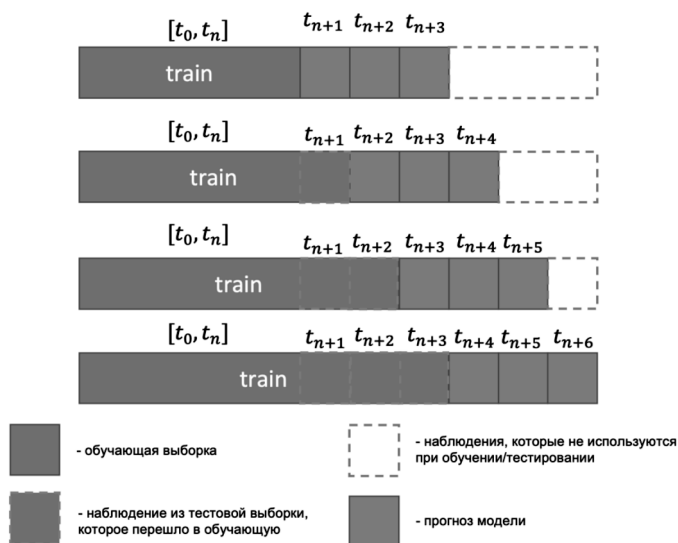


Рис. 5. Схема кросс-валидации

На первом окне модель обучается на тренировочной выборке, делает предсказания на три шага вперед (размер тестовой выборки фиксирован), рассчитываются метрики качества. На следующем этапе прогнозное окно смещается на одно наблюдение вперед, в тренировочную выборку попадает следующее наблюдение, модель обучается на обновленной выборке и делает предсказания на следующие три шага. Процедура повторялась в течение 10 этапов, метрики качества были усреднены по всем окнам.

В качестве метрик качества моделей были выбраны MAPE (Mean Absolute Percentage Error) и MDA (Mean Directional Accuracy). MAPE – средняя абсолютная процентная ошибка:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - a(x_i)|}{|y_i|}, \quad (20)$$

где

y_i — истинное значение,

$a(x_i)$ — предсказанное значение.

MAPE — одна из наиболее распространенных метрик, используемых в подобных задачах. Чем меньше значение метрики, тем лучше прогноз.

MDA — средняя точность направления, показывает долю случаев, когда модель правильно предсказала направление изменения целевого показателя:

$$MDA = \frac{1}{n} \sum_t 1_{sgn}(A_t - A_{t-1}) = sgn(F_t - A_{t-1}), \quad (21)$$

где

A_t — истинное значение в момент t ,

F_t — прогноз в момент t ,

1 — индикатор,

sgn — функция сигнум.

В отличие от MAPE эта метрика может быть более полезной в задачах прогнозирования экономических показателей, так как зачастую более важным является предугадать направление динамики показателя, а не его точное значение. Чем больше значение метрики, тем лучше прогноз.

Для упрощения процесса выбора лучшей модели на основе пересчитанных метрик была введена композитная метрика PROD, одновременно учитывающая величину абсолютной ошибки и точность предсказания направления:

$$PROD = \frac{(100 - MAPE) \cdot MDA}{100}, \quad (22)$$

где $MAPE$ — средняя абсолютная процентная ошибка, MDA — средняя точность направления.

3.4. Результаты эмпирического исследования

Прогнозная сила моделей была измерена с помощью метрик MAPE и MDA. В табл. 8 представлены метрики одномерных моделей N-HITS и ARIMA, построенных на первом этапе для группы индексов промышленных материалов, включающей в себя индексы производства кокса, нефтепродуктов и ядерных материалов; резиновых и пластмассовых изделий; химического производства.

При прогнозировании единственного индекса производства эконометрические модели оказываются лучше по величине абсолютной ошибки. Однако нейросетевая модель чаще предсказывает правильно направление динамики показателей.

Далее был построен многомерный прогноз для той же группы индексов производства с помощью многомерной модели N-HITS и VAR. Метрики MAPE и MDA, характеризующие прогнозную силу многомерных моделей представлены в табл. 9.

При одновременном прогнозировании 3 показателей по величине абсолютной ошибки лучшими также оказались эконометрические модели (VAR). В половине случаев нейросетевые модели лучше предсказывают направление динамики показателей.

Для выделения лучших результатов в рамках всех подходов метрики одномерных и многомерных моделей были сведены в единую табл. 10.

Среди всего набора моделей по точности предсказания направления лучшими являются:

Таблица 8. Метрики одномерных прогнозных моделей

	MAPE, %		MDA, %	
	N-HITS	ARIMA	N-HITS	ARIMA
	без новостного индекса		без новостного индекса	
Производство нефтепродуктов	4,92	3,06	55	20
Производство резины	3,25	2,24	40	35
Химическое производство	3,13	1,93	40	60
	новостной индекс BERT		новостной индекс BERT	
Производство нефтепродуктов	5,33	3,15	50	30
Производство резины	2,99	2,71	45	45
Химическое производство	3,69	1,88	55	55
	новостной индекс LLaMA		новостной индекс LLaMA	
Производство нефтепродуктов	5,02	3,08	60	25
Производство резины	2,86	2	50	40
Химическое производство	3,33	1,86	75	60
	новостной индекс DeepSeek		новостной индекс DeepSeek	
Производство нефтепродуктов	4,45	3,37	60	40
Производство резины	2,6	2,11	40	45
Химическое производство	3,63	1,93	65	60

ПРИМЕЧАНИЕ: Темно-серым и серым отмечено лучшее значение метрики MAPE/MDA среди моделей N-HITS и ARIMA.

Таблица 9. Метрики многомерных прогнозных моделей

	MAPE, %		MDA, %	
	N-HITS	VAR	N-HITS	VAR
	без новостного индекса		без новостного индекса	
Производство нефтепродуктов	3,88	2,77	40	45
Производство резины	2,52	1,59	65	60
Химическое производство	2,27	1,88	60	55
	новостной индекс BERT		новостной индекс BERT	
Производство нефтепродуктов	4,54	2,92	40	50
Производство резины	2,97	1,97	45	45
Химическое производство	2,47	2,03	50	40
	новостной индекс LLaMA		новостной индекс LLaMA	
Производство нефтепродуктов	4,35	2,71	45	50
Производство резины	3,39	1,54	45	65
Химическое производство	2,88	1,89	65	55
	новостной индекс DeepSeek		новостной индекс DeepSeek	
Производство нефтепродуктов	4,66	2,9	40	50
Производство резины	3,28	1,63	55	55
Химическое производство	3,99	1,8	70	50

ПРИМЕЧАНИЕ: темно-серым и серым отмечено лучшее значение метрики MAPE/MDA среди моделей N-HITS и VAR.

Таблица 10. Сравнение прогнозной способности одномерных и многомерных моделей

	MAPE, %				MDA, %			
	N-HITS-1	ARIMA	N-HITS-3	VAR	N-HITS-1	ARIMA	N-HITS-3	VAR
	без новостного индекса				без новостного индекса			
Производство нефтепродуктов	4,92	3,06	3,88	2,77	55	20	40	45
Производство резины	3,25	2,24	2,52	1,59	40	35	65	60
Химическое производство	3,13	1,93	2,27	1,88	40	60	60	55
	новостной индекс BERT				новостной индекс BERT			
Производство нефтепродуктов	5,33	3,15	4,54	2,92	50	30	40	50
Производство резины	2,99	2,71	2,97	1,97	45	45	45	45
Химическое производство	3,69	1,88	2,47	2,03	55	55	50	40
	новостной индекс LLaMA				новостной индекс LLaMA			
Производство нефтепродуктов	5,02	3,08	4,35	2,71	60	25	45	50
Производство резины	2,86	2	3,39	1,54	50	40	45	65
Химическое производство	3,33	1,86	2,88	1,89	75	60	65	55
	новостной индекс DeepSeek				новостной индекс DeepSeek			
Производство нефтепродуктов	4,45	3,37	4,66	2,9	60	40	40	50
Производство резины	2,6	2,11	3,28	1,63	40	45	55	55
Химическое производство	3,63	1,93	3,99	1,8	65	60	70	50

ПРИМЕЧАНИЕ: Полужирным шрифтом и курсивом отмечено лучшее значение метрики MAPE/MDA среди моделей N-HITS-1, ARIMA, N-HITS-3 и VAR.

- одномерная модель N-HITS с новостными индексами LLaMA или DeepSeek для предсказания производства кокса и нефтепродуктов (доля правильных предсказаний направлений — 60%);
- многомерная модель N-HITS без новостного индекса или VAR с новостным индексом LLaMA для предсказания производства резины (доля правильных предсказаний направлений — 65%);

Таблица 11. Метрики лучших моделей и наивного прогноза для рассматриваемых индексов промышленного производства

1. Неметаллическое производство	ARIMA	DS RER	3,69	65	62,6	2,54	50	48,73
2. Производство машин	ARIMA	REL INDIVIDUAL	7,48	55	50,88	12,07	35	30,78
3. Текстильное производство	ARIMA	REL INDIVIDUAL	3,49	60	57,9	3,18	5	4,84
4. Пищевое производство	NHITS-1	DS RAG REL	2,89	65	63,12	2,2	25	24,45
5. Производство кожи	NHITS-1	DEEP	9,86	60	54,08	7,4	20	18,52
		SEEK						
6. Производство нефтепродуктов	NHITS-1	DS RAG	4,15	65	62,31	2,92	50	48,54
7. Производство резины	NHITS-1	DS RAG REL	5,46	75	70,91	2,05	45	44,08
8. Производство электрооборудования	NHITS-1	DS SUM REL	8,89	70	63,78	10,73	15	13,39
9. Металлургическое производство	NHITS-3	DS RER	3,8	85	81,77	3,52	40	38,59
10. Обработка древесины	NHITS-3	DS RER REL	4,99	70	66,51	4,33	40	38,27
11. Производство транспорта	NHITS-3	BERT	8,4	70	64,12	8,77	35	31,93
12. Химическое производство	NHITS-3	REL INDIVIDUAL	4,15	65	62,3	2,26	25	24,43
13. Целлюлозное производство	NHITS-3	DS SUM	4,07	70	67,15	3,43	40	38,63

— одномерная модель N-HITS с новостным индексом LLaMA для химического производства (доля правильных предсказаний направлений — 75%).

Среди всего набора моделей по величине абсолютной ошибки лучшими являются:

— VAR с новостным индексом LLaMA для прогнозирования производства кокса и нефтепродуктов (ошибка составила 2,71%) и производства резины (ошибка составила 1,54%);

— VAR с новостным индексом DeepSeek для прогнозирования химического производства (ошибка составила 1,8%).

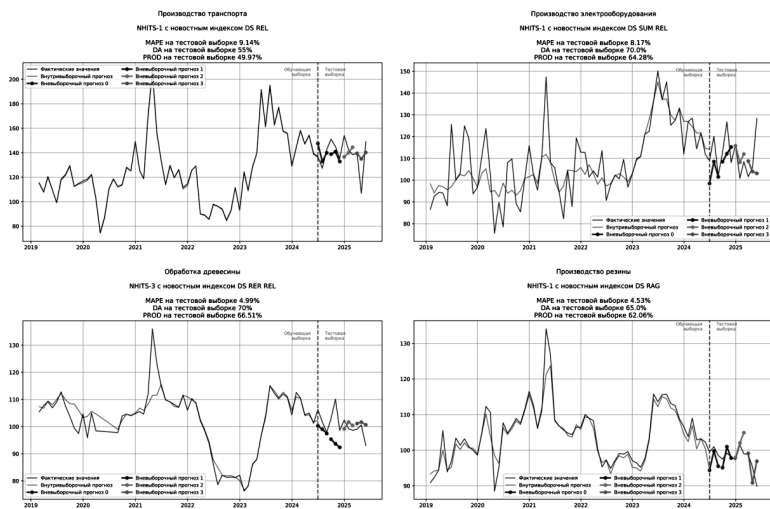


Рис. 6. Внутри- и вневыборочные прогнозы индексов производства

На следующем этапе при построении прогнозов для 13 индексов промышленного производства, объединенных в группы, были использованы модели ARIMA, одномерные и многомерные модели N-HITS. Также в качестве новостных индексов были добавлены новые новостные индексы, как общие для всех показателей, так и индивидуальные для каждой отрасли.

Для оценки качества моделей метрики MAPE, MDA и значение композитной метрики PROD лучших моделей сравнивались с метриками наивного прогноза (прогноз предыдущей точки). Модели и новостные индексы, с помощью которых были получены лучшие результаты, представлены в табл. 11. Построенные одномерные и многомерные модели с использованием новостных индексов превосходят качество наивного прогноза по композитной метрике PROD, которая одновременно учитывает величину абсолютной ошибки и точность предсказания направления.

Для всех рассматриваемых отраслей промышленного производства модель, использующая в качестве экзогенной переменной один из новостных индексов, оказывается лучшей. К тому же

новостные индексы, разработанные с применением более сложных моделей или с использованием отбора релевантной информации, превосходят базовый подход с использованием модели BERT. Для 10 из 13 индексов промышленного производства лучшее качество прогнозов было получено с помощью одномерной или многомерной нейросетевой модели.

Графики внутри- и вневыборочных прогнозов некоторых моделей продемонстрированы на рис. 6.

Выводы

Для решения основной исследовательской задачи, направленной на улучшение качества краткосрочных прогнозов отраслевых показателей индекса промышленного производства (ИПП) на горизонте до трех месяцев вперед, в ходе проведенного исследования были выполнены следующие ключевые этапы:

- разработка новостных индексов¹⁷ с использованием классического подхода (нейросетей трансформеров RuBERT) и больших языковых моделей с механизмом дополнения контекста (LLM, LLM+RAG);
- анализ способности LLM фиксировать экономический смысл новостей глубже, чем классические модели определения тональности;
- сравнительный анализ многофакторных эконометрических и нейросетевых моделей прогнозирования (ARIMAX, NHITS) с их расширенными версиями, включающими новостные индексы.

Основная исследовательская гипотеза заключалась в следующем: современные языковые модели позволяют извлекать из новостных текстов более информативные показатели тональности, способствующие повышению точности прогнозов индексов промышленного производства по сравнению с традиционными подходами. Для проверки данной гипотезы в рамках исследова-

¹⁷ На основе новостей ТАСС, The Bell, а также пресс-релизов ЦБ. Корпус новостей охватывает период с 2015 г. по август 2025 г. и содержит ~1,6 млн новостей.

ния были реализованы три класса методов оценки тональности, на основе которых формировались соответствующие новостные индексы:

- Классический. В его рамках использовалась предобученная модель RuBERT, дообученная на размеченных по тональности русскоязычных экономических новостях. Преимущества: высокая интерпретируемость, возможность тонкой настройки, низкие требования к ресурсам. Ограничения: фиксированная архитектура, отсутствие гибкой адаптации под разные макроэкономические показатели без повторного обучения.
- Большая языковая модель (LLM). Использовалась Llama 3.1-8B в режиме few-shot (техника дообучения модели без изменения весов). Ключевые особенности: возможность формулировать задания на естественном языке («определи sentiment новости относительно влияния на ИПП»), отсутствие необходимости обучать модель, более глубокое улавливание контекста и экономического смысла текста.
- Рассуждающая LLM (DeepSeek R1) с дополненным контекстом (LLM + RAG). Для каждой новости система RAG подбирала семантически и лексически близкие тексты (новости и пресс-релизы ЦБ). Эти дополнительные тексты включались в промпт, что существенно увеличивало качество интерпретации LLM. Также были реализованы дополнительные механизмы улучшения качества контекста:
- Relating-модуль, определяющий релевантность новости для сектора экономики.
- Rerank-модуль, осуществляющий повторное ранжирование найденных в RAG новостей с использованием более строгих семантических критериев и тем самым исключая шумовые новости контекста.
- Модуль суммаризации, формирующий компактное и информативное резюме найденных новостей, позволяя LLM учитывать полный контекст без превышения ограничений по длине входной последовательности при оценке тональности определенной новости.

По результатам эмпирического тестирования модели, включающие новостные индексы, продемонстрировали более высокое

качество прогнозирования по всем рассматриваемым индексам промышленного производства по сравнению с моделями без новостных переменных при прочих равных условиях, что подтверждается результатами теста Диболда—Мариано и снижением значения MAPE на тестовой выборке. В частности, использование новостных индексов снижает MAPE от 0,01 п.п. до 4,41 п.п. и повышает MDA (процент верно предсказанных направлений изменения показателя) от 5 п.п. до 30 п.п.

При этом классические подходы к определению тональности — RuBERT и RuBERT SetFit в основе новостных индексов улучшают прогноз, но уступают LLM и особенно LLM+RAG. Чистые LLM дают стабильное качество, но без расширенного контекста проигрывают RAG. Наиболее значимый эффект — при RAG-обогащении запроса, когда LLM получает больше экономического контекста с поправкой на Relating (релевантность новости для сектора). В части рассматриваемых моделей непосредственно для прогнозирования для 9 из 13 индексов производства лучшие метрики показал нейросетевой подход (модель Nhits).

Также важно отметить, что, согласно анализу предметной области, в российской научной литературе подобные подходы пока практически не представлены, тогда как в зарубежных исследованиях они активно развиваются. Таким образом, работа вносит вклад в адаптацию и развитие современных LLM-методов в отечественной практике макроэкономического прогнозирования.

Результаты проведенного исследования позволяют сделать следующие выводы:

- Новости содержат опережающую информацию для прогнозов индексов промышленного производства, а машинные методы позволяют количественно извлечь этот сигнал. При этом подходы на основе больших языковых моделей существенно расширяют возможности анализа, поскольку способны улавливать сложные экономические взаимосвязи, которые остаются недоступными для классических моделей сентимента.
- RAG-система, обогащающая новости дополнительными текстами, играет ключевую роль: она расширяет возможности «понимания» моделью контекста, что в итоге значительно улучшает прогноз.

- Подходы на основе LLM+RAG — новый этап развития методов оценки тональности и построения опережающих индикаторов в экономике.

Полученные результаты показывают, что современные LLM и особенно LLM+RAG позволяют перейти от поверхностного лексического анализа новостей к семантическому и экономически осмысленному анализу. Это открывает возможность создания нового поколения новостных опережающих индикаторов, способных улучшать прогнозы индексов промышленного производства.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Hochreiter S., Schmidhuber J. Long short-term memory // Neural computation. 1997. Vol. 9. No. 8. С. 1735–1780.
2. Olah C. Understanding LSTM Networks [Электронный ресурс] // colah's blog. 2015. URL: [http:// colah.github.io/posts/2015-08-Understanding-LSTMs](http://colah.github.io/posts/2015-08-Understanding-LSTMs)
3. Challu C. et al. Nhits: Neural hierarchical interpolation for time series forecasting // Proceedings of the AAAI conference on artificial intelligence. 2023. Vol. 37. No. 6. С. 6989–6997.
4. Costola M. et al. Machine learning sentiment analysis, COVID-19 news and stock market reactions // Research in International Business and Finance. 2023. Vol. 64. P. 101881. DOI: 10.1016/j.ribaf.2023.101881
5. Fazlija B., Harder P. Using financial news sentiment for stock price direction prediction // Mathematics. 2022. Vol. 10. No. 13. С. 2156. DOI: 10.3390/math10132156
6. Kim J., Kim H. S., Choi S. Y. Forecasting the S&P 500 index using mathematical-based sentiment analysis and deep learning models: a FinBERT transformer model and LSTM // Axioms. 2023. Vol. 12. No. 9. P. 835. DOI: 10.3390/axioms12090835
7. Li X., Tang Z. Sentiment Analysis on Inflation after COVID-19 // Applied Economics and Finance. 2023. Vol. 10. No. 1. P. 1023–1023. DOI: 10.11114/aef.v10i1.5821
8. Kuratov Y., Arkhipov M. Adaptation of deep bidirectional multilingual transformers for Russian language // arXiv preprint. 2019. DOI: 10.48550/arXiv.1905.07213
9. Smetanin S., Komarov M. Deep transfer learning baselines for sentiment analysis in Russian // Information Processing & Management. 2021. Vol. 58. No. 3. P. 102484. DOI: 10.1016/j.ipm.2020.102484
10. Saha P., Reimer J., Byrns B., Burbridge C., Dhar N., Chen J., Rayan S., Broderick G. Reconstructing Biological Pathways by Applying Selective Incremental Learning to (Very) Small Language Models [Электронный ресурс] // arXiv. 2025. URL: [https:// arxiv.org/abs/2507.04432](https://arxiv.org/abs/2507.04432).
11. Zhang Y. Interpretable Machine Learning for Macro Alpha: A News Sentiment Case Study [Электронный ресурс] // arXiv. 2025. URL: [https:// arxiv.org/abs/2505.16136](https://arxiv.org/abs/2505.16136)

12. Zakir U., Daykin E., Diagne A., Faile J. Advanced Deep Learning Techniques for Analyzing Earnings Call Transcripts: Methodologies and Applications [Электронный ресурс]//arXiv. 2025. URL: <https://arxiv.org/abs/2503.01886>
13. Kirtac K., Germano G. Sentiment trading with large language models// Finance Research Letters. 2024. Vol. 62. P. 105227. DOI: [10.1016/j.frl.2024.105227]. URL: <https://doi.org/10.1016/j.frl.2024.105227>
14. Zafar M.B. FinAI-BERT: A Transformer-Based Model for Sentence-Level Detection of AI Disclosures in Financial Reports [Электронный ресурс]//arXiv. 2025. URL: <https://arxiv.org/abs/2507.01991>
15. Lim B., Arik S. O., Loeff N., Pfister T. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting [Электронный ресурс]//arXiv. 2020. URL: <https://arxiv.org/abs/1912.09363>
16. Sun Y., Yuan H., Xu F. Financial sentiment analysis for pre-trained language models incorporating dictionary knowledge and neutral features//Natural Language Processing Journal. 2025. Vol. 11. Art. 100148. ISSN 2949-7191. DOI: [10.1016/j.nlp.2025.100148]. URL: <https://www.sciencedirect.com/science/article/pii/S294971912500024X>
17. Бондаренко В. И., Елисеев В. О., Ермоленко Т. В. Анализ эффективности глубоких языковых моделей для задачи определения тональности русскоязычного текста//Проблемы искусственного интеллекта. 2024. No. 1 С. 32.
18. Sun X. et al. Sentiment analysis through llm negotiations//arXiv preprint. 2023. DOI: 10.48550/arXiv.2311.01876
19. Thomas G. J. Enhancing TinyBERT for Financial Sentiment Analysis Using GPT-Augmented FinBERT Distillation //arXiv preprint. 2024. DOI: 10.48550/arXiv.2409.18999
20. Zhang B. et al. Enhancing financial sentiment analysis via retrieval augmented large language models//Proceedings of the fourth ACM international conference on AI in finance. 2023. P. 349–356. DOI: 10.1145/3604237.362686
21. Muhammad I., Rospocher M. On Assessing the Performance of LLMs for Target-Level Sentiment Analysis in Financial News Headlines//Algorithms. 2025. Vol. 18. No. 1. Art. 46. ISSN 1999-4893. DOI: 10.3390/a18010046. URL: <https://www.mdpi.com/1999-4893/18/1/46>
22. Mun Y., Kim N. Leveraging Large Language Models for Sentiment Analysis and Investment Strategy Development in Financial Markets// Journal of Theoretical and Applied Electronic Commerce Research. 2025. Vol. 20, no. 2. Art. 77. ISSN 0718-1876. DOI: 10.3390/jtaer20020077. URL: <https://www.mdpi.com/0718-1876/20/2/77>

23. Nasiopoulos D. K., Roumeliotis K. I., Sakas D. P., Toudas K., Reklitis P. Financial Sentiment Analysis and Classification: A Comparative Study of Fine-Tuned Deep Learning Models // International Journal of Financial Studies. 2025. Vol. 13, no. 2. Art. 75. ISSN 2227-7072. DOI: 10.3390/ijfs13020075. URL: <https://www.mdpi.com/2227-7072/13/2/75>
24. Fatouros G., Soldatos J., Kouroumali K., Makridis G., Kyriazis D. Transforming sentiment analysis in the financial domain with ChatGPT // Machine Learning with Applications. 2023. Vol. 14. Article 100508. ISSN 2666-8270. DOI: 10.1016/j.mlwa.2023.100508. URL: <https://www.sciencedirect.com/science/article/pii/S2666827023000610>
25. Breitung C., Kruthof G., Müller S. Contextualized Sentiment Analysis using Large Language Models [Электронный ресурс]. 2023. URL: <https://ssrn.com/abstract=4615038>

Научный труд № 186

**Макроэкономическая политика
и методы экономического прогнозирования**

Выпускающий редактор *Елена Попова*

Редактор *Ольга Черкасова*

Верстка *Ярослав Агеев*

Подписано в печать 22.10.2026.

Тираж 300 экз. 1-й завод 70 экз.

125993 г. Москва,

Газетный пер., д. 3–5, стр. 1

Тел. (495) 629-6736, fax (495) 691-3594

E-mail: info@iep.ru

www.iep.ru



ИНСТИТУТ ЭКОНОМИЧЕСКОЙ ПОЛИТИКИ имени Е. Т. ГАЙДАРА

**ИНСТИТУТ ГАЙДАРА — один из ведущих российских
научно-исследовательских и учебно-методических центров
в области экономических наук**

Миссия Института — содействие долгосрочному устойчивому развитию
экономики, разработке и продвижению научно обоснованных подходов
к формированию экономической политики

НАПРАВЛЕНИЯ ДЕЯТЕЛЬНОСТИ



■ НАУКА И КОНСАЛТИНГ

- Макроэкономика и финансы
 - Реальный сектор
 - Политическая экономия и институциональное развитие
- Интернет-сайт: www.iep.ru



■ ОБРАЗОВАТЕЛЬНАЯ ДЕЯТЕЛЬНОСТЬ

Аспирантура Института Гайдара создана в 1998 году



■ НАУЧНО-ПРОСВЕТИТЕЛЬСКАЯ ДЕЯТЕЛЬНОСТЬ

Телеграм-канал Института Гайдара: t.me/institute_gaidara



■ ИЗДАТЕЛЬСКАЯ ДЕЯТЕЛЬНОСТЬ

Издательство Института Гайдара основано в 2010 году

**Присоединяйтесь к нам, чтобы расширить свои горизонты
в области экономической науки и политики**

**Институт Гайдара — ваш партнер
в мире экономических знаний!**