ISSN 1813-8691 (Print)
ISSN 1813-8705 (Online)

ЭКОНОМИЧЕСКИЙ ЖУРНАЛ

Высшей школы экономики

Higher School of Economics Economic Journal

Том 29, № 1 2025

В НОМЕРЕ:

ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

ЭКОНОМИЧЕСКИЙ ЖУРНАЛ ВШЭ

Том 29, № 1 ***2025***

«Экономический журнал ВШЭ» является рецензируемым научным журналом, публикующим статьи на русском и английском языках. Со дня своего основания в 1997 г. «Экономический журнал ВШЭ» стремится к более глубокому пониманию рыночной, и особенно российской, экономики. Журнал публикует наиболее интересные работы в различных областях экономической теории и практики, по экономико-математическому моделированию и прикладным методам исследований. Редколлегия и Совет журнала состоят из известных российских и иностранных ученых, деятельность которых способствует интеграции мирового научного сообщества. Целевой аудиторией журнала являются исследователи, преподаватели вузов, аспиранты и студенты.

Принимаются статьи, соответствующие рубрикатору Journal of Economic Literature, посвященные исследованию широкого спектра вопросов, касающихся как мировой экономики в целом, так и отдельных стран, в области микро- или макроэкономики, экономической политики, эконометрики, рынка труда, социальной политики и др.

Наряду с содействием научным исследованиям, миссия журнала заключается в поиске новых авторов и новых научных идей.

Непременным условием для публикации является анонимное рецензирование представленных статей.

Журнал стремится соблюдать высокие стандарты публикационной этики.

Журнал зарегистрирован 21 июля 2016 г. Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций.
Свидетельство о регистрации средства массовой информации ПИ № ФС 77-66577

Журнал входит в перечень ведущих рецензируемых научных журналов и изданий ВАК, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени доктора и кандидата наук, и в международную базу данных публикаций на базе Web of Science – Russian Science Citation Index (RSCI), индексируется реферативной базой данных Scopus

Распространяется по подписке:
Объединенный каталог «Пресса России» – подписной индекс 79264

Адрес редакции:
101000, Москва, ул. Мясницкая, 20
Телефон: (495) 772-95-90 доб. 11874
e-mail: redact@hse.ru

Полнотекстовая электронная версия журнала доступна по адресу:
http://ej.hse.ru
Плата с авторов за публикацию рукописей не взимается
Рукописи не возвращаются

Адрес издателя и распространителя:
101000, Москва, ул. Мясницкая, 20
Издательский дом Высшей школы экономики
Телефон: (495) 772-95-90, доб. 15298, e-mail: id@ hse.ru

В НОМЕРЕ:

Стр.

HSE Economic Journal

Volume 29, No 1 **2025**

The HSE Economic Journal publishes refereed papers both in Russian and English. It has perceived better understanding of the market economy, the Russian one in particular, since being established in 1997. It disseminated new and diverse ideas on economic theory and practice, economic modeling, applied mathematical and statistical methods. Its Editorial Board and Council consist of prominent Russian and foreign researchers whose activity has fostered integration of the world scientific community. The target audience comprises researches, university professors and graduate students.

Submitted papers should match JEL classification and can cover country specific or international economic issues, in various areas, such as micro- and macroeconomics, econometrics, economic policy, labor markets, social policy.

Apart from supporting high quality economic research and academic discussion the Editorial Board sees its mission in searching for the new authors with original ideas.

The journal follows international reviewing practices – at present submitted papers are subject to single blind review of two reviewers.

The journal stands for meeting the highest standards of publication ethics.

The journal is registered July 21, 2016 in the Federal Service for Supervision in the Area of Telecom, Information Technologies and Mass Communications.
Certificate of registration of mass media PI No. FS 77-66577

The journal is indexed by the Higher Attestation Commission (VAK), Web of Science – Russian Science Citation Index (RSCI), Scopus.

Distributed by subscription:
United catalog «Russian Press» – subscription index 79264

Address:
20, Myasnitskaya str.,
Moscow, 101000, Russian Federation
*Tel.: +7 (495) 772-95-90 *11874*
E-mail: redact@hse.ru

Full-text electronic version of the journal is available at:
http://ej.hse.ru/en/

Publisher and Distributor Address:
20, Myasnitskaya str.,
Moscow, 101000, Russian Federation
Publishing House of the Higher School of Economics
Tel.: +7 (495) 772-95-90, ext. 15298, E-mail: id @hse.ru

Contens:

Page

УДК 339.5, 330.3

DOI: 10.17323/1813-8691-2025-29-1-9-41

Macroeconomic Analysis of the Impact of Economic Complexity on Income Inequality: Does Institutional Quality Matter?[1]

Natalia Davidson[1], Ekaterina Magon[2], Oleg Mariev[3]

[1] Ural Federal University,
19, Mira str., Sverdlovsk Region, Ekaterinburg, 620002, Russian Federation.
E-mail: natalya.davidson@gmail.com

[2] Ural Federal University,
19, Mira str., Sverdlovsk Region, Ekaterinburg, 620002, Russian Federation.
E-mail: Volkova2016consta@gmail.com

[3] Ural Federal University,
19, Mira str., Sverdlovsk Region, Ekaterinburg, 620002, Russian Federation.
E-mail: o.s.mariev@urfu.ru

This paper sheds light on the relationship between economic complexity and income inequality considering the role of institutions based on data over the period 1996–2020 across 52 developed and developing countries from Europe and Central Asia, and the Middle East and North Africa. Our contribution to the existing literature is twofold. First, we analyse the relationship between economic complexity and income inequality considering the institutional dimension and studying various components of institutions. Second, we take into account the non-linear form of relationship between economic complexity and income inequality, as well as heterogeneity of this relationship across groups of countries. We address endogeneity by employing a fixed effect two stage least squares model with instrumental variables. Our results demonstrate that for the overall sample of countries, an increase in a country's economic complexity results in higher level of income inequality. However, the impact of economic complexity on income inequality is heterogene-

[1] The study was supported by the grant of the Russian Science Foundation (Code: 23-18-01065).

Natalia Davidson – Associate Professor, Graduate School of Economics and Management.
Ekaterina Magon – Student, Graduate School of Economics and Management.
Oleg Mariev – Associate Professor, PhD, Graduate School of Economics and Management.

The article was received: 05.11.2024/The article is accepted for publication: 06.02.2025.

ous across groups of countries, with a U-inverted relationship in countries of Europe and Central Asia. Moreover, economic complexity combined with the high level of institutional quality can reduce income inequality. Therefore, we conclude that the improvement of all components of institutional structure will facilitate a decrease in income disparities. Our analysis shows that better educational level leads to lower income inequality. Besides, our findings emphasise the need for policy ensuring more equal gains from economic development and international trade.

***Key words*:** economic complexity; income inequality; institutional quality; economic development; instrumental variables (IV) estimation; economic policy.

JEL Classification: E02, O15, O33.

For citation: Davidson N., Magon E., Mariev O. Macroeconomic Analysis of the Impact of Economic Complexity on Income Inequality: Does Institutional Quality Matter? *HSE Economic Journal.* 2025; 29(1): 9–41.

1. Introduction

Nowadays the major challenges to global and national socio-economic development pivot on achieving sustainable economic growth and reducing income inequality, but inequalities within most countries have been significantly deepening and widening [Chancel et al., 2022]. Economic growth is still perceived as the fundamental determinant of inequality reduction, as enshrined by Kuznets (1955), who suggested that economic growth initially causes increasing inequality, which eventually evens out into lower income inequality. However, questioning the conclusions of Kuznets and subsequent works, Piketty (2014) proposed another view on the link between economic growth and inequality, claiming that when an economy reaches higher income levels, inequality tends to continue to increase, and the latter paradigm is supported by trends in income inequality dynamics. In fact, inequalities have increased consistently in advanced economies over the last decades [Nolan et al., 2019; Malla, Pathranarakul, 2022]. Such outcomes also run contrary to the view that advanced economies tend to have a higher *economic complexity* (i.e., economic diversification and sophistication of productive capabilities), which can potentially improve the standard of living for all social strata [Hidalgo, Hausman, 2009].

Therefore, the aim of this paper is to shed light on the relationship between economic complexity and income inequality considering the role of *institutions* (i.e., institutional quality measured as an average of the World Governance Indicators which include Voice and Accountability, Political Stability and Absence of Violence/Terrorism, Government Effectiveness, Regulatory Quality, Rule of Law, Control of Corruption. The Institutional quality variable ranges from –2 to 2, with 2 indicating strong governance performance.

The issue of rising income inequality is important for several reasons, including that it can undermine socio-economic gains that form the political and ontological basis for economic growth at the societal level. Persistently high inequality is related to lower and less durable economic growth in the long run, as well as to potential food insecurity and political instability [Alesina, Rodrik, 1994; Lakner, Milanović, 2016]. Evidence suggests that inequality can dampen economic growth by restraining investments and consumption [Acemoglu et al., 2012; Carvalho, Rezai, 2014;

Kumhof et al., 2015]. Countries with high inequality levels are vulnerable to economic, financial and political instability, and are less resistant to crises [Rajan, 2010; Cingano, 2014; Berg, Ostry, 2017]. Moreover, rising inequality is also fraught with fall in human capital investment and decrease in innovative activities [Topuz, 2022].

High inequality is also associated with unbalanced redistribution of economic gains among individuals, rent-seeking behaviour, and excessive concentration of resources [Mihályi, Szelényi, 2019]. Therefore, while shifting to a sustainable and inclusive economic growth paradigm [Zhu, 2022], it is essential to analyse factors determining inequality as well as develop policies to handle the increasing rates of income disparities and ensure prosperity of societies. In this framework, the concept of *economic complexity*, as a novel perspective on socio-economic development, gave rise to a burgeoning line of income inequality studies. However, literature on *the link between economic complexity and inequality* is currently limited to a dozen works with contradictory results. There are four scenarios documented in the existing literature: a negative relationship [Hartmann et al., 2017; Lee, Vu, 2019], a positive relationship [Lee, Vu, 2019; Chu, Hoang, 2020; Sepehrdoust et al., 2021]; a non-linear U-inverted relationship [Sbardella et al., 2017; Chu, Hoang, 2020; Zhu et al., 2020; Morais et al., 2021; Amarante et al., 2023; Nguyen et al., 2023]; and a non-linear U-shaped relationship [Nguyen et al., 2023; Pham et al., 2023].

The relationship between economic complexity and income inequality was found to be non-homogenous across countries with different income and development levels by many recent investigations [Lee, Vu, 2019; Chu, Hoang, 2020; Amarante et al., 2023; Nguyen et al., 2023]. Moreover, using interaction terms, authors explored factors that can mediate the impact of economic complexity on inequality, among which they cited education, government spending, institutions, and trade openness [Lee, Vu, 2019; Chu, Hoang, 2020]. Specifically, there is evidence that government policies and institutions can affect the nature of the economic complexity and inequality nexus [Chu, Hoang, 2020].

Institutions tend to co-evolve with the sophistication of an economic system [Hidalgo, Hausmann, 2009; Hartmann et al., 2017; Vu, 2022], as transformation and improvement of institutions is essential to ensure the functioning of the economic system in general and the redistribution of income in particular. Gaps in governance, weakness, or a lack of inclusive institutions can exacerbate inequalities while a country experiences economic growth. This is a consequence of the accumulation and use by a few corporations (and even individuals) of new productive capabilities, diversification, and sophistication of production, growing competitiveness, and strengthening of the country's position in the world market [Acemoglu et al., 2005; Balland et al., 2022].

Overall, theoretical predictions suggest a potentially powerful role for institutions, but in praxis their instrumentality in the relationship between economic complexity and income inequality has not received commensurate attention. Our work is aimed at filling this gap by considering the role of institutions in the nexus of economic complexity and income inequality. Our contribution to the existing literature is twofold. First, we analyse the relationship between economic complexity and income inequality considering the institutional dimension and studying various components of institutions. Second, we address the non-linear form of relationship between economic complexity and income inequality, as well as heterogeneity of this relationship across groups of countries. Besides, we account for economic development and human capital levels, and countries' participation in international trade. In addition, we address the issue of endogeneity by employing a fixed effect two stage least squares (2SLS) model with instrumental variables (IVs).

This paper is structured as follows. Section 2 reviews previous findings on the relationship between economic complexity and income inequality. Section 3 describes the empirical modelling, economic complexity measurement, econometric methods, and data used in this paper. Section 4 presents the results of our analysis and discusses the main findings. Section 5 finalises our work and provides policy implications.

2. Literature Review

2.1. Economic complexity: theoretical and empirical background

Economic complexity refers to the diversity and sophistication of a country's productive capabilities, i.e., all inputs available to the country, such as technologies, unique productive know-how and ideas that allow an economy to produce a wide range of goods [Hidalgo, Hausman, 2009]. The more inputs a country has, the more diversified and sophisticated its production structure. Therefore, economic complexity can be conceptualized as a measure of the *knowledge accumulated in a society* expressed in the products it makes [Hausmann et al., 2011]. Moreover, knowledge, local and non-local, is essential for innovation, which, in turn, creates opportunities for economic diversification and complexity [Gao, Rai, 2023].

Economic complexity is an accurate predictor of economic growth [Hidalgo, Hausmann, 2009; Chávez et al., 2017; Tacchella et al., 2018; Bustos, Yıldırım, 2022]. The sophistication of the economy (i.e., increasing economic complexity) was traditionally assumed to be associated with a commensurate decrease in income inequality, as proclaimed by the pioneering study in this field by Hartmann et al. (2017). However, there is extensive empirical evidence questioning the supposed role of economic complexity in improved income distribution and reduced inequality problems, as mentioned previously. Nevertheless, studies have highlighted several mechanisms through which economic complexity can potentially improve the distribution of income in a country.

Firstly, complex economic systems are associated with better quality institutions and higher unionization, which tend to reduce income inequality. In fact, appropriate institutions can prevent increasing inequalities in the individuals' capabilities and skills, while strong unionization provides workers with more bargaining power enabling them to earn higher wages [Hartmann, 2014; Le Caous, Huarng, 2020].

Secondly, economic diversification broadens employment opportunities, because a complex economy, due to its specific demands, requires more labour with different skill levels, resulting in lower inequality [Hartmann, 2014].

Finally, individuals living in a complex economy possess an access to a greater diversity of skills and knowledge, and, more importantly, to a larger pool of social capital (social contacts, communities, and networks); while social capital is not captured by aggregate measures of human capital, it is assumed to intrinsically reduce income inequality [Caldarelli et al., 2012; Hartmann 2014]. Besides, a complex economy creates incentives for firms to search for new knowledge, and, therefore, to look for collaborations, which also enhances social links and provides more opportunities for employees [Bernal et al., 2022].

Therefore, this theoretical paradigm posits that structural changes in the economy, accompanied by sophistication and diversification, are not only able to foster economic growth *per se*, but also facilitate a reduction of inequality. However, recent studies of the link between economic complexity and inequality present contradictory or ambiguous results, and there is a need for more research exploring the relationships between associated variables in various socio-economic contexts worldwide.

2.2. The link between economic complexity and income inequality

The exploration of the link between economic complexity and income inequality resulted in the findings suggesting that increasing economic complexity can lead to lower inequality, as reported by Hartmann et al. (2017). Their study based on panel regression with country fixed effects revealed that countries with higher levels of economic complexity have lower levels of income inequality and tend to be more inclusive. Meanwhile, a country's productive structure is affected by the interaction of various factors, from the education level to the institutional quality, which co-evolve along with the country's mix of exported goods and the economy's inclusiveness.

However, subsequent research by Lee and Vu (2019) employing dynamic panel data analysis, namely a system GMM estimator, demonstrated a positive relationship between economic complexity and inequality that contradicts the results obtained by Hartmann et al. (2017). Lee and Vu (2019) argued that an increase in economic complexity (i.e., sophistication and diversification of production and export) is associated with higher income inequality in both the short and long term. Furthermore, the presence of a positive link was evidenced by more recent research [Chu, Hoang, 2020; Sepehrdoust et al., 2021].

Several studies point out non-linear relationships between economic complexity and income inequality and underscore several determinants of income inequality along with economic complexity. Chu and Hoang (2020), exploring a positive link between economic complexity and inequality, mentioned that this relationship is far more complicated than suggested by the binary of negative or positive. They observed that economic complexity can facilitate a decrease in inequality in countries endowed with better human capital, higher institutional quality, efficient public spending, and economic freedom. Conversely, in less favourable environments, it fails to reduce income inequality.

This controversial effect of economic complexity on income inequality implies the possibility of a non-linear relationship, which can be interpreted in the framework of the Kuznets (1955) curve hypothesis. This concept implies the adverse effects of structural changes on income distribution in the initial period of economic growth and, after a certain point, the levelling of income inequality by public finance investments, including public education, the social safety net, and health care. Specifically, economic complexity initially can increase income inequality, but once a country reaches a certain level of development or complexity, income inequality starts reducing [Chu, Hoang, 2020]. In line with this framework, current empirical studies reveal a U-inverted relationship between economic complexity and inequality [Sbardella et al., 2017; Chu, Hoang, 2020; Zhu et al., 2020; Morais et al., 2021; Amarante et al., 2023; Nguyen et al., 2023]. Although the existence of a threshold of economic complexity has been addressed in most recent studies, the specific characteristics of this threshold have not yet been extensively researched.

However, there is also evidence of a U-shaped relationship between economic complexity and inequality [Nguyen et al., 2023; Pham et al., 2023]. The U-inverted effect of economic complexity on inequality was discovered in non-high-income countries, whereas in high income countries an opposite effect was revealed (i.e., the latter exhibited a U-shaped curve); the U-shaped relationship was also reported for most countries in such regions as the Middle East and North Africa (MENA) and South Asia [Nguyen et al., 2023].

Overall, there is evidence that the economic complexity effect on inequality is not homgenous across countries [Sbardella et al., 2017; Lee, Vu, 2019; Amarante et al., 2023], which, in combination with mixed results, suggests the existence of a non-linear relationship.

Moreover, some studies using interaction terms revealed factors that can moderate the impact of increased economic complexity. For instance, better education and (*ipso facto*) developed human capital have the potential to enhance the negative correlation between economic complexity and inequality [Lee, Vu, 2019; Chu, Hoang, 2020]; effective government spending and trade openness have moderating impacts on the effect of economic complexity [Chu, Hoang, 2020].

Thus, we assume that the sophistication of the economic system, along with its positive effects, can bring the aggravation of inequality. However, these adverse effects could be mediated by institutional factors. For instance, when national institutions are inclusive, they ensure a fair distribution across society of wealth and gains from increasing economic complexity and consequent economic growth. Moreover, strong institutions along with factors like human capital can enhance positive impacts of economic sophistication thus decreasing inequality, and, at the same time, mitigate negative ones, preventing further exacerbation of income disparities and the existing socio-economic problems.

2.3. Determinants of income inequality on the regional and country level

Along with research based on the country-level data, a number of studies consider the nexus between economic complexity and income inequality on the regional level. For instance, Sbardella et al. (2017) found that wage inequality increases in US counties with growing economic complexity; however, the relation between economic complexity and income inequality has an inverted U-shaped pattern consistent with the Kuznets hypothesis. Morais et al. (2021) analysed Brazilian states and also documented the U-inverted shape of this relationship, having displayed that growing levels of economic complexity first worsen and then improve income distribution in Brazil, with a more distinct pattern in highly urbanized and developed states. Hence, a certain level of economic development must be achieved before the regional production structure begins to reduce income inequality.

The abovementioned relation between a higher economic complexity and urbanization level as well as overall regional development was supported by recent research on inequality in diverse economies, including China [Zhu et al., 2020] and Romania [Le Caous, Huarng, 2020]. In urban areas, a more sophisticated industrial structure provides a wider range of occupational opportunities and greater resilience to shocks. Besides, in urban areas, workers are more skilled and have more complex networks, which increases their bargaining power on the labour market and eventually reduces inequality. At the same time, rural areas suffer from higher income inequality, which is partly explained by inequalities in opportunities for education, training, and working, less developed infrastructure, and limited social networks.

The results of regional level studies are comparable to those of national-level research [Chu, Hoang, 2020; Sepehrdoust et al., 2021]: countries that are generally wealthier and endowed with more developed human capital, institutions, and economic freedom may be likely to exhibit a reduction in inequality when their economies become more complex and diversified, while countries with the opposite characteristics may not exhibit such trends. Such non-homogenous and uneven development is difficult to control due to the self-reinforcing nature of complexity [Balland et al., 2022].

2.4. Summary of economic complexity-income inequality relationships and research hypotheses

To summarize, during recent years four types of relationships were found in empirical research: positive and negative relationships, «U-inverted», and conversely «U-shaped» non-linear relationships.

The hypothesis of economic complexity as a negative predictor of income inequality [Hartmann et al., 2017] is supported by the difference in the knowledge diversification levels typical for high- and low-complexity economies. Developing highly sophisticated industries is impossible without a proper level of knowledge diversification. This, in turn, leads to a relatively flat occupational structure, broadly distributed knowledge, and a wide range of demanded skills, thus decreasing income inequality [Constantine, Khemraj, 2019].

In contrast, low-complexity economies producing simpler and more widespread products usually depend mainly on low-skilled labour. This happens because a low value-added production requires much less advanced technology, competence, or product knowledge [Sepehrdoust et al., 2021]; as a result, the range of available occupational opportunities is constrained. Consequently, only a limited group of people benefits economically from such a production structure. This leads to a significant income discrepancy when a small number of individuals receives the largest portion of generated income, while the most numerous middle- and low- income classes have low salaries and very few opportunities for moving up the social ladder [Chu, Hoang, 2020].

At the same time, the evidence about the positive link between economic complexity and income inequality is often explained by technological changes, for instance, by the skills-based technological change theory [Violante, 2008]. The shift in production technologies required for an economy to become more complex results in an increasing demand for skilled labour over unskilled labour; the demand for highly qualified personnel grows further with the emergence of new technologies, widening the skills and income gap.

Inequality seems to be associated with growth and concentration of economic complexity, since more complex systems, by their nature, tend to be more unequal. This happens since complex adaptive systems are characterized by the preferential attachment, self-reinforcing feedback loops, and other multiplicative processes that lead to increasing inequality [Balland et al., 2022]. The example of urban areas demonstrates that inequality also rises when some individuals, corporations, and even locations occupy privileged positions, accumulate benefits from the growing diversification and sophistication of an economy, and therefore have access to a more significant share of income redistribution associated with growing economic complexity [Sbardella et al., 2017; Zhu et al., 2020; Morais et al., 2021]. Others are less fortunate due to many factors, from being in the «wrong» place to having an unsuitable skill set or level.

In turn, technological gaps, and the monopolization of advanced knowledge by technological giants [Rikap, Lundvall, 2020; Feldman et al., 2021] can potentially explain why the increasing complexity of productive capabilities and economic diversification do not reduce inequality [Balland et al., 2022].

Based on the reviewed theoretical and empirical literature on economic complexity and income inequality, we propose the following hypotheses:

H1: *The effect of economic complexity on income inequality is non-linear.*

The rationale for this hypothesis is that income inequality behaves differently at various stages of economic diversification [Sbardella et al., 2017; Chu, Hoang, 2020; Zhu et al., 2020; Morais et al., 2021; Amarante et al., 2023; Nguyen et al., 2023; Pham et al., 2023].

H2: *Institutional quality has a mediating role in the relationship between economic complexity and income inequality.*

The basis for this hypothesis is that well-functioning institutions can strengthen positive effects of economic complexity and level out its negative consequences [Lee, Vu, 2019; Chu, Hoang, 2020].

3. Methodology and Data

3.1. Measurement of economic complexity

Firstly, the interpretation of the economic complexity index (ECI) is the following. The intuition behind the concept of economic complexity proposed by Hidalgo and Hausmann (2009) is that the productive capabilities of each economy can be characterized by the range of productive knowledge that it possesses and by the number of ways in which individual knowledge can be combined to produce various goods. In other words, every good produced by a country contains information about the knowledge used for its production. Therefore, by combining data on all the products created by the economy, we can assess its level of knowledge advancement and productive capabilities. Hidalgo and Hausmann (2009) emphasize that countries' productivity depends on the diversity of non-tradable capabilities, such as institutional and human capital characteristics, as well as their interactions, that determine economic complexity.

The ECI is calculated using export data that connects countries to the products in which they have revealed comparative advantages (RCA) [Hidalgo, Hausmann, 2009; Hausmann et al., 2011; Caldarelli et al., 2012; Kemp-Benedict, 2014; Hartmann et al., 2017; Lee, Vu, 2019]. Thus, it goes beyond the idea of production diversity associated with the range of productive knowledge described above. ECI reflects not only the *diversity* of a country's economy (the number of products it produces and exports) but also the *ubiquity* of products (other countries' ability to produce and export a particular type of a product). Hence, *complex economies* are diverse and can export products with low ubiquity, meaning knowledge-intensive (sophisticated) products that only a few diverse countries are able to export [Hartmann et al., 2017; Balland et al., 2022].

Secondly, the construction of ECI implies the following concepts and stages. Hidalgo and Hausmann (2009) proposed the Method of Reflections to reveal production capabilities of countries based on trade data by calculating ECI. The title of the method is due to a symmetric set of variables for countries and products (two types of nodes in the network) that it generates. This method brings information about the capabilities available in a country based on knowledge of the measurable capabilities required for producing a specific product.

This method was further elaborated and applied in numerous papers, including Hausmann et al. (2011) and Hartmann et al. (2017). The method involves the concept of revealed comparative advantage (RCA); RCA of a country *c* in a product *p* is the share of product *p* in the export of country c to the share of product p in world export [Hidalgo, Hausmann, 2009; Hartmann et al., 2017]. An RCA greater than 1 indicates that a country has a comparative advantage in a particular product, which means that the export of a product from a country is larger than what would be expected based on the size of the country's exports and the global market for the product.

The RCA is used to determine elements of a discrete matrix M_{cp}, equal to 1 if country *c* has a revealed comparative advantage in product *p* and 0 if it does not, as shown in Eq. 1:

$$(1) \qquad M_{cp} = \begin{cases} 1 \text{ if } RCA_{cp} \geq 1 \\ 0 \text{ if } RCA_{cp} < 1. \end{cases}$$

where M_{cp} is a matrix in which rows represent different countries and columns represent different products.

A country *c* is considered to be a significant exporter of a product *p* in world trade if its RCA is greater than 1. The matrix M_{cp} allows to define the diversity of a country (Eq. 2) and the ubiquity of a product (Eq. 3), respectively. The number of products exported by a country with comparative advantage, and the number of countries exporting a product with comparative advantage are described by the following equations:

$$(2) \qquad Diversity = k_{c,0} = \sum_p M_{cp}.$$

$$(3) \qquad Ubiquity = k_{p,0} = \sum_c M_{cp}.$$

where $k_{c,0}$ and $k_{p,0}$ stand for diversity and ubiquity (respectively), measured by summing the rows and columns of the matrix M_{cp}.

Furthermore, to generate a more accurate measure of economic complexity, these indicators are jointly corrected for each other. This adjustment is needed as a country may export a wide variety of goods because of its economic size, whereby the information on a country's capabilities contained in the diversity indicator could be biased [Lee, Vu, 2019]. The Method of Reflection iteratively calculates the mean value of the previous-level diversity and ubiquity [Hidalgo, Hausmann, 2009; Kemp-Benedict, 2014; Lee, Vu, 2019].

Therefore, the adjusted matrix connects countries exporting similar products, weighted by the inverse of the ubiquity of a product, discounting common products, and normalized by the diversity of a country (Eq. 4):

$$(4) \qquad \tilde{M}_{CC'} = \frac{1}{k_{c,0}} \sum_p \frac{M_{cp} M_{c'p}}{k_{p,0}}.$$

Lastly, the ECI is formulated as shown below (Eq. 5):

$$(5) \qquad ECI_c = \frac{\vec{K}_c - \langle \vec{K} \rangle}{std(\vec{K})}.$$

where $\langle \vec{K} \rangle$ represents an average, and $std(\vec{K})$ stands for a standard deviation of a K_c – the eigenvector of $\tilde{M}_{CC'}$, associated with the second largest eigenvalue [Hausmann et al., 2011; Caldarelli et. al., 2012; Kemp-Benedict, 2014; Hartmann et al., 2017].

Thus, the Method of Reflections breaks down a country's trade into separate industries and products, making it possible to analyse the relationships between individual units in the system and ultimately draw a conclusion about the overall complexity level of an economy [Kemp-Benedict, 2014].

The limitations of ECI are the following. First, product classification is quite detailed [Hidalgo, Hausmann, 2009], however, it might not cover all firms' activities limiting itself by the main activity specified by a firm. Secondly, not all economic activities are correctly registered, especially in developing countries, including some of the ECA and MENA countries in our sample. Third, although RCA index is quite informative, its classical version used by Hidalgo and Hausmann (2009), mentioned above, does not include countries' imports, unlike, for example, Lafay index, limiting the overall understanding of a country's role in international trade. Besides, Hidalgo and Hausmann (2009) point out that the method does not capture differences among countries in capabilities utilized in production, although it is able to capture the correspondence between variety of capabilities in a country (such as employment categories), and diversity and ubiquity of products. Finally, Kemp-Benedict (2014) points out that ECI measure suggested by Hidalgo and Hausmann (2009) provides more information on export basket of a country than on a country's diversity of export.

3.2. Data

This work employs data over the period 1996-2020 across 52 developed and developing countries from Europe and Central Asia (ECA), the Middle East and North Africa (MENA). The list of countries in the sample is presented in Appendix A. The data on income inequality is obtained from the Standardized World Income Inequality Database (SWIID) [Solt, 2020]. The ECI is obtained from the Atlas of Economic Complexity, provided by the Growth Lab at Harvard University. We also include in our analysis a set of control variables established in the literature, using data from the World Development Indicators (WDI) database. Table 1 describes the variables and data sources.

Most previous studies analysed the relationship between economic complexity and inequality in large samples of countries [Hartman et al., 2017; Sbardella et al., 2017; Lee, Vu, 2019; Chu, Hoang, 2020; Amarante et al., 2023; Nguyen et al., 2023], but Lee and Wang (2021) focused on small groups of developed and developing countries. Most researchers emphasize the heterogeneity in results for countries with various income and development levels as well as for countries in different regions [Lee, Vu, 2019; Chu, Hoang, 2020; Amarante et al., 2023; Nguyen et al., 2023], but extensive sub-group and regional analyses are lacking. Moreover, recent works have questioned previous results such as those reported by Hartman et al. (2017), suggesting that they may be influenced either by the patterns typical for countries with high income levels and above average economic complexity [Amarante et al., 2023], or by patterns of development in lower-income countries [Lee, Vu, 2019].

Table 1.

Variables and data sources

Variable	Description	Source
GINI	The Gini index is a measure of inequality in equivalized (square root scale) disposable (post-tax, post-transfer) household income. GINI ranges from approximately 23.0 (more equal income distribution) to 46.0 (more unequal income distribution)	The Standardized World Income Inequality Database (SWIID) by Solt (2020)
ECI	The Economic Complexity Index measures the amount of productive knowledge belonging to each country. ECI ranges from approximately –2.0 (low complexity) to 2.0 (high complexity)	The Atlas of Economic Complexity provided by the Growth Lab at Harvard University
GDP pc	Estimate of Gross Domestic Product per capita using the WB data on GDP (constant 2015 US$) and population, total	World Development Indicators (WDI)
Gov	General government final consumption expenditure (% of GDP), which includes all government current expenditures for purchases of services and goods. It includes most national defence and security expenditures but excludes government military expenditures (they are part of government capital formation)	
Trade	Trade (% of GDP) presents the sum of exports and imports of goods and services measured as a share of gross domestic product	
Schooling	School enrolment, tertiary (% gross) is the ratio of total enrolment, regardless of age, to the population of the age group that officially corresponds to the level of education shown. *Note: tertiary education, whether or not it is an advanced research qualification, normally requires, as a minimum condition of admission, the successful completion of education at the secondary level*	
Institute	Institutional quality is an average of the World Governance Indicators which include Voice and Accountability, Political Stability and Absence of Violence/Terrorism, Government Effectiveness, Regulatory Quality, Rule of Law, Control of Corruption. The Institutional quality variable ranges from –2 to 2, with 2 indicating strong governance performance	The Worldwide Governance Indicators, 2022

These contradictory findings lead us to the idea of estimating the inequality effects of economic complexity by groups of countries. Therefore, we propose two *classifications of countries*, based on (1) their income levels, and (alternatively) (2) based on geography. According to income level, countries can be subdivided into high, upper-middle, and lower-middle income groups. Concerning classification based on geography, the first group of countries includes ECA, and the second group consists of MENA countries.

The descriptive statistics for the main variables are displayed in Table 2. The GDP per capita, government expenditure, trade and school enrolment rate are transformed into the natural logarithmic form. The Gini coefficient takes values from 22.0 to 47.0, meaning that our sample contains countries with relatively more equal income distributions, which are mostly located in Europe; and countries with relatively high levels of inequality, mainly located in MENA and Central Asia. The economic complexity values are in the interval ranging from –1.9 to 2.5, which includes countries with various levels of complexity.

Table 2.

Variable descriptive statistics (N = 1352)

Variable	Mean	Std. Dev.	Min	Max
ECI	0.63	0.86	–1.85	2.45
GINI	32.59	5.65	21.90	47.00
lnGDPpc	9.24	1.22	5.92	11.38
lnSchooling	3.81	0.54	2.08	5.02
Institute	0.36	0.92	–1.66	1.95
lnGov	2.87	0.26	2.04	3.50
lnTrade	4.42	0.39	2.59	5.53

3.3. Empirical model

Several studies mentioned endogeneity concerns pertaining to the relation between economic complexity and inequality [Lee, Vu, 2019; Chu, Hoang, 2020; Amarante et al., 2023; Nguyen et al., 2023]. In particular, Lee and Vu (2019) questioned the pioneering influential research in the field by Hartmann et al. (2017). Lee and Vu (2019) argued that the results on the negative relationship between economic complexity and inequality may be biased due to the potential endogeneity of ECI, since there is a significant difference in coefficients of economic complexity under linear pooled OLS and system GMM estimations. Chu and Hoang (2020) also addressed the potential endogeneity issue by using pooled 2SLS and system-GMM; however, they found no differences between the estimation results. Similarly, Amarante et al. (2023) incorporated lagged regressors into the fixed effect estimation to control for endogeneity, but also found no difference between these results and those obtained from simple fixed effect estimation. Since the endogeneity problem is not sufficiently covered in the existing literature on the ECI – inequality relationship, with only a few papers employing instrumental variables, there is a need to further explore this issue [Lee, Vu, 2019; Amarante et al., 2023].

We address the potential endogeneity issue by employing a fixed-effect two stage least squares (2SLS) model with instrumental variables (IVs), which provides more consistent results in comparison with simple OLS estimation [Bollen, 1989; Foster, McLanahan, 1996; Maydeu-Olivares et al., 2020]. Moreover, this allows for comparability of results with previous studies.

At the first stage of 2SLS model we use lagged ECI as IVs and assess the impact of lagged ECI on current ECI, as we assume that ECI is an endogenous variable. For robustness check we also assume endogeneity of GDP pc and include lagged GDP pc as IVs in addition to lagged ECI. The results prove to be robust (Appendix B: Table 7).

At the second stage of 2SLS model we assess the impact of economic complexity on income inequality based on the following model (Eq. 6):

$$GINI_{it} = \beta_0 + \beta_1 ECI_{it} + \beta_2 \ln GDP_{it} + \beta_3 \ln GDP_{it}^2 + \beta_4 \ln Schooling_{it} + \beta_5 Institute_{it} + \\ + \beta_6 \ln Gov_{it} + \beta_7 \ln Trade_{it} + \nu_i + \gamma_t + \varepsilon_{it}. \quad (6)$$

where *i* stands for a country and *t* stands for a time period; *ECI* and *Institute* are the economic complexity and institutions indexes; *lnGDP*, *lnSchooling*, *lnGov*, *lnTrade* are the natural logarithms of the GDP per capita, school enrolment (tertiary), government expenditures, and trade, respectively; ν_i is an unobserved effect that does not change over time (country fixed effect); γ_t – time fixed effect; and ε_{it} stands for the error term.

To check the validity of the results obtained from the instrumental regression, we use the Sargan-Hansen J-statistic (since the conventional R-squared is no longer valid in the 2SLS model). This statistic is used to test whether the instruments are uncorrelated with the error term [Baum, 2003]. The acceptance of the null hypothesis suggests that instruments are generally valid, and hence the results are reliable. The model was estimated using xtivreg2 command developed for Stata by Schaffer (2010).

Additionally, to consider a scenario of non-linear relationships between GDP per capita and income inequality [Kuznets, 1955], as well as that of *economic complexity* and income inequality [Sbardella et al., 2017; Chu, Hoang, 2020; Zhu et al., 2020; Morais et al., 2021; Amarante et al., 2023; Nguyen et al., 2023; Pham et al., 2023], we introduce quadratic terms for GDP per capita and ECI in the model. Both the baseline model and the models with quadratic terms are estimated using 2SLS with country and time fixed effects.

4. Results

4.1. Economic complexity impacts on inequality: regional features

Table 3 presents the estimation results of the impact of ECI on income inequality for all countries in our sample based on 2SLS method with country and time fixed effects. Columns 1 and 5 in Table 3 demonstrate the estimation results of the baseline model. Columns 2–4 and 6–8 provide estimation results of the models with quadratic terms of ECI and institutions that were added to the baseline model to test for non-linear relationships between variables and the robustness of the results. Models 1–4 are estimated with country fixed effects, and Models 5–8 are estimated with country and time fixed effects (FE). Heteroskedasticity-robust standard errors are applied.

Table 3.

Estimation results for all countries in the sample

Dependent variable: Gini coefficient

	1	2	3	4	5	6	7	8
ECI	1.134*** (0.355)	1.212*** (0.369)	1.370*** (0.363)	1.398*** (0.374)	1.270*** (0.369)	1.311*** (0.381)	1.504*** (0.376)	1.514*** (0.385)
lnGDP	4.510** (2.138)	4.775** (2.170)	3.444 (2.218)	3.589 (2.258)	4.197* (2.151)	4.315** (2.176)	2.796 (2.259)	2.841 (2.290)
lnGDP2	–0.261** (0.125)	–0.274** (0.126)	–0.203 (0.129)	–0.210 (0.131)	–0.278** (0.127)	–0.281** (0.128)	–0.192 (0.134)	–0.193 (0.135)
lnSchooling	–0.620*** (0.184)	–0.643*** (0.187)	–0.666*** (0.185)	–0.676*** (0.186)	–0.794*** (0.206)	–0.797*** (0.207)	–0.785*** (0.204)	–0.785*** (0.205)
Institut	–0.540* (0.282)	–0.542* (0.284)	–0.480* (0.278)	–0.482* (0.279)	–0.358 (0.275)	–0.367 (0.278)	–0.330 (0.274)	–0.334 (0.276)
lnGov	–0.026 (0.320)	–0.055 (0.323)	0.009 (0.321)	–0.006 (0.322)	–0.168 (0.318)	–0.180 (0.321)	–0.083 (0.321)	–0.088 (0.323)
lnTrade	0.982*** (0.247)	0.978*** (0.248)	0.893*** (0.251)	0.893*** (0.252)	0.799*** (0.259)	0.802*** (0.260)	0.704*** (0.265)	0.706*** (0.266)
ECI2		–0.102 (0.111)		–0.047 (0.115)		–0.060 (0.115)		–0.019 (0.120)
Institut2			–0.622*** (0.199)	–0.608*** (0.202)			–0.651*** (0.205)	–0.646*** (0.208)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	No	No	Yes	Yes	Yes	Yes
Observations	1040	1040	1040	1040	1040	1040	1040	1040
S-H test (p-value)	0.9041	0.8923	0.7475	0.7382	0.9008	0.8932	0.7249	0.7218
Countries	52	52	52	52	52	52	52	52

Notes: *, ** and *** denote a significance level at 10%, 5%, and 1%, respectively; heteroskedasticity-robust standard errors are in parentheses; IVs: lagged ECI.

The results with country fixed effects and with both country and time fixed effects proved to be consistent. Moreover, adding quadratic terms also confirmed robustness of the results. For models in Tables 3–8 the Sargan-Hansen test (S-H test) of overidentifying restrictions demonstrates that the instruments are valid (uncorrelated with the error term) and that the excluded instruments are correctly excluded from the estimated equation (p-values are in Tables 3–8). Besides, Underidentification test (LM statistic) showed that the model is identified (the null hypothesis of underidentification was rejected). Instruments were checked for redundancy, and the null hypothesis of redundancy was rejected.

Our findings reveal a positive and statistically significant effect of economic complexity on income inequality, meaning that an increase in a country's economic complexity results in a higher level of income inequality. This is consistent with previous works, which also found that rising complexity enhances income inequality [Lee, Vu, 2019; Chu, Hoang, 2020; Amarante et al., 2023]. However, in contrast to previous studies [Sbardella et al., 2017; Chu, Hoang, 2020; Zhu et al., 2020; Morais et al., 2021; Amarante et al., 2023; Nguyen et al., 2023], we observe that the quadratic term of economic complexity is insignificant. Hence, we cannot confirm the presence of a nonlinear relationship in an overall sample of ECA and MENA countries.

Meanwhile, the results demonstrate that higher education significantly contributes to the reduction of inequality, consistent with previous findings [Lessmann, Seidel, 2017; Chu, Hoang, 2020]. This observation is reinforced by the fact that income inequality is directly related to educational inequality [Lee et al., 2018]. Therefore, when higher education becomes more accessible, and its attainment rate consequently increases, income inequality tends to decrease.

At the same time, we revealed that the overall quality of institutions has a significant role in reduction of inequality, corroborating recent studies featuring the overall quality of institutions or specific aspect of institutions such as democracy [Hartmann et al., 2017; Ouechtati, 2022; Oyèkọ́lá, 2023]. Thus, development of inclusive institutions can promote equal access to gains from economic growth thus decreasing income disparities.

An opposite effect is observed for trade. Particularly, specific features of a country's international trade may be associated with increasing inequalities, in line with previous research [Bergh, Nilsson, 2010; Dorn et al., 2021; Adão et al., 2022]. This effect also coincides with the ambiguous theoretical predictions of the neoclassical trade theory on the effect of trade openness on income inequality for developing and developed countries. Namely, this theory implies that trade openness has a potential to decrease inequality in developing countries due to narrowing the income gap between skilled and unskilled labour, and providing the poorest people with larger gains; simultaneously, it tends to increase inequality in developed countries, by providing upper middle class with disproportionately more gains [Stolper, Samuelson, 1941].

At the second stage, driven by the fact that previous works reported heterogeneous results for various groups of countries and different regions [Lee, Vu, 2019; Chu, Hoang, 2020; Amarante et al., 2023; Nguyen et al., 2023], we divided our sample of 52 countries into two subsamples based on their regional belonging, as described previously (ECA and MENA). The model estimation by regions is shown in Table 4.

After dividing the sample into subgroups of ECA (Models 1–3 and 7–9) and MENA (Models 4–6 and 10–12) countries, the results with country fixed effects and with both country and time fixed effects are consistent. Adding quadratic terms also confirmed robustness of the results. For the ECA group, the sign of ECI remains statistically significant and positive. Moreover, the coefficient of its quadratic term becomes significantly negative. Hence, we conclude that the U-inverted relationship between economic complexity and income inequality is typical for ECA. This implies that the increased sophistication of the economy and diversification of production initially lead to rising disparities in the region; however, after reaching a certain level of development, ECA countries start to experience a decrease in income inequality, with the further increase in the complexity of their economies. In contrast, in MENA countries, both the ECI and its square have no significant impact on income inequality dynamics.

Table 4.

Estimation results by regions

Dependent variable: Gini coefficient

	ECA	ECA	ECA	MENA	MENA	MENA	ECA	ECA	ECA	MENA	MENA	MENA
	1	2	3	4	5	6	7	8	9	10	11	12
ECI	1.470***	1.998***	2.174***	−0.302	−0.678	−0.827	1.688***	2.096***	2.273***	−0.163	−0.413	−0.538
	(0.380)	(0.509)	(0.521)	(0.657)	(1.104)	(1.120)	(0.425)	(0.568)	(0.576)	(0.637)	(0.965)	(1.014)
lnGDP	4.593**	6.129***	5.150**	16.90***	17.97***	15.20***	3.887	5.101*	3.960	16.11***	16.91***	13.75**
	(2.055)	(2.209)	(2.241)	(4.566)	(5.238)	(5.033)	(2.485)	(2.632)	(2.718)	(4.815)	(5.199)	(5.337)
lnGDP2	−0.268**	−0.344***	−0.293**	−1.013***	−1.069***	−0.88***	−0.274*	−0.320**	−0.249	−0.929***	−0.966***	−0.751**
	(0.119)	(0.126)	(0.127)	(0.258)	(0.292)	(0.282)	(0.146)	(0.151)	(0.157)	(0.276)	(0.290)	(0.313)
lnSchoo-ling	−0.0260	−0.0768	−0.116	−1.237***	−1.248***	−1.30***	−0.272	−0.247	−0.231	−1.047***	−1.051***	−1.08***
	(0.227)	(0.229)	(0.230)	(0.270)	(0.261)	(0.259)	(0.245)	(0.248)	(0.247)	(0.339)	(0.331)	(0.321)
Institut	−0.737**	−0.779**	−0.645**	0.796	0.755	−0.179	−0.491	−0.597*	−0.504	0.716	0.709	−0.259
	(0.311)	(0.314)	(0.319)	(0.534)	(0.538)	(0.795)	(0.315)	(0.328)	(0.325)	(0.585)	(0.593)	(0.844)
lnGov	0.962**	0.886**	0.976**	−1.192**	−1.066*	−0.958*	0.795*	0.816*	0.993**	−1.330***	−1.193**	−1.056*
	(0.422)	(0.422)	(0.425)	(0.471)	(0.558)	(0.573)	(0.430)	(0.448)	(0.453)	(0.460)	(0.598)	(0.633)
lnTrade	0.852***	0.793***	0.739***	−0.169	−0.200	−0.353	0.474	0.505	0.485	−0.441	−0.539	−0.740
	(0.281)	(0.283)	(0.285)	(0.383)	(0.392)	(0.415)	(0.331)	(0.335)	(0.335)	(0.570)	(0.622)	(0.637)
ECI2		−0.450**	−0.423**		−0.281	−0.325		−0.394*	−0.389*		−0.185	−0.227
		(0.176)	(0.176)		(0.473)	(0.474)		(0.205)	(0.209)		(0.396)	(0.408)
Institut2			−0.483**			−0.956			−0.511**			−0.972
			(0.215)			(0.585)			(0.223)			(0.626)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observa-tions	860	860	860	180	180	180	860	860	860	180	180	180
S-H test (p-value)	0.7043	0.7132	0.7837	0.8714	0.9102	0.9967	0.8916	0.8851	0.9026	0.9531	0.9713	0.9754
Countries	43	43	43	9	9	9	43	43	43	9	9	9

Notes: *, ** and *** denote a significance level at 10%, 5%, and 1%, respectively; heteroskedasticity-robust standard errors are in parentheses; IVs: lagged ECI.

The effects of the control variables are consistent with the baseline model (Table 3), except for a significantly positive effect of government expenditures in ECA, and, at the same time, a significantly negative effect of this variable in the MENA countries. In other words, an effective general government expenditure demonstrates a potential to reduce inequality in MENA, while in the ECA it is associated with higher inequality. Therefore, for the ECA, we can confirm an existence of the inverted-U-shaped pattern of relations between economic complexity and inequality, in line with previous studies [Sbardella et al., 2017; Chu, Hoang, 2020; Zhu et al., 2020; Morais et al., 2021; Amarante et al., 2023; Nguyen et al., 2023]. At the same time, an absence of any significant effect of economic complexity on inequality in the MENA countries does not allow us to draw final conclusions. Meanwhile, it partly corresponds to the study by Nguyen et al. (2023), who did not find a U-inverted relationship in MENA and South Asian countries.

These results reiterate that the impact of economic complexity differs across regions of the world. ECA economies are mainly dominated by developed upper middle- and high-income countries, wherein increasing economic complexity can potentially reduce inequality after reaching a certain level of development. Conversely, the nature of this relationship in MENA is not clear and requires further investigation.

Our results also support the opinion that the results for all countries may be driven by the patterns established for countries with high income or, instead, with non-high income [Lee, Vu, 2019; Amarante et al., 2023]. Consequently, all four relationships between economic complexity and inequality discovered in previous studies are possible, although they may be specific to different groups of countries.

In our case, for the overall sample of countries we found only the dominant patterns of relation between economic complexity and income inequality, probably affected by the patterns prevailing in the developed countries. Overall, ***H1*** is partly supported, as the effect of economic complexity on income inequality turned out to be non-linear, taking a U-inverted shape, in ECA countries.

We perform the following robustness check. First, we check the robustness of the results by including the IVs for GDP p.c. for the overall sample, for ECA and MENA countries, and for estimation with interaction terms. We provide the estimation results with IVs for GDP p.c. only for the overall sample (initial estimation results are presented in Table 3) due to the size limit and format of the paper (Appendix B: Table 7). Second, we check the robustness of the results by excluding GDP p.c. and GDP p.c. squared from the model (Appendix B: Table 8). The results proved to be robust.

4.2. Institutional quality mediating effect

Addressing ***H2***, we delved into the study of the role that institutional factors play in reducing inequality. Previous works on the nexus between economic complexity and income inequality employed an aggregate measure of the institutional quality across countries. This is usually estimated as the average of the Worldwide Governance Indicators (WGIs), which aim to measure the quality of governance and institutions in a country across six dimensions: voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and control of corruption [World Bank, 2022].

There are suggestions in the literature that institutions can play the mediating role in the relationship between economic complexity and income inequality by alleviating an increase in inequality that arises due to growing complexity, but the indirect impacts of institutions on income inequality have not been studied extensively [Lee, Vu, 2019; Chu, Hoang, 2020]. Seeking to close this research gap in the literature, we consider the interaction of ECI with the institutionnal indicator and its components (Table 5) while analysing an impact of economic complexity on income inequality. This approach enables us to capture an effect of economic complexity conditional on the quality of institutions.

The model in this case has undergone some changes; in our base-line equation (Eq. 6), we alternately add WGI governance components and their interaction terms with the economic complexity index. We also considered the interaction term of an aggregate measure of institutions as in the above models (Tables 3 and 4). The estimation results are displayed in Table 5 and indicate that institutional quality and all its components are vital for decreasing income inequality when economic complexity grows.

For instance, among other institutional indicators, we included in the model the variable «control of corruption» (*Corrupt*), which captures the extent to which public power is exercised for private gains, and state administration is subordinated to the private interests of the elites. Corruption has a number of consequences, including poor tax administration, tax evasion, reduced spending on education and healthcare, and a decline in the bargaining power of workers and trade unions. Besides, corruption diminishes the quality of public services and hinders people's access to them, leading to a lower standard of living in a country; hence, the control of corruption is an important leverage for the decrease in inequality. Moreover, corruption was found to have an adverse impact on a country's development by limiting economic growth through numerous channels, among which are discouraging investments and altering the composition of public spending.

Table 5.

Estimation results with interaction terms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ECI	1.142***	1.291***	1.117***	1.639***	1.279***	1.323***	1.173***
	(0.355)	(0.356)	(0.352)	(0.393)	(0.358)	(0.377)	(0.351)
lnGDP	4.072*	4.973**	4.211**	5.153**	4.669**	4.378**	3.796*
	(2.098)	(2.097)	(2.138)	(2.126)	(2.026)	(2.017)	(2.165)
lnGDP2	–0.234*	–0.284**	–0.245**	–0.307**	–0.281**	–0.269**	–0.232*
	(0.123)	(0.123)	(0.125)	(0.124)	(0.119)	(0.119)	(0.127)
lnSchooling	–0.640***	–0.661***	–0.554***	–0.658***	–0.656***	–0.619***	–0.538***
	(0.178)	(0.180)	(0.178)	(0.182)	(0.181)	(0.180)	(0.179)
lnGov	0.0547	–0.0357	–0.0650	–0.282	0.0749	0.126	–0.0382
	(0.325)	(0.329)	(0.318)	(0.325)	(0.318)	(0.319)	(0.317)
lnTrade	0.982***	0.872***	0.914***	0.948***	1.069***	1.108***	0.964***
	(0.248)	(0.249)	(0.244)	(0.250)	(0.245)	(0.244)	(0.248)
Corrupt	–0.503***	–0.229					
	(0.193)	(0.203)					
ECI_Corrupt		–0.600***					
		(0.148)					
Gov_Effect			–0.376*	0.0549			
			(0.199)	(0.211)			
ECI_Gov_Effect				–0.774***			
				(0.162)			

Continuation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Accountability					–0.690*** (0.176)	–0.727*** (0.180)	
ECI_Accountability						0.203 (0.160)	
Polit_Stability							0.0499 (0.118)
ECI_Polit_Stability							
Rule_of_Law							
ECI_Rule_of_Law							
Reg_Quality							
ECI_Reg_Quality							
Institut							
ECI_Institut							
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	No	No	No	No	No
Observations	1040	1040	1040	1040	1040	1040	1040
S-H test (p)	0.8914	0.8659	0.8727	0.5913	0.8927	0.8982	0.8335
Countries	52	52	52	52	52	52	52

Continuation

	(8)	(9)	(10)	(11)	(12)	(13)	(14)
ECI	1.223***	1.165***	1.317***	1.176***	1.245***	1.134***	1.260***
	(0.350)	(0.354)	(0.364)	(0.353)	(0.359)	(0.355)	(0.356)
lnGDP	3.947*	3.930*	4.516**	3.848*	3.628*	4.510**	4.904**
	(2.141)	(2.116)	(2.134)	(2.098)	(2.108)	(2.138)	(2.136)
lnGDP2	−0.243*	−0.232*	−0.262**	−0.238*	−0.224*	−0.261**	−0.282**
	(0.125)	(0.124)	(0.125)	(0.123)	(0.124)	(0.125)	(0.124)
lnSchooling	−0.571***	−0.585***	−0.619***	−0.528***	−0.535***	−0.620***	−0.663***
	(0.180)	(0.181)	(0.182)	(0.179)	(0.179)	(0.184)	(0.185)
lnGov	−0.0317	0.00337	−0.0363	−0.0718	−0.0774	−0.0260	−0.0947
	(0.324)	(0.326)	(0.333)	(0.320)	(0.322)	(0.320)	(0.325)
lnTrade	0.992***	0.975***	0.953***	0.947***	0.894***	0.982***	0.934***
	(0.248)	(0.249)	(0.251)	(0.249)	(0.252)	(0.247)	(0.248)
Corrupt							
ECI_Corrupt							
Gov_Effect							
ECI_Gov_Effect							
Accountability							
ECI_Accountability							
Polit_Stability	0.164						
	(0.125)						
ECI_Polit_Stability	−0.318**						
	(0.124)						
Rule_of_Law		−0.248	−0.0438				
		(0.231)	(0.246)				
ECI_Rule_of_Law			−0.506***				
			(0.176)				
Reg_Quality				0.115	0,181		
				(0.190)	(0.195)		
ECI_Reg_Quality					−0.273*		
					(0.149)		
Institut						−0.540*	−0.326
						(0.282)	(0.294)

Continuation

	(8)	(9)	(10)	(11)	(12)	(13)	(14)
ECI_Institut							–0.519*** (0.181)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	No	No	No	No	No
Observations	1040	1040	1040	1040	1040	1040	1040
S-H test (p)	0.7461	0.8629	0.8095	0.8292	0.8162	0.9041	0.8542
Countries	52	52	52	52	52	52	52

Notes: *, ** and *** denote a significance level at 10%, 5%, and 1%, respectively; heteroskedasticity-robust standard errors are in parentheses; IVs: lagged ECI; dependent variable: Gini coefficient.

Table 6.

Estimation results with interaction terms: time fixed effects are added

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ECI	1.266*** (0.370)	1.369*** (0.368)	1.255*** (0.366)	1.738*** (0.400)	1.422*** (0.372)	1.481*** (0.396)	1.366*** (0.367)
lnGDP	3.905* (2.112)	4.446** (2.104)	4.043* (2.148)	4.604** (2.134)	4.630** (2.018)	4.445** (2.004)	3.334 (2.150)
lnGDP2	–0.256** (0.126)	–0.272** (0.125)	–0.268** (0.127)	–0.294** (0.126)	–0.313*** (0.120)	–0.311*** (0.120)	–0.248* (0.127)
lnSchooling	–0.817*** (0.204)	–0.746*** (0.205)	–0.755*** (0.202)	–0.760*** (0.205)	–0.888*** (0.205)	–0.877*** (0.204)	–0.788*** (0.203)
lnGov	–0.0808 (0.324)	–0.0973 (0.329)	–0.194 (0.316)	–0.353 (0.327)	–0.0580 (0.316)	–0.0161 (0.316)	–0.139 (0.318)
lnTrade	0.807*** (0.258)	0.739*** (0.258)	0.743*** (0.255)	0.806*** (0.261)	0.877*** (0.256)	0.915*** (0.254)	0.749*** (0.261)
Corrupt	–0.430** (0.191)	–0.204 (0.199)					
ECI_Corrupt		–0.576*** (0.157)					
Gov_Effect			–0.320* (0.193)	0.0771 (0.209)			

Continuation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ECI_Gov_Effect				-0.768^{***} (0.167)			
Accountability					-0.701^{***} (0.182)	-0.742^{***} (0.187)	
ECI_Accountability						0.234 (0.169)	
Polit_Stability							0.204^{*} (0.120)
ECI_Polit_Stability							
Rule_of_Law							
ECI_Rule_of_Law							
Reg_Quality							
ECI_Reg_Quality							
Institut							
ECI_Institut							
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1040	1040	1040	1040	1040	1040	1040
S-H test (p)	0.8905	0.8697	0.8864	0.6259	0.8949	0.9032	0.7962
Countries	52	52	52	52	52	52	52

Continuation

	(8)	(9)	(10)	(11)	(12)	(13)	(14)
ECI	1.391*** (0.366)	1.302*** (0.369)	1.446*** (0.378)	1.329*** (0.368)	1.389*** (0.373)	1.270*** (0.369)	1.370*** (0.368)
lnGDP	3.341 (2.129)	3.858* (2.122)	4.223** (2.139)	3.783* (2.115)	3.455 (2.128)	4.197* (2.151)	4.346** (2.148)
lnGDP2	–0.243* (0.126)	–0.262** (0.126)	–0.273** (0.127)	–0.273** (0.126)	–0.249** (0.126)	–0.278** (0.127)	–0.275** (0.127)
lnSchooling	–0.771*** (0.204)	–0.803*** (0.207)	–0.790*** (0.208)	–0.762*** (0.204)	–0.743*** (0.206)	–0.794*** (0.206)	–0.772*** (0.208)
lnGov	–0.115 (0.325)	–0.148 (0.320)	–0.154 (0.329)	–0.229 (0.317)	–0.212 (0.322)	–0.168 (0.318)	–0.192 (0.325)
lnTrade	0.785*** (0.261)	0.786*** (0.260)	0.776*** (0.263)	0.735*** (0.262)	0.689*** (0.266)	0.799*** (0.259)	0.773*** (0.260)
Corrupt							
ECI_Corrupt							
Gov_Effect							
ECI_Gov_Effect							
Accountability							
ECI_Accountability							
Polit_Stability	0.282** (0.126)						
ECI_Polit_Stability	–0.264** (0.126)						
Rule_of_Law		–0.226 (0.227)	–0.0333 (0.240)				
ECI_Rule_of_Law			–0.515*** (0.179)				
Reg_Quality				0.173 (0.186)	0.227 (0.191)		
ECI_Reg_Quality					–0.263* (0.153)		
Institut						–0.358 (0.275)	–0.192 (0.286)
ECI_Institut							–0.512*** (0.190)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Continuation

	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1040	1040	1040	1040	1040	1040	1040
S–H test (p)	0.7242	0.8768	0.8332	0.8291	0.8203	0.9008	0.8535
Countries	52	52	52	52	52	52	52

Notes: *, ** and *** denote a significance level at 10%, 5%, and 1%, respectively; heteroskedasticity-robust standard errors are in parentheses; IVs: lagged ECI; dependent variable: Gini coefficient

Tables 5 and 6 present results with interaction terms. Table 5 contains results with country fixed effects and Table 6 – with country and time fixed effects. The results, especially concerning interaction terms of ECI and institutional variables remain consistent confirming robustness of the model.

The control of corruption in a country demonstrates a significantly negative effect on inequality (i.e., the higher the control of corruption, the less income inequality there is). As for its mediating effect, we observe that the sign of its interaction term with ECI is significantly negative; in other words, the impact of economic complexity on income inequality is conditional on the control of corruption in a country, with lower corruption and higher economic complexity leading to lower inequality.

Overall, the moderating variables «control of corruption» (*Corrupt)*, «government effectiveness» (*Gov_effect*), «political stability» (*Polit_Stability*), «rule of law» (*Rule_of_Law*) and «regulatory quality» (*Reg_quality*) demonstrate negative and statistically significant coefficients of the interaction terms with the *ECI*. In other words, these factors contribute to the reduction of income inequality in a country by mitigating the adverse effect of economic complexity. Meanwhile, higher values of the indicator «Voice and Accountability» (*Accountability*) reduce income inequality, though interaction term turned out to be insignificant.

Finally, the interaction term between aggregate measures of institutions and economic complexity displays a similar effect, suggesting that the high quality of institutions and governance in a country can reverse the pattern of economic complexity increasing inequality. According to this observation, a country with effective governance and redistribution systems, low levels of corruption, and political stability may experience a reduction in income inequality when its level of complexity increases; conversely, low-quality national institutions can inhibit the otherwise positive effects of increasing economic complexity and exacerbate inequality.

Thus, we can conclude that well-functioning institutions can play a crucial role in shaping the impact of growing sophistication in the economy, and economic complexity can reduce income inequality if institutions are well-functioning. Institutional factors such as political stability; effective governance, high quality of public and civil service, sound policy formulation and implementation, private sector regulation and development; rule of law (e.g., expressed by the quality of contract enforcement and property rights, effective policing, and independent and competent courts); and control of corruption create favourable conditions for more equal benefits from growing economic complexity. The improvement of institutional conditions through any of these channels can facilitate a decrease in income disparities.

5. Conclusion

This paper provides new insights into understanding the essence of the relationship between economic complexity and income inequality. To address heterogeneity across groups of countries, and identify possible non-linear patterns in this relationship, we analysed the impact of economic complexity on income inequality in two regions: ECA, and MENA. There is a significant gap in levels of inequality and economic complexity between these two regions, and our analysis accounts for these diverse economic conditions. Our results underscore that the effect of increasing economic complexity may vary depending on countries' individual characteristics, such as the level of economic development, regional affiliation, human capital, and quality of institutions.

Supporting previous findings, we argue that the overall relationship between economic complexity and income inequality is positive. In other words, generally increasing economic complexity leads to higher inequality. As for a non-linear relationship, there is a lack of reliable evidence to confirm it for a sample of all countries. However, the regional sub-groups analysis indicates a U-inverted relationship between economic complexity and inequality in ECA countries, while in MENA countries there are no significant results, partially confirming ***H1*** on the non-linear relationship between economic complexity and income inequality.

This result implies that economic sophistication and diversification of production initially increase income disparities in ECA economies, but after reaching a certain level of economic complexity, these countries start to experience a decrease in income inequality alongside further increases in economic complexity. Meanwhile, we highlight the importance of considering the institutional and socio-economic environment in a particular region or country, since the beneficial effects of increasing economic complexity are associated with a more conducive environment (as explained previously), in line with previous studies.

Our finding that education, as a proxy for human capital, reduces income inequality, reflects that an increase in economic complexity is closely related to technological progress and development of productive knowledge. Therefore, sustainable economic development requires education and training of personnel that would enable employees to adapt to new knowledge and technologies and interact in the complex economic system. Thus, government should both provide equal opportunities for education, and ensure the quality and relevance of education and training to address the real challenges in the economy. In addition, it is necessary to develop and support institutions that help enhance the quality of education and its compliance with the standards as well as challenges of the modern world.

Estimation results also indicate an adverse effect of international trade on income disparities. On one hand, the economic openness and integration into international trade can contribute to the formation of competitive markets and attract capital and human resources to industries, providing jobs for employees with various skill levels, and thus decreasing income inequality. On the other hand, the gains from international trade can be distributed disproportionately, thus increasing wage differentiation between groups of workers, as well as enlarging a gap between employees and individuals in charge of business or involved in policy making. When top management, business owners, government officials, or workers from specific industries benefit from openness relatively more than the other population groups, this can deepen income inequality.

The finding that higher institutional quality decreases income inequality implies that institutional reforms are needed to create effective checks and balances and thus reduce inequality and prevent further distortions in income distribution. However, while carrying out these reforms, it is essential to avoid restricting economic freedom and hindering innovations and entrepreneurial activity. An increase in economic complexity along with the transformation and strengthening of a country's institutions will facilitate a sustainable economic growth beneficial for all social strata, and imply a fairer distribution of income in society. This type of economic development will help reduce inequality or at least will not lead to its egregious exacerbation with increasing economic complexity.

Besides, in line with previous research, our results suggest that the quality of institutions has a significant mediating role in the relationship between economic complexity and income inequality, thus confirming ***H2***. Indeed, well-functioning inclusive institutions can strengthen the positive effects of economic complexity and level-out its negative consequences. Accordingly, we claim that the relationship between economic complexity and income inequality can be influenced through policy measures aimed at promoting sound and effective regulations, providing people with high quality public services, lowering corruption, preventing disproportionately high gains for some income groups, ensuring law enforcement, maintaining the independence of the judiciary branch of power, and preserving political stability.

We believe that to maximize benefits from the increasing economic complexity accompanied by technological progress and the transition to the knowledge economy, it is vital to prioritize the development of institutions as well as implement national systems and programs for monitoring their quality and improving the most vulnerable areas.

Overall, determining policy implications should concern not only increasing economic complexity, as if it were a single causal factor for inequality, since the economic complexity indicator is determined by a combination of unknown causes and factors [Hidalgo, 2022]. Instead, policies should focus on considering the underlying factors and mechanisms that allow society to benefit from economic complexity. Consequently, further research is required to explore the relationship between economic complexity and factors that can play a mediating role between it and inequality, and also to develop rational policies addressing these factors. Besides, differences across countries and patterns of their development deserve further research to shed new light on the nexus between economic complexity and inequality for various groups of countries, and to provide new insights into the impact of economic complexity on income inequality.

Appendix A

Sampled countries

High income	GINI 2020	ECI rank 2020	Upper middle income	GINI 2020	ECI rank 2020	Lower middle income	GINI 2020	ECI rank 2020
Slovakia	22.3	13	Belarus	24.6	31	Ukraine	27.0	49
Slovenia	24.3	11	Kazakhstan	27.7	81	Kyrgyzstan	32.0	54
Czech Republic	24.4	6	Azerbaijan	30.6	121	Algeria	32.4	108
Finland	26.0	14	North Macedonia	32.0	58	Uzbekistan	35.9	78
Belgium	26.1	22	Russia	32.8	51	Tunisia	37.9	44
Norway	26.2	37	Serbia	33.4	38	Iran	38.4	85
Sweden	26.7	8	Armenia	36.3	77	Egypt	40.0	69
Denmark	26.9	23	Jordan	37.2	59	Morocco	40.6	80
Netherlands	27.4	28	Albania	37.3	74	Tajikistan	45.3	94
Austria	27.5	7	Georgia	37.5	68			
Hungary	28.3	9	Bosnia and Herzegovina	38.6	36			
Ireland	28.6	15	Bulgaria	38.9	39			
Croatia	29.5	32	Turkey	39.9	41			
Poland	29.7	26						
Germany	29.8	3						
France	30.0	18						
Cyprus	30.1	45						
Switzerland	30.9	2						
Estonia	30.9	27						
Greece	30.9	50						
United Kingdom	31.4	10						
Spain	32.0	33						
Portugal	32.1	34						
Italy	33.7	16						
Romania	34.1	19						
Israel	34.4	21						
Latvia	35.0	35						
Lithuania	35.4	29						
Qatar	40.2	71						
Saudi Arabia	46.7	42						

Appendix B

Robustness check

Table 7.

Estimation results for all countries in the sample: with lagged ECI and lagged GDP p.c. as IVs

Dependent variable: Gini coefficient

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ECI	1.196***	1.294***	1.449***	1.498***	1.386***	1.464***	1.632***	1.680***
	(0.352)	(0.368)	(0.361)	(0.374)	(0.369)	(0.386)	(0.379)	(0.392)
lnGDP	5.051**	5.420**	4.046*	4.286*	4.628**	4.866**	3.258	3.420
	(2.187)	(2.225)	(2.255)	(2.301)	(2.204)	(2.240)	(2.295)	(2.338)
lnGDP2	−0.295**	−0.313**	−0.240*	−0.252*	−0.302**	−0.311**	−0.217	−0.224
	(0.128)	(0.129)	(0.132)	(0.134)	(0.130)	(0.131)	(0.136)	(0.137)
lnSchooling	−0.601***	−0.627***	−0.644***	−0.658***	−0.787***	−0.790***	−0.765***	−0.768***
	(0.186)	(0.188)	(0.186)	(0.187)	(0.211)	(0.212)	(0.208)	(0.209)
Institut	−0.569**	−0.574**	−0.511*	−0.516*	−0.393	−0.413	−0.375	−0.387
	(0.284)	(0.286)	(0.280)	(0.281)	(0.279)	(0.283)	(0.277)	(0.280)
lnGov	−0.0470	−0.0852	−0.0168	−0.0394	−0.187	−0.206	−0.100	−0.112
	(0.322)	(0.326)	(0.322)	(0.324)	(0.321)	(0.326)	(0.325)	(0.329)
lnTrade	0.999***	0.997***	0.910***	0.911***	0.815***	0.821***	0.723***	0.728***
	(0.248)	(0.249)	(0.252)	(0.253)	(0.261)	(0.262)	(0.266)	(0.268)
ECI2		−0.123		−0.0699		−0.0926		−0.0542
		(0.113)		(0.118)		(0.120)		(0.125)
Institut2			−0.642***	−0.623***			−0.685***	−0.677***
			(0.200)	(0.204)			(0.207)	(0.211)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	No	No	Yes	Yes	Yes	Yes
N	1035	1035	1035	1035	1035	1035	1035	1035
S-H test (p–value)	0.8586	0.8992	0.7939	0.8185	0.3024	0.3446	0.3034	0.3180
Countries	52	52	52	52	52	52	52	52

Notes: *, ** and *** denote a significance level at 10%, 5%, and 1%, respectively; heteroskedasticity-robust standard errors are in parentheses; IVs: lagged ECI and lagged GDP p.c.

Table 8.

Estimation results for all countries in the sample:
GDP per capita and GDP per capita squared are excluded

Dependent variable: Gini coefficient

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ECI	1.012*** (0.343)	1.045*** (0.351)	1.314*** (0.351)	1.319*** (0.357)	1.173*** (0.367)	1.221*** (0.376)	1.483*** (0.376)	1.503*** (0.383)
lnSchooling	–0.569*** (0.165)	–0.568*** (0.166)	–0.656*** (0.166)	–0.655*** (0.166)	–0.740*** (0.204)	–0.736*** (0.204)	–0.759*** (0.202)	–0.756*** (0.202)
Institut	–0.434* (0.249)	–0.418* (0.247)	–0.424* (0.241)	–0.419* (0.240)	–0.493** (0.243)	–0.469* (0.240)	–0.463** (0.235)	–0.451* (0.233)
lnGov	0.0455 (0.305)	0.0230 (0.309)	0.0830 (0.308)	0.0769 (0.309)	0.0601 (0.314)	0.0265 (0.318)	0.105 (0.316)	0.0863 (0.319)
lnTrade	0.774*** (0.245)	0.768*** (0.245)	0.726*** (0.246)	0.725*** (0.245)	0.657*** (0.252)	0.648** (0.252)	0.613** (0.255)	0.609** (0.255)
ECI2		–0.0561 (0.107)		–0.0148 (0.110)		–0.0814 (0.112)		–0.0430 (0.116)
Institut2			–0.677*** (0.191)	–0.674*** (0.192)			–0.739*** (0.196)	–0.730*** (0.197)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	No	No	Yes	Yes	Yes	Yes
Observations	1040	1040	1040	1040	1040	1040	1040	1040
S-H test (p)	0.9379	0.9369	0.7686	0.7676	0.9070	0.9061	0.7118	0.7093
Countries	52	52	52	52	52	52	52	52

Notes: *, ** and *** denote a significance level at 10%, 5%, and 1%, respectively; heteroskedasticity-robust standard errors are in parentheses; IVs: lagged ECI.

* *
*

References

Acemoglu D., Egorov G., Sonin[2] K. (2012) Dynamics and Stability of Constitutions, Coalitions, and Clubs. *The American Economic Review*, 102, 4, pp. 1446–1476. http://dx.doi.org/10.1257/aer.102.4.1446

[2] Recognized as a foreign agent.

Acemoglu D., Johnson S., Robinson J.A. (2005) Institutions As a Fundamental Cause of Long-Run Growth. *Handbook of Economic Growth 1(A)* (eds. P. Aghion, S.N. Durlauf). Elsevier., pp 385–472. https://doi.org/10.1016/S1574-0684(05)01006-3

Adão R., Carrillo P., Costinot A., Donaldson D., Pomeranz D. (2022) Imports, Exports, and Earnings Inequality: Measures of Exposure and Estimates of Incidence. *The Quarterly Journal of Economics*, 137, 3, pp. 1553–1614. https://doi.org/10.1093/qje/qjac012

Alesina A., Rodrik D. (1994) Distributive Politics and Economic Growth. *The Quarterly Journal of Economics*, 109, 2, pp. 465–490. https://doi.org/10.2307/2118470

Amarante V., Lanzilotta B., Torres J. (2023) *Inequality and Productive Structure: New Evidence at the world Level.* UNU-WIDER Working Paper, 2023/9. https://doi.org/10.35188/UNU-WIDER/2023/317-8

Balland P.-A., Broekel T., Diodato D., Giuliani E., Hausmann R., O'Clery N., Rigby D. (2022) Reprint of the New Paradigm of Economic Complexity. *Research Policy*, 51, 8, pp. 104568. https://doi.org/10.1016/j.respol.2022.104568

Baum C.F., Schaffer M.E., Stillman S. (2003) Instrumental Variables and GMM: Estimation and Testing. *The Stata Journal*, 3, 1, pp. 1–31. https://doi.org/10.1177/1536867X0300300101

Berg A.G., Ostry J.D. (2013) Inequality and Unsustainable Growth: Two Sides of the Same Coin? *International Organization Research Journal*, 8, 4, pp. 77–99. http://dx.doi.org/10. 1057/s41308-017-0030-8

Bergh A., Nilsson Th. (2010) Do Liberalization and Globalization Increase Income Inequality? *European Journal of Political Economy*, 26, 4, pp. 488–505. https://doi.org/10.1016/j.ejpoleco.2010.03.002

Bernal P., Carree M., Lokshin B. (2022) Knowledge Spillovers, R&D Partnerships and Innovation Performance. *Technovation*, 115, pp. 102456. https://doi.org/10.1016/j.technovation.2022.102456

Bollen K.A. (1989) *Structural Equations with Latent Variables.* New York: John Wiley and Sons, Inc. https://doi.org/10.1002/9781118619179

Bustos S., Yıldırım M.A. (2022) Production Ability and Economic Growth. *Research Policy*, 51, 8, pp. 104153. https://doi.org/10.1016/j.respol.2020.104153

Caldarelli G., Cristelli M., Gabrielli A., Pietronero L., Scala A., Tacchella A. (2012) A Network Analysis of Countries' Export Flows: Firm Grounds for the Building Blocks of the Economy. *PLoS One*, 7, 10. https://doi.org/10.1371/journal.pone.0047278

Carvalho L., Rezai A. (2014) *Personal Income Inequality and Aggregate Demand*. Working Paper, 2014–23. Department of Economics, University of São Paulo. São Paulo.

Chancel L., Piketty T., Saez E., Zucman G. et al. (2022) *World Inequality Report.* World Inequality Lab.

Chávez J.C., Mosqueda M.T., Gómez-Zaldívar M. (2017) Economic Complexity and Regional Growth Performance: Evidence from the Mexican Economy. *The Review of Regional Studies*, 47, 2, pp. 201–219. https://doi.org/10.52324/001c.8023

Chu L.K., Hoang D.P. (2020) How Does Economic Complexity Influence Income Inequality? New Evidence from International Data. *Economic Analysis and Policy*, 68(C), pp. 44–57. https://doi.org/10.1016/j.eap.2020.08.004

Cingano F. (2014) *Trends in Income Inequality and its Impact on Economic Growth.* OECD Social, Employment and Migration Working Papers, 163. https://doi.org/10.1787/5jxrjncwxv6j-en

Constantine C., Khemraj T. (2019) Geography, Economic Structures And Institutions: A Synthesis. *Structural Change and Economic Dynamics*, 51(C), pp. 371–379. https://doi.org/10.1016/j.strueco.2019.01.001

Dorn D., Levell P. (2021) *Trade and Inequality in Europe and the US.* IZA Discussion Papers 14914, Institute of Labor Economics (IZA).

Feldman M., Guy F., Iammarino S. (2021) Regional Income Disparities, Monopoly and Finance. *Cambridge Journal of Regions, Economy and Society*, 14, 1, pp. 25–49. https://doi.org/10.1093/cjres/rsaa024

Foster E.M., McLanahan S. (1996) An Illustration of the Use of Instrumental Variables: Do Neighborhood Conditions Affect a Young Person's Chance of Finishing High School? *Psychological Methods*, 1, 3, pp. 249–260. https://doi.org/10.1037/1082-989X.1.3.249

Gao X., Rai V. (2023) Knowledge Acquisition and Innovation Quality: The Moderating Role of Geographical Characteristics of Technology. *Technovation*, 125, pp. 102766. https://doi.org/10.1016/j.technovation.2023.102766

Gómez-Zaldívar M., Osorio-Caballero M.I., Saucedo-Acosta E.J. (2022) Income Inequality and Economic Complexity: Evidence from Mexican States. *Regional Science Policy & Practice*, 6, 14, pp. 344–363. https://doi.org/10.1111/rsp3.12580

Hartmann D. (2014) *Economic Complexity and Human Development: How Economic Diversification and Social Networks Affect Human Agency and Welfare*. New York: Routledge.

Hartmann D., Guevara M.R., Jara-Figueroa C., Aristarán M., Hidalgo C.A. (2017) Linking Economic Complexity, Institutions, and Income Inequality. *World Development*, 93, pp. 75–93. https://doi.org/10.1016/j.worlddev.2016.12.020

Hausmann R., Hidlago C.A., Bustos S., Coscia M., Chung S., Jimenez J., Simoes A., Yildirim M.A. (2011) *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. Massachusetts: Puritan Press.

Hidalgo C.A., Hausmann R. (2009) The Building Blocks of Economic Complexity. *Proceedings of the National Academy of Science of the United States of America*, 106, 26, pp. 10570–5. https://doi.org/10.1073/pnas.0900943106

Kemp-Benedict E. (2014) *An Interpretation and Critique of the Method of Reflections*. MPRA Paper, 60705, pp. 1–16.

Kumhof M., Rancière R., Winant P. (2015) Inequality, Leverage, and Crises. *The American Economic Review*, 105, 3, pp. 1217–1245. http://dx.doi.org/10.1257/aer.20110683

Kuznets S. (1955) Economic Growth and Income Inequality. *The American Economic Review*, 45, 1, pp. 1–28.

Lakner C., Milanovic B. (2016) Global Income Distribution: From the Fall of the Berlin Wall to the Great Recession. *The World Economic Review*, 30, 2, pp. 203–232.

Le Caous E., Huarng F. (2020) Economic Complexity and the Mediating Effects of Income Inequality: Reaching Sustainable Development in Developing Countries. *Sustainability*, 12, 5, pp. 2089. https://doi.org/10.3390/su12052089

Lee C.C., Wang E.Z. (2021) Economic Complexity and Income Inequality: Does Country Risk Matter? *Social Indicators Research*, 154, pp. 35–60. https://doi.org/10.1007/s11205-020-02543-0

Lee K.K., Vu T.V. (2020) Economic Complexity, Human Capital and Income Inequality: A Cross-Country Analysis. *The Japanese Economic Review*, 71, 4, pp. 695–718. https://doi.org/10.1007/s42973-019-00026-7

Lessmann Ch., Seidel A. (2017) Regional Inequality, Convergence, and its Determinants – A View from outer Space. *European Economic Review*, 92(C), pp. 110–132. https://doi.org/10.1016/j.euroecorev.2016.11.009

Malla M.H., Pathranarakul P. (2022) Fiscal Policy and Income Inequality: The Critical Role of Institutional Capacity. *Economies*, 10, 5, pp. 115. https://doi.org/10.3390/economies10050115

Maydeu-Olivares A., Shi D., Fairchild A.J. (2020) Estimating Causal Effects in Linear Regression Models with Observational Data: The Instrumental Variables Regression Model. *Psychological Methods*, 25, 2, pp. 243–258. https://doi.org/10.1037/met0000226

Mihályi P., Szelényi I. (2019) The Place of Rent-Seeking and Corruption in Varieties of Capitalism Models. *Market Liberalism and Economic Patriotism in the Capitalist World-System* (eds. T. Gerőcs, M. Szanyi) International Political Economy Series, Palgrave Macmillan, Cham, pp. 67–97. https://doi.org/10.1007/978-3-030-05186-0_5

Morais B.M., Swart J., Jordaan J.A. (2021) Economic Complexity and Inequality: Does Regional Productive Structure Affect Income Inequality in Brazilian States? *Sustainability*, 13, 2, pp. 1–23. https://doi.org/10.3390/su13021006

Nguyen C.P., Nguyen B.Q., Le Tran D.T. (2023) Economic Complexity and Income Inequality: New Evidence of a Nonlinear Effect. *Social Science Quarterly,* 104, 4, pp. 829–868. https://doi.org/10.1111/ssqu.13281

Nolan B., Richiardi M.G., Valenzuela L. (2019) The Drivers of Income Inequality in Rich Countries. *Journal of Economic Surveys*, 33, 4, pp. 1285–1324.

Ouechtati I. (2023) Financial Inclusion, Institutional Quality, and Inequality: An Empirical Analysis. *Journal of the Knowledge Economy*, 14, 2, pp. 620–644. https://doi.org/10.1007/s13132-022-00909-y

Oyèkọ́lá Ọ. (2023) Life May be unfair, but Do Democracies Make it Any Less Burdensome? *The Journal of Economic Inequality*. https://doi.org/10.1007/s10888-023-09607-4

Pham M.H., Truong H.D.H., Hoang D.P. (2023) Economic Complexity, Shadow Economy, and Income Inequality: Fresh Evidence from Panel Data. *Research Square*. https://doi.org/10.21203/rs.3.rs-2452093/v1

Piketty T., Goldhammer A. (2014) *Capital in the Twenty-First Century*. Harvard University Press.

Rajan R.G. (2010) *Fault Lines: How Hidden Fractures Still Threaten the world Economy*. Princeton: Princeton University Press.

Rikap C., Lundvall B.-A. (2020) Big Tech, Knowledge Predation and the Implications for Development. *Innovation and Development*, 12, 3, pp. 389–416. https://doi.org/10.1080/2157930X.2020.1855825

Sbardella A., Pugliese E., Pietronero L. (2017) Economic Development and Wage Inequality: A Complex System Analysis. *PLOS ONE*, 12, 9, pp. e0182774. https://doi.org/10.1371/journal.pone.0182774

Sepehrdoust H., Tartar M., Gholizadeh A. (2021) Economic Complexity, Scientific Productivity and Income Inequality in Developing Economies. *Economics of Transition and Institutional Change*, 30, 4, pp. 737–752. https://doi.org/10.1111/ecot.12309

Schaffer M.E. (2010) *xtivreg2: Stata Module to Perform Extended IV/2SLS, GMM and AC/HAC, LIML and k-class Regression for Panel Data Models*. http://ideas.repec.org/c/boc/bocode/s456501.html

Solt F. (2020) Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database. *Social Science Quarterly*, 101(3), 1183–1199. https://doi.org/10.1111/SSQU.12795

Stolper W.F., Samuelson P.A. (1941) Protection and Real Wages. *The Review of Economic Studies*, 9, 1, pp. 58–73. https://doi.org/10.2307/2967638

Tacchella A., Cristelli M., Caldarelli G., Gabrielli A., Pietronero L. (2012) A New Metrics for Countries' Fitness and Products' Complexity. *Scientific Reports*, 2, 1, pp. 723. https://doi.org/10.1038/srep00723

Topuz S.G. (2022) The Relationship between Income Inequality and Economic Growth: Are Transmission Channels Effective? *Social Indicators Research*, 162, pp. 1177–1231. https://doi.org/10.1007/s11205-022-02882-0

Violante G.L. (2008) Skill-Biased Technical Change. *The New Palgrave Dictionary of Economics*. London: Palgrave Macmillan. https://doi.org/10.1057/978-1-349-95121-5_2388-1

Vu T. (2022) Does Institutional Quality Foster Economic Complexity? The Fundamental Drivers of Productive Capabilities. *Empirical Economics*, 63, pp. 1571–1604. https://doi.org/10.1007/s00181-021-02175-4

World Bank (2022) *Worldwide Governance Indicators*. https://info.worldbank.org/governance/wgi/

Zhu S., Yu C., He C. (2020) Export Structures, Income Inequality and Urban-Rural Divide in China. *Applied Geography*, 115, pp. 102150. https://doi.org/10.1016/j.apgeog.2020.102150

Zhu Ch. (2022) Conceptualising and Evaluating Inclusive Economic Development: A Productivity Perspective. *Development Studies Research*, 9, 1, pp. 219–229. https://doi.org/10.1080/21665095.2022.2112729

Макроэкономический анализ влияния экономической сложности на неравенство доходов: Какова роль качества институтов?

Давидсон Наталья Борисовна[1], Магон Екатерина Александровна[2], Мариев Олег Святославович[3]

[1] К.э.н., доцент кафедры международной экономики и менеджмента, Институт экономики и управления, Уральский федеральный университет, 19, ул. Мира, Екатеринбург, 620002, Россия.
E-mail: natalya.davidson@gmail.com

[2] Студентка магистратуры, Институт экономики и управления, Уральский федеральный университет
19, ул. Мира, Екатеринбург, 620002, Россия.
E-mail: Volkova2016consta@gmail.com

[3] К.э.н., доцент, заведующий кафедрой экономики, Институт экономики и управления, Уральский федеральный университет, 19, ул. Мира, Екатеринбург, 620002, Россия.
E-mail: o.s.mariev@urfu.ru

В статье рассматривается взаимосвязь между экономической сложностью и неравенством доходов с учетом роли институтов на основе данных за 1996–2020 гг. по 52 развитым и развивающимся странам Европы и Центральной Азии, а также Ближнего Востока и Северной Африки. Вклад исследования состоит, во-первых, в том, что проанализировано влияние экономической сложности на неравенство доходов с учетом качества институтов в целом и отдельных аспектов институционального развития. Во-вторых, учтена нелинейная форма зависимости между экономической сложностью и неравенством доходов, а также разнородность данной взаимосвязи для разных групп стран. Проблема эндогенности решена при помощи двухшагового метода наименьших квадратов с фиксированными эффектами и инструментальными переменными. Полученные результаты свидетельствуют о том, что для общей выборки стран рост экономической сложности в стране приводит к росту неравенства доходов. Вместе с тем влияние экономической сложности на неравенство варьируется по группам стран, а для стран Европы и Центральной Азии эта зависимость имеет форму перевернутой U. При этом в условиях развитости институтов экономическая сложность способствует снижению неравенства. Можно сделать вывод о том, что развитие всех аспектов институтов приводит к снижению неравенства доходов. Также результаты исследования свидетельствуют о том, что снижению уровня неравенства способствует развитие сферы образования. Кроме того, полученные результаты позволяют сделать вывод о необходимости экономической политики, способствующей более равномерному распределению выгод от экономического развития и международной торговли.

Ключевые слова: экономическая сложность; неравенство доходов; качество институтов; экономическое развитие; инструментальные переменные; экономическая политика.

УДК 330.42

DOI: 10.17323/1813-8691-2025-29-1-42-71

О решении детерминированной и стохастической задачи домашнего хозяйства с конечным горизонтом планирования

Пильник Н.П.

В статье на примере оптимизационной задачи домохозяйства, которое принимает решение об объемах потребления и инвестирования, показано, какие сложности возникают в детерминированных и стохастических постановках на конечном временном интервале. В задаче для ее разрешимости на конечном временном интервале добавляется специальное терминальное условие на собственный капитал агента, обобщающее стандартные варианты таких условий.

В статье рассматриваются две постановки. Первая постановка – детерминированный случай, предполагающий, что домохозяйству известны траектории всех экзогенных переменных на всем рассматриваемом временном интервале. Найдено аналитическое решение этой задачи и показано, что за счет выбора параметра терминального ограничения в задаче на конечном временном интервале всегда можно получить траекторию потребления из решения аналогичной задачи, поставленной для бесконечного горизонта планирования. Если же выбирать коэффициент терминального условия так, чтобы оптимальная траектория потребления продолжала предыдущее историческое значение, то при определенном сочетании начальных условий задача домохозяйства может быть либо разрешима только до определенного горизонта планирования, либо быть вообще неразрешима.

Вторая постановка – стохастический случай, когда домохозяйство знает только закон распределения экзогенных переменных. Полное аналитическое решение в этом случае представить не удается, однако предлагается последовательный алгоритм, который позволяет получить пошаговое описание расчета такого решения. Исследование свойств построенной модели позволяет показать, насколько отличается работа со стохастическими оптимизацион-

Пильник Николай Петрович – к.э.н., доцент, Национальный исследовательский университет «Высшая школа экономики»; с.н.с., Физический институт имени П.Н. Лебедева РАН; с.н.с., Научно-исследовательский финансовый институт Минфина России. E-mail: npilnik@hse.ru

Статья поступила: 17.11.2024/Статья принята: 06.02.2025.

ными задачами для анализа отклонений от некоторой выделенной траектории состояний (сбалансированного роста) в ответ на реализацию других состояний (шоков) от задачи анализа конкретных реализовавшихся траекторий переменных агента.

Ключевые слова: терминальные условия; оптимизационная задача домохозяйства; конечный горизонт планирования; условия оптимальности в форме Лагранжа; сбалансированный рост.

Для цитирования: Пильник Н.П. О решении детерминированной и стохастической задачи домашнего хозяйства с конечным горизонтом планирования. *Экономический журнал ВШЭ*. 2025; 29(1): 42–71.

For citation: Pilnik N.P. On the Solution of a Deterministic and Stochastic Household Problem with a Finite Planning Horizon. *HSE Economic Journal.* 2025; 29(1): 42–71. (In Russ.)

1. Введение

Описание поведения экономических агентов в динамических моделях с помощью принципа оптимальности является одним из основных способов объяснения закономерностей изменения количественных переменных в современной экономической теории. Задачи такого типа используются как отдельные блоки более общих моделей – моделей экономического равновесия, представляющих собой популярный инструмент анализа и прогнозирования развития национальной экономики. Чаще всего речь идет о динамических стохастических моделях общего равновесия (DSGE), в которых оптимизационные задачи агентов формулируются в дискретном времени с бесконечным горизонтом планирования, а часть переменных, которые не являются управлением агента, описываются как случайные процессы.

Такие оптимизационные задачи кроме непосредственно функционала, который агент максимизирует, начиная от текущего (в терминах модели) момента времени до горизонта планирования, как правило, включают набор ограничений в виде равенств или неравенств, связывающих переменные модели в один и тот же или в соседние периоды времени. Отдельно следует сказать о граничных условиях таких динамических задач. Начальные условия выглядят вполне естественно с экономической точки зрения: значения переменных типа «запас» наследуются из периода, предшествующего текущему. Конечные (или терминальные) условия, как правило, являются необходимыми с математической точки зрения для обеспечения разрешимости задачи, но оказываются сложнее для экономической интерпретации.

Специфика использования терминальных условий в значительной степени зависит от способа описания времени в модели: является ли рассматриваемый интервал конечным или бесконечным, дискретным или непрерывным. В случае конечного временного множества (и соответственно конечного горизонта планирования) терминальное условие сводится к одному или нескольким соотношениям на значения переменных агента в последний момент времени. В случае бесконечного временного множества терминальное условие приходится формулировать через предельные соотношения.

Задачи агентов в дискретном времени в детерминированной постановке рассматриваются, как правило, на бесконечном временном интервале (см.: [Chang, 1998; Kamihigashi, 2001; Blot, Hayek, 2014]. Главное преимущество таких задач в возможности получения полного аналитического решения. Для этого чаще всего используются условия оптимальности в форме Лагранжа, а терминальное условие, сформулированное как ограничения на возможный рост переменных типа «запас» на бесконечности, позволяет найти начальное значение переменной типа «поток», используемой в качестве аргумента функционала агента. Аналогичный формат терминальных условий используется и при решении оптимизационных задач через уравнение Беллмана как, например, в работах [Wiszniewska-Matyszkiel, 2011; Da Lio, 2000; Tkachev, Abate, 2012].

Задачи агентов в непрерывном времени оказываются сложнее из-за необходимости корректного описания пространства функций, в котором ищется решение. Тем не менее в детерминированной постановке здесь также иногда удается найти решение в аналитическом виде. Задачи с конечным горизонтом планирования удобнее решать через условия оптимальности в форме Лагранжа, как в [Андреев и др., 2007; Pospelov, 2013]. В этом случае оказывается возможным поставить единое терминальное условие на линейную комбинацию фазовых переменных модели, коэффициенты которой определяются в процессе решения. Задачи с бесконечным горизонтом планирования удобнее решать через принцип максимума Понтрягина. И хотя по форме эти условия достаточно схожи с их аналогами в дискретном времени, однако при их использовании возникает целый ряд технических проблем, требующий аккуратного изучения получаемых условий оптимальности, как показано в работах [Aseev et al., 2012; Carlsonet et al., 2012; Лобанов, 1999].

Добавление в задачу агента стохастических переменных существенно усложняет ее исследование и делает практически невозможным получение аналитического решения. При этом терминальное условие остается ее обязательным элементом, однако теперь нужно уточнять, в каком смысле оно должно выполняться (на каждой возможной траектории, в среднем или, например, с заданной вероятностью). Сама формулировка терминального условия в стохастических задачах на бесконечном временном горизонте оказывается весьма нетривиальным вопросом, как показано в [Ljungqvist, 2018; Benveniste, 1982; Brock, 1982].

Большинство современных динамических моделей общего равновесия используют именно стохастическую версию оптимизационных задач. Варианты включаемых в них терминальных условий можно найти в [Ljungqvist, Sargent, 2018; Candless, 2009; Fernández-Villaverde, 2016]. Тем не менее важно отметить, что в большинстве прикладных DSGE-моделей вопрос терминальных условий отдельно не рассматривается в предположении, что используется их стандартная версия, корректность применения которой не всегда очевидна. Это связано со стандартной практикой исследования не исходной, а линеаризованной вокруг траектории сбалансированного роста версии модели. Как мы покажем далее на относительно простом примере, такие переходы могут порождать свои проблемы и не всегда их следует считать корректными.

В этой статье мы рассмотрим оптимизационную задачу домохозяйства, которое принимает решение об объемах потребления и инвестирования на конечном временном интервале, в двух постановках. Первая постановка – детерминированный случай, предполагающий, что домохозяйству известны траектории всех экзогенных переменных на всем рассматриваемом временном интервале. Вторая постановка – стохастический случай, когда домохозяйство знает только закон распределения экзогенных переменных. Детерминированная постановка позволит получить аналитическое решение и подробно проанализировать свойства, которые привносит в модель специфическое условие трансверсальности, изначально описанное в работе [Пильник, Поспелов, 2007]. На примере стохастической постановки модели будет показано, какие сложности, связанные с использованием терминальных условий, могут возникать в задачах стохастической оптимизации и DSGE-моделях.

В статье рассматривается задача только одного экономического агента – домохозяйства. В большинстве современных DSGE-моделей именно этот блок (или один из блоков этого типа) описывается наиболее детально и представляет основную сложность. В этом плане выводы, полученные в статье, относятся не только к таким оптимизационным задачам, но и моделям равновесия в целом.

2. Детерминированный случай

2.1. Постановка задачи домохозяйства

Перейдем к рассмотрению детерминированной постановки оптимизационной задачи. Рассмотрим домохозяйство, которое максимизирует полезность собственного потребления на протяжении последовательности моментов времени из конечного набора $t = 0,\ldots,T-1$.

$$\sum_{t=0}^{T-1} U\left(C_t\right)\beta^t \to \max \tag{2.1}$$

за счет выбора траекторий потребления C_t, объема купленных инвестиционных товаров I_t, остатков денежных средств M_t, объема накопленных сбережений B_t и вложений в капитал K_t в рамках финансового баланса

$$M_{t+1} - M_t = W_t - p_t^C C_t - p_t^I I_t - B_{t+1} + \left(1 + r_t^B\right) B_t + r_t^K p_{t-1}^I K_t,\ t = 0,\ldots,T-1, \tag{2.2}$$

где W_t – экзогенный входящий денежный поток, который можно интерпретировать и как оплату труда, и как объем получаемых дивидендов; p_t^C, p_t^I – цены потребительских и инвестиционных товаров; r_t^B, r_t^K – доходности накопленных сбережений и вложений в капитал. Отметим, что, в отличие от задач в непрерывном времени, при формулировке задач в дискретном времени могут использоваться различные обозначения индекса времени для запасов и потоков. В настоящей статье используется система обозначений, аналогичная [McCandless, 2009; Costa, 2018], которая предполагает, что для запасов индекс времени указывает на начало периода. Таким образом, домохозяйство выбирает переменные типа «поток» C_t, I_t для моментов времени $t = 0,\ldots,T-1$, а переменные типа «запас» M_t, B_t, K_t для моментов времени $t = 1,\ldots,T$, причем значения M_0, B_0, K_0 считаются известными. Отдельно заметим, что в списке начальных условий только запасы, решение задачи агента, как будет показано далее, не будет зависеть от начальных условий потоков.

Например, M_t – объем денежных средств на начало периода t, а M_{t+1} – объем денежных средств на начало периода $t+1$, что эквивалентно объему денежных средств на конец периода t. В изменение этого запаса вносят изменения потоки периода t, например, входящий денежный поток W_t. В этой логике потоки периода t, которые измеряются в ценах базового периода, умножаются на средние индексы цен этого же периода как, например, $p_t^C C_t$, а проценты по запасам начисляются на значения соответствующих запасов на начало периода, т.е. $r_t^B B_t$.

Сложнее обстоит ситуация с расчетом доходности капитала r_t^K, поскольку начисления происходят не на объем капитала K_t, а на стоимость капитала в предыдущем периоде $p_{t-1}^I K_t$, поскольку именно эта сумма отражает объем вложений в этот актив со стороны домохозяйства. При этом поскольку (2.2) может быть записано в том числе для момента $t = 0$, то в задаче появляется произведение $p_{-1}^I K_0$, относящееся не к стартовому периоду $t = 0$, а скорее к предыдущему периоду, предшествующему интервалу планирования, значения переменных в котором считаются заданными. Таким образом, K_0 – начальное значение капитала в рассматриваемой задаче, а p_{-1}^I – начальное значение цен, в которых номинирован этот капитал.

В качестве функции полезности будет использоваться CRRA-функция в виде

$$U(C_t) = \frac{C_t^{1-\alpha}}{1-\alpha}. \tag{2.3}$$

Заметим, что далее, если не уточняется иное, предполагается, что в используемых соотношениях параметр времени t изменяется от 0 до $T-1$, как в (2.2).

Также в задаче ставится динамическое ограничение, связывающее инвестиции и изменение вложений в капитал

$$K_{t+1} - K_t = I_t - \delta K_t, \tag{2.4}$$

где δ – норма амортизации капитала. Наконец, в задачу домохозяйства добавляется ограничение ликвидности, определяющее его спрос на деньги для работы с другими активами

$$M_t \geq \tau^B B_t + \tau^K p_{t-1}^I K_t, \ \ t = 1, \ldots, T-1. \tag{2.5}$$

Поскольку задача решается на конечном временном интервале, то дополнительно ставится терминальное ограничение в виде

$$M_T + a_T^B B_T + a_T^K K_T \geq \gamma\left(M_0 + a_0^B B_0 + a_0^K K_0\right). \tag{2.6}$$

В такой постановке предполагается, что домохозяйство решает свою задачу при известных для всех моментов времени экзогенных переменных $W_t, p_t^C, p_t^I, r_t^B, r_t^K$, причем если переменные W_t, p_t^C, r_t^B, r_t^K известны для всех моментов времени 0 до $T-1$, то для p_t^I в силу (2.2) должно быть задано кроме значений для всех моментов времени 0 до $T-1$ еще и начальное значение, обозначаемое p_{-1}^I.

Заметим, что использование в (2.2) вместо слагаемого $r_t^K p_{t-1}^I K_t$ его аналога $r_t^K p_t^I K_t$ может показаться более естественным, однако это не так. Кроме содержательного соображения, что доходность рассчитывается от объема вложений, т.е. стоимости капитала в предыдущем периоде, задаваемого величиной $p_{t-1}^I K_t$, обратим также внимание на то, что при использовании в этом выражении переменной p_t^I возникает проблема оценки стоимости капитала в последний момент времени K_T, поскольку, как будет показано далее, эта оценка связана с переменной p_T^I, которая относится к моменту времени после рассматриваемого интервала. И если для детерминированной постановки знание этой переменной агентом еще можно предположить, то для рассматриваемой далее стохастической постановки это приводит к явному противоречию.

Отдельно остановимся на терминальном условии (2.6), общую идею постановки которого, основанную на описании динамики первого интеграла поля экстремалей оптимизационной задачи, можно найти в работе [Пильник, Поспелов, 2007]. Там же представлено исследование свойств этого условия в рамках моделей общего равновесия в непрерывном времени. Это условие связывает значение запасов агента (в непрерывной постановке – фазовых переменных) в начальный и конечный моменты времени через заданный коэффициент γ, допустимые значения которого, обеспечивающие разрешимость исходной задачи, обсуждаются в подразделе 2.3.2. В свою очередь коэффициенты $a_T^B, a_T^K, a_0^B, a_0^K$ определяются в процессе решения задачи агента и не требуют задания на этапе ее постановки. В работе [Андреев и др., 2007] показано, что наличие в задаче одного условия такого вида оказывается достаточным для существования решения на конечном временном интервале.

Также отметим, что полученные далее выводы относительно свойств решения оптимизационной задачи домашнего хозяйства верны не только для представленного в рамках соотношений (2.1)–(2.6) наборе переменных. Вообще говоря, эта задача может быть существенно изменена или дополнена другими потоками и запасами (налоги, показатели рынка труда, валютные активы или пассивы и так далее). В рассматриваемом случае для наглядности выбраны наиболее часто используемые в моделях общего равновесия переменные.

2.2. Решение задачи домохозяйства

Для оптимальности траекторий C_t, I_t, M_t, B_t, K_t достаточно, чтобы они доставляли максимум функции Лагранжа

$$
\begin{aligned}
&\sum_{t=0}^{T-1} U(C_t)\beta^t + \xi_t\left(-M_{t+1} + M_t + W_t - p_t^C C_t - p_t^I I_t - B_{t+1} + \left(1 + r_t^B\right)B_t + r_t^K p_{t-1}^I K_t\right) + \\
&+\psi_t\left(-K_{t+1} + K_t + I_t - \delta K_t\right) + \varphi_t\left(M_t - \tau^B B_t - \tau^K p_{t-1}^I K_t\right) + \\
&+\Phi\left(M_T + a_T^B B_T + a_T^K K_T - \gamma\left(M_0 + a_0^B B_0 + a_0^K K_0\right)\right)
\end{aligned}
\tag{2.7}
$$

при выполненных условиях дополняющей нежесткости для учета неравенств (2.5) и (2.6).

(2.8)
$$\varphi_t\left(M_t - \tau^B B_t - \tau^K p_{t-1}^I K_t\right) = 0,\ \varphi_t \ge 0,\ M_t \ge \tau^B B_t + \tau^K p_{t-1}^I K_t,\ t = 1,\ldots,T-1,$$
$$\Phi\left(M_T + a_T^B B_T + a_T^K K_T - \gamma\left(M_0 + a_0^B B_0 + a_0^K K_0\right)\right),$$

(2.9)
$$\Phi \ge 0,\ M_T + a_T^B B_T + a_T^K K_T \ge \gamma\left(M_0 + a_0^B B_0 + a_0^K K_0\right).$$

Выражение (2.7) достигает максимума по траекториям C_t, M_t, K_t, I_t, B_t тогда и только тогда, когда в каждой точке обращаются в ноль производные по C_t, I_t, $t = 0,\ldots,T-1$, производные по M_t, K_t, B_t, $t = 1,\ldots,T-1$, а также производные по M_T, K_T, B_T (момент времени T выделяется отдельно в силу специфичности условий оптимальности для конца интервала планирования). Тогда получим следующие три группы условий, отличающихся моментами времени, для которых они актуальны. Первая группа – условия оптимальности по C_t, I_t, записанные для $t = 0,\ldots,T-1$,

(2.10)
$$U'(C_t)\beta^t - \xi_t p_t^C = 0,$$

(2.11)
$$\psi_t - \xi_t p_t^I = 0.$$

Вторая группа – условия оптимальности по M_t, K_t, B_t, записанные для $t = 0,\ldots,T-2$,

(2.12)
$$-\xi_t + \xi_{t+1} + \varphi_{t+1} = 0,$$

(2.13)
$$\xi_{t+1} r_{t+1}^K p_t^I - \psi_t + \psi_{t+1}(1+\delta) - \varphi_{t+1}\tau^K p_t^I = 0,$$

(2.14)
$$-\xi_t + \xi_{t+1}\left(1 + r_{t+1}^B\right) - \varphi_{t+1}\tau^B = 0.$$

Третья группа – условия оптимальности по M_T, K_T, B_T

(2.15)
$$-\xi_{T-1} + \Phi = 0,$$

(2.16)
$$-\psi_{T-1} + \Phi a_T^K = 0,$$

(2.17)
$$-\xi_{T-1} + \Phi a_T^B = 0.$$

Соотношения (2.2), (2.4), (2.8)–(2.17) образуют полную систему условий для определения решения задачи домохозяйства.

В силу свойств функции полезности $U(C_t)$ (бесконечному значению производной в нуле и положительности при любом конечном значении аргумента) и положительности цен p_t^C из (2.10) следует, что переменная ξ_t положительна при всех $t = 0,\ldots,T-1$. Тогда из (2.15) следует, что

(2.18)
$$\Phi = \xi_{T-1} > 0.$$

В свою очередь из условия (2.11) и пары условий (2.16), (2.17) могут быть найдены коэффициенты терминального ограничения (2.6) перед переменными в последний момент времени T

(2.19) $$a_T^B = 1, \quad a_T^K = p_{T-1}^I.$$

Представляется естественным аналогичным образом задать и значения коэффициентов терминального ограничения (2.6) перед переменными в начальный момент времени

(2.20) $$a_0^B = 1, \quad a_0^K = p_{-1}^I.$$

Отметим, что в соотношении (2.20) появляется индекс $t = -1$, обозначающий, что используется последнее известное значение цен инвестиционных товаров из предыстории. Тогда в силу условия дополняющей нежесткости (2.9) с учетом строго неравенства (2.18) и значений на коэффициенты (2.19) и (2.20) можем переписать терминальное условие (2.6) в виде

(2.21) $$M_T + B_T + p_{T-1}^I K_T = \gamma\left(M_0 + B_0 + p_{-1}^I K_0\right).$$

В силу неотрицательности двойственной переменной φ_t из (2.12) следует, что ξ_t не может возрастать по времени, т.е.

(2.22) $$\varphi_t = \xi_{t-1} - \xi_t \geq 0.$$

Тогда согласно (2.18) и (2.22) можем утверждать, что для всех моментов времени

$$\xi_t > 0.$$

Поэтому для удобства дальнейших преобразований введем переменную ρ_t, равную темпу падения двойственной переменной к финансовому балансу ξ_t, в виде

(2.23) $$\rho_t = -\frac{\xi_t - \xi_{t-1}}{\xi_t}.$$

Кроме того, также для компактности используемых соотношений введем показатели инфляции для инвестиционного и потребительского товара

(2.24) $$i_t^I = \frac{p_t^I - p_{t-1}^I}{p_{t-1}^I}, \; i_t^C = \frac{p_t^C - p_{t-1}^C}{p_{t-1}^C}.$$

Тогда из (2.12), (2.14) и (2.23) получим

(2.25) $$\rho_t = \frac{r_t^B}{1+\tau^B}.$$

В свою очередь из (2.12), (2.13), (2.11) с учетом введенных обозначений (2.23) и (2.24) следует

(2.26) $$\rho_t = \frac{r_t^K + i_t^I - \left(1 + i_t^I\right)\delta}{1 + \tau^K}.$$

Из (2.25) и (2.26) можно видеть, что между переменными r_t^B, r_t^K и i_t^I должна существовать постоянная линейная связь, в противном случае задача домохозяйства окажется неразрешима. Технически такая особенность решения связана с тем, что мы не ставим ограничения на переменные B_t и K_t, содержательно – с тем, что мы изначально хотим найти внутреннее решение, предполагающее одновременную положительность обоих активов. Задачи с ограничениями (причем это могут быть не только требования неотрицательности, но и более сложные ограничения) представляют собой особый интерес, но в данном случае рассматриваться не будут.

Будем считать, что процентная ставка r_t^B строго положительна для всех моментов времени. Тогда из (2.25) следует, что $\rho_t > 0$, а поэтому из (2.23) и (2.22) получаем, что $\varphi_t > 0$. Следовательно, в силу условия дополняющей нежесткости (2.8) ограничение (2.5) будет выполнено как строгое равенство.

(2.27) $$M_t = \tau^B B_t + \tau^K p_{t-1}^I K_t,\ t = 1,\ldots,T-1.$$

В работах [Андреев и др., 2007; Pospelov, 2013] для более общего аналога рассматриваемой задачи показано, что для описания решения задачи агента удобно ввести переменную, представляющую собой линейную комбинацию запасов, находящихся в управлении агента, в виде

(2.28) $$\Omega_t = M_t + B_t + p_{t-1}^I K_t.$$

Как показано в работе [Пильник, Поспелов, 2007], переменная Ω_t может быть интерпретирована как собственный капитал агента. Возможна также ее интерпретация как объем чистых активов, но в случае рассматриваемой задачи это не так наглядно, поскольку в ней отсутствуют пассивы (хотя в качестве кредитов можно рассмотреть случай отрицательной B_t). Тогда финансовый баланс (2.2) с учетом условий оптимальности (2.25), (2.26) и (2.27) может быть записан в весьма компактом виде

(2.29) $$\Omega_{t+1} - \Omega_t = \rho_t \Omega_t + W_t - p_t^C C_t.$$

Уравнение (2.29) описывает причины изменения собственного капитала агента во времени. В случае рассматриваемой задачи увеличение собственного капитала происходит за счет доходности ρ_t и внешнего притока W_t, а уменьшение – за счет расходов на потребление $p_t^C C_t$. Выше мы уже говорили о том, что рассматриваемая задача может быть расширена за счет добавления других переменных. Удобство замены (2.28), основанной на (2.6) и условиях оптимальности (2.19), позволяет и в более общих случаях получить соотношение вида (2.29) и далее работать с ним, а не с исходным набором активов и пассивов агента.

Замена (2.28) также позволяет переписать и терминальное условие (2.21) в виде

(2.30) $$\Omega_T = \Omega_0 \gamma.$$

Таким образом терминальное ограничение (2.6) может интерпретироваться как условие на рост собственного капитала агента в течение периода его планирования.

Условие (2.10) в предположении о виде функции полезности (2.3) позволяет получить динамическое уравнение для потребления

(2.31) $$C_t = C_{t-1} \beta^{\frac{1}{\alpha}} \left(\frac{1+\rho_t}{1+i_t^C} \right)^{\frac{1}{\alpha}}, \ t = 1, \ldots, T-1$$

и выразить траекторию потребления через параметры, экзогенные и двойственные переменные задачи домохозяйства как

(2.32) $$C_t = \tilde{C} \beta^{\frac{t+1}{\alpha}} \left(\prod_{i=0}^{t} \frac{1+\rho_i}{1+i_i^C} \right)^{\frac{1}{\alpha}}.$$

Отдельно следует остановиться на константе $\tilde{C}$. Если бы задача ставилась в непрерывном времени (как, например, в работе [Пильник, Поспелов, 2007]), то константа $\tilde{C}$ соответствовала бы начальному значению управления, которое тоже определяется из задачи агента. В случае задачи в дискретном времени более естественно рассматривать величину C_0 в качестве объема потребления в первом периоде, которая, как можно видеть из (2.32), выражается через константу $\tilde{C}$ следующим образом:

(2.33) $$C_0 = \tilde{C} \beta^{\frac{1}{\alpha}} \left(\frac{1+\rho_0}{1+i_0^C} \right)^{\frac{1}{\alpha}}.$$

Обратим внимание, что если известна предыстория поведения домохозяйства на каком-то временном интервале, предшествующем рассматриваемому в задаче (например, анализируется случай последовательного решения задач аналогичных рассматриваемой), то может быть известна величина C_{-1} – уровень потребления в момент $t = -1$, непосредственно предшествующий определяемому домохозяйством в рамках задачи. Переменная C_t является управлением типа «поток», поэтому для нее значение C_{-1} и константа $\tilde{C}$ из (2.33) совершенно не обязаны совпадать, поэтому отождествлять константу $\tilde{C}$ с начальным значением переменной в случае задачи в дискретном времени, вообще говоря, неверно.

Для того чтобы определить $\tilde{C}$, разрешим (2.29) относительно начального значения собственного капитала

(2.34) $$\Omega_{t+1} = \Omega_0 \prod_{i=0}^{t}(1+\rho_i) + \sum_{i=0}^{t}\left(W_i - p_i^C C_i\right) \prod_{j=i+1}^{t}\left(1+\rho_j\right).$$

Аналогично [Пильник, Поспелов, 2007] введем для удобства параметр $\overline{\gamma}$ в виде

(2.35) $$\overline{\gamma} = 1 - \gamma \prod_{t=0}^{T-1}\frac{1}{1+\rho_t}.$$

Тогда, переписав выражение (2.34) в момент времени $t = T-1$ с учетом (2.30), (2.32) и (2.35), получим

(2.36) $$\tilde{C} = \frac{\Omega_0 \overline{\gamma} + \sum_{t=0}^{T-1} W_t \prod_{i=0}^{t}\frac{1}{1+\rho_i}}{p_{-1}^C \sum_{t=0}^{T-1} \beta^{\frac{t+1}{\alpha}} \left(\prod_{i=0}^{t}\frac{1+\rho_i}{1+i_i^C}\right)^{\frac{1-\alpha}{\alpha}}}.$$

2.3. Исследование решения детерминированной задачи домохозяйства

2.3.1. Задача с бесконечным горизонтом планирования как предельный случай задачи домохозяйства

В этом разделе мы покажем несколько свойств рассматриваемой детерминированной оптимизационной задачи, поставленной на конечном временном интервале, связанных с использованием терминального условия в виде (2.6). Во-первых, покажем его достаточную общность, позволяющую получать как предельный случай граничное условие из задач с бесконечным горизонтом планирования, описанных в работах [McCandless, 2009; Chang, 1998; Kamihigashi, 2001; Blot, Hayek, 2014].

Для этого заметим, что из (2.34) может быть получена удобная форма записи условия трансверсальности для задачи на бесконечном временном горизонте $T \to +\infty$. Для этого перепишем в виде

(2.37) $$\sum_{i=0}^{T-1}\left(p_i^C C_i - W_i\right)\prod_{j=0}^{i}\frac{1}{1+\rho_j} = \Omega_0 - \Omega_T \prod_{i=0}^{T-1}\frac{1}{1+\rho_i}.$$

Аналогично [McCandless, 2009] перейдем к пределу $T \to +\infty$ и потребуем, чтобы рост Ω_T был ограничен в смысле

(2.38) $$\lim_{T \to +\infty} \Omega_T \prod_{t=0}^{T-1}\frac{1}{1+\rho_t} = 0.$$

При этом задача будет разрешима только тогда, когда выполнено неравенство, следующее из (2.34) и неотрицательности расходов на потребление $p_t^C C_t$,

$$\sum_{t=0}^{T-1} W_t \prod_{i=0}^{t} \frac{1}{1+\rho_i} \geq -\Omega_0 \overline{\gamma}, \tag{2.39}$$

которое задает соотношение между начальным условием на Ω_0 и траекторией W_t. Аналог (2.39) в задаче с бесконечным временным горизонтом $T \to +\infty$ будет иметь вид

$$\sum_{t=0}^{+\infty} W_t \prod_{i=0}^{t} \frac{1}{1+\rho_i} \geq -\Omega_0. \tag{2.40}$$

Таким образом, можно утверждать, что задача с бесконечным горизонтом планирования является предельным случаем рассматриваемой задачи с учетом терминального условия(2.38). Важно отметить, что даже в задаче с несколькими запасами (в случае задачи в непрерывном времени – фазовыми переменными) для существования решения, как показано в [Пильник, Поспелов, 2007], оказывается достаточно одного терминального условия.

2.3.2. Воспроизводимость в задаче с конечным горизонтом траекторий из задачи с бесконечным горизонтом

Второй особенностью рассматриваемого терминального условия является возможность в задаче, поставленной на конечном временном интервале, воспроизводить часть оптимальной траектории потребления из задачи с бесконечным горизонтом планирования за счет выбора параметра терминального условия.

Покажем, при каких значениях параметров терминального условия (2.6) в задаче с конечным горизонтом T за счет выбора параметра γ может быть воспроизведена часть траектории управления C_t, полученная в задаче с бесконечным горизонтом. Для этого приравняем соотношение (2.36), записанное для конечного значения T и для $T \to +\infty$ при выполнении условия трансверсальности (2.38). Выразив параметр $\overline{\gamma}$, получим

$$\overline{\gamma}^* = A_T + \frac{1}{\Omega_0}\left(A_T \sum_{t=0}^{+\infty} W_t \prod_{i=0}^{t} \frac{1}{1+\rho_i} - \sum_{t=0}^{T-1} W_t \prod_{i=0}^{t} \frac{1}{1+\rho_i} \right), \tag{2.41}$$

где

$$A_T = \sum_{t=0}^{T-1} \beta^{\frac{t+1}{\alpha}} \left(\prod_{i=0}^{t} \frac{1+\rho_i}{1+i_i^C} \right)^{\frac{1-\alpha}{\alpha}} \Bigg/ \sum_{t=0}^{+\infty} \beta^{\frac{t+1}{\alpha}} \left(\prod_{i=0}^{t} \frac{1+\rho_i}{1+i_i^C} \right)^{\frac{1-\alpha}{\alpha}}. \tag{2.42}$$

Причем выражение (2.41) может принимать и отрицательные значения при соответствующих значениях W_t.

Однако нужно учитывать, что особенностью терминального условия (2.6) являются ограничения на возможные значения параметра γ, при которых задача домохозяйства в принципе разрешима. Заметим, что в том числе ставится вопрос о возможности выбора параметра терминального условия в соответствии с (2.41)–(2.42). Определим, какие допустимые значения может в принципе принимать параметр $\overline{\gamma}$. Минимальное значение $\overline{\gamma}$, соответствующее максимальному значению γ, согласно (2.35), может быть получено из условия, что $\tilde{C} = 0$ (и тем самым согласно (2.32) оптимальный уровень потребления $C_t = 0$) и записывается в виде пары эквивалентных условий

$$\overline{\gamma}_{\min} = -\frac{1}{\Omega_0}\sum_{t=0}^{T-1} W_t \prod_{i=0}^{t} \frac{1}{1+\rho_i}, \tag{2.43}$$

$$\gamma_{\max} = \prod_{t=0}^{T-1}(1+\rho_t) + \frac{1}{\Omega_0}\sum_{t=0}^{T-1} W_t \prod_{i=t+1}^{T-1}(1+\rho_i). \tag{2.44}$$

Минимальное значение γ, соответствующее максимальному значению $\overline{\gamma}$, может быть найдено из условия (2.34), но записанного не от стартового момента ноль, а от конечного момента T так, чтобы $\lim\limits_{t\to+\infty} \Omega_t$ удовлетворял условию (2.38). То есть фактически мы тем самым находим минимальное (возможно отрицательное) значение собственного капитала Ω_T, которое при минимальном потреблении $C_t = 0$ за горизонтом T способно обеспечить выполнение условия трансверсальности (2.38). Таким образом можем записать

$$\gamma_{\min} = -\frac{1}{\Omega_0}\sum_{t=T}^{+\infty} W_t \prod_{i=T}^{t} \frac{1}{1+\rho_i}, \tag{2.45}$$

$$\overline{\gamma}_{\max} = 1 + \frac{1}{\Omega_0}\left(\prod_{i=0}^{T-1}\frac{1}{1+\rho_i}\right)\sum_{t=T}^{+\infty} W_t \prod_{i=T}^{t}\frac{1}{1+\rho_i}. \tag{2.46}$$

Заметим, что если входящий поток W_t отсутствует, то согласно соотношениям (2.43) и (2.46) $\overline{\gamma}$ может принимать значения в отрезке $[0,1]$, что позволяет ее интерпретировать как долю максимально возможного собственного капитала Ω_T (которую можно найти из (2.34) при $C_t = 0$), использованную для потребления.

Покажем, что $\overline{\gamma}^*$ из (2.41) всегда может быть реализована. Заметим, что из (2.42) $A_T \geq 0$. Тогда неравенство $\overline{\gamma}_{\min} \leq \overline{\gamma}^*$ следует из (2.40), (2.41) и (2.43). Если же учесть, что из (2.42) также следует, что $A_T \leq 1$, то из (2.41) и (2.46) следует, что $\overline{\gamma}^* \leq \overline{\gamma}_{\max}$. Таким образом, в рассматриваемой задаче домохозяйства с конечным горизонтом планирования T при любых начальных условиях (при которых существует решение, т.е. выполнено (2.40)) можно подобрать параметр γ так, что траектории потребления C_t и собственного капитала Ω_t (что следует из (2.34)) совпадут с частями траекторий этих же переменных, полученных в задаче с бесконечным горизонтом планирования.

Кроме того, можно сформулировать и более сильное утверждение, следующее из предыдущего вывода. Рассмотрим набор дискретных временных интервалов $\left[\tau_0,T_0\right],\left[\tau_1,T_1\right],\left[\tau_2,T_2\right],\ldots,\left[\tau_n,T_n\right]$, где $\tau_0=0$ и для всех i верно, что $\tau_i\le\tau_{i+1}\le T_i$, т.е. каждый следующий интервал начинается внутри предыдущего. Домохозяйство в каждый момент τ_i решает свою задачу для интервала $\left[\tau_i,T_i\right]$ при заданном значении запасов (или собственного капитала Ω_{τ_i}) и следует полученному решению вплоть до момента τ_{i+1}. В момент τ_{i+1} домохозяйство фиксирует значения своих запасов (или собственного капитала $\Omega_{\tau_{i+1}}$) и перерешивает свою задачу для интервала $\left[\tau_{i+1},T_{i+1}\right]$, после чего последовательность действий повторяется.

В такой постановке оказывается возможным подобрать такие значения параметров терминальных условий $\gamma_0,\gamma_1,\gamma_2,\ldots,\gamma_n$ для каждого из рассматриваемых интервалов, что на отрезке $\left[0,T_n\right]$ будет реализована траектория потребления C_t и собственного капитала Ω_t, совпадающая с решением задачи домохозяйства, изначально сформулированной на отрезке $\left[0,T_n\right]$ с учетом (2.41), (2.42). То есть хотя отдельные участки траектории C_t были получены в разных оптимизационных задачах, связанных между собой только через начальные значения запасов, итоговый результат будет удовлетворять (2.31) для всех $t\in\left[0,T_n\right]$.

2.3.3. Продолжимость предыдущей траектории потребления

Третья особенность рассматриваемой задачи связана с возможностью продолжения предыдущей траектории потребления. В отличие от переменных, характеризующих уровень накопленных запасов M_t,B_t,K_t, значения которых наследуются из предыдущих периодов через задание начальных значений M_0,B_0,K_0, значение потребления из предшествующего планируемому интервалу момента времени в решении вообще не используется. Возникает естественный вопрос, насколько описанное выше свойство преемственности по времени может быть воспроизведено для произвольного стартового значения потребления.

Пусть в периоде, предшествующем периоду ноль, в результате решения домохозяйством своей задачи на интервале $\left[t',T'\right]$, таком что $t'<0\le T'$, было получено значение потребления C_{-1}. Тогда может быть решена обратная задача: из соотношения (2.36) может быть найдено значение параметра $\overline{\gamma}$, обеспечивающее преемственность траектории потребления в смысле $\tilde{C}=C_{-1}$,

$$\overline{\gamma}^{**}=\frac{1}{\Omega_0}\left(C_{-1}p_{-1}^{C}\sum_{t=0}^{T-1}\beta^{\frac{t+1}{\alpha}}\left(\prod_{i=0}^{t}\frac{1+\rho_i}{1+i_i^C}\right)^{\frac{1-\alpha}{\alpha}}-\sum_{t=0}^{T-1}W_t\prod_{i=0}^{t}\frac{1}{1+\rho_i}\right). \quad (2.47)$$

Проверим продолжимость произвольного значения C_{-1}. Для этого согласно (2.43)–(2.46) должно быть выполнено двойное неравенство

(2.48) $$\overline{\gamma}_{\min} \le \overline{\gamma}^{**} \le \overline{\gamma}_{\max}.$$

Левая часть (2.48) в силу неотрицательности C_{-1} по описанному выше построению выполнена всегда. В свою очередь, правая часть (2.48) согласно (2.47) и (2.46) выполняется, только если

(2.49) $$C_{-1} p_{-1}^{C} \sum_{t=0}^{T-1} \beta^{\frac{t+1}{\alpha}} \left(\prod_{i=0}^{t} \frac{1+\rho_i}{1+i_i^C} \right)^{\frac{1-\alpha}{\alpha}} \le \Omega_0 + \sum_{t=0}^{+\infty} W_t \prod_{i=0}^{t} \frac{1}{1+\rho_i}.$$

В силу (2.49) значение C_{-1} может попасть в одну из трех групп. Если

(2.50) $$C_{-1} \le \frac{\Omega_0 + \sum_{t=0}^{+\infty} W_t \prod_{i=0}^{t} \frac{1}{1+\rho_i}}{p_{-1}^{C} \sum_{t=0}^{+\infty} \beta^{\frac{t+1}{\alpha}} \left(\prod_{i=0}^{t} \frac{1+\rho_i}{1+i_i^C} \right)^{\frac{1-\alpha}{\alpha}}},$$

то может быть построена траектория потребления, продолжающая значение C_{-1}, т.е. в смысле (2.33) выполняется $\tilde{C} = C_{-1}$ при любом горизонте T. Заметим, что правая часть неравенства совпадает с (2.36) при $T \to +\infty$.

Если же соотношение (2.50) не выполнено, то это означает, что значение C_{-1} не может быть продолжено на бесконечном горизонте и существует некоторый T^*, при котором (2.49) еще выполняется, а при $T^* + 1$ нарушается. Причем, если выполняется неравенство

$$C_{-1} > \frac{\Omega_0 + \sum_{t=0}^{+\infty} W_t \prod_{i=0}^{t} \frac{1}{1+\rho_i}}{p_{-1}^{C} \beta^{\frac{1}{\alpha}} \left(\frac{1+\rho_0}{1+i_0^C} \right)^{\frac{1-\alpha}{\alpha}}},$$

то это означает, что $T^* = 0$, т.е. такой уровень потребления не может быть реализован ни на одном шаге.

Таким образом, при нарушении соотношения (2.50) в задаче с бесконечным горизонтом планирования ни при каких условиях не может выполняться соотношение (2.31) для момента времени $t = 0$, связывающее решение задачи C_0 с заданным значением C_{-1}, которое могло быть унаследовано от решения аналогичной задачи домохозяйства, но на более «раннем» интервале.

По результатам этого раздела мы можем сделать заключение, что используемое терминальное условие (2.6) в детерминированной оптимизационной задаче не только позволяет рассматривать как частный случай наиболее часто используемую постановку на бесконечном временном интервале, но позволяет воспроизводить в задаче с конечным горизонтом часть будущей траектории потребления из задачи с бесконечным горизонтом за счет выбора параметра терминального условия. Причем такой выбор параметра возможен всегда в отличие от выбора параметра для продолжения предыдущей траектории потребления. Далеко не все начальные значения потребления можно реализовать как решение оптимизационной задачи на бесконечном временном интервале. Более того, существуют такие начальные значения потребления, которые нельзя реализовать даже при решении задачи на один шаг вперед.

3. Стохастический случай

3.1. Постановка задачи домохозяйства

Найдя аналитическое решение задачи домохозяйства в детерминированной постановке и исследовав свойства оптимальных траекторий в зависимости от выбора параметров временного множества и терминального условия, перейдем к рассмотрению стохастической постановки. В целом, техника решения и отдельные выводы будут достаточно сильно перекликаться с предыдущим разделом, однако добавление случайных величин, как будет показано ниже, делают задачу домохозяйства существенно более сложной с точки зрения поиска аналитического решения.

Для описания случайности в оптимизационной задаче будем считать, что в каждый момент времени t случайным образом реализуется состояние s_t из множества возможных вариантов S. Последовательность реализовавшихся состояний от момента 0 до момента t будем обозначать как $s^t = \left(s_0, s_1, \ldots, s_t\right)$, а вероятность ее реализации как $\pi_t\left(s^t\right)$. Тогда можно считать аналогично [Ljungqvist, Sargent, 2018; Fernández-Villaverde et al., 2016], что домохозяйство максимизирует математическое ожидание полезности собственного потребления, которое может быть записано в виде

$$\sum_{t=0}^{T-1}\sum_{s^t} U\left(C_t\left(s^t\right)\right)\pi_t\left(s^t\right)\beta^t \to \max, \tag{3.1}$$

за счет выбора траекторий потребления $C_t\left(s^t\right)$, остатков денежных средств $M_t\left(s^{t-1}\right)$, объема накопленных сбережений $B_t\left(s^{t-1}\right)$ и вложений в капитал $K_t\left(s^{t-1}\right)$ в рамках финансового баланса

$$\begin{aligned} M_{t+1}\left(s^t\right) - M_t\left(s^{t-1}\right) = W_t\left(s^t\right) - p_t^C\left(s^t\right)C_t\left(s^t\right) - p_t^I\left(s^t\right)I_t\left(s^t\right) - \\ -B_{t+1}\left(s^t\right) + \left(1 + r_t^B\left(s^t\right)\right)B_t\left(s^{t-1}\right) + r_t^K\left(s^t\right)p_{t-1}^I\left(s^{t-1}\right)K_t\left(s^{t-1}\right). \end{aligned} \tag{3.2}$$

Аналогично детерминированному случаю ставится динамическое ограничение, связывающее инвестиции и изменение вложений в капитал

(3.3) $$K_{t+1}\left(s^t\right)-K_t\left(s^{t-1}\right)=I_t\left(s^t\right)-\delta K_t\left(s^{t-1}\right),$$

и требование на минимальный объем денежных средств

(3.4) $$M_t\left(s^{t-1}\right)\geq \tau^B B_t\left(s^{t-1}\right)+\tau^K p_{t-1}^I\left(s^{t-1}\right)K_t\left(s^{t-1}\right).$$

Начальные условия для M_0, B_0, K_0 считаются известными и ставится терминальное ограничение в виде

(3.5) $$M_T\left(s^{T-1}\right)+a_T^B\left(s^{T-1}\right)B_T\left(s^{T-1}\right)+a_T^K\left(s^{T-1}\right)K_T\left(s^{T-1}\right)\geq \gamma\left(M_0+a_0^B B_0+a_0^K K_0\right).$$

Обратим внимание, что коэффициенты терминального условия $a_T^B\left(s^{T-1}\right), a_T^K\left(s^{T-1}\right)$ зависят от реализовавшегося состояния, что может показаться несколько странным. Выше в соотношении (2.19) для детерминированной версии модели было показано, что эти коэффициенты при решении задачи приобретают смысл цен соответствующих запасов. Вполне естественно, что в разных состояниях эти цены приобретают разные значения (как, например, в ограничении 3.2)), а устойчивой остается только функциональная связь между этими коэффициентами и экзогенными переменными задачи. В этом смысле ограничение (3.5) можно воспринимать не как одно ограничение, которое должно быть выполнено в каждом состоянии в последний момент времени, а как семейство ограничений, в котором найдется свое для каждого состояния. При этом правая часть (3.5) не зависит от реализовавшегося состояния, поэтому можно считать, что это ограничение представляет собой требование на минимальный объем собственного капитала агента при всех реализуемых в различных состояниях мира значений экзогенных переменных.

В такой постановке предполагается, что домохозяйство решает свою задачу при известных для каждого состояния экзогенных переменных $W_t\left(s^t\right), p_t^C\left(s^t\right), p_t^I\left(s^t\right)$, $r_t^B\left(s^t\right), r_t^K\left(s^t\right)$, но не зная, о конкретном реализовавшемся состоянии, ориентируясь только на его вероятность.

3.2. Решение задачи домохозяйства

3.2.1. Достаточные условия оптимальности задачи

Для решения сформулированной задачи далее будут использоваться условия в форме Лагранжа, адаптированные для задач со стохастическими переменными. Эта техника, в целом, является достаточно стандартной и используется, например, при решении задач агентов в DSGE-моделях (подробнее она описана в [Ljungqvist, Sargent, 2018]). Тем не менее в большинстве современных работ пропускается этап работы с отдельными траекториями состояний мира и выписываются сразу условия оптимальности в терминах математических ожиданий. Но поскольку в настоящей статье рассматривается задача на конечном временном интервале со специфическими терминальными условиями, то далее будет использовано именно подробное описание по траекториям состояний мира.

Необходимые (и в силу вогнутости задачи достаточные) условия могут быть получены за счет вариации Лагранжиана вида (2.7), записанного с учетом двойного суммирования как в (3.1) и учета влияния реализовавшихся состояний на выполнение соотношений (3.2), (3.3), (3.4), а также терминального условия (3.5), по управлениям C_t, M_t, K_t, I_t, B_t. Эти условия для каждой последовательности состояний s^t записываются в виде следующего набора уравнений. Первая группа – условия оптимальности по C_t, I_t, записанные для $t = 0,\ldots,T-1$,

(3.6) $$U'\left(C_t\left(s^t\right)\right)\pi_t\left(s^t\right)\beta^t - \xi_t\left(s^t\right)p_t^C\left(s^t\right) = 0,$$

(3.7) $$\psi_t\left(s^t\right) - \xi_t\left(s^t\right)p_t^I\left(s^t\right) = 0.$$

Вторая группа – условия оптимальности по M_t, K_t, B_t, записанные для $t = 0,\ldots,T-2$

(3.8) $$-\xi_t\left(s^t\right) + \sum_{s^{t+1}|s^t}\xi_{t+1}\left(s^{t+1}\right) + \varphi_{t+1}\left(s^t\right) = 0,$$

(3.9) $$\sum_{s^{t+1}|s^t}\left(\xi_{t+1}\left(s^{t+1}\right)r_{t+1}^K\left(s^{t+1}\right)p_t^I\left(s^t\right) - \psi_{t+1}\left(s^{t+1}\right)\right) + \\ + \psi_t\left(s^t\right)(1+\delta) - \varphi_{t+1}\left(s^t\right)\tau^K p_t^I\left(s^t\right) = 0,$$

(3.10) $$-\xi_t\left(s^t\right) + \sum_{s^{t+1}|s^t}\xi_{t+1}\left(s^{t+1}\right)\left(1 + r_{t+1}^B\left(s^{t+1}\right)\right) - \varphi_{t+1}\left(s^t\right)\tau^B = 0.$$

Третья группа – условия оптимальности по M_t, K_t, B_t

(3.11) $$-\xi_{T-1}\left(s^{T-1}\right) + \Phi\left(s^{T-1}\right) = 0,$$

(3.12) $$-\psi_{T-1}\left(s^{T-1}\right) + \Phi\left(s^{T-1}\right)a_T^K\left(s^{T-1}\right) = 0,$$

(3.13) $$-\xi_{T-1}\left(s^{T-1}\right) + \Phi\left(s^{T-1}\right)a_T^B\left(s^{T-1}\right) = 0.$$

Соотношения (3.2), (3.3), (3.6)–(3.13), а также аналоги условий дополняющей нежесткости (2.8), (2.9) образуют полную систему условий для определения решения задачи домохозяйства в этой постановке.

Аналогично рассуждениям (2.18)–(2.20) можем на основании (3.11), (3.12) и (3.13) последовательно записать

(3.14) $$\Phi\left(s^{T-1}\right) = \xi_{T-1}\left(s^{T-1}\right) > 0,$$

(3.15) $$a_T^B\left(s^{T-1}\right) = 1,\ a_T^K\left(s^{T-1}\right) = p_{T-1}^I\left(s^{T-1}\right).$$

И как следствие из (3.14) и (3.15) получим аналогичное (2.21) терминальное условие, которое должно быть выполнено для всех возможных конечных состояний s^{T-1}

$$M_T\left(s^{T-1}\right)+B_T\left(s^{T-1}\right)+p_{T-1}^I\left(s^{T-1}\right)K_T\left(s^{T-1}\right)=\gamma\left(M_0+B_0+p_{-1}^I K_0\right). \tag{3.16}$$

Заметим, что соотношение (3.16) выполняется в любом случае, независимо от предыдущих состояний. Дело в том, что в момент $T-1$ агент уже знает значения всех экзогенных переменных, поэтому для него уже не остается неопределенности, и он направит все свои свободные в смысле (3.16) средства на увеличение потребления.

Аналогичное (2.23) выражение для доходности в случае стохастической задачи может быть записано для каждой пары последовательных состояний

$$\rho_t\left(s^t\right)=-\frac{\xi_t\left(s^t\right)-\xi_{t-1}\left(s^{t-1}\right)}{\xi_t\left(s^t\right)}. \tag{3.17}$$

Тогда с учетом (3.8), (3.9), (3.7) и (3.17) можно записать

$$\rho_t\left(s^t\right)=\frac{r_t^B\left(s^t\right)}{1+\tau^B}. \tag{3.18}$$

А с учетом (2.24), (3.8), (3.7), (3.10) и (3.17)

$$\rho_t\left(s^t\right)=\frac{r_t^K\left(s^t\right)+i_t^I\left(s^t\right)-\left(1+i_t^I\left(s^t\right)\right)\delta}{1+\tau^K}. \tag{3.19}$$

Таким образом, аналогично (2.25) и (2.26) из соотношений (3.18) и (3.19) снова для разрешимости задачи агента требуется связь между переменными r_t^B, r_t^K и i_t^I, выполненная отдельно для каждого состояния s^t.

С учетом использованной ранее замены (2.28) на основании соотношений (3.16), (3.18) и (3.19) финансовый баланс (3.2) может быть записан как уравнение на собственный капитал домохозяйства

$$\Omega_{t+1}\left(s^t\right)=\left(1+\rho_t\left(s^t\right)\right)\Omega_t\left(s^{t-1}\right)+W_t\left(s^t\right)-p_t^C\left(s^t\right)C_t\left(s^t\right), \tag{3.20}$$

а терминальное условие (3.5) переписано в виде

$$\Omega_T\left(s^{T-1}\right)=\Omega_0\gamma. \tag{3.21}$$

3.2.2. Алгоритм решения задачи

Далее опишем алгоритм получения полного решения задачи домохозяйства. Рассмотрим уравнение (3.20), записанное в последний момент $T-1$

$$(3.22)\quad \Omega_T(s^{T-1}) = \left(1+\rho_{T-1}\left(s^{T-1}\right)\right)\Omega_{T-1}\left(s^{T-2}\right) + W_{T-1}\left(s^{T-1}\right) - p^C_{T-1}\left(s^{T-1}\right)C_{T-1}\left(s^{T-1}\right).$$

Из (3.22) может быть получено выражение для величины собственного капитала $\Omega_{T-1}\left(s^{T-2}\right)$ через следующее значение $\Omega_T\left(s^{T-1}\right)$, которое с учетом (3.21) записывается в виде

$$(3.23)\qquad \Omega_{T-1}\left(s^{T-2}\right) = \frac{\gamma\Omega_0 + W_{T-1}\left(s^{T-1}\right) - p^C_{T-1}\left(s^{T-1}\right)C_{T-1}\left(s^{T-1}\right)}{1+\rho_{T-1}\left(s^{T-1}\right)}.$$

Соотношение (3.23) может быть записано для разных состояний s^{T-1}, произвольную несовпадающую пару из которых для примера обозначим как s_j^{T-1} и s_k^{T-1}. Тогда в предположении, что каждому из этих состояний предшествовало одно и то же состояние s^{T-2} из (3.23), можем записать связь

$$(3.24)\qquad \begin{gathered} \frac{\gamma\Omega_0 + W_{T-1}\left(s_j^{T-1}\right) - p^C_{T-1}\left(s_j^{T-1}\right)C_{T-1}\left(s_j^{T-1}\right)}{1+\rho_{T-1}\left(s_j^{T-1}\right)} = \\ = \frac{\gamma\Omega_0 + W_{T-1}\left(s_k^{T-1}\right) - p^C_{T-1}\left(s_k^{T-1}\right)C_{T-1}\left(s_k^{T-1}\right)}{1+\rho_{T-1}\left(s_k^{T-1}\right)}. \end{gathered}$$

И таким образом из (3.24) можем выразить

$$(3.25)\qquad \begin{aligned} C_{T-1}\left(s_j^{T-1}\right) &= \frac{\left(1+\rho_{T-1}\left(s_j^{T-1}\right)\right)p^C_{T-1}\left(s_k^{T-1}\right)}{\left(1+\rho_{T-1}\left(s_k^{T-1}\right)\right)p^C_{T-1}\left(s_j^{T-1}\right)}\cdot C_{T-1}\left(s_k^{T-1}\right) + \\ &+ \frac{\rho_{T-1}\left(s_k^{T-1}\right) - \rho_{T-1}\left(s_j^{T-1}\right)}{1+\rho_{T-1}\left(s_k^{T-1}\right)}\cdot\frac{\gamma\Omega_0}{p^C_{T-1}\left(s_j^{T-1}\right)} + \\ &+ \frac{1}{p^C_{T-1}\left(s_j^{T-1}\right)}\cdot\left(W_{T-1}\left(s_j^{T-1}\right) - \frac{1+\rho_{T-1}\left(s_j^{T-1}\right)}{1+\rho_{T-1}\left(s_k^{T-1}\right)}\cdot W_{T-1}\left(s_k^{T-1}\right)\right). \end{aligned}$$

Соотношение (3.6) с учетом (3.8) и (3.17) может быть переписано в виде

$$(3.26)\qquad \begin{gathered} \frac{U'\left(C_{T-2}\left(s_i^{T-2}\right)\right)\pi_{T-2}\left(s_i^{T-2}\right)}{p^C_{T-2}\left(s_i^{T-2}\right)} = \\ = \sum_{s_j^{T-1}}\left(1+\rho_{T-1}\left(s_j^{T-1}\right)\right)\frac{\beta U'\left(C_{T-1}\left(s_j^{T-1}\right)\right)\pi_{T-1}\left(s_j^{T-1}\right)}{p^C_{T-1}\left(s_j^{T-1}\right)}. \end{gathered}$$

Или с учетом (2.3) можно переписать (3.26) как

$$C_{T-2}\left(s^{T-2}\right)=\beta^{-\frac{1}{\alpha}}\left(\sum_{s_j^{T-1}}\frac{1+\rho_{T-1}\left(s_j^{T-1}\right)}{1+i_{T-1}^{C}\left(s_j^{T-1}\right)}\pi_{T-1}\left(s_j^{T-1}\mid s^{T-2}\right)\left(C_{T-1}\left(s_j^{T-1}\right)\right)^{-\alpha}\right)^{-\frac{1}{\alpha}}, \tag{3.27}$$

где

$$\pi_{T-1}\left(s_j^{T-1}\mid s^{T-2}\right)=\frac{\pi_{T-1}\left(s_j^{T-1}\right)}{\pi_{T-2}\left(s^{T-2}\right)}. \tag{3.28}$$

После этого на основе (3.27) с учетом (3.25) и (3.28) получается набор соотношений, который при заданных для каждого состояния s_j^{T-1} при фиксированном s^{T-2} значениях $\rho_{T-1}\left(s_j^{T-1}\right)$, которые определяются из экзогенных переменных согласно (3.18) и (3.19), позволяет при известном $C_{T-2}\left(s^{T-2}\right)$ найти каждое последующие $C_{T-1}\left(s_j^{T-1}\right)$. Кроме того, с использованием тех же (3.25), (3.27) и (3.28) соотношение (3.23) может быть переписано как зависимость $\Omega_{T-1}\left(s^{T-2}\right)$ от $C_{T-2}\left(s^{T-2}\right)$, экзогенных переменных $W_{T-1}\left(s^{T-1}\right)$, $p_{T-1}^{C}\left(s^{T-1}\right)$ и уже определившейся согласно (3.18) $\rho_{T-1}\left(s^{T-1}\right)$.

Далее эта процедура повторяется последовательно для каждого предыдущего момента времени. Аналогично (3.22) для момента $T-2$ выписывается соотношение на собственный капитал

$$\begin{aligned}\Omega_{T-1}\left(s^{T-2}\right)=&\left(1+\rho_{T-2}\left(s^{T-2}\right)\right)\Omega_{T-2}\left(s^{T-3}\right)+\\&+W_{T-2}\left(s^{T-2}\right)-p_{T-2}^{C}(s^{T-2})C_{T-2}\left(s^{T-2}\right),\end{aligned} \tag{3.29}$$

из которого аналогично (3.23) выражается $\Omega_{T-2}\left(s^{T-3}\right)$, позволяющее аналогично (3.24) записать соотношение, связывающее $C_{T-2}\left(s^{T-2}\right)$ для всех вариантов s^{T-2}. А соотношение, аналогичное (3.26), позволяет выразить каждое из них через $C_{T-3}\left(s^{T-3}\right)$. Причем благодаря полученному на предыдущем этапе соотношению, которое выражает $\Omega_{T-1}\left(s^{T-2}\right)$ через $C_{T-2}\left(s^{T-2}\right)$, после преобразования (3.29) в нем останется только значение собственного капитала в последний момент времени, значение которого известно из (3.21).

После повторения описанных выше итераций до момента 1 получим соотношение

$$\Omega_1\left(s^0\right)=\left(1+\rho_0\left(s^0\right)\right)\Omega_0+W_0\left(s^0\right)-p_0^{C}\left(s^0\right)C_0\left(s^0\right), \tag{3.30}$$

в котором $\Omega_1\left(s^0\right)$, согласно проведенным на предыдущих этапах преобразованиям, выражается через $C_0\left(s^0\right)$. Полученное из (3.30) итоговое соотношение для $C_0\left(s^0\right)$ будет включать начальное значение собственного капитала Ω_0, параметр γ и траектории $W_t\left(s^t\right)$, $p_t^C\left(s^t\right)$, $r_t^B\left(s^t\right)$ для всех возможных последовательностей состояний $s^0,\ldots,s^{T-1}$.

В итоге решением задачи агента является программа действий агента, которая в каждый момент времени t в зависимости от текущего состояния мира и значения запасов (или собственного капитала Ω_t) определяет уровень потребления $C_t\left(s^t\right)$.

3.3. Исследование решения стохастической задачи домохозяйства

3.3.1. Компактная запись условий оптимальности

Заметим, что умножив записанное для каждой комбинации состояний соотношение (3.26) на соответствующую вероятность и сложив все полученные соотношения, можно записать часто используемое выражение

$$\frac{U'(C_t)}{p_t^C}=\beta\,\mathrm{E}_t\frac{(1+\rho_{t+1})U'(C_{t+1})}{p_{t+1}^C}. \tag{3.31}$$

Аналогичный прием для уравнения (3.20) позволяет переписать через математические ожидания уравнения на динамику собственного капитала агента

$$\mathrm{E}_t\Omega_{t+2}=\Omega_{t+1}\mathrm{E}_t\left(1+\rho_{t+1}\right)+\mathrm{E}_t\left(W_{t+1}-p_{t+1}^C C_{t+1}\right). \tag{3.32}$$

Если к соотношениям (3.31) и (3.32) добавить терминальное условие (3.21), считая, что переменная ρ_t известна из условий оптимальности, то мы получим систему, полностью описывающую динамику потребления. Однако в таком виде исследовать ее оказывается крайне неудобно из-за того, что в (3.31) под математическим ожиданием с учетом (2.3) в знаменателе стоит степенная функция от переменной C_{t+1}, а в (3.32) – линейная функция от той же переменной. Поэтому в DSGE-моделях (см.: [McCandless, 2009; Fernández-Villaverde et al., 2016]), как правило, переходят к линейным приближениям исходных уравнений.

3.3.2. Аналог траектории сбалансированного роста в задаче домохозяйства

Линеаризация, как правило, проводится вокруг так называемой траектории сбалансированного роста, предполагающей, что в модели отсутствуют шоки, а все агенты решают свои задачи с учетом этой информации. То есть в терминах рассматриваемой нами задачи потребителя множество состояний в каждый момент времени t сокращается до одного (будем обозначать его $\bar{s}_t$), а сама задача решается исходя из предположения о том, что последовательность реализованных состояний $\bar{s}^{T-1}$ известна.

Заметим, что на выбор состояний, описывающих сбалансированный рост, часто накладываются дополнительные требования. Например, в случае рассматриваемой задачи утверждение об отсутствии шоков может быть записано как набор условий на математическое ожидание экзогенных переменных:

$$W_t\left(\overline{s}^t\right) = E_0 W_t\left(s^t\right), \tag{3.33}$$

$$p_t^C\left(\overline{s}^t\right) = E_0 p_t^C\left(s^t\right),\ p_t^I\left(\overline{s}^t\right) = E_0 p_t^I\left(s^t\right), \tag{3.34}$$

$$r_t^K\left(\overline{s}^t\right) = E_0 r_t^K\left(s^t\right),\ r_t^B\left(\overline{s}^t\right) = E_0 r_t^B\left(s^t\right). \tag{3.35}$$

При этом, естественно, предполагается, что в каждый момент времени t среди исходного набора состояний существует последовательность состояний $\overline{s}^t$, удовлетворяющая соотношениям (3.33)–(3.35).

Сопоставим решение детерминированной задачи при последовательности реализованных состояний $\overline{s}^{T-1}$ и решение стохастической задачи. С учетом соотношений (3.25) и (3.27), записанных для произвольно момента времени t, можем записать для стохастической задачи

$$C_{t-1}\left(\overline{s}^{t-1}\right) = \beta^{-\frac{1}{\alpha}}\left(\sum_{s_j^t}\frac{1+\rho_t\left(s_j^t\right)}{1+i_t\left(s_j^t\right)}\pi_t\left(s_j^t\,|\,\overline{s}^{t-1}\right)\left(\begin{array}{l}\dfrac{\left(1+\rho_t\left(s_j^t\right)\right)p_t^C\left(\overline{s}^t\right)}{\left(1+\rho_t\left(\overline{s}^t\right)\right)p_t^C\left(s_j^t\right)}\cdot C_t\left(\overline{s}^t\right)+\\ +\dfrac{\rho_t\left(\overline{s}^t\right)-\rho_t\left(s_j^t\right)}{1+\rho_t(\overline{s}^t)}\cdot\dfrac{\gamma\Omega_0}{p_t^C(s_j^t)}+\\ +\dfrac{1}{p_t^C\left(s_j^t\right)}\cdot\left(W_t\left(s_j^t\right)-\dfrac{1+\rho_t\left(s_j^t\right)}{1+\rho_t\left(\overline{s}^t\right)}\cdot W_t\left(\overline{s}^t\right)\right)\end{array}\right)^{-\alpha}\right)^{-\frac{1}{\alpha}}. \tag{3.36}$$

Аналогичное соотношение для детерминированной задачи, следующее из (2.31), записывается в виде

$$C_{t-1}\left(\overline{s}^{t-1}\right) = C_t\left(\overline{s}^t\right)\beta^{-\frac{1}{\alpha}}\left(\frac{1+\rho_t\left(\overline{s}^t\right)}{1+i_t^C\left(\overline{s}^t\right)}\right)^{-\frac{1}{\alpha}}. \tag{3.37}$$

Выполнение одного из соотношений из пары (3.36) и (3.37) совершенно не означает выполнения второго, поскольку (3.36) содержит большее число произвольных значений экзогенных переменных. Тем не менее можно показать, что существуют такие комбинации этих экзогенных переменных, когда соотношения (3.36) и (3.37) окажутся эквиваленты. Рассмотрим частный случай, предполагающий

$$W_t\left(s_j^t\right)=W_t\left(\overline{s}^t\right)=0,\ \gamma=0. \tag{3.38}$$

Тогда (3.36) перепишется в виде соотношения

$$C_{t-1}\left(\overline{s}^{t-1}\right)=C_t\left(\overline{s}^t\right)\beta^{-\frac{1}{\alpha}}\left(\sum_{s_j^t}\frac{1+\rho_t\left(s_j^t\right)}{1+i_t\left(s_j^t\right)}\pi_t\left(s_j^t\mid s^{t-1}\right)\left(\frac{\left(1+\rho_t\left(s_j^t\right)\right)p_t^C\left(\overline{s}^t\right)}{\left(1+\rho_t\left(\overline{s}^t\right)\right)p_t^C\left(s_j^t\right)}\right)^{-\alpha}\right)^{-\frac{1}{\alpha}}, \tag{3.39}$$

которое с учетом (3.37) позволяет записать

$$\left(\frac{1+\rho_t\left(\overline{s}^t\right)}{1+i_t^C\left(\overline{s}^t\right)}\right)^{1-\alpha}=\sum_{s_j^t}\left(\frac{1+\rho_t\left(s_j^t\right)}{1+i_t^C\left(s_j^t\right)}\right)^{1-\alpha}\pi_t\left(s_j^t\mid s^{t-1}\right) \tag{3.40}$$

или в терминах математического ожидания

$$\left(\frac{1+\rho_t\left(\overline{s}^t\right)}{1+i_t^C\left(\overline{s}^t\right)}\right)^{1-\alpha}=E_{t-1}\left(\frac{1+\rho_t\left(s^t\right)}{1+i_t^C\left(s^t\right)}\right)^{1-\alpha}. \tag{3.41}$$

То есть сбалансированный рост в рамках рассматриваемой постановки может быть реализован в качестве решения исходной стохастической задачи агента только при условии, что совместное распределение экзогенных переменных удовлетворяет определенным соотношениям, частным случаем которых является набор (3.38), (3.41). Во всех остальных случаях оказывается, что при произвольном совместном распределении экзогенных переменных сбалансированный рост не является решением стохастической задачи агента, допускающей реализацию ровно тех же состояний, которые определяют сбалансированный рост. Таким образом, корректность линеаризации, вообще говоря, должна проверяться для каждой конкретной модели и каждого распределения случайных переменных отдельно.

3.3.3. Обсуждение подходов к анализу решения стохастической задачи

Для оценки воздействия отдельных шоков (т.е. реализаций состояния s_t, отличающегося от $\overline{s}_t$) обычно проводится анализ функций импульсного отклика (IRF – подробнее, например, в [Fernández-Villaverde et al., 2016]), т.е. рассчитываются отклонения траекторий, полученных при некотором произвольном наборе состояний, от траекторий сбалансированного роста при одинаковом наборе начальных значений и параметров. Если траектория сбалансированного роста может быть реализована в качестве оптимальной траектории в стохастической задаче, то этот прием позволяет игнорировать терминальное условие (3.21), определяющее начальный уровень потребления $\tilde{C}$ по аналогии с (2.33), и далее работать со стохастической динамической системой, состоящей из соотношений (3.31) и (3.32). Если же траектория сбалансированного роста не может быть реализована подобным образом, то возникнет систематическое расхождение, связанное с несовпадающими начальными значениями по потреблению $\tilde{C}$. Заметим при этом, что при такого рода анализе предполагается, что агент решает свою задачу в начальный момент времени ноль и действует в соответствии с построенной стратегией для всех моментов времени $t = 0,\ldots,T-1$.

Выше при рассмотрении детерминированной задачи мы отмечали, что уравнение (2.31), описывающее динамику потребления, верно при $t = 0,\ldots,T-1$. Но при выбранном в соответствии с (2.41) и (2.42) параметре граничного условия γ и в предположении, что до начального момента времени ноль поведение агента определялось такой же задачей с параметрами, выбранными аналогичным образом, уравнение (2.31) может быть записано и для $t = 0$. То есть именно в этом случае система, состоящая из уравнения для потребления (2.31), выполненного для $t = 1,\ldots,T-1$, уравнения для собственного капитала (2.29), выполненного для $t = 0,\ldots,T-1$, и выражения для начального значения потребления (2.36), может быть заменена динамической системой, состоящей из двух уравнений: (2.29) и (2.31), выполненных для всех $t = 0,\ldots,T-1$.

Похожий прием может быть использован при записи уравнений для сбалансированного роста на основе (3.31) и (3.32), где распространение соотношения (3.31) на момент времени $t = 0$ осуществляется в предположении, что как минимум в предыдущий момент времени траектории переменных также соответствовали сбалансированному росту. Причем поскольку для остальных состояний эту пару уравнений удобно переписать в терминах отклонений от сбалансированного роста, то и начальные условия у этой динамической системы оказываются известны и равны нулю, поскольку Ω_0 фиксирована, а $\tilde{C}$ зависит только от распределения экзогенных переменных, но не зависит от конкретной реализации. И это полностью соответствует, например, стандартной практике исследования сбалансированного роста и функций импульсного отклика в DSGE-моделях, как, например, в [Fernández-Villaverde et al., 2016], соотношения которых могут быть переписаны в виде VAR-моделей.

Однако предположение о том, что до момента $t = 0$ траектория потребления соответствовала сбалансированному росту, не выглядит естественно, когда речь заходит о повторном решении своей задачи домохозяйством. Если потребление в предыдущий момент времени (выше мы его обозначали $t = -1$) не совпадает со значением на сбалансированном росте (а вероятность этого ненулевая, причем в случае бесконечного множества возможных состояний в этот момент она будет равна единице), то записывать для момента $t = 0$ соотношение (2.31) для сбалансированного роста или (3.31) для стохастической постановки некорректно. Причем, как показано в разделе 2.3, подбор параметров граничного условия может решить эту проблему только если выполнено (2.50), что, вообще говоря, совершенно неочевидно для произвольного распределения шоков, а тем более в постановках, где шоки описываются нормальным распределением и могут принимать произвольные значения от минус до плюс бесконечности.

Еще более сложной становится проблема связи начального модельного и последнего исторического значения потребления в ситуации, когда домохозяйство заново решает свою задачу каждый раз, когда поступает новая информация о реализации случайных экзогенных величин. Из-за того что начальное значение потребления не будет удовлетворять ни уравнению (2.31), ни уравнению (3.31), мы получим, что на фактической реализации эти уравнения не будут выполнены ни для одной точки. Таким образом, переход от системы соотношений, включающей (3.31), (3.32) и начальное условие, получаемое согласно алгоритму из раздела 2.2, к динамической системе, состоящей только из (3.31), распространенного на момент $t = 0$, и (3.32), является удобным приемом для анализа воздействия отдельных шоков на отклонение от сбалансированного роста, но порождает содержательное противоречие при анализе конкретных реализаций переменных модели, что становится особенно важно, например, для решения задач оценки параметров или прогнозирования.

4. Заключение

В статье на примере оптимизационной задачи домохозяйства, которое принимает решение об объемах потребления и инвестирования, показано, какие сложности возникают в детерминированных и стохастических постановках на конечном временном интервале. В задаче для ее разрешимости на конечном временном интервале добавляется специальное терминальное условие на собственный капитал агента, обобщающее стандартные варианты таких условий.

Первая постановка – детерминированный случай, предполагающий, что домохозяйству известны траектории всех экзогенных переменных на всем рассматриваемом временном интервале. Найдено аналитическое решение этой задачи и показано, что за счет выбора параметра терминального ограничения в задаче на конечном временном интервале всегда можно получить траекторию потребления из решения аналогичной задачи, поставленной для бесконечного горизонта планирования. Если же выбирать коэффициент терминального условия так, чтобы оптимальная траектория потребления продолжала историческую траекторию, то при определенном сочетании начальных условий задача домохозяйства может быть либо разрешима только до определенного горизонта планирования, либо быть вообще неразрешима.

Вторая постановка – стохастический случай, когда домохозяйство знает только закон распределения экзогенных переменных. Полное аналитическое решение в этом случае представить не удается, однако предлагается последовательный алгоритм, который дает пошаговое описание расчета такого решения. Показано, что сбалансированный рост в рамках рассматриваемой постановки может быть реализован в качестве решения исходной стохастической задачи агента только при условии, что совместное распределение экзогенных переменных удовлетворяет определенным соотношениям. Во всех остальных случаях оказывается, что при произвольном совместном распределении экзогенных переменных сбалансированный рост может не являться решением стохастической задачи агента, что позволяет сделать вывод, что корректность линеаризации вокруг сбалансированного роста, вообще говоря, должна проверяться для каждой конкретной модели и каждого распределения случайных переменных отдельно.

Еще более сложной становится проблема связи начального модельного и последнего исторического значения потребления в ситуации, когда домохозяйство заново решает свою задачу каждый раз, когда поступает новая информация о реализации случайных экзогенных величин. Показано, что переход от системы уравнений, описывающих решение задачи домохозяйства, к динамической системе, является удобным приемом для анализа воздействия отдельных шоков на отклонение от сбалансированного роста, но порождает содержательное противоречие при анализе конкретных реализаций переменных модели, что становится особенно важно, например, для решения задач оценки параметров или прогнозирования.

Исследование свойств решений обеих оптимизационных задач позволяет поставить вопрос о том, что дает учет стохастики в экономических моделях, использующих принцип оптимальности, если их аналитическое решение оказывается невозможным, а методы приближенного анализа (линеаризации) настолько модифицируют исходную модель, что из нее пропадают нелинейные эффекты, являющиеся главным следствием добавления случайных величин в модель.

* *
*

СПИСОК ЛИТЕРАТУРЫ

Андреев М.Ю., Поспелов И.Г., Поспелова И.И., Хохлов М.Ю. Новая технология моделирования экономики и модель современной экономики России. М.: МИФИ, 2007.

Лобанов С.Г. К теории оптимального экономического роста // Экономический журнал ВШЭ. 1999. Т. 3. № 1. С. 28–41.

Пильник Н.П., Поспелов И.Г. О естественных терминальных условиях в моделях межвременного равновесия // Экономический журнал ВШЭ. 2007. Т. 11. № 1. С. 3–34.

Aseev S.M., Besov K.O., Kryazhimskii A.V. Infinite-Horizon Optimal Control Problems in Economics // Russian Mathematical Surveys. 2012. Vol. 67. № 2. P. 195.

Benveniste L.M., Scheinkman J.A. Duality Theory for Dynamic Optimization Models of Economics: The Continuous Time Case // Journal of Economic Theory. 1982. Vol. 27. № 1. P. 1–19.

Blot J., Hayek N. Infinite-Horizon Optimal Control in the Discrete-Time Framework. New York: Springer, 2014.

Brock W.A. Asset Prices in a Production Economy // The Economics of Information and Uncertainty. University of Chicago Press, 1982. P. 1–46.

Carlson D.A., Haurie A.B., Leizarowitz A. Infinite Horizon Optimal Control: Deterministic and Stochastic Systems. Springer Science & Business Media, 2012.

Chang R. Credible Monetary Policy in an Infinite Horizon Model: Recursive Approaches // Journal of Economic Theory. 1998. Vol. 81. № 2. P. 431–461.

Costa C. Understanding dsge Models: Theory and Applications. Vernon Press, 2018.

Da Lio F. On the Bellman Equation for Infinite Horizon Problems with Unbounded Cost Functional // Applied Mathematics and Optimization. 2000. Vol. 41. P. 171–197.

Ekeland I., Scheinkman J.A. Transversality Conditions for Some Infinite Horizon Discrete Time Optimization Problems // Mathematics of Operations Research. 1986. Vol. 11. № 2. P. 216–229.

Fernández-Villaverde J., Rubio-Ramírez J.F., Schorfheide F. Solution and Estimation Methods for DSGE Models // Handbook of Macroeconomics. Elsevier, 2016. Vol. 2. P. 527–724.

Kamihigashi T. Necessity of Transversality Conditions for Infinite Horizon Problems // Econometrica. 2001. Vol. 69. № 4. P. 995–1012.

Ljungqvist L., Sargent T.J. Recursive Macroeconomic Theory. MIT Press, 2018.

McCandless G. The ABCs of RBCs: An Introduction to Dynamic Macroeconomic Models. Harvard University Press, 2009.

Pospelov I.G. Intensive Quantities in an Economy and Conjugate Variables // Mathematical Notes. 2013. Vol. 94. P. 146–156.

Tkachev I., Abate A. Regularization of Bellman Equations for Infinite-Horizon Probabilistic Properties // Proceedings of the 15th ACM international conference on Hybrid Systems: computation and control. 2012. P. 227–236.

Wiszniewska-Matyszkiel A. On the Terminal Condition for the Bellman Equation for Dynamic Optimization with an Infinite Horizon // Applied Mathematics Letters. 2011. Vol. 24. № 6. P. 943–949.

On the Solution of a Deterministic and Stochastic Household Problem with a Finite Planning Horizon

Nikolay Pilnik

National Research University Higher School of Economics,
11, Pokrovsky Boulevard, Moscow, 109028, Russian Federation.
E-mail: npilnik@hse.ru

The article uses the example of an optimization problem of a household that makes a decision on the volumes of consumption and investment to show what difficulties arise in deterministic and stochastic formulations on a finite time interval. In order to make the problem solvable on a finite time interval, a special terminal condition on the agent's equity capital is added, generalizing the standard versions of such conditions.

The article considers two settings. The first setting is a deterministic case, assuming that the household knows the trajectories of all exogenous variables over the entire time interval under consideration. An analytical solution to this problem is found and it is shown that by choosing the parameter of the terminal constraint in the problem on a finite time interval, it is always possible to obtain a consumption trajectory from the solution of a similar problem set for an infinite planning horizon. If the coefficient of the terminal condition is chosen so that the optimal consumption trajectory continues the previous value, then with a certain combination of initial conditions, the household's problem can either be solvable only up to a certain planning horizon, or be completely unsolvable.

The second statement is a stochastic case, when the household knows only the distribution law of exogenous variables. In this case, it is not possible to provide a complete analytical solution, but a sequential algorithm is proposed that allows one to obtain a step-by-step description of the calculation of such a solution. The study of the properties of the constructed model allows one to show how different the work with stochastic optimization problems for the analysis of deviations from a certain selected trajectory of states (balanced growth) in response to the implementation of other states (shocks) is from the problem of analyzing specific realized trajectories of the agent's variables.

Key words: terminal conditions; household optimization problem; finite planning horizon; optimality conditions in Lagrange form; balanced growth.

JEL Classification: C61.

* *
*

References

Andreev M.Yu., Pospelov I.G., Pospelova I.I., Khokhlov M.Yu. (2007) *New Technology for Modeling the Economy and a Model of the Modern Economy of Russia*. M.: MEPhI.

Aseev S.M., Besov K.O., Kryazhimskii A.V. (2012) Infinite-Horizon Optimal Control Problems in Economics. *Russian Mathematical Surveys*, 67, 2, pp. 195.

Benveniste L.M., Scheinkman J.A. (1982) Duality Theory for Dynamic Optimization Models of Economics: The Continuous Time Case. *Journal of Economic Theory*, 27, 1, pp. 1–19.

Blot J., Hayek N. (2014) *Infinite-Horizon Optimal Control in the Discrete-Time Framework*. New York: Springer.

Brock W.A. (1982) Asset Prices in a Production Economy. *The Economics of Information and Uncertainty*. University of Chicago Press, pp. 1–46.

Carlson D.A., Haurie A.B., Leizarowitz A. (2012) *Infinite Horizon Optimal Control: Deterministic and Stochastic Systems*. Springer Science & Business Media.

Chang R. (1998) Credible Monetary Policy in an Infinite Horizon Model: Recursive Approaches. *Journal of Economic Theory*, 81, 2, pp. 431–461.

Costa C. (2018) *Understanding dsge Models: Theory and Applications*. Vernon Press.

Da Lio F. (2000) On the Bellman Equation for Infinite Horizon Problems with Unbounded Cost Functional. *Applied Mathematics and Optimization*, 41, pp. 171–197.

Ekeland I., Scheinkman J.A. (1986) Transversality Conditions for Some Infinite Horizon Discrete Time Optimization Problems. *Mathematics of Operations Research*, 11, 2, pp. 216–229.

Fernández-Villaverde J., Rubio-Ramírez J.F., Schorfheide F. (2016) Solution and Estimation Methods for DSGE Models. *Handbook of Macroeconomics*. Elsevier, 2, pp. 527–724.

Kamihigashi T. (2001) Necessity of Transversality Conditions for Infinite Horizon Problems. *Econometrica*, 69, 4, pp. 995–1012.

Ljungqvist L., Sargent T.J. (2018) *Recursive Macroeconomic Theory*. MIT Press.

Lobanov S.G. (1999) On the Theory of Optimal Economic Growth. *HSE Economic Journal*, 3, 1, pp. 28–41.

McCandless G. (2009) *The ABCs of RBCs: An Introduction to Dynamic Macroeconomic Models*. Harvard University Press.

Pilnik N.P., Pospelov I.G. (2007) On Natural Terminal Conditions in Models of Intertemporal Equilibrium. *HSE Economic Journal*, 11, 1, pp. 3–34.

Pospelov I.G. (2013) Intensive Quantities in an Economy and Conjugate Variables. *Mathematical Notes*, 94, pp. 146–156.

Tkachev I., Abate A. (2012) *Regularization of Bellman Equations for Infinite-Horizon Probabilistic Properties*. Proceedings of the 15th ACM international conference on Hybrid Systems: computation and control, pp. 227–236.

Wiszniewska-Matyszkiel A. (2011) On the Terminal Condition for the Bellman Equation for Dynamic Optimization with an Infinite Horizon. *Applied Mathematics Letters*, 24, 6, pp. 943–949.

УДК 338.1, 338.2, 332.1

DOI: 10.17323/1813-8691-2025-29-1-72-102

The Impact of Carbon Tax and Research Subsidies on Economic Growth in Japan[1]

Galina Besstremyannaya[1], Richard Dasher[2], Sergei Golovan[3]

[1] National Research University Higher School of Economics,
11, Pokrovsky Blvd., Moscow, 109028, Russian Federation.
E-mail: gbesstremyannaya@hse.ru

[2] Stanford University,
521, Memorial Way, Stanford, California, 94305, USA.
E-mail: rdasher@stanford.edu

[3] New Economic School,
3, Nobelya st., Skolkovo Innovation Center, Moscow, 121205, Russian Federation.
E-mail: sgolovan@nes.ru

A considerable amount of work has shown that a carbon tax combined with research subsidies may be regarded as effective policy for encouraging the spread of low-carbon technologies for the benefit of society. This paper exploits the macroeconomic approach of endogenous growth models with technological change in order to make a comparative assessment of the impact of such policy measures on economic growth in the US and Japan in the medium and long term. Our estimates

[1] The article was prepared with the support of the Basic Research Program of HSE University, Moscow, Russia. We are indebted to the Editor, an anonymous referee, Petros Sofronis, Dmitry Veselov, Konstantin Styrin, Pavel Krivenko, and participants of the 2019 EEA – ESEM Congress and of public seminars at Stanford University's US-Asia Technology Management Center (July 2023 and March 2024), at HSE University's International Laboratory for Macroeconomic Analysis (April-May 2019) and at the International Institute for Carbon-Neutral Energy Research (Kyushu University, Japan, May 2017) for helpful comments. Path dependence as one of the key features of the [Acemoglu et al., 2016] model, the need to touch upon the contribution of several countries to carbon emissions, measured at any meteorological station, and potential technology adoption across countries were kindly noted by the anonymous referee.

Galina Besstremyannaya – Senior researcher at the International Laboratory for Macroeconomic Analysis, Professor at the Department of Applied Economics.
Richard Dasher – Director, US-Asia Technology Management Center at Stanford University.
Sergei Golovan – Docent.

The article was received: 11.06.2024/The article is accepted for publication: 06.02.2025.

with the micro and macro data reveal similarities among Japanese and US energy firms as regards the elasticity of the innovation production function in R&D expenditure and the probability of radical innovation. However, according to energy patent statistics, clean innovation is not as wide-spread in Japan as it is in the US. This may explain our quantitative findings of the need for a stronger reliance on a carbon tax in Japan as opposed to the US.

Key words: endogenous growth; technological change; innovation; carbon tax; energy.

JEL Classification: O11, O13, O47, Q43, Q49.

For citation: Besstremyannaya G., Dasher R., Golovan S. The Impact of Carbon Tax and Research Subsidies on Economic Growth in Japan. *HSE Economic Journal.* 2025; 29(1): 72–102.

1. Introduction

Confronting pollution and mitigating climate change has long been on the agenda in many developed countries, notably in the Nordic part of the EU and Japan [Gokhale, 2021; Fujii, Managi, 2016; International Energy Agency, 2016]. In particular, Japan might be viewed as a country with a long history of public and private environmental initiatives in support of eco-innovation which led to the decrease of carbon emissions from 1990 to 2022 [Green House Inventory Office, 2024]. Since 2003, Japan has been implementing a strategic energy policy, which addresses various technology issues related to energy efficiency as well as concerns about emissions and the environment [Ministry of Economy, Trade and Industry, 2014]. In 2012, Japan was among the first Asian countries to introduce a carbon tax on consumers as a part of the concept of "greening the Japanese tax system" within the fourth energy plan [Ministry of the Environment, 2017]. The tax is intended to encourage the use of green technologies by households and firms. Revenues from the carbon tax and other energy taxes are used to provide subsidies to develop environmentally-friendly (clean) technologies [Ministry of Finance, 2010, 2015; Wakiyama & Zusman, 2016]. However, Japan continues to contribute a third of the world's carbon emissions, and the size of the carbon tax is considerably lower than in other developed countries or than the IMF recommended value [Gokhale, 2021].

The use of a carbon tax as the sole policy instrument causes only the slow propagation of carbon-neutral technologies [Popp et al., 2010]. Accordingly, there is a need for other governmental policies (most commonly, in the form of research subsidies) that help offset the cost of firm's R&D targeted at eco-innovation. Since the 1990s, several generations of R&D-based endogenous growth models have been applied for the analyses of the impact of the carbon tax and research subsidies in Japan. Firstly, these are models with carbon-emitting (dirty) technology in the energy sector [Goto & Sawa, 1993; Goto, 1995; Mizunoya & Higano, 2000]. Variations of the approach include models with many industries [Lee et al., 2012; Takeda & Arimura, 2021; Matsumura et al., 2024] or with monetary policy [Hamaguchi, 2024]. The models do not explicitly include both carbon-neutral (clean) and carbon-emitting (dirty) sectors but may use the concept of energy efficiency (e.g. as in [Goto, 1995]) which should lead to a decrease in energy use and hence a fall in carbon emissions. Secondly, there are general equilibrium models with clean and dirty sectors [Lee at al., 2022; Silva Herran & Fujimori, 2021] but there exists no competition between clean and dirty technologies. Finally, there are approaches to studying the effect of carbon taxes in Japan through computable equilibrium models with aggregate-level regression analysis [Ministry of the Environment, 2017].

To the best of our knowledge, the richer, more realistic and more modern model of monopolistic competition [Klett, Kortum, 2004] as well as the concept of competing carbon-emitting and carbon-neutral technologies [Acemoglu et al., 2012; Aghion et al., 2016] have not yet been applied to the analysis of Japanese environmental policy. The purpose of this paper is to provide a quantitative estimate of the effects of carbon emissions and research subsidies on economic growth in Japan within such a framework.

We employ the Acemoglu et al.'s (2016) model which is unique in the class of the Klette & Kortum (2004) endogenous growth models with clean and dirty sectors in four aspects: 1) the potential use of clean or dirty technology for each product of a firm, 2) technological change in the clean and dirty sectors, where innovation may be radical (breakthrough) or incremental, 3) habit formation (path dependence) in the firm's choice of clean or dirty technology, and 4) richness in the use of microdata, including firm's financials, the R&D production function and patent data. Specifically, the elasticity of the R&D production function, quality differences between carbon-emitting and carbon-neutral technologies, and various parameters on firm dynamics are taken from real world data on companies and their patents.

The empirical analysis goes beyond the traditional assessment of macroeconomic policy in the Japanese energy sector, as the methodology of the [Acemoglu et al., 2016] model, which we use, uniquely allows for technological changes within the clean and dirty sectors. We exploit large datasets on Japanese manufacturing corporations and national data on their patents in clean and dirty technologies over the last quarter century to numerically evaluate the size of the clean and dirty sectors. Next, we follow the endogenous growth model by [Acemoglu et al., 2016] and empirically estimate the scenario with a combination of a carbon tax and research subsidies as regards the impact of these policy instruments on innovation rates and economic output in the carbon-emitting and carbon-neutral sectors. The results of our micro analysis with firm-level and patent-level data reveal many similarities between Japanese and US innovative energy firms as regards the elasticity of output with respect to R&D inputs, the probability of radical innovation and the number of patents per firm product in the energy sector, the share of R&D expenditure in firm sales, and the share of R&D labor in total labor. However, there exists an important difference between innovative firms in the two countries: in Japan there is a higher labor productivity with dirty technology in comparison with clean technology (the difference in productivity is called the technology gap in [Acemoglu et al., 2016]). This may explain our quantitative finding of a stronger reliance on the carbon tax in Japan in comparison with the US and the relatively longer period of a temporary fall in overall economic output due to the introduction of policy measures in Japan.

The paper builds on several streams of literature dealing with clean and dirty technologies in Japan[2]. Microeconomic analyses have studied the impact of the carbon tax (and research

[2] Concerning the microeconomic literature in the international context, evidence on the impact of policy instruments on innovation in the energy sector as well as a meta-review of research focused on carbon emissions and technological change in the energy sector are given in [Popp et al., 2010]. A few analyses suggest that the choice of environmentally friendly technologies are linked to energy prices and a history of a firm's innovative activity [Aghion et al., 2016; Popp, 2006; Popp & Newell, 2012]. As for policy instru-

subsidies) on production and emissions in Japan, and on energy use and investment in energy efficiency [Nakata & Lamont, 2001; Wakiyama & Zusman, 2016; Aden & Dirir, 2023]. Another focus of the microeconomic research is the impact of green innovation and R&D expenditure on firm growth [Kimura, 2023; Iino et al., 2021; Adebayo & Kirikkaleli, 2021; Hosono et al., 2022; Suzuki & Takemura, 2016]. Consumer willingness to pay for green products and the search for the optimal mix of clean/dirty energy, especially after the 2011 Great East Japan Earthquake are also on the agenda of microeconomic research in Japan [Komiyama & Fujii, 2017; Nomura &Akai, 2004]. Studies in the microeconomic context show a behavioral response of firms and consumers to market mechanisms and regulatory actions in the field of energy economics [De Groot et al., 2001; Tanikawa, 2004].

Concerning the preceding macroeconomic literature, here we list the streams of earlier studies along the lines of the reviews in Aghion et al. (2016) and Acemoglu et al. (2016), and add applications of such approaches to research on Japan. Firstly, there are models with the carbon cycle and the cost of carbon emissions [Nordhause, 2008; Golosov et al., 2014], as well as models with technological change in the energy sector [Smulders & de Nooji, 2003; Hassler et al., 2012] that enable researchers to assess the impact of a carbon tax on economic growth. There are similar models applied to the Japanese economy [Matsumura et al., 2024; Hamaguchi, 2024; Goto & Salva, 1993]. The second stream of research are models with clean and dirty sectors and technological change in each sector [Acemoglu et al., 2012; Gans, 2012; Golosov et al., 2014]. Analyses of the Japanese economy under such an approach may be found in Lee et al. (2022) and Silva Herran & Fujimori (2021). Thirdly, there are models allowing for competition between clean and dirty technologies [Acemoglu et al., 2012; Aghion et al., 2016].

Overall, the findings of macroeconomic analyses show that regulations aimed at reducing carbon emissions lead to a decline of gross domestic product and/or its growth rate in many countries [Metz et al., 2007, Table 3.12; Jorgenson & Wilcoxen, 1990]. Using revenues obtained from carbon taxes for the development of carbon-neutral technologies may mitigate the problem of a GDP decrease. Reviews of the literature on links between economic growth, carbon emissions and government policies can be found in [Xepapadeas, 2005; Jorgenson et al., 1993].

The remainder of the paper is organized as follows. Section 2 outlines the model of Acemoglu et al. (2016) which is employed in the empirical part of the paper. Data on Japan is descrybed in Section 3. Section 4 proceeds with the quantification of the model parameters for Japan. The results of the policy analysis for the Japanese economy as well as the robustness of estimates and the limitations of the approach are given in Section 5. The final section contains a discussion of the results and overall conclusions of the analysis.

The paper is essentially a replication of Acemoglu et al.'s (2016) article for Japan. As regards the replication of the analysis with the endogenous growth model based on US data for Japan, we believe that although not focused on the clean or dirty sectors, the work closest to ours is Kodama & Li (2019), who employ Aghion et al.'s (2019) model with heterogeneous innovation.

ments, a carbon tax combined with research subsidies may be regarded as an optimal policy for minimizing carbon emissions and/or maximizing social welfare [Fischer & Newell, 2008; Gerlagh & Van der Zwaan, 2006; Popp, 2006].

2. The Acemoglu et al.'s (2016) model

2.1. Key insights into the theoretical framework

Overview of the model and research question. The [Acemoglu et al., 2016] model belongs to the literature on competing clean and dirty sectors in the economy [Acemoglu et al., 2012; Gans, 2012] but the novelty of the model is the consideration of clean and dirty technology for each product as well as the path dependence of clean or dirty innovation. The model builds on the key concepts of endogenous growth models with technological change: the firm offering the best quality owns the market for the relevant intermediate good (product) [Grossman & Helpman, 1990; Romer, 1990]; firms innovate to maximize profits by adding new products or improving the quality of existing products [Klette & Kortum, 2004; Lentz & Mortensen, 2008]. Following the seminal work by [Akcigit & Kerr, 2010], the innovation in the model is heterogeneous: it may be radical (breakthrough) or incremental. The model assumes that all innovations are patented and the economy is closed.

The [Acemoglu et al., 2016] model is used as a tool to find the optimal values for a combination of two policy instruments: subsidies for research into carbon-neutral technologies and a tax on carbon emissions. The model studies the evolution of a non-steady state equilibrium, focusing on the time profiles of economic variables across optimal policies and the *laissez-faire* (null) policy. The variables of primary interest are output by firms using carbon-neutral and carbon-emitting technologies, innovative activity by clean and dirty firms, and the overall growth of the economy. The key features of the [Acemoglu et al., 2016] model is summarized in the following.

Market competition between firms with clean and dirty technologies. The model employs the [Klette & Kortum, 2004] framework of the monopolistic competition of firms who produce a continuum of intermediate goods in the economy. There are dirty (carbon-emitting) and clean (carbon-neutral) technologies for each good and a quality ladder as regards the labor productivity of clean and dirty technologies. Only firms with the most advanced technology within the clean and dirty sector may produce each good. There are incumbent firms who aim at sustaining their monopoly position in the market for each intermediate good and new entrants whose purpose is to enter the market through quality competition. There is a free entry condition to the market. Entrants and incumbents have a number of goods and technologies and are engaged in monopolistic competition.

Profit-maximizing firms invest in R&D expecting a resulting innovation that will boost the quality (labor productivity) of the technology used in manufacturing the good. There is also a quality gap (difference in labor productivity) between the two technologies for each good. The "active" technology (i.e. clean or dirty) used in the market for each good is chosen based on the marginal cost which depends on taxes and the price of exhaustible resources (employed within the dirty technology). A firm's decision on R&D is influenced by the R&D subsidy from the government.

Heterogeneous innovation. Similarly to the approach of [Akcigit & Kerr, 2010], there are two types of innovation in the model. Incremental innovation leads to minor advances in the current leading-edge technology. So incremental clean innovation improves the current best clean technology and incremental dirty innovation enhances the best dirty technology. Radical (breakthrough) innovation causes an advance over the leading-edge technology, regardless of whether it is clean or dirty.

The path dependence of innovation. A key feature of the model is the path dependence of the firm's choice of each technology – namely, the fact the type of firm's innovative activity (i.e. clean or dirty) depends on the firm's past innovative activity, and that clean innovation, if sustained for a while, is self-reinforcing[3]. There is a stock of knowledge within each technology (clean or dirty) which a firm can employ for further quality improvements. It is conjectured for simplicity "that each firm specializes in either clean or dirty technologies" [Acemoglu et al., 2016, p. 63].

Although this plausible supposition about firm behavior has been already put forward in the microeconomics literature [Popp, 2006; Popp & Newell, 2012], to the best of our knowledge, the work by [Acemoglu et al., 2016] may be regarded as the first macroeconomic model with a path dependency of clean/dirty innovation[4]. The introduction of the carbon cycle in the model stems from the literature with the general equilibrium framework and most closely resembles the approach in [Golosov et al., 2014].

R&D decision of firms. The application of the [Klette & Kortum, 2004] framework means that R&D expenditure leads to a Poisson flow rate of new innovation. A subsidy for clean or dirty technology enhances a firm's investment in the corresponding type of R&D. The R&D decision of firms is path dependent.

Final good. The producer of the aggregate (final) good uses intermediate goods as inputs. The producer of the final good is a profit maximizer, so the choice of a clean or dirty intermediate good is conducted according to their relative quality (labor productivity) and the size of the carbon tax on the dirty intermediate good.

Policy instruments and their impact. The government collects carbon taxes, imposes taxes on consumers to balance its budget and provides R&D subsidies to clean/dirty technologies. The policy instruments are the carbon tax and research subsidies. The research subsidy enhances R&D investment into technology. The carbon tax affects the choice of technology used by firms that produce intermediate goods as well as impacting the choice of clean or dirty intermediate goods by the final good producer. There is also the welfare damage of carbon emissions as they increase the amount of carbon in the atmosphere.

The welfare effect of dirty technology. The use of dirty technologies leads to carbon emissions that impact production and social welfare. Specifically, carbon-emissions cause economic damage, decreasing the productivity of the final good.

2.2. Formal description of the model

This section borrows from pp. 57–71 of Acemoglu et al. (2016) to provide a list of the key equations of the model which point to the key parameters used in the empirical estimation.

Consumption. The utility function of a representative household is $U_0 = \int_0^{\infty} e^{-\rho t} \ln C_t dt$, where C_t is consumption at time t and $\rho > 0$ is the social discount rate. The household works in all the firms in the economy, as well as owning them, so the budget constraint becomes

[3] We are grateful to the anonymous referee for highlighting path dependency as one of major features of the model [Acemoglu et al., 2016].

[4] Path dependency is a key element of another seminal macroeconomic model which also appeared in 2016 [Aghion et al., 2016].

$w_t^u + w_t^s L^s + \Pi_t - T_t \geq C_t$, where Π_t is net profit, w_t^u and w_t^s are the wage rates for unskilled (with measure one) and skilled (with measure L^s) labor respectively (skilled labor does R&D), and T_t is transfers.

Intermediate goods. Each intermediate good $y_{i,t}$ can be produced with either dirty or clean technology, denoted by $j \in \{c,d\}$. (Below the terms "dirty/clean technology" and "dirty/clean sector" are often used interchangeably.) The production function of firm f for clean technology is

(1) $$y_{i,t}^c(f) = q_{i,t}^c(f) l_{i,t}^c(f),$$

where $l_{i,t}^c(f)$ is labor and $q_{i,t}^c(f)$ is labor productivity.

The production function for dirty technology uses not only labor but also an exhaustible resource $e_{i,t}$:

(2) $$y_{i,t}^d(f) = q_{i,t}^d(f) l_{i,t}^d(f)^{1-\nu} e_{i,t}(f)^{\nu}$$

where $e_{i,t} = \zeta l_{i,t}^e$, $\nu \in (0,1)$, $\zeta > 0$ and the stock of the exhaustible resource is given by

(3) $$\dot{R}_t = -\int_0^1 e_{i,t} di.$$

The solution of the cost minimization problem by producers within the dirty sector yields

(4) $$e_{i,t} = \left(\frac{\nu}{1-\nu}\frac{w_t^u}{p_{e,t}}\right) l_{i,t}^d.$$

Denote $\tilde{P}_{e,t}$ as the normalized price of the exhaustible resource:

(5) $$\tilde{P}_{e,t} \equiv (1-\nu)^{\nu-1}\left(\frac{p_{e,t}}{w_t^u \nu}\right)^{\nu}.$$

Carbon cycle. The atmospheric carbon concentration S_l is due to carbon emissions accumulating in the atmosphere. It negatively affects the aggregate production Y_t as follows:

(6) $$lnY_t = -\gamma\left(S_t - \overline{S}\right) + \int_0^1 lny_{i,t} di,$$

where $\overline{S}$ is the preindustrial level of the atmospheric carbon concentration, γ is a parameter with positive values reflecting the negative impact, and $y_{i,t}$ is the amount of intermediate good i. The amount of atmospheric carbon concentration S_t, if $t = T$ is the date when emission began is expressed as:

(7) $$S_t = \int_0^{t-T} (1 - d_l) K_{t-l} dl ,$$

where carbon emission K_t is proportionate to the total output of dirty sectors $Y_t^d = \int_0^1 y_{i,t}^d di$:

(8) $$K_t = \kappa Y_t^d$$

and $(1 - d_l)$ is the share of a unit of carbon, emitted l years ago and left in the atmosphere, and:

(9) $$d_l = (1 - \varphi_P)(1 - \varphi_0 e^{-\varphi l}) ,$$

with φ_P denoting the fraction of emission permanently remaining in the atmosphere; $(1 - \varphi_P)$ being the fraction of the transitory component in the first period; φ indicating the rate of decay of the carbon concentration.

Heterogeneous innovation. R&D expenditure targeted at the clean or dirty sectors may lead to successful innovation. The aggregate innovation rate by entrants and incumbents is denoted as z_t^j. The innovation can be of two types: an incremental innovation (which occurs with probability $1 - \alpha$) and a breakthrough innovation (with probability α).

An incremental innovation implies an improvement of one step in the quality ladder and this is modelled as a proportional improvement of quality by $\lambda > 1$. An incremental innovation during Δt causes the creation of a new technology $q_{i,t+\Delta t}^c = \lambda q_{i,t}^c$, but a breakthrough innovation in sector j that is behind sector $-j$ leads to a new technology with $q_{i,t+\Delta t}^j = \lambda q_{i,t}^{-j}$, meaning that it builds on the more advanced technology level of sector $-j$. An incremental innovation in the clean sector implies that labor productivity is $q_{i,t}^c = \lambda^{n_{i,t}^c}$, where a positive integer $n_{i,t}^c$ denotes the number of steps that this technology has taken since time $t = 0$. A similar equation holds for the dirty sector, $q_{i,t}^d = \lambda^{n_{i,t}^d}$.

The relative productivity of the dirty to clean technology is given as $\frac{q_{i,t}^d}{q_{i,t}^c} = \lambda^{n_{i,t}}$, where $n_{i,t} \equiv n_{i,t}^d - n_{i,t}^c$ is the technology gap between the dirty and clean sectors for intermediate good i at time t.

The price-adjusted policy gap m_t is derived as

(10) $$m_t \equiv \frac{1}{\ln\lambda} \left[\ln\left(\frac{1 + \tau_t^d}{1 + \tau_t^c} \tilde{P}_{e,t} \right) \right],$$

where $\tilde{P}_{e,t}$ is the price for exhaustible resource and τ_t^j are taxes.

Price and quantity of intermediate goods. Using the final good as a numeraire, the demand for the intermediate good can be expressed as:

(11) $$y_{i,t} = \frac{Y_t}{p_{i,t}}.$$

Let $\tilde{q}_{i,t}^{d}$ and $\tilde{q}_{i,t}^{c}$ be defined as the qualities in the dirty (clean) sector, which are adjusted by taxes and the price of the exhaustible resource in the case of the dirty (clean) sector as follows:

(12) $$\tilde{q}_{i,t}^{d} \equiv \frac{q_{i,t}^{d}}{\left(1+\tau_t^{d}\right)\tilde{P}_{e,t}} \text{ and } \tilde{q}_{i,t}^{c} \equiv \frac{q_{i,t}^{c}}{1+\tau_t^{c}}.$$

The profit maximization problem of the producers of intermediate goods in the dirty (clean) sector yields:

(13) $$p_{i,t}^{j} = min\left\{\frac{\lambda w_t^{u}}{\tilde{q}_{i,t}^{j}}, \frac{w_t^{u}}{\tilde{q}_{i,t}^{-j}}\right\} \text{ and } y_{i,t}^{j} = max\left\{\frac{\tilde{q}_{i,t}^{j}}{\lambda w_t^{u}}, \frac{\tilde{q}_{i,t}^{-j}}{w_t^{u}}\right\} \cdot Y_t.$$

Denote $\bar{Q}_t \equiv \exp\left(\int \ln \tilde{q}_{it} di\right)$ as the quality index of tax-adjusted labor productivities, and $\Lambda_t \equiv \prod_n \lambda\left(n-m_t\right)^{-\mu_{n,t}}$ as an inverse function of equilibrium markups.

Plugging the optimal values of the intermediate goods in the production function of the final good yields:

(14) $$w_t^{u} = \bar{Q}_t \Lambda_t.$$

R&D production function. Denote u_f^{j} as the number of intermediate goods in which firm f has the leading method of production in sector j; u_f^{j} is also considered as the stock of knowledge which may be used for subsequent innovations. Assume that each firm specializes in either clean or dirty technologies. The Poisson flow rate of new innovations is

(15) $$X_f^{j} = \theta\left(H_f^{j}\right)^{\eta}\left(u_f^{j}\right)^{1-\eta},$$

where u_f^{j} is the knowledge stock, H_f^{j} denotes researchers (labor in the R&D sector), η is the R&D elasticity with respect to researchers, and $\theta > 0$. Define $x^{j} \equiv X^{j}/u^{j}$ as innovation per intermediate good. So the flow rate of innovation is $X^{j} = u^{j}x^{j}$.

The demand for researchers is

(16) $$h^{j}\left(x^{j}\right) = \left(\frac{x^{j}}{\theta}\right)^{1/\eta}.$$

An innovation of incumbents requires $F_{I,i}$ researchers per intermediate good and entrants have to engage $F_E > F_I$ researchers. The mass of entrants who conduct R&D is denoted E_i^j.

Policy instrument. The policy instrument is a proportional subsidy from the government to R&D in the clean or dirty sector: the subsidy rate is $s_t^j \in [0,1]$ and the government budget becomes

$$(1+\chi)S_t = T_t. \tag{17}$$

Here parameter χ measures the wastage of the subsidy in the course of R&D research, and hence the difference D_t (distortions) between output and consumption becomes

$$D_t = \chi S_t. \tag{18}$$

Value function. The value of a given firm is given as

$$\begin{aligned} rV_{\vec{n}^j}^j - \dot{V}_{\vec{n}^j}^j = \sum_{i=1}^{u} \Big\{ & \pi_{n_i}^j + z^j \left(V_{\vec{n}_{-i}^j}^j - V_{\vec{n}^j}^j \right) + \\ & + z^{-j} \left[1 - \alpha + \mathbb{I}_{(n_i^j \le 0)} \alpha \right] \left(V_{\vec{n}_{-i}^j \cup \{n_i^j - 1\}}^j - V_{\vec{n}_i^j}^j \right) + \\ & + \mathbb{I}_{\left(n_i^j > 0\right)} z^{-j} \alpha \left(V_{\vec{n}_i^j \cup \{-1\}}^j - V_{\vec{n}_i^j}^j \right) \Big\} + \\ & + \int \max_{x^j \ge 0} \Big\{ u^j x^j \left(\mathbb{E}_n V_{\vec{n}^j \cup \left\{n_{u+1}^j\right\}}^j - V_{\vec{n}^j}^j \right) - \\ & - \left(1 - s^j\right) u^j w^s \left[\left(x^j\right)^{1/n} \theta^{-1/\eta} + \mathbb{I}_{\left(x^j > 0\right)} F_I \right] \Big\} dF_I , \end{aligned} \tag{19}$$

where $\vec{n}^j \equiv \left[n_1^j, \ldots, n_u^j \right]$ is the vector of intermediate goods for which this firm holds the leading-edge technology of type j, n_i^j is the technology gap between technologies j and $-j$ for the same intermediate good, and $\vec{n}_{-i}^j$ is $\vec{n}^j$ without its *i*-th element n_i^j.

Here $\pi_{n_i}^j$ denotes the profits generated from u^j intermediate goods.

The technology gap declines by one step to $n_i^j = n_i^j - 1$ owing to an incremental innovation (with probability $1-\alpha$) or if technology j was already behind $\left(n_i^j \le 0\right)$. The firm falls behind by one step to $n_i^j = -1$ in the case of a breakthrough innovation (probability α) and when technology j was leading $\left(n_i^j > 0\right)$.

The form of the value function implies that innovation is path dependent. Indeed, when the use of clean technology is profitable for certain intermediates, then the average profit for producers who use clean technology is $\Gamma_t^c \equiv \sum_{n<m} \mu_{n,t} \pi_t (m_t - n)$, where $\mu_{n,t}$ denotes the share of intermediate goods for which the clean technology is exactly n steps behind the dirty one at time t. The sequence $\left\{\Gamma_t^c\right\}_{l=0}^{\infty}$ affects incentives through the expected per product value of innovation which (owing to the form of the value function, as is shown in Lemma 1 of [Acemoglu et al., 2016]) is proportionate to $\mu_{n,t}$. After clean innovation has been sustained for a certain period of time, the share of intermediates with markup larger than or equal to the quality gap ($n \le m$) goes up, and this increases the probability of a successful clean innovation for a profitable intermediate good.

Free entry, labor market clearing conditions and production. The free-entry condition for technology j is expressed as;

(20) $$\max_{x_{E,t}^j \ge 0} \left\{ x_{E,t}^j \overline{v}_t^{-j} Y_t - \left(1 - s_t^j\right) w_t^s \left[h\left(x_{E,t}^j\right) + F_E \right] \right\} \le 0,$$

and it becomes an equality when $E_t^j > 0$.

The labor market-clearing condition for skilled workers includes demand from incumbent and entrant firms as follows:

(21) $$L^s = \sum_j \left\{ \mathbb{I}_{\left(x_{E,t}^j > 0\right)} \left[h\left(x_{E,t}^j\right) + F_E \right] E_t^j + \int_0^1 \mathbb{I}_{\left(x_{i,t}^j > 0\right)} \left[h\left(x_{I,t}^j\right) + F_{I,i,t} \right] di \right\},$$

where the expression for $x_{E,t}^j$ and $x_{I,t}^j$ as functions of the normalized skilled wage $\tilde{w}_t^s$ is:

(22) $$x_t^j = I_{\left(x_{i,t}^j > 0\right)} \left[\frac{\eta \theta^{1/n} \overline{v}_t^j}{\left(1 - s_t^j\right) \tilde{w}_t^s} \right]^{\eta/(1-\eta)}$$

The labor market-clearing condition for unskilled workers is:

(23) $$1 = \frac{Y_t}{w_t^u} \cdot \left\{ \sum_{n \le m_t} \frac{\mu_{n,t}}{\left(1 + \tau_t^c\right) \lambda\left(m_t - n\right)} + \left[\nu + (1 - \nu) \left(\frac{1/\zeta}{\tilde{P}_{e,t}} \right) \right] \sum_{n > m_t} \frac{\mu_{n,t}}{\left(1 + \tau_t^d\right) \lambda\left(n - m_t\right)} \right\}.$$

From this condition it follows that the aggregate output may be expressed as the function of tax-adjusted labor productivities as follows:

(24) $$Y_t = \bar{Q}_t \Lambda_t \Omega_t^{-1} exp\left[-\gamma \left(S_t - \overline{S} \right) \right],$$

where

$$\Omega_t \equiv \sum_{n\le m_t}\frac{\mu_{n,t}}{\left(1+\tau_t^c\right)\lambda\left(m_t-n\right)}+\left[\nu+\left(1-\nu\right)\left(\frac{1/\zeta}{\tilde{P}_{e,t}}\right)\right]\sum_{n>m_t}\frac{\mu_{n,t}}{\left(1+\tau_t^d\right)\lambda\left(n-m_t\right)}.$$

Then production in the dirty sector is:

(25) $$Y_t^d=\frac{Y_t}{\left(1+\tau_t^d\right)w_t^u\tilde{P}_{e,t}}\left[\frac{1}{2}Q_{m,t}^d+\frac{1}{\lambda\left(n-m\right)}\sum_{n>m_t}Q_{n,t}^d\right],$$

where the quality index $Q_{n,t}^d \equiv \int_{i\in\mu_n} q_{i,t}^d di$.

The solution of the intertemporal maximization problem of the consumer leads to the Euler equation:

(26) $$g_{C,t}=r_t-\rho.$$

Owing to the Hoteling rule, the price of the exhaustible resource satisfies the equation:

(27) $$\frac{p_{e,t}}{w_t^u}=\left(\frac{p_{e,0}}{w_0^u}-\zeta\right)\int_0^t e^{r_s-g_{w,s}}ds+\zeta.$$

The technology gaps evolve according to the following system of equations:

(28) $$\dot{\mu}_{n>1,t}=z_t^d\mu_{n-1,t}+\left(1-\alpha\right)z_t^c\mu_{n+1,t}-z_t\mu_{n,t},$$

when $n>1$ and when $n\le 1$:

(29) $$\begin{aligned}\dot{\mu}_{1,t}&=z_t^d\mu_{0,t}+\left(1-\alpha\right)z_t^c\mu_{2,t}+\alpha z_t^d\sum_{n<0}\mu_{n,t}-z_t\mu_{1,t},\\ \dot{\mu}_{-1,t}&=z_t^c\mu_{0,t}+\left(1-\alpha\right)z_t^d\mu_{-2,t}+\alpha z_t^c\sum_{n>0}\mu_{n,t}-z_t\mu_{-1,t},\\ \dot{\mu}_{n<-1,t}&=z_t^c\mu_{n+1,t}+\left(1-\alpha\right)z_t^d\mu_{n-1,t}-z_t\mu_{n,t}.\end{aligned}$$

Here z_t^j denotes the aggregate innovation rate by entrant and incumbent firms (see subsections "Heterogeneous innovation" and "The R&D production function" above).

Dynamic equilibrium path. For any given time path of taxes and subsidies $\left[\tau_t^j,s_{I,t}^j,s_{E,t}^j\right]_{t=0}^{\infty}$, a dynamic equilibrium path is the time path of the following variables:

$$\left[y_{i,t}^j,p_{i,t}^j,x_{I,t}^j,x_{E,t}^j,Y_t,w_t^s,w_t^u,e_{i,t},p_{e,t},R_t,E_t^j,\left\{\mu_{n,t}\right\}_{n=-\infty}^{\infty},\left\{Q_{n,t}^d\right\}_{n=-\infty}^{\infty},r_t,S_t\right]_{t=0}^{\infty}$$

where $y_{i,t}^j$ and $p_{i,t}^j$ maximize profits in (13); $x_{I,t}^j$ and $x_{E,t}^j$ come from (22); w_t^u is given by (14); aggregate output Y_t is determined by (24); w^s comes from the free-entry condition (20) under positive entry and from skilled labor market clearing (21) under the absence of positive entry; E_t^j is set by the skilled labor market clearing (21) under positive entry; $\{\mu_{n,t}\}_{n=-\infty}^{\infty}$ follow (28) and (29); $Q_{n,t}^d$ satisfies (20) and (22); the interest rate comes from the Euler equation (26); the quantity and price of exhaustible resource are given by (4) and (27); R_t is set by (3); and S_t is expressed by (7)–(9) with Y_t^d coming from (25).

2.3. Empirical strategy

The [Acemoglu et al., 2016] model is used as a theoretical tool to find the optimal values for a combination of two policy instruments: subsidies for research into carbon-neutral technologies and a tax on carbon emissions. The model studies the evolution of a non-steady state equilibrium, focusing on the time profiles of economic variables across optimal policies (policies with optimal values of carbon tax rate and research subsidies) and the laissez-faire (null) policy. The variables of primary interest are output by firms using carbon-neutral and carbon-emitting technologies, innovative activity by clean and dirty firms, and the overall growth of the economy. The model assumes that all innovations are patented.

The model has 18 parameters $\{\rho, \bar{S}, \gamma, \varphi, \varphi_0, \varphi_P, \kappa, \nu, \chi, L^s, \alpha, \eta, \theta, \lambda, F_l, F_E, R_0, \zeta\}$, and the distribution of $\{\mu_{0,t}\}_{n=-\infty}^{\infty}$ that needs to be estimated. The values of the initial level of carbon concentration $\bar{S}$ and the parameter for emission damage to production γ (both come from equation (6)), of the three parameters of the carbon cycle $\{\varphi, \varphi_0, \varphi_P\}$ of Equation (9) – can be set as exogenous and are not involved in calibrating the other parameters of the model.

The estimation of the model is conducted in several steps. Firstly, there are exogenously set parameters: the value of the social discount rate ρ, the parameter κ in (8) – the share of the dirty output that leads to carbon emissions, the parameter ν in (1) – the elasticity of the dirty output in its production function, and the parameter χ in (18) – the distortionary effects of the R&D subsidies.

These values are set as follows: $\rho = 0.01$, $\nu = 0.04$, $\bar{S} = 581 GtC$, and $\gamma = 5.3 \cdot 10^{-5} GtC^{-1}$ – in accordance with Golosov et al. (2014); φ_P comes from the Intergovernmental Panel on Climate Change [2007] and equals 20 percent; the parameter κ is set to match the World data on carbon emissions and the amount of dirty output in the US in (8), $\{\varphi, \varphi_0\}$ are estimated based on the US data for carbon emissions and the data on carbon concentrations from the Hawaii-based Mauna Loa observatory, χ is set at 10 percent.

Secondly, the values of L^s, α, and η are computed from microdata on enterprises, their financials, and R&D activity. Thirdly, the distribution of technology gaps $\{\mu_{0,t}\}_{n=-\infty}^{\infty}$ is determined with the use of the data on patents pertaining to clean or dirty technologies. Fourthly, six parameters are estimated via the method of moments. These are: θ in (15) – the elasticity of the

R&D production function with respect to researches, λ of subsection "Heterogeneous innovation" – a proportional improvement of quality due to incremental innovation, F_l in subsection "The R&D production function" and in the value function specified in (19) – the number of researches per intermediate good in incumbent firms, R_0 in (3) – the stock of the exhaustible resource, and ζ in (2) – the labor productivity of the extraction of the exhaustible resource. Specifically, theoretical moments for the four variables must be close to their empirical counterparts. The four moments to match the model and the data for the energy sector are: per worker growth rates of the sector (based on (24)–(25)), entry/exit rates of firms (can be computed from (19)), the share of the R&D expenditure in output (where R&D expenditure is used as a proxy for R&D labor in (8)). R_0 and ζ are then estimated to correspond to the national data for the level and the rate of the growth of emissions. The remaining variables are estimated from the model (e.g., the number of researchers in old and new firms).

Finally, the optimal values of the policy instruments are estimated using the calibrated model. The objective function for the social planner is welfare[5] at time *T* which is the sum of the production and quality increase less distortions and emission damage, where emission damage combines the amounts of emissions permanently remaining in the atmosphere S_T^{perm} and transitory S_T^{trans} and Y_T^{base} implies baseline economic production:

$$W = \underbrace{\int_0^T \ln Y_t e^{-\rho t} dt + e^{-\rho T}\Bigg[\ln Y_T^{\text{base}}}_{\text{Production less Distortions}} + \underbrace{\frac{g_T}{\rho}}_{\text{Growth Potential}} \underbrace{- \frac{\gamma}{\rho}\left(S_T^{\text{perm}} + S_T^{\text{trans}} \frac{\rho}{\rho+\phi} - \overline{S}\right)}_{\text{Emission Damage}}\Bigg].$$

The time profiles of the main economic and climate variables are then contrasted between the *laissez-faire* and the optimal policies. The findings using data for the US energy sector and energy patents from 1975 to 2004 reveal that a non-trivial combination of the two policy measures is optimal for maximizing social welfare and has the following economic effects: an increase in innovation and quality (labor productivity) in the carbon-neutral sector; a redirection of production to the carbon-neutral sector; and long-term economic growth, but a decrease of growth in the short term. The decrease of growth in the short (and possibly medium) term is explained by the superiority of the existing dirty technologies, which can be seen from the micro data on quality in the carbon-neutral and carbon-emitting sectors.

As shown in Section VI of Acemoglu et al. (2016) on the robustness of the results of the research with respect to the exogenously set values of the aforementioned parameters, the qualitative and, to a large extent, quantitative conclusions about the optimal policy mix of the carbon tax and research subsidies are robust to the choice of the exogenously set parameters as follows: the social discount rate may be changed from 1 percent as in Golosov et al. (2014) to lower values such as 0.1 percent or 0.5 percent as in Stern (2007); instead of setting ν equal 4 percent as in

---

[5] The explicit formula is inferred from lines 770-784 of the code of infinite_weave.py which supplements the [Acemoglu et al., 2016] paper.

Golosov et al. (2014), it can be assigned lower values of 0 or 2 percent; the parameter γ can be increased two- or five-fold compared to the value in Golosov et al. (2014); χ can be either lowered to 0 (no distortions) or increased to 20 percent. The conclusions are also robust to the values of parameters α (0.3 or 0.5) and η (0.35 or 0.65).

3. Data on Japan

We use several blocks of data on the Japanese economy for our quantification. Firstly, we take meteorological data from two sources. National carbon emissions per capita come from the World Bank, which collects estimates from the Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory (Tennessee, US). We use data from the Japan Meteorological Agency on atmospheric carbon concentrations, which are measured at three stations: Ryori (120 km from Sendai on the Pacific coast of Honshu island, in the Tohoku area), Minamitorishima (an island 1848 km southeast of Tokyo in the North Pacific Ocean) and Yonagunijima (an island in the East China Sea in the Pacific Ocean, 108 km from Taiwan). The values of carbon concentration demonstrate similar seasonality and are generally close across the three stations. However, the history of observations is the longest at the Ryori station.

Secondly, we exploit several databases on Japanese companies. The Nikkei NEEDS database contains the financial and administrative data for 6,500 companies from the late 1960s to the early 1970s and onwards. Most of the companies are large corporations, and they account for 50–80 percent of production in their industries. The Nikkei NEEDS data are manually matched to non-anonymous company data from the Japan National Innovation Survey [2015] and to Orbis, Bureau van Djik data with NACE/US SIC industrial classifications for firm products (with an overlap for about 80 percent of Nikkei NEEDS firms).

Thirdly, patent statistics are calculated using the Tokyo Institute of Intellectual Property Patent Database [2020]. This is a recently created NBER-like database [Goto & Motohashi, 2007], which contains Japan's domestic patent applications submitted since 1964.

In the robustness section of this paper, we employ data for Japanese firms from Orbis, Bureau van Djik (2009–2019) which is linked by Bureau van Djik to the European Patent Office database on firm-level patent statistics (available till 2019).

Finally, we use aggregate data on R&D labor in various sectors of the Japanese economy from the Japan Statistical Agency (2023).

4. Quantification for Japan

4.1. Carbon cycle

We fit the exponential (geometric) equation for the carbon cycle [Acemoglu et al., 2016; Golosov et al., 2014], as described in (7)–(9), which draws on the approach used by [Archer, 2005] on the existence of a transitory carbon component in the atmosphere. We use carbon concentration data from the Ryori meteorological station, the World Bank data on carbon emissions by Japan, and the value of the share of emissions, permanently remaining in the atmosphere, from the Intergovernmental Panel on Climate Change [2007]. We fit equation (7) using Japan's data for 1986–2008, so that the final time period was comparable to the US estimates and find that

$\hat{\phi} = 0.0202$ and $\hat{\phi}_0 = 0.4173$. The values of the rate of decay are close to the parameter estimates for the US economy during a similar time period: 0.0313 as reported in [Acemoglu et al., 2016] and 0.0228 in [Golosov et al., 2014]. The share of the transitory component is close to the estimate in [Golosov et al., 2014], but differs from the value in [Acemoglu et al., 2016]. See Table 1 for a detailed comparison. We discuss the approach in the robustness section of this paper.

Table 1.

Contrasting parameters of the carbon cycle in Japan and the US

Parameter	Definition	US		Japan
		Acemoglu et al., 2016	Golosov et al., 2014	
ϕ_p	share of emissions permanently remaining in atmosphere	0.2	0.2	0.2
ϕ	rate of decay of carbon concentration	0.0313	0.0228	0.0202
ϕ_0	$(1-\phi_p)\phi_0$ share of transitory component in period 0	0.7661	0.3930	0.4173

4.2. Carbon-neutral and carbon-emitting technology

Our definitions of carbon-neutral technologies combine the approaches of the three sources. Firstly, we exploit the [OECD, 2009] methodology on use of patent classes for environmentally friendly technologies, as described in the patent search strategy for the identification of selected "environmental" technologies, developed as part of the OECD project on "Environmental Policy and Technological Innovation". Secondly, we supplement this list of patent classes using the International Patent Classification (IPC) Green inventory of the World International Property Organization [WIPO, 2017]. Finally, we add the patent classes for the energy sector from the corresponding appendix to [Popp & Newell, 2012]. The groups of patent classes used in our analysis for the definition of carbon-neutral technologies are summarized in Table 2.

Table 2.

Carbon-neutral technologies, according to the International Patent Classification

Clean/green technology classes	Source
Air, water and waste related technologies	OECD/WIPO/Popp and Newell (2012)
Alternative energy production	WIPO/Popp and Newell (2012)
Transportation	WIPO
Energy conservation	WIPO
Agriculture/forestry (e.g. alternative irrigation techniques)	WIPO
Nuclear power generation	WIPO
Administrative, regulatory or design aspects (e.g. carbon-emissions trade)	WIPO

4.3. Energy sector

We use the UN International Industrial Classification codes to define energy sector firms according to the approach of the United Nations Industrial Development Organization [Upadhyaya, 2010]. Our analysis also considers the manufacture of motor vehicles and of general-purpose machinery, following [Acemoglu et al., 2016]. The full list of energy sector NACE Revision 2 codes is given in Table 3.

Following [Acemoglu et al., 2016] we examine energy patents in estimating the probability of breakthrough innovation. For this purpose, the paper uses patent classes in the International Patent Classification[6] which correspond to groups of energy goods and services in Table 3. These codes are: B21-B23 (metal working, metallurgy), B60-B64 (vehicles and motor vehicles), C08 (organic macromolecule), C21-C30 (metallurgy, electrolytic processes, crystal growth), F21-F28 (lighting, heating), G21 (nuclear physics and nuclear engineering), H01-H05 (electricity).

Our empirical analysis focuses on the time period after 1989, in order to include the years after the 1988 revision of the Japan Patent Law. The revision allowed multiple claims and may have influenced the strength of Japanese patents, especially in their applicability across industrial fields.

Our sample, which is an overlap between the Nikkei NEEDS, the Japan National Innovation Survey and Orbis database contains 1178–2565 manufacturing firms from 1989 to 2013. There are 303–589 energy firms each year, according to our definition. The share of energy firms is stable at 23–25% of all firms.

Table 3.

Energy sector, according to the UN International Industrial Classification

Industry name	NACE	Source
Mining of coal and lignite; extraction of peat	05	UNIDO [Upadhyaya, 2010]
Extraction of crude petroleum and natural gas	06	UNIDO [Upadhyaya, 2010]
Mining of uranium and thorium ores	07	UNIDO [Upadhyaya, 2010]
Manufacture of coke, refined petroleum products and nuclear fuel	19	UNIDO [Upadhyaya, 2010]
Electricity, gas, steam and air conditioning supply	35	UNIDO [Upadhyaya, 2010]
Manufacture of motor vehicles	29	[Acemoglu et al., 2016]
Manufacture of general purpose machinery	28	[Acemoglu et al., 2016]

4.4. Technology gaps

Following [Acemoglu et al., 2016], we define a clean firm as a firm, whose share of clean patents in all its patents exceeds a certain threshold. However, instead of using the [Acemoglu et al., 2016] threshold of 25% (which gives 11% of clean firms from the US data), we choose a lower value of 5% for our sample. The empirical distribution for the share of clean patents differs between American and Japanese firms. In Japan there is only a negligible number of firms with

[6] https://www.wipo.int/classifications/ipc/en/

over a quarter of clean patents. If we wanted to establish the size of the clean sector as 10–11% of producers (to make Japan's economy comparable to the US), it would require an extremely loose definition, by which just 1% of clean patents would suffice to make a company clean. As a compromise, we choose a threshold of 5% of clean patents for a firm to be regarded as environmentally friendly. The value is supported by micro evidence on the relative weight of environmentally friendly initiatives in the behavior of Japanese firms, which are attentive to their social responsibility regarding the environment [Tanikawa, 2004]. The threshold of 5% means that the share of clean firms in Japan is on average 3% of all firms (varying from 1 to 5% in different years).

According to the model of [Acemoglu et al., 2016], the technology gap between dirty and clean technologies for each product is defined as the difference in the number of innovation steps (see subsection "Heterogenous innovation"). Formally, this is given by $n_{i,t} \equiv n_{i,t}^d - n_{i,t}^c$, where $n_{i,t}^d$ and $n_{i,t}^c$ are innovation steps of *dirty* and *clean* technology for product *i* at time *t*, respectively. The technology gap is employed in the value function (19) and in the derivation of the dynamic equilibrium in the model.

Following the empirical strategy in [Acemoglu et al., 2016], we compute the cumulative number of patents for clean and dirty Japanese incumbent firms at the US SIC3 level (and for robustness check at the US SIC4 level). Then this innovation flow of patents for clean and dirty technologies is normalized by the mean patent flow (i.e., the annual number of energy patents per product by all firms). The resulting distribution of the technology gap is given on Figure 1. As shown by the distribution, dirty technology is one to four steps ahead for most products, although dirty technology leads 10 to 120 steps for few products. The shape of the distribution is generally close to that in the US. However, clean technology is up to 10 steps ahead of dirty technology for a few products in the US according to [Acemoglu et al., 2016], but we found no similar pattern for Japan.

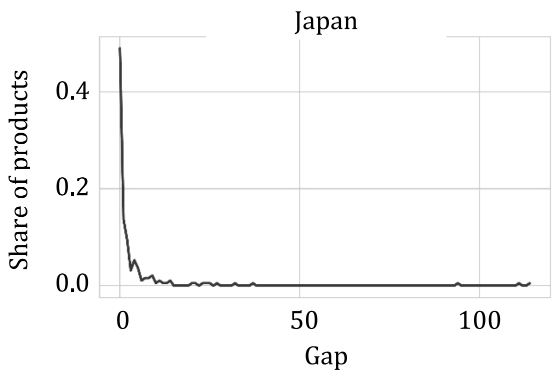


Fig. 1. Technology gap between carbon-emitting and carbon-neutral sectors across products

4.5. Parameters for the Japanese energy sector

To replicate the analysis in [Acemoglu et al., 2016] in estimating moment targets and other parameters related to technological change in the energy sector, this paper considers energy firms with high R&D expenditure (we set the threshold of 1 mln. yen, which is a reasonable country-level analogue to the value of 1 mln. USD used for the US despite being much lower).

The resulting parameters for the energy sector that are listed in Table 4 may be divided into several groups. One group is linked to quality changes through innovation. As innovations are quantified through patents, the quality evaluations are based on patent citations. To compute the probability of radical innovation [Acemoglu et al., 2016] compares citations for energy patents within three years after patenting to citations within 10 years. Patents are defined as 'major entrants' if their citations in the three years exceed the 90th percentile (i.e. a reasonable threshold value) of the citations for patents as old as 10 years. The share of major entrants, which equals 0.04 for the US energy sector, is regarded as an empirical estimate of the probability of radical innovation. Our use of the patent data for Japan's economy with a similar approach produces a comparable estimate of 0.038.

Another variable on innovation outcomes is mean patent flow, which is defined in [Acemoglu et al., 2016] as the annual number of citation-weighted patents per product. While the US estimate for the energy sector is 43 patents, our calculations give a corresponding value of 40 patents for Japan.

The second group of parameters relates to the R&D production function. The [Acemoglu et al., 2016] strategy follows the microeconomic approach to proxy R&D output by patents and takes R&D expenditure as an input. The regression analysis exploits pooled data with firm-level clustered standard errors and adds annual dummies to the right-hand side of the equation. The resulting value for R&D elasticity is 0.5 for the US data: it is the mean estimate across the models in levels and in the first differences and across the two specifications (the normalization of input and output by product counts or by domestic sales). Our calculations with the data for Japan's energy sector give a range of elasticity [0.2, 0.6], so the mean estimate is 0.4. The share of R&D labor in unskilled labor is 0.055 in the US, as estimated in [Acemoglu et al., 2016] using micro data. We infer the value of 0.059 using the data from the Japan Statistical Agency (2023).

The third group of parameters are moment targets – the mean values of the four key variables, which are used in model calibration through a simulated method of moments. The variables relate to microdata on company history and financials, and are comparable across innovative energy firms in the US and Japan: the entry rate and exit rate of firms, mean R&D expenditure per sales, and the growth of sales per worker.

The values of the parameters in Table 4 do not differ appreciably between the innovative energy firms in the US and Japan.

However, the US-Japan differences in the gaps between dirty and clean technologies, which we reported in section 4.4, are likely to result in a higher reliance on carbon tax rather than on research subsidies in the context of the models of [Acemoglu et al., 2016] and [Golosov et al., 2014].

Table 4.

Parameters for the energy sector in Japan and the US

	US	Japan
Patents		
Probability of radical innovation	0.04	0.038
Patents per product (citation weighted)	43	40
R&D		
Share of R&D labor	0.055	0.059
Elasticity of innovation output in R&D expenses	0.5	0.4
Production (moments for calibration)		
Entry rate of firms	0.013	0.017
Exit rate of firms	0.018	0.012
Growth of sales per worker	0.012	0.029
Share of R&D expenditure in sales	0.066	0.056

Note: The US data for the energy sector in 1975–2004 come from [Acemoglu et al., 2016]. Japanese estimates for the energy sector (unless otherwise stated) are based on our data for 1989–2012. Regarding the entry rate of firms, [Acemoglu et al., 2016] use the labor share of entrants, while we use the number of firms with the Japanese data. The table reports the share of R&D labor of all specialties in the sector "Electrical machinery and equipment" using the plant-level data on R&D employment from the Japan Statistical Agency (2023), File 4, columns 1,3,15. (The average share across all industries and plants of all size in Japan is 0.037, and the shares in the other subsectors of the energy sector, as classified in Table 3 in this paper are in the range of 0.02 to 0.2).

5. Results

5.1. Carbon tax, research subsidy, innovation rates, and outputs in the clean and dirty sector

Our computations use Python codes from [Acemoglu et al., 2016][7]. While [Acemoglu et al., 2016] analyze various ways to parametrize the time profiles for policy instruments, we focus

[7] The codes are available as supplementary material to the [Acemoglu et al., 2016] paper on the *Journal of Political Economy* website: https://www.journals.uchicago.edu/doi/suppl/10.1086/684511. Firstly, the data are loaded into load_data.py and the parameters for the carbon cycle, energy sector, and R&D are entered into infinite_weave.py. Then the estimation_weave.py is used to calibrate the remaining six parameters (described in section 2.3 of this paper in the fourth set of estimated parameters), and the calibration starts from the manually entered initial values for the parameters. Our investigation of the algorithm hints at the fact that the model seems to have a knife-edge solution (see [Posch, 2011; Growiec, 2007; Ellison & Fudenberg, 2003]) and no convergence can be achieved under entering the initial values for the six parameters that appreciably differ from the equilibrium ones. So in comparing the impact of different policy measures for the Japanese economy, using the code generate_policy.py, we employ moment targets obtained from the US data. This may be justified by the fact that parameters for the energy sectors of Japan and the US are very similar (Table 4). However, we enter the Japan-specific technology gaps in load_data.py in conducting the estimations with generate_policy.py.

on the two scenarios, which are most realistic to implement and are often analyzed in the Japanese context [Ministry of the Environment, 2017]. Optimal constant policies imply fixed values of research subsidies and carbon taxes over the whole period, while optimal three-step policies allow for step-wise changes in the course of adapting policy instruments.

The results for the Japanese economy as regards carbon taxes and research subsidies are given in Fig. 2 and can be contrasted to those reported by [Acemoglu et al., 2016, Fig. 10] for the US in the case of the three-step policy. The carbon tax is around 0.05 during the first period in the US, while it is twice as much and is close to 0.1 in Japan. The value of research subsidies is close to 0.8 in the US during the first period, while it is below 0.8 in Japan. Similarly, there is higher reliance on carbon taxes and lower reliance on research subsidies in Japan relative to the US in the second period.

The combination of carbon tax with research subsidies switches innovation in Japan from the carbon-emitting to the carbon-neutral sector (Fig. 3). Innovation in the carbon-emitting sector vanishes after 50 years of policy implementation. Similarly, there is a redirection of production from the dirty to the clean sector: the output of the dirty sector steadily declines, while production in the clean sector gradually increases (Fig. 4–5). The results reveal that the carbon-neutral sector would disappear in the medium-run under the laissez-faire, i.e. without the optimal policies.

The optimal policy instruments not only sustain the growth of clean production, but lead to overall economic growth in the long term (Fig. 6). The number of years, during which aggregate output in Japan declines as incentives to use clean technologies are applied, is comparable to the 20 years estimated by [Golosov et al., 2014] for reaching the *laissez-faire* level of production in the US with the application of similar incentives. However, the period is longer in Japan, which may be explained by more distortions due to the relatively slower advance of clean technologies. The environmental effects of policy instruments are similar to those in [Acemoglu et al., 2016]: decrease of national carbon emissions and a limited contribution by the country to temperature increases.

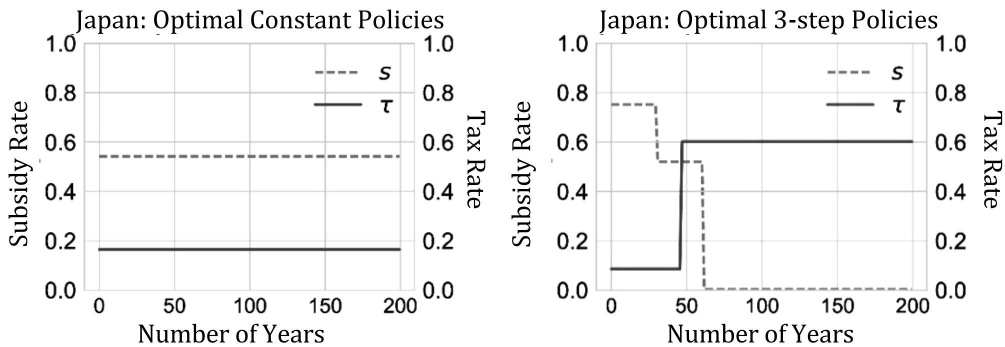


Fig. 2. Tax rate and research subsidies under the optimal policies

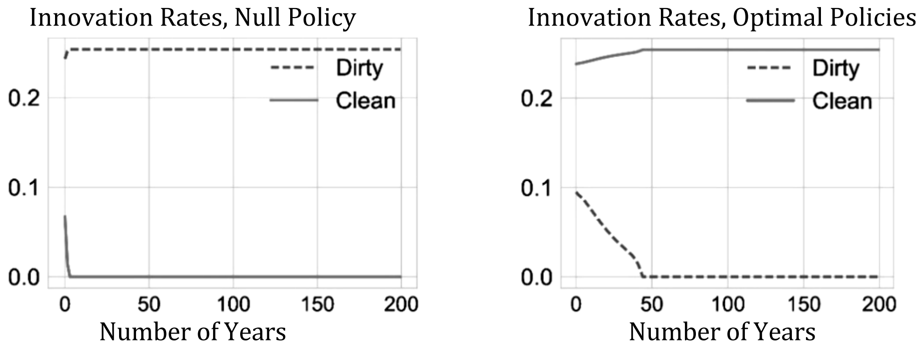


Fig. 3. Innovation rates under the laissez-faire and optimal policies

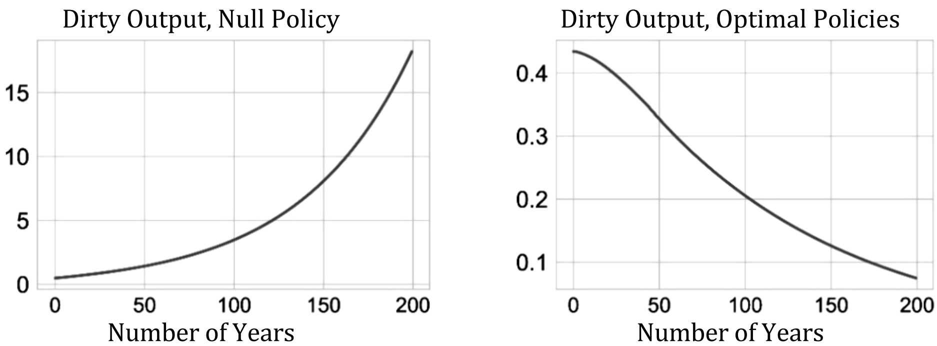


Fig. 4. Output in the carbon-emitting sector under the laissez-faire and optimal policies

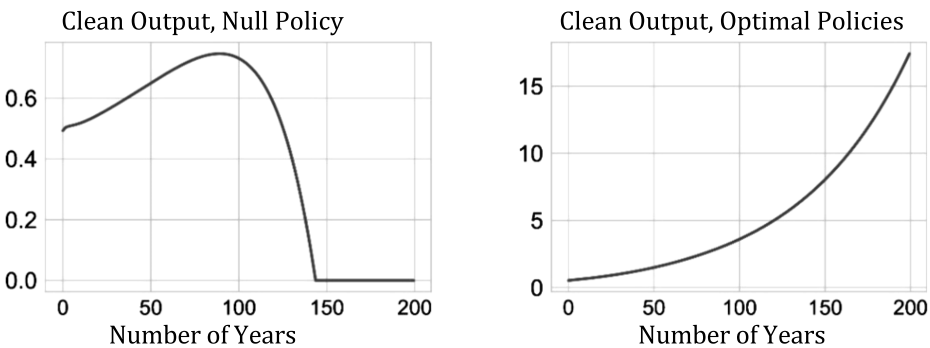


Fig. 5. Output in the carbon-neutral sector under the laissez-faire and optimal policies

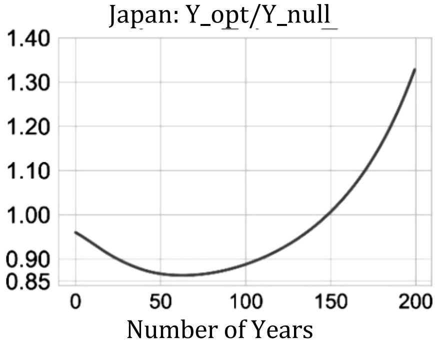


Fig. 6. Ratio of aggregate output in the economy (clean and dirty) under optimal policies to output under the laissez-faire

5.2. Robustness of the estimates and limitations of the analysis

Four types of robustness checks may be justified in evaluating the results of the empirical analysis. Firstly, the question is: how the results within the framework of the [Acemoglu et. al., 2016] model (applied to the data of a given country) are robust to changes in the values of the key parameters. The issue was briefly touched upon at the end of Section 2.3 of this paper and may be expanded here as follows. The results, for instance, as regards the values of the carbon tax rate and research subsidies and the time profiles of production, are generally robust to a range of the social discount rate in the utility function of the representative household (see Fig. 6, 10 and 11 in [Acemoglu et al., 2016]). More specifically, if there is a minor increase in the size of the carbon tax rate due to lowering the social discount rate from 1 to 0.5 percent and there is a noticeable rise of the carbon tax rate when the social discount rate goes down from 1 to 0.1 percent. Arguably, the change of the social discount rate in the opposite direction: from 1 to 2 percent would lead to a negligible fall in the carbon tax rate.

According to Section VI in [Acemoglu et al., 2016], the results are robust to the values of the elasticity of dirty output with respect to exhaustible resources in Equation (2): 0, 2 or 4 percent; to the values of the parameter for the negative impact of the carbon concentration on the overall economic output in the economy in Equation (6); to the share of research subsidies that are lost, i.e. the distortions parameter in Equation (17); to the probability of a breakthrough innovation (0.3, 0.4 or 0.5); and to the elasticity of the R&D production function with respect to labor inputs in Equation (15).

Secondly, an important issue for the robustness check deals with the comparison of the findings for the US and Japan under the change of one of the key exogenous parameters across the two countries. As the national economies of the US and Japan at the end of the 20th century and at the beginning of the 21st century are very similar as regards institutional infrastructure, economic incentives and the principles of political economy, a reasonable candidate for such cross-country differences is social discount rates, which is largely explained by the psychological aspects of consumer behavior. Arguably, the social discount rate in Japan is likely to be lower than in the US, but as is shown in the analysis in Figures 6 and 11 in [Acemoglu et al., 2016], the size of the carbon tax rate does not change during the first 50 years under lowering the social discount rate from 1 to 0.5 percent. Under the lowering of the social discount rate from 1 to 0.1 percent, the carbon tax rate goes up in the first 50 years. So our conclusion about the higher values of the carbon tax rate in Japan in comparison to the US would only be strengthened under the assumption that the social discount rate in Japan is much lower than in the US.

The third question is how the results are robust to alternative ways of computing the technology gap, i.e. the gap between the labor productivity of dirty and clean technologies. Here we follow the approach of [Acemoglu et al., 2016] to consider different ways of computing the technology gap: based on the US SIC3 codes (as reported in Section 4.4 of this paper) or based on SIC4 codes; using the data for energy patents in the whole economy, only in the energy sector or in the manufacturing sector. For this purpose, we employ both Nikkei NEEDS data (which overlap with Japan Innovative Survey) for large innovative Japanese firms from 1989 to 2012 and the data of Orbis, Bureau van Djik for large and medium-sized Japanese firms from 2009 to 2019. In the baseline analysis in this paper, we linked the NIKKEI NEEDS firms to their patents using the Tokyo Institute of Intellectual Property database (2020). Our robustness check employs the European Patent Office data (coming from the Japan Patent Office) and available within the Orbis database with firm – patent level links till 2019.

Our robustness check hints at the key difference between the technology gap in the manufacturing and energy sectors of the US and Japan – generally, there is an absence of products for which the number of innovation steps in clean technology exceeds the number of innovation steps in dirty technology. Exceptions are only a handful of products – 50–100 non-energy sector products in the selected years with a negative technology gap: i.e., the clean technology is ahead of the dirty. However, the value of the lead of clean technology is only 1 step. Yet, the findings for the US economy show that the technology gap is negative for over 9 percent of products, and its absolute values are larger than 1 for 7 percent of products, see Section VI.H in [Acemoglu et al., 2016] and page A-13 in the online supplement to [Acemoglu et al., 2016]. As regards energy sector or energy patents, our robustness checks with the 3-digit or 4-digit code products in 1989–2012 with Nikkei NEEDs firms or for 2009–2019 with Orbis firms (about 1300 patenting firms and overall about 28,000 firms in the energy sector) shows that the technology gap is always positive: there are more innovation steps in the dirty sector than in the clean sector for each product. The fourth issue concerning the robustness of the estimates is the measurement of the parameters for the carbon cycle. In this paper, we used the carbon concentration from the Ryori station owing to it having the longest time series across the three Japanese meteorological stations. In reality, the curves on carbon concentration based on the data from the three stations are very similar in years when the data are available for each station, despite the fact that the geographical proximity to China and other Asian countries varies from 100 to 3,000 kilometers across the three stations. Arguably, carbon emissions from several countries contribute to carbon concentration measured at any of the three Japanese stations. More generally, our empirical exercise fitting the carbon cycle showed that the shape of the data series for Japan (available in Fig. 1 on the working paper version of this paper [Besstremyannaya, Dasher, Golovan, 2019]) does not differ from that for the US in [Acemoglu et al, 2016]. Moreover, the inability to exclude the impact of carbon emissions from several other countries is likely to apply to the measurement of carbon concentration data according to Mauna Loa station in Hawaii[8]. So our estimations, that were targeted at the completeness of using the Japanese data (as regards meteorological figures), may serve as a proof of the robustness of estimating the carbon cycle with data from

[8] The fact was noted by the reviewer to our paper.

different meteorological stations across the world. Moreover, using the estimates of [Acemoglu et al. 2016] or [Golosov et al., 2014] would not change the quantitative or qualitative results of our analysis. In fact, borrowing the US parameters for the carbon cycle is an existing approach in macroeconomic analysis with Japan-based data. For example, model calibration in [Matsumura et al., 2024] employs the carbon cycle parameters from [Golosov et al., 2014].

Finally, we note the remaining limitations of the analysis. The [Acemoglu et al., 2016] model assumes that all innovations are patented. The prevalence of patenting may be considered comparable across the US and Japan [Cohen et al., 2002] and hence the assumption is unlikely to lead to an appreciable overestimation of the technology gap or an underestimation of radical innovation in Japan. However, the model considers a closed economy, so the approach disregards technology adoption across countries[9]. For instance, the spillover effect of clean innovation in the direction from the US to Japan may cause lower values of the lead between dirty and clean technologies in Japan than those estimated with patent data. Concerning the external validity of the analysis, patents may not offer a full representation of clean innovation in developing and emerging countries (e.g. BRICS countries), where clean innovation may appear in the forms of frugal innovation [Besstremyannaya and Dasher, 2024]. Therefore the distribution of the technology gap in emerging countries based on patent data could overestimate the distance between dirty and clean technologies, and lead to the overestimation of the value of the carbon tax in the optimal policies.

Thirdly, the model is of a general equilibrium kind but the calibration of the model deals only with the data for the energy sector and energy patents. This limitation has led to the development of the models in various sectors of the economy which depend on energy to a different extent. However, the existing models, as noted in the Introduction to the present paper, are not as rich in employing microeconomic data or using the principles of quality competition. The fourth limitation of the model concerns the estimate of the elasticity of R&D output with respect to R&D labor or R&D expenses. According to our companion paper on the heterogeneity of the growth of Japanese innovative firms [Besstremyannaya, Dasher & Golovan, 2022], the successful use of R&D expenditure in terms of producing innovation output depends to a large extent on R&D management, which is commonly disregarded in the macroeconomic approach. Moreover, the focus on only patenting firms (and arguably, high-growth firms as such firms are likely to be more successful in innovation) may lead to an overestimation of the elasticity parameter if the analysis is to be expanded to R&D-investing but non-patenting firms.

6. Discussion and Conclusion

Endogenous growth models with technological change assume that competitive firms conduct R&D to raise profits through improving their technology [Klette & Kortum, 2004]. Stemming from the Schumpeterian concept of creative destruction and the [Arrow & Debreu, 1954] general equilibrium framework, the models account for the actions of the main economic agents in the market and the actions of government as a social planner. Not only are the models rich in the explanations they offer of numerous regularities in company growth [Acemoglu et al., 2013; Lentz & Mortensen, 2008], but they also make it possible to incorporate various economic externalities such as carbon emissions. This paper uses the framework of such a model [Acemoglu et al., 2016] and applies the approach for the analysis of the Japanese economy.

[9] The concern was noted by a reviewer of our paper.

The policies considered in this paper are a combination of carbon taxes and research subsidies. The analysis demonstrates that there is a temporary decline of economic output associated with the development of carbon-neutral technologies (both in applications to the US and Japan), and we conjecture that it can be explained by technology costs. For example, empirical microeconomic analyses show that technology costs negatively affect individual decisions to use thermal insulation technologies, and the scope of the effect is several times larger than the effect of energy prices [Hassett & Metcalf, 1995; Jaffe & Stavins, 1995]. Inadequate access to financing may be an impediment to introducing clean technologies at small firms [Jaffe et al., 2003]. However, financial impediments may be of secondary importance in comparison with alternative investment choices, capital depreciation and energy prices[10]. As regards the overall social welfare function, judged from a macroeconomic perspective, the costs of clean technologies (borne by the government through research subsidies) can be offset against economic gains. The gains can be measured in terms of economic growth or an increase of social welfare thanks to the prevention of carbon emissions.

Concerning the comparison of the values of the optimal policy instruments in the US and Japan, this paper finds that the carbon tax rate in Japan should not be below that in the US. Yet, the current value of the carbon tax rate in Japan (as adopted in 2012) was one of the lowest across OECD countries and is several times lower than the US carbon tax rate. Moreover, while the US adopted a policy of increasing its carbon tax rate from 8 to 51 USD per ton of CO2 (which is within the range of international recommendations of 35 to 70 USD), Japan has kept its carbon tax rate constant at the 2012 value of about 2.6 USD [Gokhale, 2021]. Another key finding of the paper is the fact that the length of the period for the decline of aggregate output in Japan owing to introduction of the carbon tax and research subsidies is longer than that in the US. Both findings are driven by the relatively low spread of clean technologies in Japan compared to in the US.

Along with the cross-country analysis, the present paper follows [Acemoglu et al., 2016] to show that the optimal values of both the carbon tax and research subsidies are non-zero for Japan. In fact, market mechanisms, such as the introduction of carbon taxes or theincrease of energy prices, can be viewed as an economic incentive for firms and households to employ carbon-neutral technologies [Jaffe et al., 2003; Sanstad et al., 1995]. However, market forces alone cause only the slow propagation of carbon-neutral technologies and diminish the potential for reducing emissions [Popp et al., 2010][11]. Accordingly, there is a need for research subsidies that stimulate the diffusion of currently existing green technologies.

---

10 See the socio-economic analysis for the Dutch firms in [Nijkamp et al., 2001]. The incentives of Japanese firms in their voluntary adoption of environmental technologies are analyzed in similar qualitative research by [Tanikawa, 2004].

11 In fact, there is a certain "habit-formation" in the decision by a firm regarding technology development. For instance, econometric estimates show that R&D can be viewed as a function of a firm's past history in terms of its clean/dirty innovation [Aghion et al., 2016].

* *
*

References

Acemoglu D., Aghion P., Bursztyn L., Hemous D. (2012) The Environment and Directed Technical Change. *The American Economic Review*, 102, 1, pp. 131–166.

Acemoglu D., Akcigit U., Bloom N., Kerr W.R. (2013) *Innovation, Reallocation and Growth*. National Bureau of Economic Research, w18993.

Acemoglu D., Akcigit U., Hanley D., Kerr W. (2016) Transition to Clean Technology. *Journal of Political Economy*, 124, 1, pp. 52–104.

Adebayo T.S., Kirikkaleli D. (2021) Impact of Renewable Energy Consumption, Globalization, and Technological Innovation on Environmental Degradation in Japan: Application of Wavelet Tools. *Environment, Development and Sustainability*, 23, 11, pp. 16057–16082.

Aden K., Dirir S.A. (2023) Unraveling the Relationship between Environmental Taxation and Trade: The Moderating Role of Judicial System within the Japanese Context. *Journal of International Trade, Logistics and Law*, 9, 1, pp. 203–215.

Aghion P., Bergeaud A., De Ridder M., Van Reenen J. (2024) *Lost in Transition: Financial Barriers to Green Growth.* INSEAD Working Paper no 2024/16/EPS. Available at: ssrn_id4771433_code865831.pdf

Aghion P., Bergeaud A., Boppart T., Klenow P.J., Li H. (2019) Missing Growth from Creative Destruction. *The American Economic Review*, 109, 8, pp. 2795–2822.

Aghion P., Dechezleprêtre A., Hemous D., Martin R., Van Reenen J. (2016) Carbon Taxes, Path Dependency, and Directed Technical Change: Evidence from the Auto Industry. *Journal of Political Economy*, 124, 1, pp. 1–51.

Akcigit U., Kerr W.R. (2010) *Growth Through Heterogeneous Innovations*. National Bureau of Economic Research. Working paper 16443.

Archer D. (2005) Fate of Fossil Fuel CO2 in Geologic Time. *Journal of Geophysical Research: Oceans*, 110, C9, pp. 1–6.

Arrow K.J., Debreu G. (1954) Existence of an Equilibrium for a Competitive Economy. *Econometrica: Journal of the Econometric Society*, 22, 3, pp. 265–290.

Besstremyannaya G., Dasher R., Golovan S. (2022) Quantifying Heterogeneity in the Relationship between R&D Intensity and Growth at Innovative Japanese Firms: A Quantile Regression Approach. *Applied Econometrics*, 67, pp. 27–45.

Besstremyannaya G., Dasher R. (2024) Financing and Management of Innovation in India: New Paths for Green Innovation. *Journal of Corporate Finance Research,* 18, 3, pp. 97–109.

Besstremyannaya G., Dasher R., Golovan S. (2019) *Technological Change, Energy, Environment and Economic Growth in Japan*. Ruhr Economic Papers. No. 797, RWI – Leibniz-Institut für Wirtschaftsforschung, Ruhr-University Bochum, TU Dortmund University, University of Duisburg-Essen. Available at: https://ideas.repec.org/p/zbw/rwirep/797.html

Cohen W.M., Goto A., Nagata A., Nelson R.R., Walsh J.P. (2002) R&D Spillovers, Patents and the Incentives to Innovate in Japan and the United States. *Research Policy*, 31, 8–9, pp. 1349–1367.

Dasgupta P., Mäler K.-G. (2000) Net National Product, Wealth, and Social Well-being. *Environment and Development Economics*, 5, 1, pp. 69–93.

De Groot H.L., Verhoef E.T., Nijkamp P. (2001) Energy Saving by Firms: Decision-Making, Barriers and Policies. *Energy Economics*, 23, 6, pp. 717–740.

Ellison G., Fudenberg D. (2003) Knife-edge or Plateau: When Do Market Models Tip? *The Quarterly Journal of Economics*, 118, 4, pp. 1249–1278.

Fischer C., Newell R.G. (2008) Environmental and Technology Policies for Climate Mitigation. *Journal of Environmental Economics and Management*, 55, 2, pp. 142–162.

Fujii H., Managi S. (2016) Research and Development Strategy for Environmental Technology in Japan: A Comparative Study of the Private and Public Sectors. *Technological Forecasting and Social Change*, 112, pp. 293–302.

Gans J.S. (2012) Innovation and Climate Change Policy. *American Economic Journal: Economic Policy*, 4, 4, pp. 125–145.

Gerlagh R., Van der Zwaan B. (2006) Options and Instruments for a Deep Cut in CO2 Emissions: Carbon Dioxide Capture or Renewables, Taxes or Subsidies? *The Energy Journal*, 27, 3, pp. 25–48.

Gokhale H. (2021) Japan's Carbon Tax Policy: Limitations and Policy Suggestions. *Current Research in Environmental Sustainability*, 3, 100082.

Golosov M., Hassler J., Krusell P., Tsyvinski A. (2014) Optimal Taxes on Fossil Fuel in General Equilibrium. *Econometrica*, 82, 1, pp. 41–88.

Goto A., Motohashi K. (2007) Construction of a Japanese Patent Database and a First Look at Japanese Patenting Activities. *Research Policy*, 36, 9, pp. 1431–1442.

Goto N. (1995) Macroeconomic and Sectoral Impacts of Carbon Taxation: A Case for the Japanese Economy. *Energy Economics*, 17, 4, pp. 277–292.

Goto N., Sawa T. (1993) An Analysis of the Macro-Economic Costs of Various CO2 Emission Control Policies in Japan. *The Energy Journal*, 14, 1, pp. 83–110.

Green House Inventory Office, National Institute of Environmental Studies (2024) National GHG Inventory. Available at: https://www.nies.go.jp/gio/en/aboutghg/index.html

Grossman G.M., Helpman E. (1990) Trade, Innovation, and Growth. *The American Economic Review*, 80, 2, pp. 86–91.

Growiec J. (2007) Beyond the Linearity Critique: The Knife-Edge Assumption of Steady-State Growth. *Economic Theory*, 31, 3, pp. 489–499.

Hamaguchi Y. (2024) Comparative Study of Environmental and Monetary Policies in an R&D-based Growth Model. *Journal of Environmental Management*, 360, 121199.

Hassett K.A., Metcalf G.E. (1995) Energy Tax Credits and Residential Conservation Investment: Evidence from Panel Data. *Journal of Public Economics*, 57, 2, pp. 201–217.

Hassler J., Krusell P., Olovsson C. (2012) *Energy-Saving Technical Change*. National Bureau of Economic Research. Working paper 18456.

Hosono K., Takizawa M., Yamanouchi K. (2022) *Firm Age, Productivity, and Intangible Capital*. Discussion paper 20-e-001, RIETI. Available at: https://www.rieti.go.jp/jp/publications/dp/20e001.pdf

Iino T., Inoue H., Saito Y.U., Todo Y. (2021) How Does the Global Network of Research Collaboration Affect the Quality of Innovation? *The Japanese Economic Review*, 72, pp. 5–48.

International Energy Agency (2016) *Energy Policies of IEA Countries, Japan*. Available at: https://www.iea.org/publications/freepublications/publication/EnergyPoliciesofIEACountriesJapan2016.pdf

Jaffe A.B., Newell R.G., Stavins R.N. (2003) Technological Change and the Environment. *Handbook of Environmental Economics*, 1, pp. 461–516.

Jaffe A.B., Stavins R.N. (1995) Dynamic Incentives of Environmental Regulations: The Effects of Alternative Policy Instruments on Technology Diffusion. *Journal of Environmental Economics and Management*, 29, 3, pp. S43–S63.

Japan Statistical Agency (2023) Kagaku gijutsu kenkyū chōsa 2023-nen (reiwa 5-nen). (Science and Technology Research Survey 2023 (Reiwa 5)). Available at: https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200543&tstat=000001032090&cycle=0&year=20230&month=0&tclass1=000001211180&tclass2val=0 File 4 Sangyō, kenkyūsha kibo betsu kenkyū kankei jūgyōshasū, shanai shiyō kenkyū-hi, ukeire kenkyū-hi oyobi shagai shishutsu kenkyū-hi (Number of research-related employees by industry and researcher size, in-house research expenses, incoming research expenses, and external research expenses) Available at: https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200543&tstat=000001032090&cycle=0&year=20230&month=0&tclass1=000001211180&stat_infid=000040126403&tclass2val=0

Jorgenson D.W., Wilcoxen P.J. (1990) Environmental Regulation and US Economic Growth. *The RAND Journal of Economics*, 21, pp. 314–340.

Jorgenson D., Wilcoxen P., Kneese A., Sweeney J. (1993) Energy, the Environment, and Economic Growth. *Handbook of Energy and Natural Resource Economics*.

Kimura Y. (2023) *Green Transformation Innovation*. RIETI Discussion Paper Series 23-E-086. Available at: https://www.rieti.go.jp/jp/publications/dp/23e086.pdf

Klette T.J., Kortum S. (2004) Innovating Firms and Aggregate Innovation. *Journal of Political Economy*, 112, 5, pp. 986–1018.

Kodama N., Li H. (2019) *Missing Growth in the Lost Decade*. RIETI Discussion Paper Series 19-E-026. Available at: https://www.rieti.go.jp/jp/publications/dp/19e026.pdf

Komiyama R., Fujii Y. (2017) Assessment of Post-Fukushima Renewable Energy Policy in Japan's Nation-Wide Power Grid. *Energy Policy*, 101, pp. 594–611.

Lee S., He Y., Suk S., Morotomi T., Chewpreecha U. (2022) Impact on the Power Mix and Economy of Japan under a 2050 Carbon-Neutral Scenario: Analysis Using the E3ME Macro-Econometric Model. *Climate Policy*, 22, 7, pp. 823–833.

Lee S., Pollitt H., Ueta K. (2012) An Assessment of Japanese Carbon Tax Reform Using the E3MG Econometric Model. *The Scientific World Journal*, 1, 835917.

Lentz R., Mortensen D.T. (2008) An Empirical Model of Growth through Product Innovation. *Econometrica*, 76, 6, pp. 1317–1373.

Maehara Y., Kuku A., Osabe Y. (2022) Macro Analysis of Decarbonization-Related Patent Technologies by Patent Domain-Specific BERT. *World Patent Information*, 69, 102112.

Matsumura K., Naka T., Sudo N. (2024) Analysis of the Transmission of Carbon Taxes Using a Multi-Sector DSGE. *Energy Economics*, 107642.

Metz B., Davidson O., Bosch P., Dave R., Meyer L. (2007) *IPCC, 2007: Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom; New York, NY, USA: Cambridge University Press.

Ministry of Economy, Trade and Industry (2014) *Strategic Energy plan 2014*. Available at: http://www.enecho.meti.go.jp/en/category/others/basic plan/pdf/4th strategic energy plan.pdf

Ministry of Finance. (2010) *Handbook of Japanese Taxes 2010*. Available at: https://www.mof.go.jp/english/tax policy/publication/taxes2010e/taxes2010e.pdf

Ministry of Finance (2015) *Tokubetsu kaikei gaido bukku, heisei 27 nenhan (Guidebook on special account, 2015, in Japanese)*. Available at: https://www.mof.go.jp/budget/topics/special_account/fy2016/tokka27togo.pdf

Ministry of the Environment (2017) *Greening of the Whole Tax System and Carbon Tax in Japan*. Available at: https://www.env.go.jp/en/policy/tax/20170130 greening.pdf

Mizunoya T., Higano Y. (2000) Air Pollutants Emission Reduction and the Optimal Tax-Subsidy Policy for Japan. *Studies in Regional Science*, 30, 3, pp. 85–104.

Nakata T., Lamont A. (2001) Analysis of the Impacts of Carbon Taxes on Energy Systems in Japan. *Energy Policy*, 29, 2, pp. 159–166.

Nijkamp P., Rodenburg C.A., Verhoef E.T. (2001) The Adoption and Diffusion of Environmentally Friendly Technologies among Firms. *International Journal of Environmental Technology and Management*, 1, 1–2, pp. 87–103.

Nomura N., Akai M. (2004) Willingness to Pay for Green Electricity in Japan As Estimated through Contingent Valuation Method. *Applied Energy*, 78, 4, pp. 453–463.

OECD (2009) *Patent Search Strategy for the Identification of Selected "Environmental" Technologies Developed As Part of the OECD Project on "Environmental Policy and Technological Innovation"*. Available at: https://www.oecd.org/env/consumption-innovation/44387109.pdf.

Popp D. (2006) R&D Subsidies and Climate Policy: Is There a "Free Lunch"? *Climatic Change*, 77, 3, pp. 311–341.

Popp D., Newell R. (2012) Where Does Energy R&D Come from? Examining Crowding out from Energy R&D. *Energy Economics*, 34, 4, pp. 980–991.

Popp D., Newell R.G., Jaffe A.B. (2010) Energy, the Environment, and Technological Change. *Handbook of the Economics of Innovation*, 2, pp. 873–937.

Posch O. (2011) Risk Premia in General Equilibrium. *Journal of Economic Dynamics and Control*, 35, 9, pp. 1557–1576.

Romer P.M. (1986) Increasing Returns and Long-Run Growth. *Journal of Political Economy*, 94, 5, pp. 1002–1037.

Romer P.M. (1990) Endogenous Technological Change. *Journal of Political Economy*, 98, 5, Part 2, pp. S71–S102.

Sanstad A.H., Blumstein C., Stoft S.E. (1995) How High Are Option Values in Energy-Efficiency Investments? *Energy Policy*, 23, 9, pp. 739–743.

Silva Herran D., Fujimori S. (2021) Beyond Japanese NDC: Energy and Macroeconomic Transitions Towards 2050 in Emission Pathways with Multiple Ambition Levels. *Sustainability Science*, 16, 2, pp. 489–501.

Smulders S., De Nooij M. (2003) The Impact of Energy Conservation on Technology and Economic Growth. *Resource and Energy Economics*, 25, 1, pp. 59–79.

Suzuki S., Takemura K. (2016) *The Effects of Diversity in Innovation: the Moderating Role of Universal-Diverse Leaders*. RIETI Discussion Paper Series 16-E-086. Available at: https://www.rieti.go.jp/jp/publications/dp/16e086.pdf

Takeda S., Arimura T.H. (2021) A Computable General Equilibrium Analysis of Environmental Tax Reform in Japan with a Forward-Looking Dynamic Model. *Sustainability Science*, 16, 2, pp. 503–521.

Tanikawa H. (2004) *Incentive Schemes for Japanese Companies' Voluntary Environmental Initiatives-Empirical Analysis Based on Case Studies and Questionnaires (Japanese)*. Research Institute of Economy, Trade; Industry (RIETI), 04-J-030.

Upadhyaya S. (2010) *Compilation of Energy Statistics for Economic Analysis*. United Nations Industrial Development Organization (UNIDO), Vienna. WP 01/2010. Available at: https://downloads.unido.org/ot/99/28/9928770/WP1.pdf

Wakasugi R., Koyata F. (1997) R&D, Firm Size and Innovation Outputs: Are Japanese Firms Efficient in Product Development? *Journal of Product Innovation Management*, 14, 5, pp. 383–392.

Wakiyama T., Zusman E. (2016) What Would Be the Effects of a Carbon Tax in Japan: An Historic Analysis of Subsidies and Fuel Pricing on the Iron & Steel, Chemical, and Machinery Industries. *AIMS Energy*, 4, 4, pp. 606–632.

WIPO (2017) *International Patent Classification (IPC) Green Inventory*. Available at: http://www.wipo.int/classifications/ipc/en/est/.

Xepapadeas A. (2005) Economic Growth and the Environment. *Handbook of Environmental Economics*, 3, pp. 1219–1271.

Влияние налога на выбросы углерода и субсидий на НИОКР на экономический рост в Японии

Бесстремянная Галина Евгеньевна[1], **Дашер Ричард**[2], **Головань Сергей Витальевич**[3]

[1] Старший научный сотрудник Международной лаборатории макроэкономического анализа, профессор департамента прикладной экономики,
Национальный исследовательский университет «Высшая школа экономики»,
11, Покровский бульвар, Москва, 101000, Россия.
E-mail: gbesstremyannaya@hse.ru

[2] Директор, Американо-азиатский центр управления технологиями,
Стэнфордский университет
521, Мэмориал Уэй, Стэнфорд, Калифорния, 94305, США.
E-mail: rdasher@stanford.edu

[3] Доцент, Российская экономическая школа,
3, ул. Нобеля, Инновационный центр Сколково, Москва, 121205, Россия.
E-mail: sgolovan@nes.ru

Множество исследований свидетельствует о том, что налог на выбросы углерода в сочетании с субсидиями на НИОКР могут рассматриваться как эффективная форма государственной политики для поощрения распространения низкоуглеродных технологий на благо общества. В данной статье используется макроэкономический подход моделей эндогенного роста с технологическими изменениями для проведения оценки влияния таких мер на экономический рост в Японии в среднесрочной и долгосрочной перспективе, и для сравнения наших результатов с существующими выводами о подобных мерах для экономики США. Наши оценки с использованием микро- и макроданных выявляют сходство между японскими и американскими энергетическими компаниями в отношении эластичности функции производства инноваций по расходам на НИОКР и в отношении вероятности радикальных инноваций. Однако, согласно проанализированным нами данным об энергетических патентах в Японии, чистые инновации не так широко распространены в этой стране, как в США. Это может объяснить наши количественные выводы о необходимости более сильной опоры на налог на выбросы углерода в Японии по сравнению с США.

Ключевые слова: эндогенный рост; технологические изменения; инновации; налог на выбросы; энергетика.

УДК 336.762.1

DOI: 10.17323/1813-8691-2025-29-1-103-131

Гибридные подходы к прогнозированию реализованной волатильности ETF: глубокое обучение и теорема восстановления

Патласов Д.А.

В данной статье исследуется задача многошагового прогнозирования реализованной волатильности. В работе вводится модификация функции потерь вида квантильный лог-гиперболический косинус (quantile log-cosh), также в качестве экзогенных факторов используется информация, извлекаемая из опционов с помощью теоремы восстановления [Ross, 2015] в контексте задачи прогнозирования реализованной волатильности торгуемых биржевых фондов (Exchange-Traded Fund, ETF) SPY (SPDR S&P 500 ETF Trust) и QQQ (Invesco QQQ Trust). Ставятся две гипотезы: первая предполагает, что quantile log-cosh в нейронных сетях повысит точность предиктивной модели на тестовом наборе данных по сравнению с теми же моделями, обучаемыми на других целевых функциях. Вторая гипотеза заключается в использовании информации, извлекаемой из теоремы восстановления. Данная теорема позволяет аппроксимировать истинную плотность распределения состояний SPY и QQQ в терминах марковских цепей и избавиться от предпосылок риск-нейтральной меры в финансовых моделях. Тогда по второй гипотезе ожидается, что модель с факторами, извлеченными с помощью теоремы восстановления, будет показывать более точные прогнозы на тестовой выборке по сравнению с классической моделью гетерогенной авторегрессии (Heterogeneous Autoregressive Model for Realized Volatility, HAR-RV). Для проверки гипотез используются следующие модели машинного обучения: LSTM, GRU, BiLSTM, BiGRU, FCNN и N-BEATS. Результаты показывают, что модификация quantile log-cosh позволяет улучшить точность предсказаний моделей на тестовом наборе данных. Также включение в модели прогнозирования реализованной волатильности экзогенных факторов из теоремы восстановления позволяет значительно превзойти модель HAR-RV, особенно на долгосрочном горизонте.

Патласов Дмитрий Александрович – аспирант кафедры информационных систем и математических методов в экономике, Пермский государственный национальный исследовательский университет. E-mail: dmitriypatlasov@gmail.com

Статья поступила: 24.07.2024/Статья принята: 06.02.2025.

Ключевые слова: прогнозирование; нейронные сети; реализованная волатильность; теорема восстановления.

Для цитирования: Патласов Д.А. Гибридные подходы к прогнозированию реализованной волатильности ETF: глубокое обучение и теорема восстановления. *Экономический журнал ВШЭ*. 2025; 29(1): 103–131.

For citation: Patlasov D.A. Hybrid Approaches to Predicting Realized ETF Volatility: Deep Learning and the Recovery Theorem. *HSE Economic Journal.* 2025; 29(1): 103–131. (In Russ.)

1. Введение

Фондовый рынок является одним из центральных элементов мировой экономики, оказывая значительное влияние на финансовую стабильность и экономическое развитие стран. В современном мире, характеризующемся высокой степенью неопределенности, предсказание волатильности рынка становится важнейшим аспектом для агентов финансовых рынков. В данной статье рассматривается задача прогнозирования волатильности для ETF SPY и QQQ.

Следует отметить, что ETF обладают рядом уникальных характеристик, отличающих их от традиционных активов, таких как акции или фьючерсы. Основные особенности ETF, влияющие на волатильность:

– ETF торгуются как акции, но их стоимость привязана к стоимости базовых активов через механизм создания и погашения паев (creation and redemption). Это снижает отклонение цены ETF от справедливой стоимости, но может приводить к повышенной краткосрочной волатильности в периоды рыночного стресса;

– ликвидность ETF определяется не только объемом торгов, но и ликвидностью базовых активов. В случае резких рыночных движений повышенная торговая активность может усилить волатильность ETF;

– трейдеры и маркет-мейкеры активно используют ETF, что делает их подверженными моментальным скачкам волатильности.

Наиболее распространенными подходами к моделированию волатильности в XX в. считаются модели на базе условной гетероскедастичности для ненаблюдаемой волатильности финансовых активов, а также модель стохастической волатильности для подразумеваемой волатильности опционов. Однако эти модели сильно зависят от спецификации базового процесса волатильности и требуют жестких ограничений на оцениваемые параметры. В то время считалось, что волатильность ненаблюдаема. В дальнейших работах внимание исследований по измерению волатильности переключилось на данные высокой частоты. Впервые Мертон [Merton, 1980] показал, что волатильность может быть определена как сумма квадратов доходностей на уровне высокой частоты. Авторы работ [Andersen, Bollerslev, Diebold, Ebens, 2001], а также [Andersen, Bollerslev, Diebold, Labys, 2003] указали, что суммирование квадратов внутридневных доходностей дает наблюдаемую меру дневной волатильности, названную реализованной дисперсией (Realized Variance, RV). Тогда можно сказать, что волатильность, за исключением ошибок измерения и скачков, становится «наблюдаемой».

RV измеряет внутридневную волатильность, где частота выборки может быть даже на уровне тиков. По мере уменьшения интервалов RV приближается к непрерывной волатильности. Андерсен и Боллерслев [Andersen, Bollerslev, 1998] определяют RV следующим образом:

(1) $$RV_{t+1} = \sum_{j=1,\dots,m} r^2_{m,t+\frac{j}{m}},$$

где m – количество интервалов, а непрерывно накопленные доходности определяются равенством

(2) $$r_{m,t} \equiv p_{t,m} - p_{t-1,m}, \forall m = 1,2,3,\dots,n,$$

где p_t и p_{t-1} – логарифмы цен финансовых активов.

RV может быть применена к ценам, следующим винеровскому процессу. Математическое определение винеровского процесса W_t следующее:

1. $W_0 = 0$ с вероятностью 1.

2. W_t – процесс с независимыми приращениями.

3. $W_t - W_s \sim N(0, \sigma^2_{t-s}), \forall 0 \le s < t < +\infty.$

4. Траектории W_t непрерывны с вероятностью 1.

В контексте финансовых рынков, если p_t следует винеровскому процессу, это означает, что изменения логарифма цены могут быть описаны в виде

(3) $$dp_t = \mu_t dt + \sigma_t dW_t, t \in \mathbb{Z},$$

где σ_t – мгновенная волатильность процесса или стандартное отклонение, строго положительная и суммируемая с квадратом функция $\left(\int_0^t \sigma_s^2 ds < \infty\right)$; μ_t – дрейфовая составляющая; dW_t – дифференциал винеровского процесса.

Далее рассмотрим непрерывно начисленную доходность между $t-h$ и t, при этом $0 < h \le t$:

(4) $$r_t = p_t - p_{t-h} = \int_{t-h}^{t} \mu_s ds + \int_{t-h}^{t} \sigma_s dW_s.$$

Квадратичная дисперсия (Quadratic Variance, QV) доходности – естественная мера изменчивости диффузионного процесса мартингала, согласно теореме о стохастическом интегрировании, определяется равенством

(5) $$p_t = QV_t = \int_{t-h}^{t} \sigma_s^2 ds,$$

где QV_t – квадратичная дисперсия.

Равенство (5) показывает, что инновации дрейфа не влияют на дисперсию диффузионного процесса доходностей. В этом контексте квадратичная дисперсия полностью определяется инновациями локального мартингала, она совпадает с интегрированной дисперсией (Integrated Variance, IV), тогда накопленная волатильность доходности определяется равенством

(6) $$IV_t = \int_{t-h}^{t} \sigma_s^2 ds = QV_t,$$

где IV_t – интегрированная дисперсия.

Авторы работ [Andersen, Bollerslev, Diebold, Labys, 2001; Barndorff-Nielsen, Shephard, 2002] из свойств квадратичной дисперсии отмечают, что при отсутствии неэффективности микроструктуры на рынке QV доходностей может выражена:

(7) $$p_t = p\lim_{n\to\infty} \sum_{j=1}^{n} \left[p_{s_j} - p_{s_{j-1}} \right]^2,$$

для каждой последовательности разбиения $0 = s_0 < s_1 < \ldots < s_n = t$ при $| s_j - s_{j-1} | \to 0$, когда число разбиений $n \to \infty$.

Равенство (2) можно представить в виде:

(8) $$r_t = \sum_{m=1}^{n} r_{t,m}.$$

Тогда реализованная дисперсия сходится по вероятности к квадратичной дисперсии и к интегрированной дисперсии дня t при $n \to \infty$. Пусть реализованная дисперсия будет

(9) $$RV_t = \sum_{m=1}^{n} r_{t,m}^2.$$

Отсюда следует, что

(10) $$RV_t \xrightarrow{p} \left[r, r \right]_t - \left[r, r \right]_{t-h} \equiv QV_t,$$

(11) $$RV_t \xrightarrow{p} IV_t.$$

Реализованная волатильность – это последовательная оценка условной волатильности, такая, что

(12) $$RV_t \xrightarrow{p} v_{t,h}^2,$$

где $v_{t,h}^2$ – условная дисперсия доходностей на интервале $\left[t-h, t \right]$, которая может быть оценена, например, семейством моделей GARCH.

Это подразумевает, что ожидаемая реализованная волатильность является последовательной оценкой ожидаемой условной волатильности:

$$\text{(13)} \qquad \mathrm{E}\left[RV_t \mid \Omega_{t-h}\right] \xrightarrow{p} \mathrm{E}\left[v_{t,h}^2 \mid \Omega_{t,h}\right],$$

где Ω_{t-h} – доступное информационное множество, доступное на момент времени $t-h$.

Если процесс доходности определен функцией, суммируемой с квадратом и $\mu_t \equiv 0$, то реализованная волатильность является несмещенной оценкой условной дисперсии доходности:

$$\text{(14)} \qquad \mathrm{E}\left[RV_t \mid \Omega_{t-h}\right] = \mathrm{E}\left[QV_t \mid \Omega_{t,h}\right] = \mathrm{var}\left[r_{t,h} \mid \Omega_{t,h}\right].$$

Этот эквивалент объединяет реализованную волатильность и условную дисперсию из моделей GARCH.

Из существующих стилизованных фактов волатильности отмечается, что, как правило, включение экзогенных признаков в модели прогнозирования волатильности не способно существенно повысить точность (например: [Engle, 1982; Christoffersen, Diebold, 2000; Andersen, Bollerslev, Diebold, 2003; Welch, Goyal, 2008; Bollerslev, Tauchen, Zhou, 2009; Hansen, Lunde, 2005]).

Автор настоящего исследования видит ключевую ценность в контексте задачи прогнозирования волатильности в информации, извлекаемой из применения теоремы восстановления (the recovery theorem) [Ross, 2015] к рыночным данным опционов, преобразовании этих результатов в вид временных рядов и использовании в качестве признаков в модели прогнозирования реализованной волатильности активов.

Теорема восстановления позволяет произвести переход от риск-нейтральной меры к истинной (физической) мере. Помимо смены вероятностной меры, теорема восстановления позволяет получить плотность распределения базового актива опциона в истинной мере. Следует отметить, что, несмотря на дискуссии среди авторов теоретических исследований о применении теоремы восстановления для повышения точности прогнозов различных финансовых показателей, а также улучшения торговых стратегий и управления рисками, эмпирические работы свидетельствуют о том, что даже при частичном выполнении условий применимости результаты прогнозных моделей улучшаются. Об этом говорят, например, исследования [Audrino, Huitema, Ludwig, 2021; Gagnon, Power, Toupin, 2022]. Среди теоретических работ, критикующих теорему восстановления, можно отметить исследования [Amir, 2019; Borovička, Hansen, Scheinkman, 2016].

Также в настоящем исследовании вводится модификация функции потерь quantile log-cosh. Модификация заключается в корректировке штрафа функции потерь. Предлагается увеличивать штраф за ложные прогнозы модели на данных выше 0,75-го квантиля. Так как мера волатильности является показателем, который отражает состояние нестабильности финансового рынка, то ошибки модели в наиболее высоких значениях волатильности могут быть критическими с практической точки зрения. Тогда появляются основания для улучшения качества прогнозов модели в области больших значений.

2. Данные

Для расчетов факторов из теоремы восстановления используются данные о наблюдаемой подразумеваемой волатильности колл-опционов, страйках и сроках до погашения на два ETF-инструмента: SPY – наиболее ликвидный ETF, отслеживающий индекс S&P 500, является репрезентативным для широкого фондового рынка США; QQQ – фокусируется на технологическом секторе (NASDAQ-100), что делает его более чувствительным к изменениям волатильности в периоды технологических пузырей или кризисов. Данные получены из цепочек опционов (SPX Option Chains). Следует отметить, что S&P 500 является наилучшей аппроксимацией на движение всего фондового рынка США, а NASDAQ 100 приближением на движения технологических компаний. Исходные данные являются функцией, которая позволяет оценивать уровень подразумеваемой волатильности для любой пары страйк и срок до погашения, что представляет из себя поверхность подразумеваемой волатильности (implied volatility surface). Поставщик данных: optionsDX, URL: https://www.optionsdx.com/. Расчет по методологии исследования производится на каждый торговый день с 01.01.2010 г. по 31.12.2023 г. для SPY и с 01.01.2012 г. по 31.12.2023 г. для QQQ.

В качестве безрисковой ставки рынка США используется доходность по 10-летним казначейским облигациям США. Тикер инструмента: «^TNX», поставщик информации: Yahoo Finance.

Тогда, на первом этапе были подготовлены данные по цепочкам опционов на инструменты SPY и QQQ, а далее определены матрицы поверхности подразумеваемой волатильности для обоих инструментов на конец каждой торговой сессии. Каждый элемент матрицы – это наблюдаемый на срочном рынке колл-опцион, базовыми активами которых являются SPY и QQQ со своими значениями страйка и срока до погашения. Среднее количество колл-опционов для воспроизведения поверхности подразумеваемой волатильности составило 351 для инструмента SPY и 322 для инструмента QQQ. После чего к исходным данным по опционам были применены все итерации, определенные в теореме восстановления.

Для расчета реализованной волатильности применяются исторические котировки торгуемых биржевых фондов SPY и QQQ. Частота наблюдений равна одному дню. Данные по котировкам получены с Yahoo Finance.

3. Методы

3.1. Постановка задачи

В статье рассматривается проблема прогнозирования реализованной волатильности торгуемых биржевых фондов SPY и QQQ на основе показателей из теоремы восстановления Росса. Реализованная волатильность представляет собой измеренную волатильность на основе фактических цен закрытия для акций, фьючерсов, индексов и т.д. [Corsi, 2008]).

Цель работы состоит в том, чтобы разработать архитектуру и обучить модель машинного обучения, которая наилучшим образом способна прогнозировать реализованную волатильность на несколько шагов вперед, на базе предпосылок модели HAR-RV, факторов, извлеченных после оценки теоремы восстановления и оценки модели с помощью нейронных сетей.

В HAR-модели будущая реализованная волатильность является функцией запаздывающей дневной, недельной и месячной реализованной волатильности:

$$RV_{t+k} = \alpha^d + \beta^d RV_t^d + \beta^w RV_t^w + \beta^m RV_t^m + \varepsilon_{t+k}^d, \quad (15)$$

где d – дневная частота; w – недельная частота (5 торговых дней); m – месячная частота (21 торговый день); k – шаг прогнозирования, дни (модели будут оцениваться для $k \in [5, 10, 15, 20]$.

Формула для расчета реализованной волатильности:

$$RV_t = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\ln\left(\frac{S_t}{S_{t-1}}\right)\right]^2}, \quad (16)$$

где RV_t – реализованная волатильность; n – количество наблюдений (диапазон скользящего окна); S_t – цена SPY или QQQ.

Относительно простая структура и возможность оценить модель с помощью метода наименьших квадратов (МНК) способствовали широкому использованию этой спецификации. Корси [Corsi, 2008] показал, что эта модель генерирует более точные прогнозы на тестовой выборке (out-of-sample) по сравнению с моделями краткосрочной памяти.

В рамках статьи ставится задача включения в перечень экзогенных переменных модели HAR-RV дополнительных параметров, которые способны усилить точность модели на тестовой выборке. Назовем такую модель гибридной и присвоим расширение до аббревиатуры HARx-RV. Список исследуемых зависимых переменных и признаков приведены в табл. 1.

Таблица 1.

Определение переменных, используемых в моделях прогнозирования реализованной волатильности инструментов SPY и QQQ

Обозначение переменной	Название	Описание
Зависимые переменные (эндогенные факторы)		
$RV_{t+5}; RV_{t+10}; RV_{t+15}; RV_{t+20}$	Реализованная волатильность на 5, 10, 15, 20 дней вперед	Реализованная волатильность, рассчитанная на данных SPY и QQQ: Частота данных – 1 день; Диапазон скользящего окна – 21 день
Регрессоры (экзогенные факторы)		
$RV_t^d; RV_t^w; RV_t^m$	Реализованная волатильность	–

Окончание табл. 1.

Обозначение переменной	Название	Описание
$\psi_{1,t};\psi_{1,t-5}$	Восстановленная ожидаемая доходность	Аппроксимация среднего из истинного распределения вероятностей (физической меры по теореме восстановления)
$\psi_{2,t};\psi_{2,t-5}$	Восстановленная ожидаемая волатильность	Аппроксимация стандартного отклонения из истинного распределения вероятностей (физической меры по теореме восстановления)
$\psi_{3,t};\psi_{3,t-5}$	Восстановленный эксцесс	Аппроксимация эксцесса из истинного распределения вероятностей (физической меры по теореме восстановления)
$\psi_{4,t};\psi_{4,t-5}$	Восстановленная асимметрия	Аппроксимация асимметрии из истинного распределения вероятностей (физической меры по теореме восстановления)
$\psi_{5,t};\psi_{5,t-5}$	Склонность к риску (предпочтение риска)	Разница восстановленной ожидаемой волатильности и риск-нейтральной (в риск-нейтральной мере)
$\psi_{6,t};\psi_{6,t-5}$	Ожидаемая вероятность падения инструмента на 20% в следующем месяце	Вероятность, соответствующая 20-процентному снижению базового актива из эмпирической функции плотности распределения (из теоремы восстановления)

Тогда модель HARx-RV примет вид (для каждого k):

$$RV_{t+k} = \alpha^d + \beta^d RV_t^d + \beta^w RV_t^w + \beta^m RV_t^m + \\ + \gamma_1\psi_{1,t} + \gamma_2\psi_{2,t} + \gamma_3\psi_{3,t} + \gamma_4\psi_{4,t} + \gamma_5\psi_{5,t} + \gamma_6\psi_{6,t} + \\ + \delta_1\psi_{1,t-5} + \delta_2\psi_{2,t-5} + \delta_3\psi_{3,t-5} + \delta_4\psi_{4,t-5} + \delta_5\psi_{5,t-5} + \delta_6\psi_{6,t-5} + \varepsilon_{t+k}^d. \tag{17}$$

В дальнейшем оценка модели HARx-RV будет произведена при помощи МНК с коррекцией Ньюи – Уэста [Костырка, Малахов, 2021], а также алгоритмами глубокого обучения.

Следует заметить, что выборка разделена на обучающую (с 01.01.2010 г. (для SPY) с 01.01.2012 г. (для QQQ) до 31.12.2018 г.), валидационную (с 01.01.2019 г. по 31.06.2020 г.) и тестовую (с 01.07.2020 г. по 31.12.2023 г.).

Далее формально опишем процесс извлечения из данных по опционам информации, получаемой после применения теоремы восстановления.

Теорема восстановления предоставляет метод для извлечения истинных вероятностей и ядра ценообразования (pricing kernel) из наблюдаемых рыночных данных по опционам, при условии, что известны переходные вероятности в риск-нейтральной мере. Основная идея теоремы заключается в следующем:

1) риск-нейтральная плотность распределения π^* и истинная плотность распределения f связаны через отношение маржинальных полезностей потребления;

2) риск-нейтральные вероятности используются для ценообразования финансовых инструментов в модели отсутствия арбитража;

3) истинные вероятности отражают реальное распределение вероятностей исходов в будущем с учетом субъективных предпочтений инвесторов.

Пусть $\pi^*\left(\theta_i,\theta_j\right)$ – риск-нейтральная вероятность перехода из состояния θ_i в состояние θ_j, а $f\left(\theta_i,\theta_j\right)$ – истинная вероятность этого перехода. Тогда эти вероятности связаны следующим образом [Ross, 2015]:

$$\pi^*\left(\theta_i,\theta_j\right)=e^{r(\theta_i)}\frac{U'\left(c\left(\theta_j\right)\right)}{U'\left(c\left(\theta_i\right)\right)}f\left(\theta_i,\theta_j\right), \tag{18}$$

где $r\left(\theta_i\right)$ – безрисковая ставка; $U'\left(c(\cdot)\right)$ – маржинальная полезность потребления; $c\left(\theta\right)$ – потребление в состоянии θ.

Для получения истинной и риск-нейтральной плотностей распределения используется следующая процедура:

1) определение риск-нейтральной плотности: плотность определяется через рыночные цены опционов, используя модель ценообразования опционов Блэка – Шоулза, а также процедуру извлечения цен Эрроу – Дебре;

2) восстановление истинной плотности: истинная плотность восстанавливается из π^* с использованием отношения маржинальных полезностей и безрисковой ставки, как указано в уравнении выше. Математически это реализуется через спектральное разложение матрицы переходных вероятностей в риск-нейтральной мере;

3) переходные вероятности и марковские цепи: восстановление вероятностей переходов рынка из одного состояния в другое осуществляется с использованием марковских цепей.

Из получаемой марковской цепи можно извлечь плотность распределения рынка на каждый торговый день, из которой в свою очередь считаются моменты распределения. Более подробно о полной процедуре применения и доказательстве существования теоремы восстановления можно узнать в работе Стивена Росса [Ross, 2015], а также в работах [Bakshi, Chabi-Yo, Gao, 2018; Carr, Yu, 2012; Jackwerth, Menner, 2020]. При подтверждении гипотезы о том, что прогнозная модель будет обладать достаточным уровнем обобщенности и способна объяснять динамику реализованной волатильности, предполагается проверка этой модели на устойчивость (робастность). Для проверки робастности будет произведено сравнение с базовой моделью HAR-RV с помощью теста Дибольда – Мариано.

Для достижения поставленной цели предполагается решение следующих задач:

– загрузка, предобработка исходных данных;

– отбор и обучение моделей машинного обучения;

– оценка качества обученных моделей на тестовой выборке при помощи следующих метрик качества: MSE, MAE, R^2, MAPE, QLIKE (Quasi-Likelihood), тест Дибольда – Мариано.

3.2. Модифицированная функция потерь квантильный лог-гиперболический косинус

В качестве элемента предварительного анализа данных посмотрим на режимы «низкой» и «высокой» волатильности по обоим инструментам с помощью модели скрытых марковских процессов (HMM).

На рис. 1, 2 представим разделение режимов реализованной волатильности SPY и QQQ на обучающей выборке. В рамках предварительного анализа не будем рассматривать реализованную волатильность на валидационной и тестовой выборках, чтобы не было априорного представления о динамике и распределении реализованной волатильности в данные периоды времени.

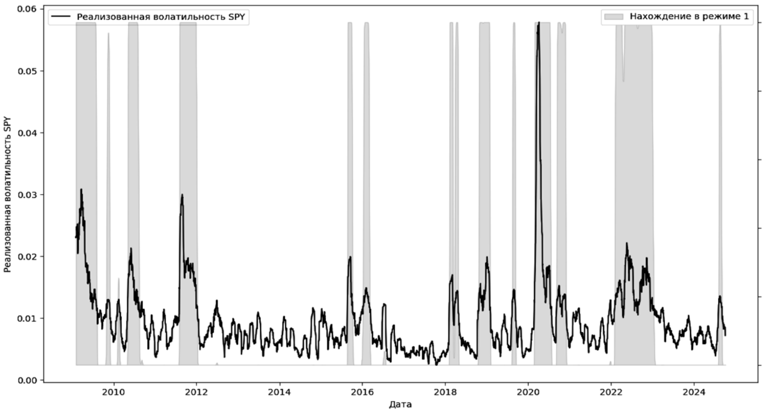


Рис. 1. HMM-модель реализованной волатильности SPY с вероятностями переходов между режимами на обучающей выборке (01.01.2010 г.–31.12.2018 г.)

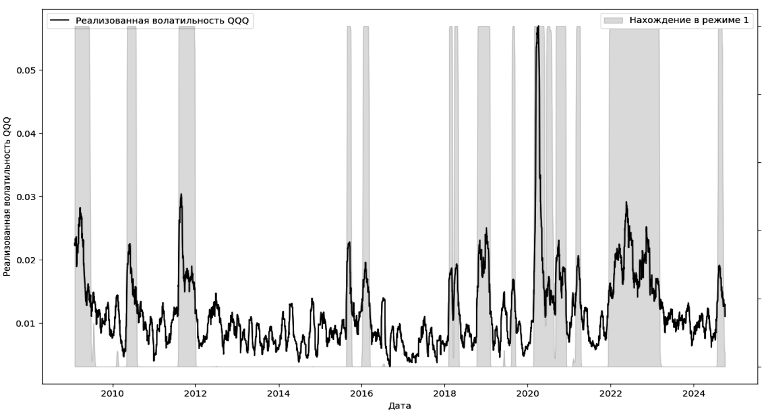


Рис. 2. HMM-модель реализованной волатильности QQQ с вероятностями переходов между режимами на обучающей выборке (01.01.2012 г.–31.12.2018 г.)

Несмотря на то, что модель HMM не будет непосредственно использоваться в задаче прогнозирования, она играет важную роль в общей постановке задачи. Рассмотрение HMM в контексте прогнозирования реализованной волатильности является основанием для формирования контекста для других моделей. Теперь допустимо сказать, что сменяемость режима волатильности сопровождается ее резким скачком (прыжком), когда значения RV переходят в диапазон высоких квантилей. Отсюда возникает необходимость разработки метода, который бы адаптировался под более точное обучение модели на данных высоких квантилей. Для решения такой проблемы в настоящем исследовании вводится модификация целевой функции нейронных сетей лог-гиперболического косинуса (log-cosh). Модификацию будем называть квантильный лог-гиперболический косинус (quantile log-cosh).

Функция потерь log-cosh является одним из распространенных выборов для задач регрессии, особенно в контексте временных рядов и данных с сильными флуктуациями. Ее преимущества можно объяснить следующим образом.

1. Робастность: log-cosh предлагает робастный подход к оценке ошибки, так как лучше учитывает большие ошибки по сравнению с квадратичной функцией потерь MSE и средней абсолютной ошибкой MAE. Следует также отметить, что функция потерь log-cosh хорошо адаптируется к особенностям данных, так как она позволяет модели оценивать как большие, так и маленькие ошибки с одинаковым весом [Barron, 2019]. В эмпирических тестах модель с log-cosh демонстрирует значительное улучшение точности прогнозирования по сравнению с MSE и MAE, особенно в условиях высокой волатильности [Barron, 2019].

2. Совместимость с оптимизационными методами: log-cosh в качестве функции потерь хорошо работает с различными оптимизационными алгоритмами, которые применяются при обучении нейронных сетей.

3. Возможности для модификации: введение штрафа за ошибки в верхнем квартиле данных (75-й квантиль) позволяет более точно прогнозировать экстремальные значения волатильности, что критически важно для управления рисками.

Далее представим предлагаемую модификацию. Дополнение заключается в добавлении штрафа λ при расчете ошибок на данных больших значений, а именно выше η-го квантиля. Тогда новая лосс-функция quantile log-cosh примет вид:

$$\ell\left(RV, \widehat{RV}\right) = \sum_{i=1}^{n} \ln\left[\cosh\left(RV_i, \widehat{RV}_i\right) \cdot \left(1 + 1_{\{RV_i > q_\eta\}} \cdot \lambda \cdot \left(RV_i - q_\eta\right)\right)\right], \quad (19)$$

где $\ell\left(RV, \widehat{RV}\right)$ – функция потерь (quantile log-cosh); q_η – значение η-го квантиля от выборки истинных значений зависимой переменной на обучающем наборе данных; $1_{\{RV_i > q_\eta\}}$ – индикаторная функция, которая принимает значение 1, когда RV_i больше q_η и 0 в ином случае; λ – множитель штрафа к функции потерь выше η-го квантиля.

3.3. Описание исходных данных

Представим в табл. 2 описательные статистики исходных данных.

Таблица 2.

Описательные статистики временных рядов признаков на обучающей выборке

Фактор	Среднее	Стандартное отклонение	Минимум	Медиана	Максимум	Асимметрия	Эксцесс
Описательные статистики обучающей выборки SPY (2010-01-04–2018-12-31)							
ψ_1	–0,00	0,01	–0,15	–0,00	0,10	–0,51	20,23
ψ_2	0,08	0,04	0,03	0,07	0,31	1,77	3,63
ψ_3	–0,67	1,24	–2,60	–0,98	4,73	1,15	1,10
ψ_4	–0,51	0,41	–1,82	–0,46	0,86	–0,41	–0,04
ψ_5	0,03	0,03	0,01	0,02	0,22	2,11	5,58
ψ_6	0,03	0,05	0,00	0,01	0,31	2,64	7,87
RV	0,01	0,00	0,00	0,01	0,03	1,64	3,68
Описательные статистики обучающей выборки QQQ (2012-01-03–2018-12-31)							
ψ_1	–0,01	0,06	–0,60	0,00	0,15	–5,14	32,14
ψ_2	0,16	0,12	0,04	0,13	2,04	5,24	49,86
ψ_3	–1,44	0,83	–2,90	–1,52	2,21	0,86	0,90
ψ_4	–0,20	0,34	–1,44	–0,19	0,87	0,11	1,17
ψ_5	0,11	0,12	0,01	0,08	1,94	5,35	52,49
ψ_6	0,11	0,21	0,00	0,07	2,19	6,32	47,66
RV	0,01	0,00	0,00	0,01	0,02	1,46	2,31

Далее проведем тесты на коинтеграцию между признаками моделей и реализованной волатильностью на 5, 10, 15 и 20 дней вперед. Для этого оценим статистики Энгла – Грейнджера (EG-тест) и Йохансена [Dwyer, 2015].

Тест Энгла – Грейнджера проверяет наличие коинтеграционных связей между временными рядами на двухэтапной основе. Нулевая гипотеза теста: остатки нестационарны. Если нулевая гипотеза отвергается, значит, остатки стационарны, что свидетельствует о наличии коинтеграции.

Тест Йохансена – более мощный метод, который позволяет выявлять наличие коинтеграции между несколькими временными рядами. Он основан на модели векторной авторегрессии (VAR), преобразованной в модель векторной коррекции ошибок. Нулевая гипотеза теста Йохансена: количество коинтеграционных векторов меньше или равно заданному значению. Представим результаты данных тестов в табл. 3.

Таблица 3.

Тесты на коинтеграцию для реализованной волатильности SPY и QQQ (критическое значение теста Йохансена (5%) равно 15,49)

Фактор	SPY		QQQ	
	p-value теста Энгла – Грейнджера	статистика теста Йохансена	p-value теста Энгла – Грейнджера	статистика теста Йохансена
Целевая переменная – RV_{t+5}				
ψ_1	0,01	524,64	0,31	446,56
ψ_2	0,00	164,95	0,34	277,58
ψ_3	0,00	378,89	0,32	113,56
ψ_4	0,02	506,77	0,33	141,93
ψ_5	0,00	167,58	0,20	265,07
ψ_6	0,00	276,29	0,21	423,18
RV_t	0,00	1055,52	0,00	544,07
RV_{t-5}	0,00	551,55	0,00	277,71
RV_{t-21}	0,00	59,52	0,01	44,81
Целевая переменная – RV_{t+10}				
ψ_1	0,00	508,80	0,20	442,60
ψ_2	0,00	169,18	0,14	274,35
ψ_3	0,00	375,82	0,10	115,02
ψ_4	0,01	506,25	0,21	141,42
ψ_5	0,00	167,74	0,26	262,04
ψ_6	0,00	274,92	0,16	421,06
RV_t	0,00	556,14	0,00	280,86
RV_{t-5}	0,00	356,06	0,00	192,51
RV_{t-21}	0,00	41,36	0,01	38,65

Окончание табл. 3.

Фактор	SPY		QQQ	
	p-value теста Энгла – Грейнджера	статистика теста Йохансена	p-value теста Энгла – Грейнджера	статистика теста Йохансена
Целевая переменная – RV_{t+15}				
ψ_1	0,01	509,81	0,14	443,72
ψ_2	0,00	171,52	0,11	276,15
ψ_3	0,00	367,21	0,07	126,75
ψ_4	0,02	502,83	0,13	144,05
ψ_5	0,00	165,29	0,14	264,18
ψ_6	0,00	263,05	0,13	421,15
RV_t	0,00	365,98	0,00	187,01
RV_{t-5}	0,00	0,00	0,00	118,84
RV_{t-21}	0,00	0,00	0,01	38,28
Целевая переменная – RV_{t+20}				
ψ_1	0,01	495,96	0,01	447,94
ψ_2	0,00	188,56	0,01	279,46
ψ_3	0,00	368,20	0,01	138,67
ψ_4	0,01	506,00	0,02	148,74
ψ_5	0,00	177,07	0,01	268,13
ψ_6	0,00	271,15	0,01	421,27
RV_t	0,00	245,00	0,00	128,90
RV_{t-5}	0,00	71,36	0,00	50,88
RV_{t-21}	0,00	43,61	0,00	44,73

Тесты на коинтеграцию показывают, что для SPY существует значительная коинтеграция между признаками и реализованной волатильностью на всех горизонтах прогноза, что указывает на наличие устойчивых долгосрочных взаимосвязей. Для QQQ коинтеграция проявляется менее выраженно по результатам теста Энгла – Грейнджера, но статистики теста Йохансена подтверждают наличие коинтеграционных связей на всех горизонтах для всех переменных.

3.4. Модели оценки

3.4.1. Рекуррентные нейронные сети LSTM и GRU

В ряде исследований, например [Rodikov, Antulov-Fantulin, 2022], было показано, что LSTM и GRU могут улучшить прогнозы реализованной волатильности по сравнению с классическими моделями GARCH и HAR-RV, особенно для краткосрочных прогнозов. Тем не менее в большинстве работ по прогнозированию реализованной волатильности отмечается, что модели машинного обучения не всегда значительно превосходят эконометрические, в том числе линейные модели для прогнозирования реализованной волатильности, особенно в случае более долгосрочных прогнозов. В работе [Branco et al., 2024] говорится о том, что линейные модели на базе гетерогенной авторегрессии часто оказываются более конкурентоспособными, особенно для прогнозов на месячных и недельных интервалах по сравнению с алгоритмами глубокого обучения. Однако в краткосрочных прогнозах (на один день) модели LSTM и GRU могут показать лучшие результаты.

Также важно отметить, что эконометрические модели имеют значительное преимущество в интерпретируемости. Например, модели HAR-RV и GARCH позволяют понимать эффекты каждого из лагов или компонентов волатильности.

В рамках настоящего исследования будет произведено сравнение точности прогнозирования реализованной волатильности на примере моделей LSTM, GRU, BiLSTM и BiGRU с линейной моделью HAR-RV с коррекцией Ньюи – Уэста.

Далее опишем некоторые математические свойства моделей глубокого обучения. На рис. 3 представим рекуррентный модуль LSTM.

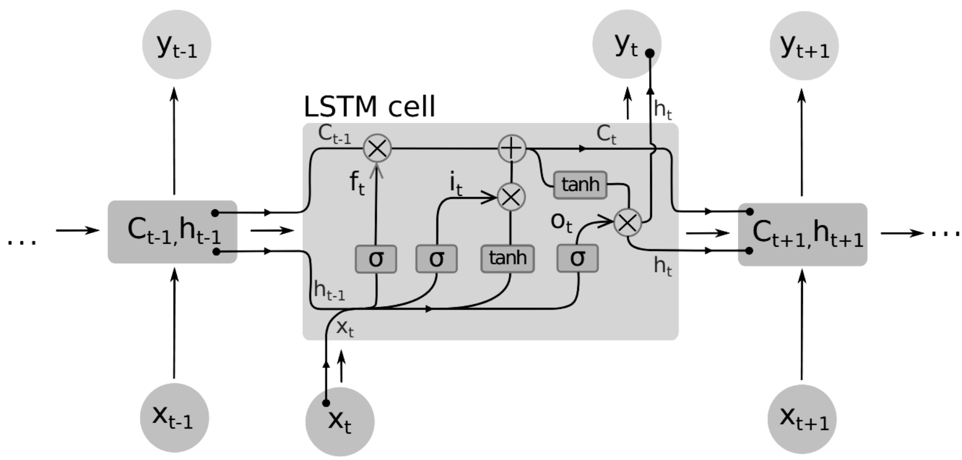


Рис. 3. Рекуррентный модуль LSTM

Вычисления на каждом временном шаге внутри LSTM-сети для одного примера последовательных данных:

$$f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right),$$
$$i_t = \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right),$$
$$\tilde{C}_t = \tanh\left(W_C\left[h_{t-1}, x_t\right] + b_C\right), \quad (20)$$
$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t,$$
$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right),$$

где f_t, i_t, o_t – вентили LSTM для забывания, входа и выхода соответственно; $\tilde{C}_t$ – новая кандидатская ячейка памяти; C_t – обновленная ячейка памяти; h_t – скрытое состояние; $\sigma, \tanh$ – функции активации сигмоида и гиперболический тангенс, соответственно; b – смещение модели.

Представим общее описание GRU-сети. Одну ячейку GRU-модуля отобразим на рис. 4.

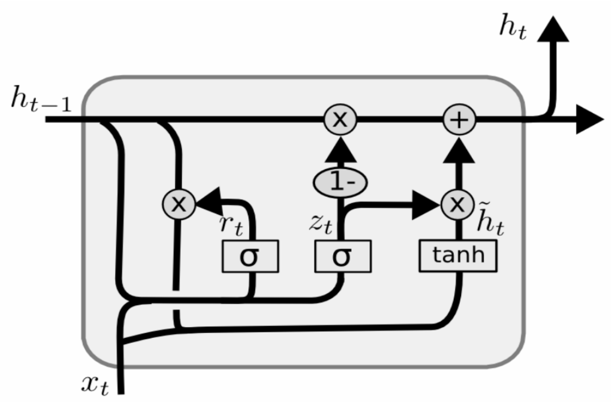


Рис. 4. Рекуррентный модуль GRU

Вычисления на каждом временном шаге внутри GRU-сети для входных последовательностей, представленных в виде тензорных представлений

$$z_t = \sigma\left(W_z\left[h_{t-1}, x_t\right]\right),$$
$$r_t = \sigma\left(W_r\left[h_{t-1}, x_t\right]\right), \quad (21)$$

где z_t – вектор обновления; r_t – вектор сброса; W_z, W_r – матрицы весов для векторов обновления и сброса, соответственно, которые подбираются в процессе обучения сети и применяются к комбинации предыдущего скрытого состояния h_{t-1} и текущего входного вектора x_t :

$$\tilde{h}_t = \tanh\left(W\left[r_t \cdot h_{t-1}, x_t\right]\right),$$
$$h_t = \left(1 - z\right) \cdot h_{t-1} + z_t \cdot \tilde{h}_t, \quad (22)$$

где $\tilde{h}_t$ – кандидат на новое скрытое состояние; h_t – новое скрытое состояние; W – матрица весов для вычисления кандидата на новое скрытое состояние $\tilde{h}_t$, также обучаемая в процессе обучения сети.

3.4.2. Блочная архитектура N-BEATS

Модель Neural Basis Expansion Analysis for Time Series Forecasting (N-BEATS) была впервые предложена в статье [Oreshkin, Carpov, Chapados, Bengio, 2019]. Расширенная архитектура под названием N-BEATSx (N-BEATS with Exogenous Variables) представляет собой усовершенствованную версию модели N-BEATS, специально разработанную для анализа временных рядов с учетом экзогенных переменных. Модель состоит из набора блоков, каждый из которых обучается выделять специфические особенности временного ряда. В N-BEATSx эти блоки могут обрабатывать как непосредственно временные ряды, так и экзогенные переменные, что значительно увеличивает гибкость и точность модели.

В работе [Souto, Moradi, 2024] исследуется применение модели N-BEATSx для прогнозирования реализованной волатильности. Это первая работа, которая использует данную архитектуру для прогнозирования волатильности, и в исследовании ставилась цель выделить преимущества архитектуры N-BEATSx перед широко используемыми моделями LSTM, TCN, HAR, GARCH и GJR-GARCH. По результатам авторов, N-BEATSx в среднем показывает на 13% более точные прогнозы на тестовой выборке для среднесрочных прогнозов и на 8% лучше для долгосрочных прогнозов по сравнению с другими моделями (TCN, GARCH и т.д.). Однако модель N-BEATSx не показала более точных прогнозов реализованной волатильности, чем LSTM-сеть. Стоит отметить, что модель оказалась на 43–60% более устойчива к изменениям параметров (включая случайность, связанную с начальной инициализацией параметров сети) в краткосрочных, среднесрочных и долгосрочных прогнозах. В рамках тестов на робастность N-BEATSx показала значительное преимущество по устойчивости к изменению параметров по сравнению с другими нейронными сетями. Это преимущество, по мнению авторов, связано с архитектурой модели, которая позволяет ей эффективно использовать большое количество параметров и избегать переобучения благодаря многослойной структуре. Интересно, как отмечают авторы, что для развивающихся рынков, таких как IBOVESPA и S&P BSE SENSEX, преимущество N-BEATSx не так очевидно, как для развитых рынков (например, S&P 500, NASDAQ).

3.4.3. Подбор гиперпараметров

Перед построением моделей необходимо отобрать наилучшую комбинацию гиперпараметров. Подбор комбинации будем производить при помощи механизма поиска по сетке с максимизацией метрики MSE на валидационной выборке.

В табл. 4 отобразим перечень используемых комбинаций гиперпараметров в нейронных сетях.

Таблица 4.

Перечень значений гиперпараметров, использованных для поиска по сетке

Гиперпараметр	LSTM, GRU, BiLSTM, BiGRU	FNN	N-BEATSx
Предобработка признаков	MinMaxScaler; StandardScaler; Yeo-Johnson transformation		
Количество:			
слоев	2; 3; 4		–
нейронов	16; 32; 64; 128; 256	16; 32; 64; 128; 256; 512	
Дропаут	0,0; 0,1; 0,2; 0,3		–
Рекуррентный дропаут	0,0; 0,1; 0,2; 0,3	–	
Размер батча	32; 64; 128; 256		
Функция активации	Tanh, Soft sign; Sigmoid	ReLU; Leaky ReLU; Linear; Tanh; SeLU	
Скорость обучения	0,001; 0,0003; 0,0001		
Количество блоков	–		2; 3; 4

В представленном перечне гиперпараметров не указано количество эпох обучения нейронных сетей. Одна эпоха обозначает, что нейронная сеть прошла по всей обучающей выборке, либо слева направо, либо справа налево, либо в обе стороны в случае Bidirectional архитектур. Механизм ранней остановки прекращает обучение, когда отслеживаемая ошибка на валидационной выборке растет n раз подряд. Установим n равное 5. Другими словами, если отслеживаемая ошибка растет 5 эпох подряд, то обучение прекратится. В данном случае будет отслеживаться метрика RMSE.

После проведения процедуры поиска по сетке с помощью кросс-валидации были отобраны наилучшие комбинации гиперпараметров нейронных сетей для инструментов SPY и QQQ. Данные комбинации обеспечивают наименьшее значение MSE на валидационной выборке. Определим отобранные комбинации (табл. 5, 6).

Таблица 5.

Используемые гиперпараметры в нейронных сетях для прогнозирования реализованной волатильности SPY

Гиперпараметр	LSTM	GRU	BiLSTM	BiGRU	FNN	N-BEATSx
Предобработка признаков	MinMaxScaler					
Количество:						
слоев	2				2; 3; 4	–
нейронов	64	32			16	128
Дропаут	0,0					–

Окончание табл. 5.

Гиперпараметр	LSTM	GRU	BiLSTM	BiGRU	FNN	N-BEATSx
Рекуррентный дропаут	0,0		0,1	0,2	–	–
Размер батча	64					
Функция активации	Tanh				ReLU	
Скорость обучения	0,0003				0,001	
λ (гиперпараметр quantile log-cosh)	5					
η (гиперпараметр quantile log-cosh)	0,75					
Количество блоков	–					2
Оптимизатор	Adam					

Таблица 6.

Используемые гиперпараметры в нейронных сетях для прогнозирования реализованной волатильности QQQ

Гиперпараметр	LSTM	GRU	BiLSTM	BiGRU	FNN	N-BEATSx
Предобработка признаков	MinMaxScaler					
Количество:						
слоев	2				2; 3; 4	-
нейронов	32					128
Дропаут	0,0					–
Рекуррентный дропаут	0,1	0,0			–	
Размер батча	64					
Функция активации	Tanh				Linear	
Скорость обучения	0,0003				0,001	
λ (гиперпараметр quantile log-cosh)	10					
η (гиперпараметр quantile log-cosh)	0,75					
Количество блоков	–					4
Оптимизатор	Adam					

4. Результаты исследования и обсуждение

4.1. Оценка качества моделей прогнозирования реализованной волатильности

Далее перейдем к результатам обучения описанных моделей на предобработанных временных рядах (табл. 7, 8).

Таблица 7.

Метрики качества моделей прогнозирования реализованной волатильности SPY на тестовой выборке

Модель	Целевая переменная											
	RV_{t+5}			RV_{t+10}			RV_{t+15}			RV_{t+20}		
	R^2, %	MAPE, %	QLIKE	R^2, %	MAPE, %	QLIKE	R^2, %	MAPE, %	QLIKE	R^2, %	MAPE, %	QLIKE
LSTM	**88**	**9**	**0,17**	**78**	**14**	**0,32**	**66**	18	**0,47**	**30**	67	**0,20**
GRU	86	12	0,20	77	15	0,34	**66**	18	0,48	22	67	0,22
BiLSTM	85	13	0,22	77	15	0,34	62	**17**	0,49	22	**65**	0,22
BiGRU	68	19	0,42	64	21	0,52	60	20	0,65	23	96	0,23
FNN	40	28	0,91	23	44	1,76	28	28	1,07	12	129	0,32
N-BEATSx	37	25	0,26	36	28	1,03	27	35	1,19	12	290	0,62
HARx-RV	**88**	10	**0,17**	77	15	0,33	**66**	18	0,47	22	67	0,22
HAR-RV	87	11	0,20	72	16	0,41	57	20	0,59	23	80	0,23

Таблица 8.

Метрики качества моделей прогнозирования реализованной волатильности QQQ на тестовой выборке

Модель	Целевая переменная											
	RV_{t+5}			RV_{t+10}			RV_{t+15}			RV_{t+20}		
	R^2, %	MAPE, %	QLIKE	R^2, %	MAPE, %	QLIKE	R^2, %	MAPE, %	QLIKE	R^2, %	MAPE, %	QLIKE
LSTM	88	**9**	**0,14**	77	**13**	**0,29**	65	**17**	0,45	46	**20**	0,68
GRU	88	10	0,17	71	14	0,36	51	18	0,68	33	21	0,79
BiLSTM	**89**	**09**	**0,14**	**78**	**13**	**0,29**	64	16	0,47	42	21	0,72
BiGRU	78	13	0,24	71	18	0,44	54	22	0,60	44	23	0,95
FNN	73	16	0,64	58	20	0,80	28	26	1,13	24	26	1,26
N-BEATSx	69	15	0,40	22	25	1,13	21	31	1,13	20	24	1,03
HARx-RV	**89**	10	0,16	**78**	16	0,32	**68**	18	**0,44**	**53**	21	**0,61**
HAR-RV	88	**09**	0,15	71	14	0,35	52	18	0,57	26	22	0,87

Результаты расчета статистик и p-value теста Дибольда – Мариано представлены в табл. 9.

Таблица 9.

Результаты теста Дибольда – Мариано на точность прогнозов реализованной волатильности SPY и QQQ

Модель	SPY			QQQ		
	статистика Дибольда – Мариано	p-value	принимаемая гипотеза	статистика Дибольда – Мариано	p-value	принимаемая гипотеза
Целевая переменная – RV_{t+5}						
LSTM	1,39	0,17	H0	0,25	0,80	H0
GRU	–1,31	0,19	H0	–0,82	0,41	H0
BiLSTM	–1,97	0,05	H0	0,70	0,48	H0
BiGRU	–4,42	0,00	H1	–2,85	0,00	H1
FNN	–6,61	0,00	H1	–4,06	0,00	H1
N-BEATSx	–6,95	0,00	H1	–4,20	0,00	H1
HARx-RV	0,70	0,49	H0	–0,16	0,88	H0
HAR-RV	–	–	–	–	–	–
Целевая переменная – RV_{t+10}						
LSTM	**2,47**	**0,01**	**H1**	**2,33**	**0,02**	**H1**
GRU	1,20	0,23	H0	0,13	0,90	H0
BiLSTM	**2,36**	**0,02**	**H1**	1,82	0,07	H0
BiGRU	–1,87	0,06	H0	–0,99	0,32	H0
FNN	–5,79	0,00	H1	–2,50	0,01	H1
N-BEATSx	–6,17	0,00	H1	–5,22	0,00	H1
HARx-RV	1,98	0,05	H0	0,59	0,56	H0
HAR-RV	–	–	–	–	–	–
Целевая переменная – RV_{t+15}						
LSTM	**3,18**	**0,00**	**H1**	**2,52**	**0,01**	**H1**
GRU	**2,72**	**0,01**	**H1**	–1,00	0,32	H0
BiLSTM	**2,73**	**0,01**	**H1**	**2,69**	**0,01**	**H1**
BiGRU	0,23	0,82	H0	–0,55	0,58	H0
FNN	–3,64	0,00	H1	–3,20	0,00	H1
N-BEATSx	–7,98	0,00	H1	–3,08	0,00	H1
HARx-RV	**2,84**	**0,00**	**H1**	1,79	0,07	H0
HAR-RV	–	–	–	–	–	–

Окончание табл. 9.

Модель	SPY			QQQ		
	статистика Дибольда – Мариано	p-value	принимаемая гипотеза	статистика Дибольда – Мариано	p-value	принимаемая гипотеза
Целевая переменная – RV_{t+20}						
LSTM	**2,92**	**0,00**	**H1**	**3,4**	**0,00**	**H1**
GRU	**2,63**	**0,01**	**H1**	**2,34**	**0,02**	**H1**
BiLSTM	**3,09**	**0,00**	**H1**	**3,29**	**0,00**	**H1**
BiGRU	–0,89	0,37	H0	0,77	0,44	H0
FNN	–2,36	0,02	H1	–0,97	0,33	H0
N-BEATSx	–3,1	0,00	H1	–1,86	0,06	H0
HARx-RV	**2,73**	**0,01**	**H1**	**3,01**	**0,00**	**H1**
HAR-RV	–	–	–	–	–	–

По результатам оценок наиболее эффективной моделью из исследуемых для прогнозирования реализованной волатильности SPY и QQQ признаем LSTM-сеть, включающую в себя экзогенные факторы из теоремы восстановления и обученную на функции потерь quantile log-cosh. Поэтому далее проведем доказательный эксперимент эффективности quantile log-cosh на примере LSTM-сети.

4.2. Оценка эффективности quantile log-cosh в качестве функции потерь на примере нейронной сети LSTM

Определим методику сравнения производительности LSTM-сети с различными функциями потерь. Сравнение с другими популярными функциями потерь (MSE, MAE, log-cosh и Huber) будет производиться по средним квантильным ошибкам (quantile error, QE):

$$QE_{t+k}(q) = q \cdot \max\left(RV_{t+k} - \widehat{RV_{t+k}}, 0\right) + (1-q) \cdot \max\left(RV_{t+k} - \widehat{RV_{t+k}}, 0\right), \quad (23)$$

где k – горизонт прогнозирования: 5, 10, 15, 20 дней; q – квантиль (0, 0,1, …, 1).

Среднее значение квантильной ошибки (mean quantile error, MQE):

$$MQE_{t+k} = \frac{1}{N}\sum_{q} QE_{t+k}(q), \quad (24)$$

где N – количество квантилей.

Обратим внимание на тот факт, что MQE_{t+k} является метрикой MAE, которая была разложена по различным квантилям исходных данных.

Существенной мотивацией введения новой функции потерь, а именно квантильного лог-гиперболического косинуса, было то, что процесс волатильности имеет склонность к резким скачкам, которые при разложении ряда попадают в диапазон высоких квантилей. Поэтому метрики MQE_{t+k} будут посчитаны для всей тестовой выборки и отдельно для данных, находящихся выше 75-го квантиля, чтобы оценить, насколько модификация вида quantile log-cosh повлияла на качество прогнозов модели в пиковых значениях волатильности.

Далее перейдем к сравнению полученных результатов оценки производительности LSTM-сети с различными функциями потерь для задачи прогнозирования реализованной волатильности SPY и QQQ (табл. 10, 11).

Таблица10.

Средние значения квантильных ошибок для различных метрик и горизонтов прогнозирования реализованной волатильности инструмента SPY

LSTM с разными функциями потерь (ℓ)	MQE_{t+5}	MQE_{t+10}	MQE_{t+15}	MQE_{t+20}	Среднее $MQE_{t+k} \times 10^2$
Квантили с 0 до 1					
ℓ – квантильный лог-гиперболический косинус $(\lambda = 10, q = 0,75)$	0,1108	0,1423	0,1796	0,2157	0,1621
ℓ – лог-гиперболический косинус	0,1009	0,1508	0,1759	0,2211	0,1622
ℓ – Huber $(\delta = 1)$	0,1000	0,1522	0,1765	0,2194	0,1620
ℓ – Huber $(\delta = 0,5)$	0,0978	0,1396	0,1762	0,2180	0,1579
ℓ – MSE	0,0992	0,1419	0,1782	0,2174	0,1592
ℓ – MAE	0,0988	0,1435	0,2070	0,3064	0,1889
Квантили с 0,75 по 1					
ℓ – квантильный лог-гиперболический косинус $(\lambda = 10, q = 0,75)$	0,2438	0,2724	0,3563	0,4241	0,3241
ℓ – лог-гиперболический косинус	0,1393	0,2088	0,3186	0,4725	0,2848
ℓ – Huber $(\delta = 1)$	0,1602	0,2009	0,3310	0,3900	0,2705
ℓ – Huber $(\delta = 0,5)$	0,1948	0,2488	0,3214	0,4330	0,2995
ℓ – MSE	0,1490	0,2833	0,3482	0,3864	0,2917
ℓ – MAE	0,1558	0,2957	0,4540	0,6658	0,3928

Таблица11.

Средние значения квантильных ошибок для различных метрик и горизонтов прогнозирования реализованной волатильности инструмента QQQ

LSTM с разными функциями потерь (ℓ)	MQE_{t+5}	MQE_{t+10}	MQE_{t+15}	MQE_{t+20}	Среднее $MQE_{t+k} \times 10^2$
Квантили с 0 до 1					
ℓ – квантильный лог-гиперболический косинус $(\lambda = 5, q = 0,75)$	0,1332	0,1939	0,2415	0,2898	0,2146
ℓ – лог-гиперболический косинус	0,1333	0,1863	0,2400	0,2946	0,2136
ℓ – Huber $(\delta = 1)$	0,1311	0,1915	0,2380	0,3016	0,2155
ℓ – Huber $(\delta = 0,5)$	0,1353	0,1908	0,2445	0,3025	0,2183
ℓ – MSE	0,1295	0,1881	0,2386	0,3068	0,2158
ℓ – MAE	0,1487	0,1960	0,2555	0,3494	0,2374
Квантили с 0,75 по 1					
ℓ – квантильный лог-гиперболический косинус $(\lambda = 5, q = 0,75)$	0,1914	0,2939	0,4274	0,5608	0,3684
ℓ – лог-гиперболический косинус	0,1983	0,3354	0,4558	0,6138	0,4008
ℓ – Huber $(\delta = 1)$	0,1885	0,3238	0,4599	0,6673	0,4099
ℓ – Huber $(\delta = 0,5)$	0,1903	0,3103	0,4896	0,6537	0,4110
ℓ – MSE	0,1964	0,3558	0,4697	0,6813	0,4258
ℓ – MAE	0,2119	0,4073	0,5701	0,8039	0,4983

Резюмируя вышесказанное, можно говорить о некоторой состоятельности использования функции потерь вида quantile log-cosh при работе на временных рядах реализованной волатильности. На квантилях выше 0,75 тестовой выборки реализованной волатильности инструмента QQQ LSTM-сеть с лосс-функцией квантильный лог-гиперболический косинус продемонстрировала наименьшее абсолютное значение отклонений от исходных данных при реализации прогнозов на 5, 10, 15 и 20 дней вперед. Функция потерь quantile log-cosh показывает потенциал для улучшения качества прогнозов в пиковых значениях волатильности, однако требует дополнительной проверки на других наборах данных.

5. Заключение

Таким образом, в данной работе была рассмотрена проблема прогнозирования реализованной волатильности ETF на основе экзогенных факторов, извлекаемых из теоремы восстановления, а также предложена модификация функции потерь quantile log-cosh для улучшения качества прогнозов в периоды высокой волатильности.

Основные выводы исследования.

1. Специфические свойства ETF предопределяют методы прогнозирования. ETF, в отличие от отдельных акций, обладают уникальными характеристиками, такими как механизм арбитража, высокая ликвидность и влияние маркет-мейкеров, что делает их волатильность отличной от традиционных активов. Эти особенности обосновывают применение сложных методов прогнозирования, включая нейросетевые модели и учет опционной информации.

2. Использование факторов, извлекаемых из теоремы восстановления, значительно повышает точность прогнозов. Теорема восстановления позволяет реконструировать плотность распределения базового актива в истинной мере, что дает дополнительные информативные признаки для прогнозирования волатильности. Эксперименты показали, что включение этих факторов в модели HAR-RV и нейросетевые архитектуры позволяет достичь существенного прироста точности.

3. Модифицированная функция потерь quantile log-cosh улучшает прогнозы в периоды высокой волатильности. Введение штрафа за ошибки в верхних квантилях распределения волатильности позволяет более точно моделировать экстремальные рыночные движения. Тестирование на ETF SPY и QQQ показало, что модели, обученные с использованием quantile log-cosh, превосходят аналоги с MSE, MAE и стандартным log-cosh, особенно на горизонтах прогнозирования 10–20 дней.

4. LSTM-сеть с quantile log-cosh и экзогенными факторами из теоремы восстановления оказалась наиболее эффективной. В сравнении с HAR-RV, стандартными рекуррентными сетями (GRU, BiLSTM) и классическими методами прогнозирования, модель LSTM с предложенной модификацией функции потерь и факторным анализом на базе теоремы восстановления продемонстрировала наивысшую точность прогнозов на тестовой выборке.

Доступность данных. Код в Python и данные, используемые для анализа, предоставляются по запросу.

* *
*

СПИСОК ЛИТЕРАТУРЫ

Костырка А.В., Малахов Д.И. А был ли сдвиг: эмпирический анализ тестов на структурные сдвиги в волатильности доходностей // Прикладная эконометрика. 2021. № 1. С. 110–139.

Amir A., Levy A., Levenstein M., Lubin R. Porat BG. Can We Recover the Cover? // Algorithmica. 2019. Vol. 81. № 7. P. 2857–2875.

Andersen T.G., Bollerslev T., Diebold F.X., Labys P. Modeling and Forecasting Realized Volatility // Econometrica. 2003. Vol. 71. № 2. P. 579–625.

Andersen T.G., Bollerslev T., Diebold F.X., Labys P. The Distribution of Realized Exchange Rate Volatility // Journal of the American Statistical Association. 2001. Vol. 96. № 453. P. 42–55.

Andersen T.G., Bollerslev T. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts // International Economic Review. 1998. P. 885–905.

Andersen T.G., Bollerslev T., Diebold F.X. Some Like it Smooth, and Some Like it Rough: Untangling Continuous and Jump Components in Measuring, Modeling, and Forecasting Asset Return Volatility // Modeling, and Forecasting Asset Return Volatility. 2003. September.

Audrino F., Huitema R., Ludwig M. An Empirical Implementation of the Ross Recovery Theorem As a Prediction Device // Journal of Financial Econometrics. 2021. Vol. 19. № 2. P. 291–312.

Bakshi G., Chabi-Yo F., Gao X. A Recovery that We Can Trust? Deducing and Testing the Restrictions of the Recovery Theorem // The Review of Financial Studies. 2018. Vol. 31. № 2. P. 532–555.

Barndorff-Nielsen O.E., Shephard N. Estimating Quadratic Variation Using Realized Variance // Applied Econometrics. 2002. Vol. 17. № 5. P. 457–477.

Barron J.T. A General and Adaptive Robust Loss Function // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. P. 4331–4339.

Bollerslev T., Tauchen G., Zhou H. Expected Stock Returns and Variance Risk Premia // The Review of Financial Studies. 2009. Vol. 22. № 11. P. 4463–4492.

Borovička J., Hansen L.P., Scheinkman J.A. Misspecified Recovery // The Journal of Finance. 2016. Vol. 71. № 6. P. 2493–2544.

Branco R.R., Rubesam A., Zevallos M. Forecasting Realized Volatility: Does Anything Beat Linear Models? // Journal of Empirical Finance. 2024. P. 101524.

Carr P., Yu J. Risk, Return, and Ross Recovery // Journal of Derivatives. 2012. Vol. 20. № 1. P. 38.

Christoffersen P.F., Diebold F.X. How Relevant Is Volatility Forecasting for Financial Risk Management? // Review of Economics and Statistics. 2000. Vol. 82. № 1. P. 12–22.

Corsi F. A Simple Approximate Long-Memory Model of Realized Volatility // Journal of Financial Econometrics. 2008. Vol. 7. № 2. P. 174–196.

Dwyer G.P. The Economics of Bitcoin and Similar Private Digital Currencies // Journal of Financial Stability. 2015. 17. P. 81–91.

Engle R.F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation // Econometrica: Journal of the Econometric Society. 1982. P. 987–1007.

Gagnon M.H., Power G.J., Toupin D. Forecasting Market Index Volatility Using Ross-Recovered Distributions // Quantitative Finance. 2022. Vol. 22. № 2. P. 255–271.

Hansen P.R., Lunde A. A Realized Variance for the whole Day Based on Intermittent High-Frequency Data // Journal of Financial Econometrics. 2005. Vol. 3. № 4. P. 525–554.

Huber P.J. Robust Estimation of a Location Parameter // The Annals of Mathematical Statistics. 1964. 35(1). P. 73–101.

Jackwerth J.C., Menner M. Does the Ross Recovery Theorem Work Empirically? // Journal of Financial Economics. 2020. Vol. 137. № 3. P. 723–739.

Merton R.C. On Estimating the Expected Return on the Market: An Exploratory Investigation // Journal of Financial Economics. 1980. Vol. 8. № 4. P. 323–361.

Oreshkin B.N., Carpov D., Chapados N., Bengio Y. N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting // Preprint arXiv:1905.10437. 2019.

Qin L., Linetsky V., Nie Y. Long Forward Probabilities, Recovery, and the Term Structure of Bond Risk Premiums // The Review of Financial Studies. 2018. Vol. 31. № 12. P. 4863–4883.

Rodikov G., Antulov-Fantulin N. Can LSTM Outperform Volatility-Econometric Models? // Preprint arXiv:2202.11581. 2022.

Ross S. The Recovery Theorem // The Journal of Finance. 2015. Vol. 70. № 2. P. 615–648.
Souto H.G., Moradi A. Introducing NBEATSX to Realized Volatility Forecasting // Expert Systems with Applications. 2024. Vol. 242. P. 122802.
Welch I., Goyal A. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction // The Review of Financial Studies. 2008. Vol. 21. № 4. P. 1455–1508.

Hybrid Approaches to Predicting Realized ETF Volatility: Deep Learning and the Recovery Theorem

Dmitry Patlasov

Perm State National Research University,
15, Bukireva str., Perm, 614068, Russian Federation.
E-mail: dmitriypatlasov@gmail.com

This article examines the task of multi-step forecasting of realized volatility. The paper introduces a modification of the loss function of the form quantile log-hyperbolic cosine (quantile log-cosh), and information extracted from options using the recovery theorem [Ross, 2015] is also used as exogenous factors in the context of predicting the realized volatility of exchange-Traded funds (ETF) SPY (SPDR S&P 500 ETF Trust) и QQQ (Invesco QQQ Trust). Two hypotheses are put forward: the first one assumes that the quantile log-cosh in neural networks will increase the accuracy of the predictive model on the test dataset compared to the same models trained on other target functions. The second hypothesis is to use information extracted from the recovery theorem. This theorem makes it possible to approximate the true distribution density of SPY and QQQ states in terms of Markov chains and get rid of the assumptions of a risk-neutral measure in financial models. Then, according to the second hypothesis, it is expected that the model with the factors extracted using the recovery theorem will show more accurate predictions on the test sample compared to the classical heterogeneous autoregression (HAR-RV) model. The following machine learning models are used to test hypotheses: LSTM, GRU, BiLSTM, BiGRU, FCNN and N-BEATS. The results show that the modification of the quantile log-cosh makes it possible to improve the accuracy of model predictions on the test dataset. Also, the inclusion of exogenous factors from the recovery theorem in the forecasting models of realized volatility makes it possible to significantly outperform the HAR-RV model, especially over the long-term horizon.

***Key words*:** forecasting; neural networks; realized volatility; the recovery theorem.

JEL Classification: G17, C45.

* *
*

References

Amir A., Levy A., Levenstein M., Lubin R. Porat BG. (2019) Can We Recover the Cover*? Algorithmica*, 81, 7, pp. 2857–2875.

Andersen T.G., Bollerslev T., Diebold F.X., Labys P. (2003) Modeling and Forecasting Realized Volatility. *Econometrica*, 71, 2, pp. 579–625.

Andersen T.G., Bollerslev T., Diebold F.X., Labys P. (2001) The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association*, 96, 453, pp. 42–55.

Andersen T.G., Bollerslev T. (1998) Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review*, pp. 885–905.

Andersen T.G., Bollerslev T., Diebold F.X. (2003) Some Like it Smooth, and Some Like it Rough: Untangling Continuous and Jump Components in Measuring, Modeling, and Forecasting Asset Return Volatility. *Modeling, and Forecasting Asset Return Volatility*. September.

Audrino F., Huitema R., Ludwig M. (2021) An Empirical Implementation of the Ross Recovery Theorem As a Prediction Device. *Journal of Financial Econometrics*, 19, 2, pp. 291–312.

Bakshi G., Chabi-Yo F., Gao X. (2018) A Recovery that We Can Trust? Deducing and Testing the Restrictions of the Recovery Theorem. *The Review of Financial Studies*, 31, 2, pp. 532–555.

Barndorff-Nielsen O.E., Shephard N. (2002) Estimating Quadratic Variation Using Realized Variance. *Applied Econometrics*, 17, 5, pp. 457–477.

Barron J.T. (2019) A General and Adaptive Robust Loss Function. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4331–4339.

Bollerslev T., Tauchen G., Zhou H. (2009) Expected Stock Returns and Variance Risk Premia. *The Review of Financial Studies*, 22, 11, pp. 4463–4492.

Borovička J., Hansen L.P., Scheinkman J.A. (2016) Misspecified Recovery. *The Journal of Finance*, 71, 6, pp. 2493–2544.

Branco R.R., Rubesam A., Zevallos M. (2024) Forecasting Realized Volatility: Does Anything Beat Linear Models? *Journal of Empirical Finance*, pp. 101524.

Carr P., Yu J. (2012) Risk, Return, and Ross Recovery. *Journal of Derivatives*, 20, 1, pp. 38.

Christoffersen P.F., Diebold F.X. (2000) How Relevant Is Volatility Forecasting for Financial Risk Management? *Review of Economics and Statistics*, 82, 1, pp. 12–22.

Corsi F. (2008) A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 7, 2, pp. 174–196.

Dwyer G.P. (2015) The Economics of Bitcoin and Similar Private Digital Currencies. *Journal of Financial Stability*, 17, pp. 81–91.

Engle R.F. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica: Journal of the Econometric Society*, pp. 987–1007.

Gagnon M.H., Power G.J., Toupin D. (2022) Forecasting Market Index Volatility Using Ross-Recovered Distributions. *Quantitative Finance*, 22, 2, pp. 255–271.

Hansen P.R., Lunde A. (2005) A Realized Variance for the whole Day Based on Intermittent High-Frequency Data. *Journal of Financial Econometrics*, 3, 4, pp. 525–554.

Huber P.J. (1964) Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35, 1, pp. 73–101.

Jackwerth J.C., Menner M. (2020) Does the Ross Recovery Theorem Work Empirically? *Journal of Financial Economics*, 137, 3, pp. 723–739.

Kostyrko A.V., Malakhov D.I. (2021) Was There a Shift: An Empirical Analysis of Tests for Structural Shifts in Yield Volatility. *Applied Econometrics*, 1, pp. 110–139.

Merton R.C. (1980) On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics*, 8, 4, pp. 323–361.

Oreshkin B.N., Carpov D., Chapados N., Bengio Y. (2019) *N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting*. Preprint arXiv:1905.10437.

Qin L., Linetsky V., Nie Y. (2018) Long Forward Probabilities, Recovery, and the Term Structure of Bond Risk Premiums. *The Review of Financial Studies*, 31, 12, pp. 4863–4883.

Rodikov G., Antulov-Fantulin N. (2022) *Can LSTM Outperform Volatility-Econometric Models?* Preprint arXiv:2202.11581.

Ross S. (2015) The Recovery Theorem. *The Journal of Finance*, 70, 2, pp. 615–648.

Souto H.G., Moradi A. (2024) Introducing NBEATSX to Realized Volatility Forecasting. *Expert Systems with Applications*, 242, pp. 122802.

Welch I., Goyal A. (2008) A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies*, 21, 4, pp. 1455–1508.

УДК 330.567.2, 366.1, 316.344.2, 338.124

DOI: 10.17323/1813-8691-2025-29-1-132-159

Changes in Out-of-Home Food and Alcohol Expenditure during the COVID-19 Pandemic[1]

Valentin Voytenkov

National Research University Higher School of Economics,
11, Pokrovsky Blvd., Moscow, 109028, Russian Federation.
E-mail: vvoytenkov@hse.ru

This study investigates changes in the expenditure on out-of-home food and alcohol during the COVID-19 pandemic. Our research specifically explores intergroup variations among households, taking into account the dissimilarity of Russian regions in light of the degree of quarantine restrictions enforced. To test the research hypotheses, we use microdata from the Household Budget Survey conducted by the Federal State Statistics Service. To compare the expenditures on out-of-home food and alcohol in regions with soft, medium, and hard restrictive measures, we employ t-test for comparing means and the Kolmogorov – Smirnov test for comparing distributions. The Tobit model is applied to compare different social groups' household spending habits. The joint analysis of out-of-home food and alcohol expenditure enables the separation of involuntary savings from coping strategies using models for censored data, thereby facilitating an in-depth assessment of household well-being in the face of shocks. Our findings show a reduction in out-of-home food expenditure across all social groups and all levels of quarantine restrictions. The share of alcohol expenditure decreased in almost all social groups in regions with soft measures but significantly increased in those with medium and hard restrictions.

[1] This article is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE University).

The author would like to express their gratitude to Professors Olga Demidova and Elena Semerikova of FES HSE for engaging in discussions at the Laboratory for Spatial Econometric Modelling of Socio-Economic Processes in Russia. The author is appreciative of the insights provided by Professor Lucia Reisch (Queens' College Cambridge) and Professor Alan Mathios (Cornell University) which have enhanced the analysis of the paper. The author wishes to express gratitude to two anonymous reviewers, whose insightful comments have contributed to a substantial revision of the text.

Valentin Voytenkov – research assistant at Laboratory for Spatial Econometric Modeling of Socio-Economic Processes in Russia.

The article was received: 12.10.2024/The article is accepted for publication: 06.02.2025.

***Key words*:** households; intergroup analysis; out-of-home food expenditure; alcohol expenditure; COVID-19; rigidity of quarantine measures.

JEL Classification: D12, H12, R29.

For citation: Voytenkov V. Changes in Out-of-Home Food and Alcohol Expenditure during the COVID-19 Pandemic. *HSE Economic Journal.* 2025; 29(1): 132–159.

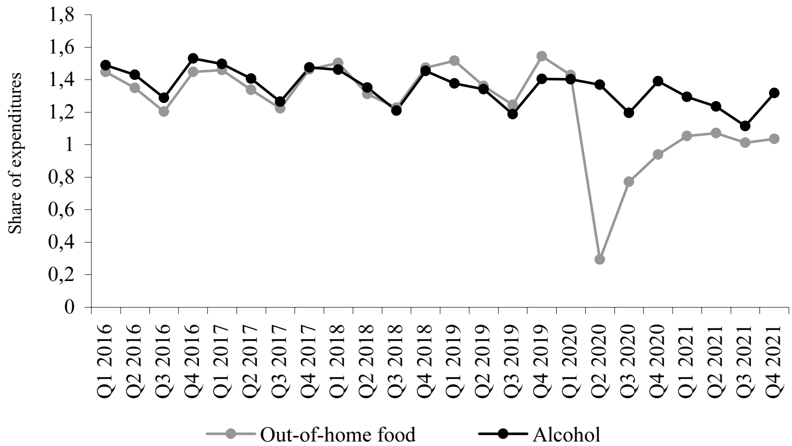
1. Introduction

Households are the economic agents most sensitive to external shocks. Numerous studies have estimated the effect of crises on households' living standards, inequality, and aggregate consumption [Alonso, Rodríguez, Rojo, 2015; Blundell et al., 2022]. The natural response to economic uncertainty and political crises is a decline in aggregate consumption, which has been the subject of numerous papers [Gautier, Ulgazi, Vertier, 2020; Grigoryev et al., 2021]. Concurrently, household responses may vary across categories of expenditures and their socio-economic characteristics.

[Barigozzi et al., 2012] note that the distribution of household expenditure is stable over time but varies across expenditure categories. In this study, we investigate the impact of the COVID-19 pandemic on two of the most elastic categories of household expenditure[2]: out-of-home food and alcohol. The study of these expenditure categories is of particular interest in the context of the COVID-19 pandemic for several reasons. *First*, before the pandemic crisis, the shares of these expenditure categories in the structure of household consumption were relatively equal, whereas during the pandemic, the dynamics of expenditure on these categories changed significantly. Figure 1 displays the proportion (in %) of out-of-home food and alcohol spending in overall household expenditures. Changes in the spending patterns on out-of-home food and alcohol expenditures, precipitated by quarantine restrictions, appear to serve as an indicator of social well-being and long-term shifts in lifestyle patterns, rather than an indicator of household welfare. The observed changes in out-of-home food expenditure may be indicative of shifts in daily practices, such as transitioning to remote work arrangements, modifying dietary habits, or avoiding public places. The increase in alcohol expenditures, on the other hand, may be a consequence of heightened stress levels and maladaptation. Despite their relatively modest contribution to the overall consumption structure, these changes offer insight into the mechanisms by which households adapt to extreme conditions. *Second*, in the context of the impact of the COVID-19 pandemic on alcohol consumption, two opposing trends can be observed: an increase in alcohol consumption due to stress and a decrease in alcohol consumption due to freeing up time for sport (see Section 2.2). This study is of particular significance in the context of Russia, where the proportion of heavy drinkers is high. In 2020, the prevalence of heavy drinking in Russia was 20% of the total number of people who consume alcohol [Antonov, 2024]. *Third*, the nature of the pandemic had a direct impact on out-of-home food expenditures due to the imposition of restrictions on movement. In this context, the response of households to such restrictions

[2] Federal State Statistics Service methodology defines the household consumption structure as comprising five groups of expenditures: food at home, food outside the home, alcohol, non-food items, and services. It should be noted that statistics on smaller expenditure groups, such as tobacco, are not kept.

is particularly worthy of examination, as compliance with the restrictions was predetermined by attitudes towards the pandemic [Sobol, Blachnio, Przepiorka, 2020]. Furthermore, it is crucial to examine the evolution of consumption patterns across diverse household groups. The estimation of the effects on different social groups has the potential to improve crisis planning and strengthen regional support measures (both financial and psychological), especially in household groups where alcohol expenditure increases *Fourthly,* despite its modest share in the structure of household consumption, expenditures on out-of-home food and alcohol have a decisive role in catering and trade. In February 2020[3], approximately 20 per cent of all small and medium-sized enterprises (SMEs) were engaged in retail and restaurant activities. Consequently, even minor reductions in household expenditure on the categories under study have the potential to adversely affect the revenues and employment of small firms, thereby hindering economic development and the recovery after COVID-19.



Note: Graph compiled from household level data from a sample household survey conducted by [Federal State Statistics Service, 2022].

Fig 1. Percentage of consumer spending on food outside the home and alcohol

We observe similar dynamics in the share of consumer spending on out-of-home food and alcohol, which fluctuate between 1.2% and 1.6% and follow seasonal patterns. The dynamics diverge during the COVID-19 pandemic's acute phase when the share of expenditure on eating out dramatically drops in the second quarter of 2020 compared to the first quarter of 2020 due to restrictive measures. Conversely, the share of expenditure on alcohol slightly declines from 1.40% in the first quarter of 2020, to 1.37% in the second quarter of 2020. This finding contradicts previous empirical research which suggests a rise in alcohol consumption during COVID-19 lockdowns [Jacob et al., 2021; Schmits, Glowacz, 2022].

The motivation for this paper is fourfold. *First*, the reaction to external shocks varies across diverse social and demographic groups [Abebe, Charlebois, Music, 2022]. During the pandemic, this may be attributed to households' diverse adaptability levels: younger households appear

---

[3] Unified register of small and medium-sized enterprises. (https://rmsp.nalog.ru/statistics.html)

to be more receptive towards delivery services and online buying as compared to their older counterparts. According to [Abebe, Charlebois, Music, 2022], an increase in educational attainment results in a greater willingness by households to use delivery services. Low-income households tend to consume their meals at home, while wealthy households often indulge in dining out. To address the varied responses of households to external shocks, we categorise them into groups based on their income, education level, and number of children.

Second, the delineation of social groups is unclear in the literature, creating ambiguity in identifying wealthy or impoverished households, high- or low-educated households, and those with few or many children. Nevertheless, household typification is a crucial element in accurately assessing the impact of external shocks on a specific group of households [Rausch, Metcalf, Reilly, 2011]. The present paper categorizes various studies and methodologies used to define household groups. Our approach to dividing households within Russian regions is modified on the basis of these frameworks (see Section 4).

Third, the COVID-19 pandemic distinguishes itself from the global financial crisis and local economic crisis of 2014 (sanction crisis) in its impact on crisis-affected regions and the economy [Kolomak, 2020]. Therefore, quantifying the impact of these crises using dummy variables requires improvement.

Fourth, one of the defining characteristics of the COVID-19 pandemic is the implementation of varying degrees of quarantine regulations. The impact of quarantine measures on household consumption has been found to vary depending on the severity of the quarantine measures [Chen, Li, Li, 2024; Yukseltan et al., 2022]. Similarly, [Baker et al., 2020] indicate that the greatest decline in consumer spending was observed among households with children and low-income groups. The frequency of the emergence of new viruses[4] and the evolution of existing ones also provide further motivation to explore the potential consequences of quarantine restrictions. Given the potential for the re-imposition of quarantine restrictions, we conduct a detailed examination of the response of households to the quarantine measures introduced during the COVD-19 pandemic. To analyse the impact of COVID-19, we utilise a measure of the strictness of quarantine restrictions, which distinguishes between regions with soft, medium, and hard restrictions. A distinctive feature of this paper is the examination of quarantine restrictions over time. Our methodology considers changes in pandemic policies across regions over time, as regions may shift between levels of restrictions. Therefore, the *purpose* of the study is to examine the impact of the COVID-19 pandemic on share of expenditures on out-of-home food and alcohol across different social groups of households in Russian regions.

This paper addresses gaps in the current literature as follows. *First*, it distinguishes various types of regions according to the level of quarantine restrictions imposed, categorising them as soft, medium, or hard. We adopt a dynamic methodology to classify each region under the specified quarantine levels. *Second*, we consider inter-group differentiation of households by identifying social groups based on income level, education level, and the number of children in the household. This paper differs from previous studies in that it employs a number of modifications of existing approaches to categorise households, taking into account the specifics of the Russian households. *Third*, we compare changes in the shares of expenditure on out-of-home food and alcohol, which have been the focus of a small body of literature.

---

[4] To illustrate, in the autumn of 2024, the World Health Organisation (WHO) declared a lethal outbreak of the Marburg virus in Rwanda. (https://www.who.int/emergencies/disease-outbreak-news/item/2024-DON537)

The paper is organised as follows: Section 2 analyses the literature on the impact of the pandemic on out-of-home food and alcohol expenditures. This section is also devoted to theoretical background and hypotheses. Section 3 describes the data and the empirical estimation framework. Section 4 is devoted to describing the approach used to divide Russian households according to education, income, and number of children, which are described and justified. This section also outlines the household profile contingent on the group. Section 5 provides empirical results on the impact of the COVID-19 pandemic on out-of-home food and alcohol for different household groups. Finally, Section 6 provides the hypothesis testing results and the conclusions of the study.

2. Literature Review

2.1. The impact of COVID-19 on household expenditures

There is extensive literature exploring the effects of the pandemic on household expenditures on out-of-home food worldwide. According to [Grigoryev et al., 2021], consumer services, which encompass cafés and restaurants, were most affected by the extensive quarantine restrictions implemented globally. [Chen, Qian, Wen, 2021] show that quarantine restrictions significantly changed the consumption basket of Chinese households. For instance, expenditure on restaurants decreased by 64–72% in 2020 compared to 2019, resulting in an overall consumption decline of 14–69% depending on the city. Similarly, [Gautier, Ulgazi, Vertier, 2020] report a 70–90% decrease during the acute phase of pandemic, compared to the previous year in expenditure on transport and out-of-home food in France. An important limitation of these studies is the concentration on national-level analyses. These papers consider households as a homogenous unit, whereas their composition is highly heterogeneous. A limited number of studies have examined intergroup disparities in response to the pandemic. For instance, research conducted in the United States suggests that there has been a decline in out-of-home food expenditure, ranging from 19.5% to 33.7%, depending on the household group [Dhakal, Acharya, Wang, 2022]. In this context, the present study addresses the literature gap by evaluating how the COVID-19 pandemic affected expenditures on out-of-home food and alcohol in Russia for different household types.

There is a large body of medical research related to alcohol consumption during the COVID-19 pandemic [Jacob et al., 2021; Marano et al.; 2022; Schmits, Glowacz, 2022]. According to the literature, two contrasting views have emerged about the effect of the pandemic regarding alcohol consumption. The first view suggests that individuals employ alcohol as a coping mechanism during these challenging times [Avery et al., 2020; Rahman et al., 2020]. The authors note that continuously staying at home and sudden changes in habits led to heightened stress levels, resulting in increased alcohol consumption. As demonstrated by [Anderson et al., 2022], a considerable escalation in alcohol expenditure was observed during periods of pub closures, notably among the most deprived households. In this context, it can be anticipated that a reduction in the stringency of quarantine restrictions may result in a decline in both alcohol consumption and alcohol expenditure. A contrasting viewpoint suggests that increased time for sports and physical activity at home has led to a reduction in alcohol consumption [Ammar et al., 2020; Pišot et al., 2020]. The authors document that individuals who exercise periodically commenced the activity due to the increased free time resulting from the elimination of commuting to work. As posited by [Acharya, Dhakal, 2022] the repercussions of the pandemic have been shown to vary across different socioeconomic demographics. The research utilised data pertaining to American households and showed that low-income households exhibited a decrease in expenditure on alcohol, whilst high-income households demonstrated an increase in this category. The ambiguity inherent in these empirical findings serves as an impetus for the employment of theoretical models and the formulation of our hypotheses.

2.2. Theoretical background and hypotheses statement

The theoretical framework of this study is predicated on several fundamental theories and concepts, including the theory of the allocation of time [Becker, 1965], mental accounting [Thaler, 1985], human capital theory [Becker, 2009], and stress and coping [Lazarus, 1984]. The consumer choice theory [Becker, 1965; Thaler, 1985], sheds light on the response of the most elastic categories of goods and services in response to external shocks. [Becker, 1965] underscores the significance of incorporating the time spent on consumption (indirect cost) in addition to the price (direct cost). Consequently, the decline in out-of-home food expenditure can be explained in terms of increased opportunity costs, as during lockdowns, additional time was required to visit a public establishment. In this context, the consumption behaviour of household groups with reduced mobility (e.g. families with children) is of particular interest. [Thaler, 1985] concept of mental accounting involves the categorisation of consumer expenditure into specific groups ("food", "entertainment", etc.). In such circumstances, households may opt to reduce expenditure categories they deem less pressing, particularly in the face of external shocks. In essence, the theory propounded by [Thaler, 1985], posits that households subjected to the pandemic may choose to spend less on "luxury" categories, such as out-of-home food and alcohol, from their consumption.

It is evident that quarantine restrictions engender considerable stress due to the disruption of routines, the curtailing of social interactions, and the limitation of personal autonomy. In response to such challenges, [Lazarus, 1984] delineated two coping strategies: emotion-focused, which entails the assessment of emotional responses, and problem-focused, which involves the identification of stressors. Given the ineffectiveness of an individual's influence on the trajectory of the pandemic, it is plausible that individuals have resorted to alcohol consumption as a coping mechanism. The use of alcohol as a coping mechanism in response to uncertainty has been demonstrated to reduce anxiety [Avery et al., 2020; Rahman et al., 2020]. However, empirical evidence suggests heterogeneity in the response of households to stress, with variations observed across different income groups [Acharya, Dhakal, 2022; Anderson et al., 2022].

The theory of human capital [Becker, 2009] underscores the pivotal function of education in shaping consumer spending patterns. The fundamental premise of this theory posits that education enhances rational decision-making by fostering heightened awareness of the long-term implications of consumption decisions. This notion finds particular relevance in the context of alcohol, given its inherent addictive nature. Empirical evidence indicates that households with higher levels of education exhibit a reduced propensity for alcohol consumption, attributable to heightened awareness of its adverse health implications [Cutler, Lleras-Muney, 2010; Yen, Jensen, 1996]. In light of the empirical and theoretical background, the present study aims to test the following hypotheses.

Hypothesis 1: In regions with severe and medium quarantine restrictions, the decrease in the share of expenditure on out-of-home food is higher among households with children compared to households without children.

Hypothesis 2: The increase in the share of expenditure on alcohol in regions influenced by strict and medium quarantine restrictions is inherent only to low-income groups of households.

Hypothesis 3. Regardless of the severity of quarantine restrictions, households with higher levels of education have a lower share of alcohol expenditure compared to less educated households.

3. Data and Empirical Setup

3.1. Data

To examine changes in consumer spending on out-of-home food and alcohol during the pandemic, we use data from the household budget survey (HBS) conducted by [Federal State Statistics Service, 2022]. The data include social and financial data on households, including the income level, educational attainment, and the number of children. The dataset comprises 576,200 observations recorded from Q1 2019 to Q4 2021, cleaned from statistical outliers[5]. The number of households surveyed is subject to change annually and within the year. As a replacement for a dropped-out household, a new one is introduced with the same identification number. Basically, this is a cross-sectional sample with rotating households, the frequency of measurement is one quarter. Hence, it is unfeasible to classify households utilising singular identification codes, and neither is it possible to ascertain the overall number of households partaking in the survey, segregated by quarter[6].

A distinctive attribute of HBS is the integration of in-person interviews and diary (and logbooks) monitoring by households. Information pertaining to a range of social characteristics (e.g., household size, place of residence, housing area) is obtained through both methods. Data on household income and expenditure is collected via the diary (and logbooks) records. Expenditure on specific categories (e.g., out-of-home food[7], alcohol) is documented in the designated sections of the household diary. It is important to note that food purchased by delivery from catering establishments is not included in the expenditure on out-of-home food. Cash income received, financial assets, and savings are also entered into the household diary. Household diaries are kept for a quarter and a household survey is conducted at the end of the quarterly cycle. After completing one cycle, a household can either drop out of the survey or remain in the survey. Thus, variable generation occurs both through the survey and through diary entries. Table 1 provides a detailed description of the variables used in the survey analysis.

The Federal State Statistics Service methodology delineates five groups of expenditures within the consumption pattern: food at home, out-of-home food, alcohol, non-food items, and services. This paper focuses on food outside the home and alcohol for several reasons. *First*, despite the evident distinction between these expenditure categories (alcohol is an addictive good, while eating out can be regarded as a luxury item), studying them reveals the heterogeneity of household responses to external shocks. Consequently, fluctuations in alcohol expenditure may signify the utilisation of coping mechanisms in response to stress, while changes in out-of-home food expenditure may denote a pragmatic reallocation of resources. The joint analysis of these

---

[5] Specifically, households with per capita household income below 1,000 rubles and above 400,000 rubles.

[6] According to Rosstat, between 48,000 and 49,000 households are surveyed annually.

[7] Only expenses in restaurants, cafes, canteens are included in this category.

two expenditure categories enables the distinction between the effects of enforced savings and the employment of coping strategies, thereby providing a comprehensive assessment of the economic and social well-being of households. *Second*, the response of all categories of household expenditures has been studied [Voytenkov, Demidova, 2023], indicating the need for separate analyses of expenditures on out-of-home food and alcohol with the application of improved methodological tools. *Third*, given the large number of zero values in the expenditures on out-of-home food and alcohol, an approach involving the use of a model that takes into account censored data is applied. Consequently, in contrast to the other expenditure categories (food at home, non-food items, services), a methodological rationale exists for the combination of expenditure on out-of-home food and alcohol.

Table 1.

Variables description

Variable	Description	Unit of measurement/interpretation
	Dependent variables	
Food out of home	Out-of-home food expenditure in total consumer spending	%
Alcohol	Alcoholic beverages expenditures in total consumer spending	%
	Independent variables	
City (X_1)	Type of locality	0 – rural, 1 – urban
Children 4-16 (X_2)	Number of children aged 4 to 16	Persons
Children under 3 (X_3)	Number of children under 3	Persons
Income (X_4)	Real income per household member, adjusted to 2016 prices and fixed set of consumer goods and services as a percentage of the national average cost[8]	Ruble
Assets (X_5)	Growth of financial assets as a percentage of income	%
Savings (X_6)	Savings as a percentage of income	%
Average age (X_7)	Average age of household members	Years
Education level (edu_k)	Maximal value of education level of household members	Scores: 1 – no basic general education 2 – basic general education 3 – secondary general education 4 – secondary vocational education 5 – higher vocational education
Housing area (X_8)	Housing area per household member	Square metres

[8] Real incomes are calculated as nominal incomes deflated by chained consumer price indices (CPI) for the period 2016–2021. Regional differences in the cost of living are then adjusted by calculating the ratio to the cost of a fixed set of consumer goods and services as a percentage of the national average cost. This indicator is calculated for each Russian region. Data on CPI and the cost of the basket are obtained from the Statistical publications titled "Regions of Russia" issued by the Russian Federal State Statistics Service (Rosstat).

Continuation

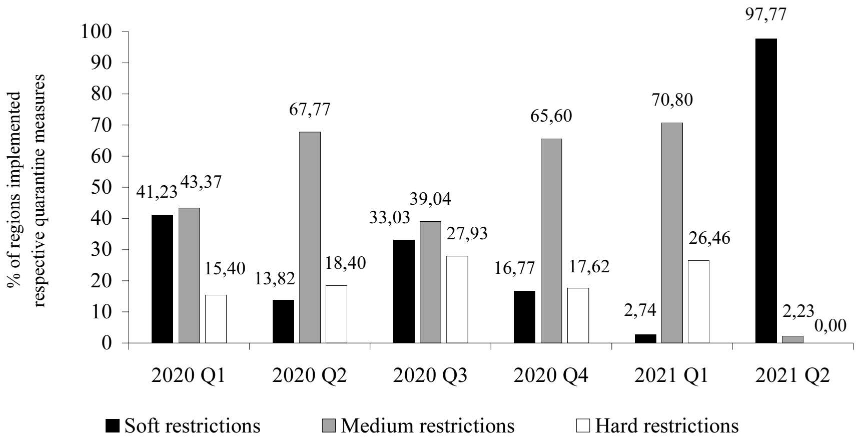
Variable	Description	Unit of measurement/interpretation
Variables, capturing the impact of COVID-19 crisis		
Dummy 2020	Dummy variable for the 2020 year	1 – values for the 2020 0 – otherwise
Isolation hard	Dummy variable reflecting the introduction of severe restrictive measures	1 – strict quarantine restrictions are introduced 0 – otherwise
Isolation hard	Dummy variable reflecting the introduction of soft restrictive measures	1 – soft quarantine restrictions are introduced 0 – otherwise

Note: Econometric modelling uses household income per capita in 2016 prices, in rubles. All variables denoted as independent are included in the regression models.

To control for household heterogeneity during the COVID-19 pandemic, we use a broad set of social and financial household characteristics. For example, as [Janssen et al., 2021] indicate, differences in consumption were also associated with socio-demographic factors, including income level and household composition. Similarly, [Xiong et al., 2021] documents the importance of family composition, as well as the age of household members, in determining consumption patterns. To illustrate, families with children have particular requirements for non-food items (toys, clothing, educational materials). Similarly, the level of education is identified as a predictor of consumer spending in the literature, as evidenced by [Maniriho et al., 2021; Soberon-Ferrer, Dardis, 1991; Varlamova, Larionova, 2015; Yen, Jensen, 1996]. In the context of consumption, education appears to serve as a proxy variable for preferences that are not directly observable. For instance, individuals with a higher level of education may exercise greater caution in their selection of goods and may also endeavour to gain insight into other cultures, which is also reflected in their consumption patterns. Furthermore, the social connections of those with higher levels of education may also be reflected in expenditure on alcohol (based on recommendations from social circles) and dining out (spending time in cafés and restaurants). In examining the financial characteristics of the household, we consider income, and savings characteristics. Income is widely regarded as the most significant predictor of consumer expenditure [Edelstein, Kilian, 2009]. [Verter, Osakwe, 2014] highlight the pivotal role of savings characteristics in shaping household consumption patterns.

To examine the effects of the pandemic, we apply a range of dummy variables, ensuring a comprehensive coverage of its impact. The St. Petersburg Policy Foundation's methodology is employed to distinguish among regions with varying degrees of quarantine restrictions [Petersburg Policy Foundation St., 2020]. This methodology for assessing the stringency of quarantine restrictions has been employed in a number of studies [Lukashina, 2020; Seliverstov et al., 2021; Voytenkov, Demidova, 2023]. As the most severe levels of quarantine measures, restrictions were imposed on visiting tourist attractions, prohibiting the operation of entertainment establishments at night, and restricting mass events. In terms of vaccination, it was mandatory for some categories of citizens. In regions with medium restrictions, vaccination was not mandatory. However, some measures were in place at the level of regions with severe restrictions. These included restrictions on the operation of catering establishments at night, as well as on groups of the population for whom self-isolation was mandatory, and a prohibition on mass events. In regions with soft quarantine restrictions, there were no mandatory vaccination requirements, restrictions on movement, or the operation of public catering enterprises.

We adopt a flexible system for determining regional quarantine restrictions, in which regions could shift from one category to another over time. Our approach involves adjusting restrictions in response to the decisions made by regional authorities. Monthly evaluations of each region's level of quarantine measures were conducted by the Petersburg Policy Foundation identifying three groups of regions: those with soft, medium, and hard restrictions. To quantify these levels of restrictions, we assigned numerical values: soft quarantine measures are denoted by 1, medium by 2, and hard by 3. We then used a simple average to aggregate the data at the regional level and quarterly dynamics. We obtained balanced panel data consisting of 510 observations across 83 Russian regions for the period from Q1 2020 to Q2 2021. Figure 2 presents the percentage of regions that have implemented the respective quarantine restrictions.



Note: Graph compiled from regional level data following methodology of [Petersburg Policy Foundation St., 2020].

Fig 2. Percentage of regions imposed different quarantine restrictions

As demonstrated in Fig. 2, the data indicates the dynamism of quarantine restrictions in Russia. It is evident that quarantine measures fluctuated throughout 2020, moving from a high proportion of regions with soft quarantine measures (approximately 41% in Q1 2020) to a predominance of medium quarantine measures (65–67% in Q2 and Q4 2020). This phenomenon can be attributed to the emergence of new strains of the virus, which led to a transition in the nature of restrictions imposed by regions. By the second quarter of 2021, the majority of quarantine restrictions had been relaxed, with soft restrictions being implemented in nearly all regions. The data accounts for quarantine restrictions in 2020, while vaccination-related restrictions (introduced in late 2021) are not considered in this paper.

The restrictions on alcohol purchase in Russia warrant particular attention. At the outset of the pandemic (Q1 2020), only a few Russian regions (Khakassia, Bashkiria, Yakutia, and Karelia) announced a ban on alcohol purchases at specific times (typically before 15:00 or 18:00 local time). This indicates that there is no correlation between the severity of quarantine measures and institutional restrictions on alcohol purchase. In contrast, there were significant institutional constraints on eating out. For example, at the beginning of the pandemic, most businesses operated takeaways, and as the Russian vaccine developed, a QR code was required to enter a restaurant.

3.2. Empirical setup

To test hypotheses 1–3 we employ the Tobit model with robust standard errors, t-test (to compare the average expenditure on alcohol and out-of-home food), and the Kolmogorov – Smirnov test (to compare the distributions of expenditures). We use Tobit models to analyse censored data. In essence, a considerable proportion of households exhibit a dependent variable value of zero, while the remainder displays positive values. The fundamental Tobit model [Tobin, 1958] takes the form (Equations 1–3):

(1) $$Y_i^* = x_i'\beta + \varepsilon_i, \; i = 1,2,\ldots N,$$

(2) $$Y_i = Y_i^* \text{ if } Y_i^* > 0,$$

(3) $$Y_i = 0 \text{ if } Y_i^* \leq 0,$$

where $\varepsilon_i \sim NID\left(0, \sigma^2\right)$ and independent of x_i. Since out-of-home food and alcohol expenditures cannot be negative, the constraint on the dependent variable can be rewritten in general terms as follows (Equations 4–5):

(4) $$Y_i = Y_i^* \text{ if } Y_i^* > 0,$$

(5) $$Y_i = 0 \text{ if } Y_i^* = 0.$$

To model the impact of COVID-19 on out-of-home food and alcohol expenditures, we employ the Engel function. Specifically, the Engel curve corresponds to the connection between household income and out-of-home food and alcohol expenditure. [Leser, 1963] suggests the incorporation of income as the main predictor of household expenditure. Several empirical studies demonstrate the influence of household socioeconomic features on household consumption trends [Edelstein, Kilian, 2009; Varlamova, Larionova, 2015]. Given this evidence, we extend the basic model by including household social characteristics (e.g., age, number of children, education level) and household financial characteristics (e.g., income). The theoretical work of [Blundell, Chen, Kristensen, 2007] highlights the significance of incorporation of demographic characteristics in Engel curve estimates. Similarly, [Hausman, Newey, Powell, 1995] indicates that one of the most effective forms of the Engel curve is the form with an income on the right-hand-side and the control of household socio-economic characteristics. The Lesser model specification differs from a number of microeconomic models, including the Almost Ideal Demand System and its modifications, in that it does not require the inclusion of prices for related goods and services. Given the peculiarity of the Rosstat sample and the impossibility of extracting the price vector (due to the unavailability of individual purchase prices), we apply this specification (Equation 6):

$$Y^*_{itjr} = \beta_{0j} + \beta_{1j}X_1 + \ldots + \beta_{8j}X_8 + \sum_{k=2}^{5}\delta_{kj}edu_k + \sum_{t=2}^{4}\gamma_{tj}q_t + \sum_{r=2}^{83}\alpha_{rj}d_r + \varphi_{softj}d_{softj} + \varphi_{hardj}d_{hardj} + \varepsilon_{ij} \quad (6)$$

where $j = 1,2$ corresponds to the expenditures on out-of-home food and alcohol respectively (dependent variables), $i = 1,\ldots,N$ is the number of households, $r = 1,\ldots,83$ is the number of the region. $X_1 - X_8$ are explanatory variables, which are explicitly stated in Table 1, $edu_k (k = 2,\ldots,5)$ is the set of dummy variables indicating the maximal level of education in household (baseline category is no basic general education, which was excluded from models). We denote dummy variables for quarters as $q_t (t = 2,\ldots,4)$, the base category is determined as the first quarter. Regional dummy variables (regional fixed effects) are denoted as $d_r (r = 2,\ldots,83)$, the base category is Altay region. In order to avoid the potential issues posed by the dummy trap, the dummy for the first quarter and dummy for Altay region were excluded from the model. To capture the impact of the COVID-19 pandemic we use d_{softj} which indicates soft measures, d_{hardj} corresponds to hard measures. Thus, as a base category, medium quarantine measures are chosen, due to the fact that this is the largest group of restrictions.

The Tobit model estimation procedure employed in this analysis adheres to established methodologies outlined in the econometric literature (see Chapter 19 from [Greene, 2012]). Models (4)–(6) were estimated using the STATA 14 statistical software, incorporating robust standard error estimators (Huber-White/sandwich estimator) to mitigate potential heteroskedasticity. Notably, the conventional White estimator assumes independence across all observations – a premise that may be excessively restrictive for datasets aggregated at the regional level, where intra-regional dependencies could plausibly exist. To address this concern, we further implemented clustered standard errors, which relax the assumption of independence within regions while maintaining independence between regions. The empirical results exhibited robustness across both estimation frameworks: key coefficients retained their statistical significance, with no substantive deviations observed.

4. Typification and Profiles of Russian Households

To conduct an intergroup analysis of Russian households, three factors are identified as the basis for the division[9]: household income level, education level, and the number of children. The selection of criteria for household typology is founded upon their theoretical (see Section 2.2 for further details) and empirical relevance in our context. Consequently, income level is a pivotal determinant of expenditure, as outlined in the Leser model, and directly influences the financial capacity of households. The level of education has been found to be related to health awareness

[9] It would appear reasonable to divide households by employment sector. However, this is not feasible, as the Rosstat questionnaires do not provide information on the employment sector of individuals, nor do they indicate who is the head of the household.

and the rationality of consumer choice [Becker, 2009], and the number of children has been found to affect expenditure patterns due to specific needs and the reduced mobility of households.

To categorise households by income we use the results of the study by [Nartikoev, Peresetsky, 2021]. In contrast to classical approaches that use the exogenous income categorization of households, Nartikoev and Peresetsky adopt an endogenous approach, which utilises log-normal distributions to identify household groups for the eight federal districts of Russia. This approach takes into account the substantial income disparity among households in Russia by segmenting into federal districts [Murashov, Ratnikova, 2017; Potapenko, Shirov, 2021]. Nartikoyev and Peresetsky's approach assumes the identification of three groups of households, whereas the current study identifies four. The upper group has been divided by income according to the median level of the upper group. The bottom 50% are categorised as "Medium high" and the top 50% as "High". The rationale for singling out the "Medium high" group is that the Household Budget Survey tends to focus on low-income groups of households. Consequently, the allocation of this group represents an attempt to identify the "middle class" of Russian society.

In contrast to income, there is no empirical evidence suggesting significant variation in the education level, and the number of children among Russian regions. Therefore, we do not consider specific regional factors for household typification. To categorise households by education level, we follow the Federal State Statistics Service classification. The first group ("Low") comprises households with basic general education. The second group ("Middle") includes households with general secondary education or secondary vocational education. The third group ("High") consists of households with tertiary education. The estimate for the educational attainment level of a household is the calculated average of the educational attainment levels of each member within the household, excluding children's levels of education from the calculation[10].

To categorise households based on the number of children, the commonly used criteria consist of childless, small (one or two children), and large families. However, there is no definition for identifying large families in Russia; instead, regional by-laws usually establish a threshold of three. The proposed criteria for household categorisation are presented in Table 2.

Table 2.

Criteria for classification of households by income, educational attainment, number of children

Variable	Household group	Description
Income	Low Medium low Medium high High	Interval boundaries are calculated based on the study of Nartikoev and Peresetsky (2021), income differences by territorial districts are taken into account
Educational attainment	Low	Households with prevalent general education
	Middle	Households with prevalent secondary and secondary vocational education
	High	Households with prevalent higher education

[10] The incorporation of the education level of children tends to result in a biased estimation of the average – consequently, the mean educational level in households with a high number of children, is substantially understated.

Continuation

Variable	Household group	Description
Number of children	No children	Households without children
	1 to 2 children	Households with 1 or 2 children
	3 or more children	Households with many children (more than 3 children)

The primary household characteristics are determined by the ratio of spending on out-of-home food, alcohol, the number of children, average household age, income, and educational level. Together, these attributes outline the household's profile. Table 3 illustrates the detailed attributes of each household group.

Table 3.

Profile of households by group

Variable	Household group	Out-of-home food, %	Alcohol, %	Children 4–16 (per 1000)	Children under 3 (per 1000)	Age, years	Income (per household member), rubles
Income	Low	1.14	1.211	78.088	744.30	38.92	10642.79
	Medium low	0.994	1.387	168.715	13.439	52.44	27159.713
	Medium high	1.177	1.372	20.648	1.680	57.41	62240.426
	High	1.57	1.588	6.021	0.441	55.30	85190.06
Educational attainment	Low	0.48	1.116	322.030	31.506	54.39	20628.52
	Middle	0.84	1.312	348.438	32.032	48.85	21032.94
	High	1.66	1.358	499.606	55.292	42.71	22204.49
Number of children	No children	0.880	1.333	0.000	0.000	56.08	29313.106
	1 to 2 children	1.688	1.264	1238.790	112.136	26.77	10831.545
	Over 3 children	1.182	0.792	2818.909	433.716	19.63	4510.503

Notes: For all variables average values are represented. The average values for the entire sample from 2019 to 2021 are presented herein. Incomes are presented per household member in 2016 prices, in rubles. The number of children is presented per 1,000 households.

Table 3 demonstrates several interesting patterns. *First*, there is an upsurge in the proportion of spending on alcohol and eating out when moving from the low-income to the high-income group. Similarly, household spending on these categories increases alongside their educational levels. *Second*, the per capita household income across education groups shows little variation, from 20,600 rubles at the "Low" education level to 22,200 rubles at the "High" education level. *Third*, the highest number of children is found in groups with low income or high education.

This section examines the typification characteristics of Russian households and delineates the attributes of each household group. Against this background, we expect households from different groups to react differently to the COVID-19 pandemic. Section 5 presents the Kolmogorov – Smirnov test and t-test results, along with Tobit model estimates for each household group.

5. Empirical Results

This section is composed of two parts, which are the statistical and econometric analyses. The statistical analysis uses t-test and Kolmogorov – Smirnov test to reveal the quantitative changes in out-of-home food and alcohol expenditures depending on the level of quarantine measures in the region. Tobit models enable us to estimate the impact of the crisis depending on individual household characteristics. Both the statistical and regression analyses investigate changes within each household group based on income, education level, and the number of children. Table 4 gives the results of the Kolmogorov-Smirnov test and t-test.

Table 4.

Kolmogorov – Smirnov test and t-test for out-of-home food and alcohol expenditure

	Income				Educational attainment			Number of children		
	Low	Medium low	Medium high	High	Low	Middle	High	No children	1 to 2 children	Over 3 children
Expenditure on food outside the home. Comparison of regions with strict quarantine measures and other regions										
t-test	–0.379*** (0.0234)	–0.163*** (0.0289)	–0.245*** (0.0425)	–0.134** (0.0633)	–0.116*** (0.0310)	–0.242*** (0.0213)	–0.354*** (0.0322)	–0.199*** (0.0188)	–0.445*** (0.0356)	–0.514*** (0.0888)
K-S test	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000
Expenditure on food outside the home. Comparison of regions with medium quarantine measures and other regions (excluding severe measures)										
t-test	–0.403*** (0.0155)	–0.308*** (0.0202)	–0.299*** (0.0297)	–0.418*** (0.0426)	–0.161*** (0.0214)	–0.283*** (0.0145)	–0.559*** (0.0216)	–0.295*** (0.0127)	–0.548*** (0.0243)	–0.254*** (0.0651)
K-S test	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Expenditure on food outside the home. Comparison of regions with soft quarantine measures and other regions (excluding severe and medium measures)										
t-test	–0.221*** (0.0189)	–0.260*** (0.0250)	–0.197*** (0.0380)	–0.259*** (0.0523)	–0.110*** (0.0257)	–0.180*** (0.0179)	–0.343*** (0.0266)	–0.181*** (0.0158)	–0.363*** (0.0298)	0.111 (0.0722)
K-S test	0.000	0.000	0.001	0.000	0.003	0.000	0.000	0.000	0.000	0.095
Expenditure on alcohol. Comparison of regions with strict quarantine measures and other regions										
t-test	0.011 (0.0155)	0.055** (0.0216)	0.024 (0.0323)	0.126*** (0.0384)	–0.026 (0.0298)	0.018 (0.0169)	0.105*** (0.0179)	0.044*** (0.0142)	0.059*** (0.0193)	–0.080 (0.0498)
K-S test	0.459	0.022	0.844	0.006	0.233	0.496	0.000	0.036	0.051	0.568

Continuation

	Income				Educational attainment			Number of children		
	Low	Medium low	Medium high	High	Low	Middle	High	No children	1 to 2 children	Over 3 children
	Expenditure on alcohol. Comparison of regions with medium quarantine measures and other regions (excluding severe measures)									
t-test	0.077*** (0.0101)	0.049*** (0.0150)	–0.006 (0.0222)	0.033 (0.0258)	0.057*** (0.0207)	0.037*** (0.0114)	0.066*** (0.0119)	0.032*** (0.0095)	0.092*** (0.0130)	0.164*** (0.0363)
K-S test	0.000	0.016	0.048	0.057	0.028	0.001	0.000	0.000	0.000	0.000
	Expenditure on alcohol. Comparison of regions with soft quarantine measures and other regions (excluding severe and medium measures)									
t-test	–0.005 (0.0117)	–0.034* (0.0176)	–0.030 (0.0267)	–0.057 (0.0306)	–0.019 (0.0237)	–0.032** (0.0134)	–0.013 (0.0139)	–0.028** (0.0112)	0.002 (0.0151)	–0.092** (0.0380)
K-S test	0.802	0.027	0.679	0.784	0.931	0.213	0.358	0.135	0.195	0.379

Note: standard errors are given in parentheses (), *** – 1% significance level, ** – 5% significance level, * – 10% significance level for t-test. For the Kolmogorov – Smirnov test, p-values are presented.

We observe a reduction in the proportion of out-of-home food, irrespective of the imposed restrictions and household characteristics. Areas under medium quarantine measures exhibit the most significant decline in spending. Conversely, regions with soft restrictions show the smallest decrease and, in some instances, no statistically significant impact on expenditure distribution.

Intergroup analysis indicates a significant decline in the proportion of spending on out-of-home food among low, medium-low-income groups, and highly educated households. In the hard quarantine regions, families with 3 or more children significantly reduced the share of their expenditure on out-of-home food compared to households with fewer children. In contrast, households with 3 or more children exhibit the smallest decline in the percentage of their expenditure on restaurants in regions with medium quarantine measures.

The proportion of spending on alcohol rose in areas with hard and medium quarantine measures, while it fell (or remained unchanged) in areas with soft quarantine measures, suggesting that the severity of restrictive measures reduced alcohol expenditure. The intergroup analysis reveals that quarantine restrictions did not affect the alcohol spending habits of those with medium-high incomes. On the other hand, in regions with hard quarantine restrictions, a considerable increase in the share of alcohol expenditures is observed among high income, highly educated households, and those with up to 3 children. In areas with medium quarantine restrictions, there is an observable rise in alcohol spending, particularly among low-income and low-educated households, and those with 3 or more children. The rationale for this is increased stress level due to the severity of restrictions, which is in line with findings of [Avery et al., 2020; Lee, 2020; Rahman et al., 2020]. In regions with soft quarantine restrictions, we find the largest decreases in the share of alcohol expenditure among middle-income, middle-educated households, and households without children. Table 5 shows the Tobit model estimations for out-of-home food and alcohol expenditures across household income groups.

Table 5.

Econometric modelling for income differentiation

	Dependent variable: the share of out-of-home food					Dependent variable: the share of alcohol				
	Low	Medium low	Medium high	High	Whole sample	Low	Medium low	Medium high	High	Whole sample
Children 4–16	0.439***	0.150	–1.558***	1.371*	0.422***	0.036**	–0.877***	–2.156***	–0.463*	–0.461***
	(0.051)	(0.113)	(0.340)	(0.766)	(0.045)	(0.018)	(0.039)	(0.120)	(0.236)	(0.015)
Children under 3	–2.543***	–3.918***	–6.129***	–11.714***	–3.036***	–0.142***	–0.856***	–2.475***	–2.877***	–0.705***
	(0.113)	(0.308)	(1.004)	(3.040)	(0.109)	(0.037)	(0.106)	(0.353)	(1.066)	(0.036)
Age	–0.205***	–0.284***	–0.336***	–0.291***	–0.271***	–0.015***	–0.043***	–0.089***	–0.057***	–0.042***
	(0.003)	(0.004)	(0.008)	(0.006)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
Log of income	3.304***	4.119***	1.566***	8.802***	4.057***	1.328***	0.351***	–0.430**	2.115***	0.870***
	(0.087)	(0.223)	(0.562)	(0.306)	(0.053)	(0.028)	(0.068)	(0.174)	(0.086)	(0.015)
Education level: basic general education	5.241	0.261	–0.497	–2.119	–0.161	0.332	0.351	–0.290	1.692***	0.872***
	(4.471)	(1.873)	(2.039)	(1.898)	(0.984)	(1.051)	(0.348)	(0.359)	(0.444)	(0.201)
Education level: secondary general education	6.827	0.853	0.540	–3.355*	0.753	0.427	1.001***	–0.069	1.307***	1.225***
	(4.463)	(1.830)	(1.974)	(1.838)	(0.966)	(1.048)	(0.336)	(0.341)	(0.426)	(0.196)
Education level: secondary vocational education	7.986*	3.170*	1.865	–2.074	2.334**	0.467	1.461***	0.543	1.507***	1.547***
	(4.462)	(1.819)	(1.956)	(1.813)	(0.961)	(1.048)	(0.333)	(0.333)	(0.419)	(0.194)
Education level: higher vocational education	10.024**	6.227***	5.918***	1.073	5.017***	0.361	1.568***	0.893***	1.072**	1.552***
	(4.462)	(1.820)	(1.959)	(1.815)	(0.962)	(1.047)	(0.333)	(0.336)	(0.420)	(0.194)
Isolation hard	–1.152***	–1.229***	–1.800***	–0.884***	–1.264***	–0.018	–0.087	–0.044	0.076	–0.027
	(0.135)	(0.186)	(0.361)	(0.325)	(0.102)	(0.040)	(0.055)	(0.105)	(0.096)	(0.030)
Isolation soft	0.570***	0.108	0.549**	0.494**	0.464***	–0.047	–0.140***	–0.022	–0.093	–0.066***
	(0.093)	(0.149)	(0.273)	(0.244)	(0.074)	(0.029)	(0.044)	(0.081)	(0.075)	(0.023)
Dummy 2020	–2.539***	–2.202***	–2.261***	–2.498***	–2.454***	0.126***	0.112***	0.087	0.132**	0.114***
	(0.073)	(0.110)	(0.204)	(0.181)	(0.057)	(0.022)	(0.033)	(0.061)	(0.055)	(0.017)
Constant	–39.192***	–43.254***	–14.929**	–96.381***	–42.113***	–12.205***	–2.291***	8.697***	–22.437***	–7.986***
	(4.552)	(2.873)	(6.282)	(4.071)	(1.124)	(1.088)	(0.746)	(1.872)	(1.104)	(0.252)
Sigma constant	10.401***	11.409***	12.657***	13.708***	11.475***	4.264***	4.916***	5.634***	5.489***	4.789***
	(0.052)	(0.096)	(0.217)	(0.185)	(0.052)	(0.018)	(0.026)	(0.049)	(0.040)	(0.014)
Observation	269914	164927	70087	70132	575060	269915	164928	70087	70132	575062

Note: *** indicates 1% significance level, ** indicates 5% significance level, * indicates 10% significance level. Robust standard errors are given in parentheses (). Each model incorporates dummy variables for quarters, with the base category assigned as the first quarter, and dummy variables for Russian regions, with the base category designated as the Altay region. I test the model specification with clustered standard errors at the regional level, which yield the same significance levels for the variables. The estimated coefficients cannot be interpreted as marginal effects, but marginal effects in case of the Tobit model have the same sign and significance. The set of estimated coefficients (control variables) also includes (but not stated in the table) type of locality, average age of household members, savings in household income, changes in financial assets in household income, housing area. Additionally, the model was tested with the inclusion of a self-isolation index from Yandex (higher index = higher restrictions) in place of the conventional restrictions. This variable enables the monitoring of actual compliance with quarantine measures, as it accounts for the actual number of individuals present in public spaces. The modelling results indicate that the obtained estimates are robust to the proxy variable used for quarantine restrictions.

We assessed the effects of the pandemic by incorporating dummy variables reflecting the extent of quarantine restrictions. The baseline category corresponds to periods when a given region implemented medium restrictions. The coefficients (*Isolation hard*, *Isolation soft*) capture how changes in restriction levels within the same region affect expenditures on out-of-home food and alcohol. The findings suggest a decline in the proportion of out-of-home food expenditures when restrictions are tightened from medium to severe within the region. However, the extent of the reduction varies between income groups, with the richest households (group "High") demonstrating a less pronounced reduction in out-of-home food expenditures compared to less affluent households. Conversely, when constraints in the region are relaxed from medium to soft, households increase the share of out-of-home expenditure, suggesting a rebound in consumption.

Expenditure on out-of-home food increases with income in all income groups. Notably, the most affluent households (group "High") demonstrate the highest growth in the proportion of expenditure on out-of-home food as income increases, while the group "Medium high" exhibits the lowest growth in the dependent variable. This phenomenon could be attributed to the fact that low and middle-low-income households receive social assistance and have a low base effect, enabling them to allocate their expenses towards eating out. Conversely, medium-high-income households do not benefit from such support programmes as found by [Mareeva, 2020], thus making it challenging to reallocate their income for out-of-home food expenditure. These results contradict studies for the US and the UK, which indicate a significant reduction in out-of-home food expenditure in the most affluent households [Hacıoğlu-Hoke, Känzig, Surico, 2021; Leone et al., 2020].

In the context of consumption, an individual's education level is a significant determinant of unobserved preferences. Similar to the effect of the income variable, it was found that the proportion of out-of-home food expenditure is elevated in households whose members have attained a higher vocational education level in comparison to households whose members have not attained basic education, in the "Low", "Medium low" and "Medium high" income groups. The effect is most pronounced among the low-income group (group "Low"), indicating the significance of higher education in shaping preferences and interests in learning about other cultures (e.g. through world cuisines). Conversely, as the income level ascends from the poorest to the wealthiest groups, the effect's magnitude and statistical significance diminish.

Statistically significant differences were not identified when strengthening quarantine measures from medium to severe within a region in terms of the share of alcohol expenditures. A decrease in the share of alcohol expenditures was found in regions with soft quarantine measures compared to regions with medium quarantine restrictions among the "Medium low" group. Given the absence of statistically significant effects in other household groups, it can be posited that alcohol habits remain relatively consistent regardless of the level of quarantine measures imposed.

As households move from poorer to richer groups, our findings demonstrate that the impact of income growth on the proportion of alcohol expenditure follows a U-shaped pattern. Therefore, the Low group households, being the poorest, spend more on alcohol as income increases than the Medium low and Medium high group households, but less than the High group households. A corresponding impact is noted by [Pu et al., 2008], where poor households spend most of their expenditure on tobacco and alcohol.

The overall impact of education level is inconsistent. The findings reveal a statistically insignificant impact of educational attainment in the 'Low' group. However, as the economic status of the households increases, a positive difference in the share of alcohol expenditure is observed between households with no basic general education and the rest of the households. We can assert that an increase in income favours a rise in alcohol expenditure. This demonstrates the significant role that education plays in influencing consumer preferences. As educational opportunities expand, low-income groups are not likely to change their expenditure on alcohol. A contrasting effect is evident among the most affluent group of households (group "High"), where heightened levels of education augment the proportion of alcohol expenditure. This phenomenon may be attributed to the notion that enhanced educational attainment fosters the establishment of social interactions, thereby contributing to an increase in the share of alcohol expenditure. Modelling results for selected expenditure groups according to household education groups are presented in Table 6.

Table 6.

Econometric modelling for education differentiation

	Dependent variable: the share of out-of-home food				Dependent variable: the share of alcohol			
	Low	Middle	High	Whole sample	Low	Middle	High	Whole sample
Children 4–16	0.310* (0.162)	0.376*** (0.071)	0.577*** (0.063)	0.422*** (0.045)	–0.754*** (0.054)	–0.498*** (0.023)	–0.280*** (0.021)	–0.461*** (0.015)
Children under 3	–3.509*** (0.444)	–3.667*** (0.192)	–2.390*** (0.138)	–3.036*** (0.109)	–0.849*** (0.132)	–0.661*** (0.059)	–0.562*** (0.046)	–0.705*** (0.036)
Age	–0.360*** (0.009)	–0.293*** (0.004)	–0.235*** (0.003)	–0.271*** (0.002)	–0.066*** (0.002)	–0.044*** (0.001)	–0.030*** (0.001)	–0.042*** (0.001)
Log of income	4.612*** (0.227)	3.883*** (0.088)	4.217*** (0.070)	4.057*** (0.053)	0.900*** (0.050)	0.943*** (0.022)	0.830*** (0.021)	0.870*** (0.015)
Education level: basic general education	–0.821 (1.173)	–	–	–0.161 (0.984)	0.596** (0.233)	–	–	0.872*** (0.201)
Education level: secondary general education	–0.152 (1.153)	–	–4.283 (4.048)	0.753 (0.966)	0.945*** (0.229)	–	–0.552 (1.676)	1.225*** (0.196)
Education level: secondary vocational education	2.011* (1.177)	–	–6.179 (6.075)	2.334** (0.961)	1.491*** (0.237)	–	1.567 (2.016)	1.547*** (0.194)
Education level: higher vocational education	8.257*** (1.334)	2.544*** (0.111)	–1.058 (2.554)	5.017*** (0.962)	2.014*** (0.328)	0.090*** (0.035)	3.802*** (1.126)	1.552*** (0.194)

Continuation

	Dependent variable: the share of out-of-home food				Dependent variable: the share of alcohol			
	Low	Middle	High	Whole sample	Low	Middle	High	Whole sample
Isolation hard	−1.858*** (0.397)	−1.406*** (0.164)	−1.103*** (0.138)	−1.264*** (0.102)	−0.063 (0.102)	−0.042 (0.045)	−0.003 (0.042)	−0.027 (0.030)
Isolation soft	−0.281 (0.296)	0.445*** (0.119)	0.570*** (0.100)	0.464*** (0.074)	0.047 (0.077)	−0.088*** (0.034)	−0.072** (0.032)	−0.066*** (0.023)
Dummy 2020	−2.178*** (0.228)	−2.303*** (0.090)	−2.627*** (0.077)	−2.454*** (0.057)	0.102* (0.059)	0.081*** (0.026)	0.152*** (0.024)	0.114*** (0.017)
Constant	−44.675*** (2.730)	−37.388*** (0.899)	−38.840*** (2.666)	−42.113*** (1.124)	−7.449*** (0.587)	−6.808*** (0.227)	−10.680*** (1.148)	−7.986*** (0.252)
Sigma constant	13.387*** (0.285)	11.743*** (0.088)	11.019*** (0.061)	11.475*** (0.052)	5.858*** (0.050)	4.924*** (0.020)	4.311*** (0.018)	4.789*** (0.014)
Observation	84809	269259	220992	575060	84809	269261	220992	575062

Note: *** indicates 1% significance level, ** indicates 5% significance level, * indicates 10% significance level. Robust standard errors are given in parentheses (). Each model specification incorporates dummy variables for quarters, with the base category assigned as the first quarter, and dummy variables for Russian regions, with the base category designated as the Altay region. I test the model specification with clustered standard errors at the regional level, which yield the same significance levels for the variables. The estimated coefficients cannot be interpreted as marginal effects, but marginal effects in case of the Tobit model have the same sign and significance. The set of estimated coefficients (control variables) also includes (but not stated in the table) type of locality, average age of household members, savings in household income, changes in financial assets in household income, housing area. Additionally, the model was tested with the inclusion of a self-isolation index from Yandex (higher index = higher restrictions) in place of the conventional restrictions. This variable enables the monitoring of actual compliance with quarantine measures, as it accounts for the actual number of individuals present in public spaces. The modelling results indicate that the obtained estimates are robust to the proxy variable used for quarantine restrictions.

As with the income-based household classification, a decrease in the share of expenditure on out-of-home food is observed with the tightening of restrictions from the average and strict level within the region. The degree of decrease in the share of food expenditures is lower among the most educated households (group 'High') compared to the other groups. An increase in the share of expenditure on food outside the home is seen when restrictions within the region are loosened from medium to soft in groups with medium and high levels of education. This can be explained by the fact that within the region, where restrictions on cafes and restaurants were relaxed, the likelihood of repeated severe restrictions was much lower. Furthermore, the absence of lockdowns meant that household purchasing power remained stable.

As income increases, the impact on expenditures on out-of-home food is positive for all groups of households. Households belonging to the "Low" education group exhibit a more substantial increase in out-of-home food expenditure. This phenomenon can be attributed to the propensity of this demographic to allocate a significant portion of their expenditure on dining out at cafes and restaurants. Furthermore, the low base effect contributes to this finding, as household expenditure profiles reveal that the "Low" education group allocates the smallest proportion of their budget to out-of-home food.

Households with higher vocational education level appear to have a higher out-of-home food expenditures compared to households with no basic education in the "Low" and " Middle " groups. At the same time, we find no statistical differences in the share of expenditure on out-of-home food between households with no education and those with higher vocational education level. These findings suggest a shift in consumer preferences during the pandemic, as well as a heightened awareness of the health risks associated with dining in cafes and restaurants during the acute phase of the pandemic.

The influence of hard quarantine measures on the proportion of spending on alcohol in all groups is statistically insignificant. In the context of the relaxation of quarantine restrictions within a region, a decline in the proportion of alcohol expenditure is observed among households with a middle or high level of education. Conversely, households with a low level of education exhibited no change in the proportion of alcohol expenditure following the relaxation of quarantine restrictions within the region. This implies that despite the leniency of restrictions, particular demographic groups chose to not to change their alcohol spending, contrary to the typical consumption pattern where more relaxed quarantine regulations tend to lead to a reduction in alcohol expenses, as indicated by the t-test findings.

The share of alcohol expenditure is positively affected by household income, irrespective of the household group. Our examination of the impact of income on household groups' expenditure highlights differences that imply contrasting consumption behaviour between education and income groups [Padel, Foster, 2005].

A positive difference is observed between households with higher vocational education levels and those with no basic education in all household groups. This indicates a higher share of alcohol expenditure as education level increases, regardless of household group. These results contradict with [Monden et al., 2003], who document that household members with low education heighten health hazards associated with excessive alcohol consumption, while those with high education mitigate such risks. The extant evidence indicates a rejection of Hypothesis 3, which posits a decrease in the proportion of alcohol expenditure with increasing educational attainment. In the context of the pandemic, alcohol has been employed as a coping mechanism by households, irrespective of their educational level (and consequently their awareness of the harms of alcohol), a phenomenon that is consistent with the findings of [Lazarus, 1984]. Table 7 shows the estimates for expenditure on alcohol and out-of-home food, corresponding to the number of children in the household.

In the context of studying the consumption structure of households, the division of families with children is of particular interest for the following reasons. Firstly, the composition of families with children differs from that of families without children, indicating potential shifts in the consumption structure. *Secondly*, children require goods and services (children's food, toys, educational services) that are not necessary for adults, contributing to consumer spending. *Thirdly*, children can influence their parents through the emotional channel of purchases, which can also cause changes in consumer spending.

In the event of a shift in quarantine restrictions from medium to severe measures within the region, a decline in the proportion of expenditure on out-of-home food is observed across all household demographics. Conversely, a transition towards milder quarantine restrictions within the region is associated with an increase in the share of out-of-home food expenditure among households without children and with one to two children. In families with three or more children, the proportion of expenditure on out-of-home food remains consistent as quarantine restrictions are relaxed. This phenomenon is consistent with Hypothesis 1, which suggests that large families may face significant mobility constraints, especially during the COVID-19 pandemic.

Table 7.

Econometric modelling for number of children differentiation

	Dependent variable: the share of out-of-home food				Dependent variable: the share of alcohol			
	No children	1 to 2 children	Over 3 children	Whole sample	No children	1 to 2 children	Over 3 children	Whole sample
Children 4–16	–	2.489*** (0.090)	0.506*** (0.025)	0.422*** (0.045)	–	0.511*** (0.030)	–0.141 (0.093)	–0.461*** (0.015)
Children under 3	–	0.843*** (0.143)	0.274*** (0.048)	–3.036*** (0.109)	–	0.571*** (0.047)	–0.077 (0.119)	–0.705*** (0.036)
Age	–0.337*** (0.003)	0.101*** (0.006)	0.173*** (0.004)	–0.271*** (0.002)	–0.054*** (0.001)	0.034*** (0.002)	0.047*** (0.012)	–0.042*** (0.001)
Log of income	4.778*** (0.073)	4.066*** (0.085)	3.231*** (0.009)	4.057*** (0.053)	0.825*** (0.018)	1.085*** (0.026)	1.424*** (0.100)	0.870*** (0.015)
Education level: basic general education	–1.657 (1.116)	5.941 (5.132)	33.617*** (0.060)	–0.161 (0.984)	0.788*** (0.217)	2.155** (0.972)	–1.737 (1.432)	0.872*** (0.201)
Education level: secondary general education	–0.948 (1.081)	6.842 (5.124)	35.104*** (0.058)	0.753 (0.966)	1.125*** (0.210)	2.451** (0.966)	–2.127 (1.416)	1.225*** (0.196)
Education level: secondary vocational education	0.993 (1.073)	7.470 (5.122)	35.649*** (0.063)	2.334** (0.961)	1.499*** (0.208)	2.555*** (0.965)	–2.056 (1.411)	1.547*** (0.194)
Education level: higher vocational education	4.498*** (1.074)	8.924* (5.121)	36.975*** (0.067)	5.017*** (0.962)	1.500*** (0.208)	2.506*** (0.965)	–2.472* (1.412)	1.552*** (0.194)
Isolation hard	–1.319*** (0.151)	–1.176*** (0.136)	–2.038*** (0.065)	–1.264*** (0.102)	–0.012 (0.039)	–0.058 (0.045)	–0.204 (0.159)	–0.027 (0.030)
Isolation soft	0.395*** (0.110)	0.495*** (0.100)	–0.078 (0.058)	0.464*** (0.074)	–0.079*** (0.030)	–0.058* (0.033)	–0.018 (0.119)	–0.066*** (0.023)
Dummy 2020	–2.440*** (0.084)	–2.473*** (0.076)	–1.824*** (0.059)	–2.454*** (0.057)	0.102*** (0.023)	0.142*** (0.026)	0.153* (0.093)	0.114*** (0.017)
Constant	–49.357*** (1.362)	–55.730*** (5.213)	–75.293*** (0.080)	–42.113*** (1.124)	–7.082*** (0.288)	–13.648*** (1.011)	–10.001*** (1.840)	–7.986*** (0.252)
Sigma constant	12.931*** (0.080)	9.626*** (0.061)	8.337*** (0.034)	11.475*** (0.052)	5.220*** (0.018)	3.799*** (0.018)	3.766*** (0.067)	4.789*** (0.014)
Observation	406509	154214	14337	575060	406510	154215	14337	575062

Note: *** indicates 1% significance level, ** indicates 5% significance level, * indicates 10% significance level. Robust standard errors are given in parentheses (). Each model specification incorporates dummy variables for quarters, with the base category assigned as the first quarter, and dummy variables for Russian regions, with the base category designated as the Altay region. I test the model specification with clustered standard errors at the regional level, which yield the same significance levels for the variables. The estimated coefficients cannot be interpreted as marginal effects, but marginal effects in case of the Tobit model have the same sign and significance. The set of estimated coefficients (control variables) also includes (but not stated in the table) type of locality, average age of household members, savings in household income, changes in financial assets in household income, housing area. Additionally, the model was tested with the inclusion of a self-Isolationlation index from Yandex (higher index = higher restrictions) in place of the conventional restrictions. This variable enables the monitoring of actual compliance with quarantine measures, as it accounts for the actual number of individuals present in public spaces. The modelling results indicate that the obtained estimates are robust to the proxy variable used for quarantine restrictions.

The impact of income growth on the share of expenditure on out-of-home food is generally positive but diminishes in absolute terms with an increase in the number of children in a household. Therefore, households without children increase their expenditure on out-of-home food significantly, in contrast with households that have children. This suggests that households without children have greater autonomy in allocating their expenditure, as well as a heightened interest in the consumption of out-of-home food. A positive difference is observed between households with higher vocational education levels and those with no basic education in all household groups. This finding suggests that as the level of education increases, there is a concomitant increase in the share of expenditure on food outside the home, irrespective of the household group.

The influence of hard quarantine measures on the proportion of spending on alcohol in all groups is statistically insignificant. The relaxation of quarantine restrictions within a region has been demonstrated to result in a decline in the proportion of alcohol expenditure among households with no children and up to two children. In contrast, the proportion of alcohol expenditure among households with three or more children remains unaltered following the relaxation of quarantine restrictions, thereby supporting Hypothesis 1. Our findings reveal that the effect of income is positive overall and that it increases with the number of children in the family. This may appear counterintuitive, as we are examining the proportion of expenditure allocated towards alcohol, rather than overall expenditures. Therefore, it is possible that adult children l-ving within the household may contribute towards this proportion.

We find that the share of alcohol expenditure is higher in households with higher vocational education level compared to households with no basic education in households without children and up to 3 children. Conversely, in large families, the correlation is inverse; that is, the proportion of alcohol expenditure declines with increasing educational attainment. This may be attributed to the fact that as household members attain higher levels of education, they become more aware of their role in their children's lives, thereby setting a positive example. Furthermore, given the labour-intensive nature of raising and educating children, educated parents may have limited time to consume alcohol. Summarising our findings, we can draw the following conclusions on consumer behaviour regarding alcohol and out-of-home food for different groups of households.

1) In the instance of enhanced quarantine restrictions within a region the proportion of expenditure on out-of-home food is observed to decline. Conversely, the relaxation of quarantine restrictions within a region is associated with an increase in the share of expenditure on out-of-home food. A similar relationship has been observed between the relaxation of quarantine measures and a decrease in the share of expenditure on alcohol, a key component in ensuring public health.

2) Income typification. Affluent Russian households faced fewer difficulties than high-income groups in other countries. The consumption behaviour of households from developed countries varies from those in Russia.

3) Education Level typification. Contrary to expectations and theory, as educational attainment rises, households increase the proportion of their expenditure on alcohol, irrespective of the severity of restrictions. The study shows that there are variations in the effects of socio-economic characteristics, indicating differences in consumption behaviour between education and income groups.

4) Number of children typification. Households without children have more autonomy in deciding their consumption habits, resulting in a more substantial increase in the portion of out-of-home food expenditure as income grows.

6. Conclusion

This paper examines the impact of the COVID-19 pandemic on out-of-home food and alcohol expenditures in Russian regions. To test the hypotheses, we use microdata from the household budget survey conducted by the Federal State Statistics Service. To obtain targeted recommendations, households are categorised according to income, education level, and the number of children. Furthermore, we consider varying degrees of quarantine measures, encompassing areas with soft, medium, and hard quarantine restrictions.

The first hypothesis is confirmed, as evidenced by the decrease in the share of out-of-home food expenditure among households with children in regions subject to severe and moderate quarantine measures. On average, the results indicate that households with children experience a decline in the share of out-of-home food expenditure that is twice as strong as that observed among households without children. The second hypothesis is partially confirmed, as high-income households in regions with strict quarantine measures increased the share of expenditure on alcohol, as indicated by the results of t-test. Overall, the results indicate a propensity to increase alcohol expenditure, specifically among low-income households. Contrary to Hypothesis 3, the t-test results indicate that in regions with both strict and medium quarantine measures, high-educated households increase the share of alcohol expenditure under quarantine. Furthermore, regression analyses demonstrate that as educational attainment rises, households, in defiance of expectations, augment their alcohol expenditure, irrespective of the severity of restrictions.

We discovered a curious effect in which the share of expenditure on out-of-home food decreases regardless of the social group and the severity of quarantine restrictions. Notably, the decline is substantially less pronounced in areas with soft quarantine restrictions, as supported by both statistical tests and the results of the regression analysis. In a similar fashion, the proportion of alcohol expenditures decreased across almost all social strata when the quarantine restrictions in the region were relaxed, whereas in regions with hard and medium quarantine measures, the share of alcohol expenditure increased significantly, implying a negative effect of quarantine restrictions on the health of household members.

We highlight the following limitations of the study. *First*, we acknowledge that attitudes towards the pandemic significantly impact the decision to visit a restaurant and consume alcohol during quarantines. However, due to the limited availability of Federal State Statistics Service data on households' intentions to comply with the quarantine, we were unable to include this aspect in our analysis. *Second*, the breakdown of alcohol expenditure facilitates specific recommendations for regulating alcohol consumption at the state level. At the household level, the Federal State Statistics Service does not provide detailed data on different types of alcohol expenditure. *Third*, the Federal State Statistics Service does not provide data on the form and place of employment of household members. Consequently, some effects of the pandemic on the structure of household consumption may not be fully captured. *Fourth,* a potential discrepancy exists between formal indicators of the severity of quarantine measures and their actual implementation in Russian regions. The available data fails to consider the degree of compliance with restrictions by the population, differences in enforcement between regions, or the dynamics of household adaptation.

* *
*

References

Abebe G.K., Charlebois S., Music J. (2022) Canadian Consumers' Dining Behaviors during the COVID-19 Pandemic: Implications for Channel Decisions in the Foodservice Industry. *Sustainability*, 14, 9, pp. 1–21.

Acharya B., Dhakal C. (2022) Risky Health Behaviors during the COVID-19 Pandemic: Evidence from the Expenditures on Alcohol, Non-alcoholic Beverages, and Tobacco Products. *PLoS One*, 17, 5, pp. 1–14.

Alonso L.E., Rodríguez C.J.F., Rojo R.I. (2015) From Consumerism to Guilt: Economic Crisis and Discourses about Consumption in Spain. *Journal of Consumer Culture*, 15, 1, pp. 66–85.

Ammar A., Brach M., Trabelsi K., Chtourou H., Boukhris O., Masmoudi L., Bouaziz B., Bentlage E., How D., Ahmed M. (2020) Effects of COVID-19 Home Confinement on Eating Behaviour and Physical Activity: Results of the ECLB-COVID19 International Online Survey. *Nutrients*, 12, 6, pp. 1–13.

Anderson P., O'Donnell A., Llopis E. J., Kaner E. (2022) The COVID-19 Alcohol Paradox: British Household Purchases during 2020 Compared with 2015–2019. *PLoS One*, 17, 1, pp. 1–13.

Antonov S. (2024) How Much Alcohol Is Drunk in Russia. *Tinkoff Journal*. Available at: https://t-j.ru/drunkards/ (Retrived 10.02.2025).

Avery A.R., Tsang S., Seto E.Y.W., Duncan G.E. (2020) Stress, Anxiety, and Change in Alcohol Use during the COVID-19 Pandemic: Findings among Adult Twin Pairs. *Frontiers in Psychiatry*, 11, pp. 1–12.

Baker S.R., Farrokhnia R.A., Meyer S., Pagel M., Yannelis C. (2020) How Does Household Spending Respond to an Epidemic? Consumption during the 2020 COVID-19 Pandemic. *The Review of Asset Pricing Studies*, 10, 4, pp. 834–862.

Barigozzi M., Alessi L., Capasso M., Fagiolo G. (2012) The Distribution of Household Consumption-Expenditure Budget Shares. *Structural Change and Economic Dynamics*, 23, 1, pp. 69–91.

Becker G.S. (1965) A Theory of the Allocation of Time. *The Economic Journal*, 75, 299, pp. 493–517.

Becker G.S. (2009) *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press.

Blundell R., Costa Dias M., Cribb J., Joyce R., Waters T., Wernham T., Xu X. (2022) Inequality and the COVID-19 Crisis in the United Kingdom. *Annual Review of Economics*, 14, pp. 607–636.

Blundell R., Chen X., Kristensen D. (2007) Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves. *Econometrica*, 75, 6, pp. 1613–1669.

Chen H., Qian W., Wen Q. (2021) The Impact of the COVID-19 Pandemic on Consumption: Learning from High-Frequency Transaction Data. *AEA Papers and Proceedings*, 111, pp. 307–311.

Chen R., Li T., Li Y. (2024) Analyzing the Impact of COVID-19 on Consumption Behaviors through Recession and Recovery Patterns. *Scientific Reports*, 14, 1, pp. 1–14.

Cutler D.M., Lleras-Muney A. (2010) Understanding Differences in Health Behaviors by Education. *Journal of Health Economics*, 29, 1, pp. 1–28.

Dhakal C., Acharya B., Wang S. (2022) Food Spending in the United States during the First Year of the COVID-19 Pandemic. *Frontiers in Public Health*, 10, pp. 1–7.

Edelstein P., Kilian L. (2009) How Sensitive Are Consumer Expenditures to Retail Energy Prices? *Journal of Monetary Economics*, 56, 6, pp. 766–779.

Federal State Statistics Service. (2022) *Household Budget Survey*. Available at: https://obdx.gks.ru/ (Retrieved: 23.10.2023).

Gautier E., Ulgazi Y., Vertier P. (2020) Inflation and Households' Inflation Expectations during the Covid-19 Pandemic. *Banque De France, Eco Notepad*, 171. pp. 1–4.

Greene W.H. (2008) Limited Dependent Variables – Truncation, Censoring and Sample Selection. *Econometric Analysis*, pp. 833–834.

Grigoryev L.M., Elkina Z.S., Mednikova P.A., Serova D.A., Starodubtseva M.F., Filippova E.S. (2021) The Perfect Storm of Personal Consumption. *Voprosy Ekonomiki*, 10, pp. 27–50.

Hacıoğlu-Hoke S., Känzig D.R., Surico P. (2021) The Distributional Impact of the Pandemic. *European Economic Review*, 134, pp. 1–19.

Hausman J.A., Newey W.K., Powell J.L. (1995) Nonlinear Errors in Variables Estimation of Some Engel Curves. *Journal of Econometrics*, 65, 1, pp. 205–233.

Jacob L., Smith L., Armstrong N.C., Yakkundi A., Barnett Y., Butler L., McDermott D.T., Koyanagi A., Shin J.Il, Meyer J. (2021) Alcohol Use and Mental Health during COVID-19 Lockdown: A Cross-Sectional Study in a Sample of UK Adults. *Drug and Alcohol Dependence*, 219, pp.1–5.

Janssen M., Chang B.P., Hristov H., Pravst I., Profeta A., Millard J. (2021) Changes in Food Consumption during the COVID-19 Pandemic: Analysis of Consumer Survey Data from the First Lockdown Period in Denmark, Germany, and Slovenia. *Frontiers in Nutrition*, 8, pp. 1–20.

Kolomak E.A. (2020) Economic Effects of Pandemic-Related Restrictions in Russia and their Spatial Heterogeneity. *R-Economy,* 6, 3, pp. 154–161.

Lazarus R.S. (1984) *Stress, Appraisal, and Coping*. Springer.

Lee S.A. (2020) Coronavirus Anxiety Scale: A Brief Mental Health Screener for COVID-19 Related Anxiety. *Death Studies*, 44, 7, pp. 393–401.

Leone L.A., Fleischhacker S., Anderson-Steeves B., Harper K., Winkler M., Racine E., Baquero B., Gittelsohn J. (2020) Healthy Food Retail during the COVID-19 Pandemic: Challenges and Future Directions. *International Journal of Environmental Research and Public Health*, 17, 20, pp. 1–14.

Leser C.E.V. (1963) Forms of Engel Functions. *Econometrica: Journal of the Econometric Society*, pp. 694–703.

Lukashina E. (2020). Strategy of Socio-Economic Development of The Kaluga Region Until 2030. Half of the Way Is Passed. *Innovacionnaya Ekonomika i Sovremennyj Menedzhment*, 5, pp. 25–30.

Maniriho A., Musabanganji E., Nkikabahizi F., Ruranga C., Lebailly P. (2021) Analysis of the Determinants of Households' Expenditures in Rwanda. *UKH Journal of Social Sciences*, 5, 1, pp. 1–10.

Marano G., Traversi G., Gaetani E., Pola R., Claro A.E., Mazza M. (2022) Alcohol Use Disorder and Liver Injury Related to the COVID-19 Pandemic. *World Journal of Hepatology*, 14, 10, pp. 1875–1883.

Mareeva S. (2020) Socio-Economic Inequalities in Modern Russia and their Perception by the Population. *The Journal of Chinese Sociology*, 7, 1, pp. 1–19.

Monden C.W.S., Van Lenthe F., De Graaf N.D., Kraaykamp G. (2003) Partner's and Own Education: Does Who You Live with Matter for Self-Assessed Health, Smoking and Excessive Alcohol Consumption? *Social Science & Medicine*, 57, 10, pp. 1901–1912.

Murashov Y., Ratnikova T. (2017) The Dynamics of Unrecorded Income of Russian Households. *Applied Econometrics*, 2, 46, pp. 30–54.

Nartikoev A.R., Peresetsky A.A. (2021) Endogenous Household Classification: Russian Regions. *Voprosy Ekonomiki*, 5, pp. 107–128.

Padel S., Foster C. (2005) Exploring the Gap between Attitudes and Behaviour: Understanding Why Consumers Buy or Do not Buy Organic Food. *British Food Journal*, 107, 8, pp. 606–625.

Pišot S., Milovanović I., Šimunič B., Gentile A., Bosnar K., Prot F., Bianco A., Lo Coco G., Bartoluci S., Katović D. (2020) Maintaining Everyday Life Praxis in the Time of COVID-19 Pandemic Measures (ELP-COVID-19 survey). *European Journal of Public Health*, 30, 6, pp. 1181–1186.

Potapenko V.V, Shirov A.A. (2021) Forecast of Russian Personal Consumption Expenditures As Function of Income Distribution and Relative Prices. *Studies on Russian Economic Development*, 1, 184, pp. 6–18.

Pu C., Lan V., Chou Y.-J., Lan C. (2008) The Crowding-Out Effects of Tobacco and Alcohol Where Expenditure Shares Are Low: Analyzing Expenditure Data for Taiwan. *Social Science & Medicine*, 66, 9, pp. 1979–1989.

Rahman M.A., Hoque N., Alif S. M., Salehin M., Islam S. M.S., Banik B., Sharif A., Nazim N.B., Sultana F., Cross W. (2020) Factors Associated with Psychological Distress, Fear and Coping Strategies during the COVID-19 Pandemic in Australia. *Globalization and Health*, 16, 1, pp. 1–15.

Rausch S., Metcalf G.E., Reilly J.M. (2011) Distributional Impacts of Carbon Pricing: A General Equilibrium Approach with Micro-data for Households. *Energy Economics*, 33, pp. 20–33.

Salehi-Isfahani D. (2023) The Impact of Sanctions on Household Welfare and Employment in Iran. *Middle East Development Journal*, 15, 2, pp. 189–221.

Schmits E., Glowacz F. (2022) Changes in Alcohol Use during the COVID-19 Pandemic: Impact of the Lockdown Conditions and Mental Health Factors. *International Journal of Mental Health and Addiction*, 20, 2, pp. 1147–1158.

Seliverstov V.E., Kravchenko N.A., Klistorin V.I., Yusupova A.T. (2021) Russian Regions and the Federal Center Against Global Threats: A Year of Fighting COVID-19. *Regional Research of Russia*, 11, pp. 405–418.

Soberon-Ferrer H., Dardis R. (1991) Determinants of Household Expenditures for Services. *Journal of Consumer Research*, 17, 4, pp. 385–397.

Sobol M., Blachnio A., Przepiorka A. (2020) Time of Pandemic: Temporal Perspectives Related to Compliance with Public Health Regulations Concerning the COVID-19 Pandemic. *Social Science & Medicine*, 265, pp. 1–5.

St. Petersburg Policy Foundation (2020) *Rating of the Petersburg Policy Foundation*. Available at: https://www.fpp.spb.ru/fpp-rating-2020-04 (Retrived 06.04.2023).

Thaler R. (1985) Mental Accounting and Consumer Choice. *Marketing Science*, 4, 3, pp. 199–214.

Tobin J. (1958) Estimation of Relationships for Limited Dependent Variables. *Econometrica: Journal of the Econometric Society*, pp. 24–36.

Varlamova J., Larionova N. (2015) Macroeconomic and Demographic Determinants of Household Expenditures in OECD Countries. *Procedia Economics and Finance*, 24, pp. 727–733.

Verter N., Osakwe C.N. (2014) A Time Series Analysis of Macroeconomic Determinants of Household Spending in the Era of Cross-Cultural Dynamics: Czech Republic As a Case Study. *Procedia Economics and Finance*, 12, pp. 733–742.

Voytenkov V., Demidova O. (2023) Impact of COVID-19 on Household Consumption in Russia. *Applied Econometrics*, 72, pp. 73–99.

Xiong J., Tang Z., Zhu Y., Xu K., Yin Y., Xi Y. (2021) Change of Consumption Behaviours in the Pandemic of COVID-19: Examining Residents' Consumption Expenditure and Driving Determinants. *International Journal of Environmental Research and Public Health*, 18, 17, pp. 1–15.

Yen S.T., Jensen H.H. (1996) Determinants of Household Expenditures on Alcohol. *Journal of Consumer Affairs*, 30, 1, pp. 48–67.

Yukseltan E., Kok A., Yucekaya A., Bilge A., Aktunc E.A., Hekimoglu M. (2022) The Impact of the COVID-19 Pandemic and Behavioral Restrictions on Electricity Consumption and the Daily Demand Curve in Turkey. *Utilities Policy*, 76, pp. 1–17.

Изменения в расходах на питание и алкоголь вне дома во время пандемии COVID-19

Войтенков Валентин Александрович

стажер-исследователь в Научно-учебной лаборатории пространственно-эконометрического моделирования социально-экономических процессов в России
Национальный исследовательский университет «Высшая школа экономики»,
11, Покровский бульвар, Москва, 109028, Россия.
E-mail: vvoytenkov@hse.ru

В данном исследовании изучаются изменения в расходах на питание и алкоголь вне дома во время пандемии COVID-19. Исследование посвящено изучению межгрупповых различий между домохозяйствами с учетом различий российских регионов по степени соблюдения карантинных ограничений. Для проверки гипотез исследования мы используем микроданные обследования бюджетов домашних хозяйств, проводимого Федеральной службой государственной статистики. Для сравнения расходов на питание и алкоголь вне дома в регионах с мягкими, средними и жесткими ограничительными мерами мы используем t-тест для сравнения средних и тест Колмогорова – Смирнова для сравнения распределений. Модель Тобита применяется для сравнения привычек домохозяйств разных социальных групп. Совместный анализ внедомашних расходов на еду и алкоголь позволяет отделить непроизвольные сбережения от стратегий преодоления, используя модели для цензурированных данных, что способствует углубленной оценке благосостояния домохозяйств в условиях потрясений. Полученные результаты свидетельствуют о сокращении расходов на продукты питания вне дома во всех социальных группах и при всех уровнях карантинных ограничений. Доля расходов на алкоголь снизилась почти во всех социальных группах в регионах с мягкими мерами, но значительно увеличилась в регионах со средними и жесткими ограничениями.

Ключевые слова: домашние хозяйства; межгрупповой анализ; расходы на питание вне дома; расходы на алкоголь; COVID-19; жесткость карантинных мер.

УДК 336.71, 336.77

DOI: 10.17323/1813-8691-2025-29-1-160-182

Does Macroprudential Policy Matter to Manage Banking Credit Risk? Evidence from Commercial Banks in Asia-Pacific Region[1]

Zainuri Zainuri[1], Sebastiana Viphindrartin[2], Regina Niken Wilantari[3], Ahmad Roziq[4]

[1] University of Jember,
Jember 68121, Indonesia.
E-mail: zainuri.feb@unej.ac.id

[2] University of Jember,
Jember 68121, Indonesia.
E-mail: sebastiana@unej.ac.id

[3] University of Jember,
Jember 68121, Indonesia.
E-mail: reginanikenw.feb@unej.ac.id

[4] University of Jember,
Jember 68121, Indonesia.
E-mail: ahmadroziq.feb@unej.ac.id

The global financial crisis triggered a debate on the pros and cons of using macroprudential policy as a prudential control tool that includes capital reserves or requirements to address systemic risk, financial credit cycles, and macroeconomic stabilization objectives. The macroprudential policy has now been establi-

[1] Thank you to the Faculty of Economics and Business and the Institute for Research and Community Service, University of Jember, who have provided encouragement, facilities, and funding to write this manuscript well.

Zainuri Zainuri – Associate Professor, Economics Department, Faculty of Economics and Business.
Sebastiana Viphindrartin – Associate Professor, Economics Department, Faculty of Economics and Business.
Regina Niken Wilantari – Associate Professor, Economics Department, Faculty of Economics and Business.
Ahmad Roziq – Professor, Accounting Department, Faculty of Economics and Business.

The article was received: 26.03.2024/The article is accepted for publication: 06.02.2025.

shed as an area of financial policy to stop excessive risk-taking in the financial sector and reduce its consequences to the real economy in response to the lessons learned from the global financial crisis. Controlling credit risk also requires a government-run fiscal sector, one of which is controlling corruption. Corruption significantly affects credit risk. This study aims to examine the effectiveness of macroprudential policy instruments and the role of institutional instruments in controlling commercial bank credit risk in the Asia Pacific region from 2012 to 2023. This study uses the generalized method of moments (GMM) as an analytical tool. The results show that loan-to-value and corruption significantly affect credit risk in Asia Pacific.

***Key words*:** capital adequacy ratio; capital conservation buffer; corruption; loan to value; credit risk, GMM.

JEL Classification: E51, E61, F55.

For citation: Z. Zainuri, S. Viphindrartin, R.N. Wilantari, A. Roziq. Does Macroprudential Policy Matter to Manage Banking Credit Risk? Evidence from Commercial Banks in Asia-Pacific Region. *HSE Economic Journal.* 2025; 29(1): 160–182.

Introduction

The active use of prudential tools to manage credit allocations and volumes was rendered obsolete in the 1990s as financial liberalization became the prevailing regulatory trend. Then, in 2008, the world financial crisis struck financially developed nations hard, revealing the limitations of the prudential framework that was in place at the time for managing systemic risk. This caused the balance to shift and the active use of prudential tools – known as macroprudential policies in the vernacular – to smooth the credit cycle and prevent significant crises. Examples of these tools include reserve requirements, loan-to-value ratios, taxes on lending, and capital requirements. Macroeconomists were compelled to examine the relationship between financial frictions and economic cycles after the global financial crisis revealed the risky relationships between micro- and macro-stability [Cordella, Pienknagura, 2020]. The most important lessons from this crisis are the necessity of strengthening the macroprudential approach to supervision and regulation, which can recognize vulnerabilities across the system and take appropriate action to ensure financial stability, and the significance of minimizing systemic financial risks.

In order to reduce systemic or system-wide financial risk and, consequently, the likelihood of disruptions in the delivery of essential financial services, which may have detrimental effects on the actual economy, macroprudential policy primarily employs prudential mechanisms. Ebrahimi Kahou and Lehar (2017) state that preserving systemic financial stability and lowering systemic risk are the two primary goals of macroprudential monitoring and regulation. The possibility of a single firm's failure causing the entire financial system to implode is known as systemic risk. According to Calmès and Théoret (2014), it is the outcome of financial institutions' undercapitalization in an increasingly interconnected market. It's critical to comprehend and measure this systemic risk to ensure our financial institutions have enough capital to survive another financial crisis. Time dimension and cross-sectional dimension are the two categories of systemic risks that macroprudential policy addresses, according to Spelta et al. (2018). The financial system's overall risk and how it changes over time are covered under the temporal dimension.

The global financial crisis triggered a debate regarding the pros and cons of using macroprudential policy as a prudential control tool that includes capital reserves or requirements to address systemic risk, financial credit cycles, and macroeconomic stabilization objectives. Nonetheless, macroprudential policies that limit loan-to-value ratios, debt-to-income ratios, credit growth restrictions, balance sheet restrictions, and capital and reserve requirements are of significant concern in reducing systemic risk in financial markets. The macroprudential policy framework has been laid out in the Bank of International Settlements (BIS) since the early 2000s and has been implemented extensively in emerging markets while developed countries have just begun to adopt it [Cerutti, Claessens, Laeven, 2015; Freixas, Laeven, Peydro, 2015; Rojas, Vegh, Vuletin, 2020], but the related models and implementation performance are still ambiguous because the policy has several goals at the beginning of this policy so that the implementation of the policy does not attract much attention in policymaking or academia [Thiemann, Aldegwy, Ibrocevic, 2018]. This ambitious policy program has been reduced to a much more scaled-back incremental approach during implementation. While focusing on increasing the system's resilience, implemented measures largely refrain from intervening in the build-up of financial risks during the upswing of the cycle. Looking at the measures introduced internationally, few, if any, bear a clear anti-cyclical character that could constrain credit in the upswing.

The macroprudential policy focuses on establishing a protective banking framework to create conditions for resilience and financial sector actors while suppressing the procyclical behavior of banks [Benigno, Chen, Otrok, Rebucci, Young, 2013; Buncic, Melecky, 2013; Claessens, Ghosh, Mihet, 2013; Freixas, Perez-Reyna, 2021]. Still, on the other hand, the policy shift towards macroprudential creates a consequence of regulatory change that can also translate into a decrease in credit supply and erosion of economic activity, with unintended consequences such as greater unemployment and decreased investment [Meuleman, Vander, 2020].

Cerutti et al. (2016) have researched using macroprudential policies through 12 macroprudential instruments to control credit growth. It is concluded that the implementation of macroprudential policies in developing countries and developed countries has differences; in developing countries, macroprudential policies related to foreign exchange are used more intensively, while in developed countries, macroprudential policies in the form of borrower-based policies are more widely used. The implementation of borrower-based policies is based on financial openness in developed countries, so fluctuations in credit growth will directly impact pro-perty price growth. The failure of monetary policy to create stability has led to an orientation toward financial stability. The shift towards macroprudential is one of the main consequences of the global financial crisis [Baker, Widmaier, 2015; Galati, Moessner, 2013]. With the policy shift towards macroprudential, central banks worldwide started implementing preventive measures to protect against systemic financial risks and utilizing non-conventional tools to ensure financial stability, where financial stability has become a key economic policy objective within the G-20 to promote international financial stability [Baker, 2013].

The literature study on the relationship between macroprudential policy and credit risk conducted by Behncke (2023) revealed that through macroprudential policy, banks limit lending and ultimately reduce banking risk directly. The spillover effect from other countries suggests that the lending limits should be carefully designed and closely monitored. The provision of capital reserves through capital buffers increases the probability of bank survival during a banking crisis [Berger, Bouwman, 2013]. In addition, the Countercyclical Capital Buffer (CCyB) policy launched can contribute to the banking system's resilience but does not prevent the credit boom and asset price collapse. According to Altunbas et al. (2018), macroprudential policies significantly affect bank risk evaluated by Z-score, especially for small, heavily funded, and undercapitalized banks. Non-performing loans (NPLs) measure credit risk in other studies [Chaibi, Ftiti, 2015]. Hence, macroprudential policies are created to support traditional macroeconomic instruments and micro-prudential regulation of financial institutions, according to academics and practitioners. By reducing the negative effects of financial volatility and mitigating excessive procrastination, such policies seek to contain (accumulated) risks to systemic stability, reducing the costs to the economy of disrupted financial services that support the operation of financial markets [Gaganis, Lozano-Vivas, Papadimitri, 2019].

Behncke (2023) carried out a research study of the effects of macroprudential measures, such as the loan-to-value (LTV) and countercyclical capital buffer (CCyB) policies, on bank len-ding and credit risk in Switzerland. Although the sectoral CCyB and the LTV ceiling have the same overall goal of protecting the mortgage market from the accumulation of systemic hazards, they do so through different mechanisms. Although it does not explicitly limit any mortgage characteristics, the LTV ceiling, and CCyB are extra capital requirements on all ongoing mortgage loans and aim to target the LTV distribution of new loans. The CCyB requires banks to increase their capital, based on the risk-weighted assets (RWA) associated with residential mortgages.

Wang & Sun (2013) the case of China and conclude that while reserve requirements and housing-related policies can help mitigate procyclicality, They are insufficient to lower systemic risks. This suggests that more carefully targeted policies may be able to contain macro-financial vulnerabilities. The study by Altunbas et al. (2018) assesses the impact of macroprudential tools on bank risk indicators across a broad range of nations. The findings indicate that these tools, particularly those intended to improve banks' resilience, have a noteworthy influence on bank risk. According to other studies that also use bank-level data; applying macroprudential rules may have unintended consequences known as leakage effects [Aiyar, Calomiris, Wieladek, 2014].

Jiménez et al. (2017) investigated the impact of countercyclical provisions on credit growth in Spain and the real effects that followed. They discovered that while these provisions effectively reduced the effects of a credit crunch (because of the accumulation of capital buffers), they were less effective in containing the credit boom that preceded the crisis. Along the same lines, Mustafa and Mirza (2022) find evidence to support the idea that marginal reserve requirements and provisions have increased bank resilience through increases in solvency and liquidity buffers, while Lopez et al. (2018) find that countercyclical provisions in Colombia effectively helped reduce the amplitude of credit cycles. Furthermore, the latter offers proof of how the abovementioned regulations reduced ex-post bank credit risk by utilizing loan-level data.

Zhang and Zoli (2014) analyzed 46 countries, 13 of which were Asian, between 2000 and 2013, concluding that housing-related policies, particularly debt-to-income (DTI) and loan-to-value (LTV) caps, were effective in limiting the growth of credit in Asian countries, while this limiting effect was relatively low for other countries. Kuttner and Shim (2016) concluded that tightening the DTI cap resulted in a 4–7% reduction in housing loans in the subsequent year. Fendoğlu (2017) demonstrated that borrower-based instruments and required reserves are useful in moderating credit cycles.

Studies on Turkey have also revealed the empirical results in the literature on the detrimental and noteworthy effects of macroprudential policy on credit development. According to Kara et al. (2013), the asymmetric interest rate corridor is one macroprudential tool that can be used to manage credit growth by influencing credit spreads. Similarly, Bulut (2015) showed that the interest rate corridor's uncertainty about the quantity of funds and fund expenses constraining impacted loans. Bumin, Taşkın and Dilvin Taşkın (2015) discovered a statistically significant negative correlation between consumer loans and the macroprudential policies implemented by BRSA. According to Ilhan et al. (2021), consumer loans were restrained by the BRSA's regulations about general requirements and maturity limitations. According to Alper et al. (2018), lending practices impact credit growth due to the reserve requirement policy.

However, further research into macroprudential regulation is needed to fully understand its implementation strategy and effectiveness. The efficiency of macroprudential policy and its underlying components in attaining the ultimate purpose of regulation, which is to improve financial stability by reducing bank systemic risk and procyclicality, has yet to provide consistent outcomes. The implementation and effectiveness of macroprudential policies have previously been studied in several articles [Altunbas et al., 2018; Apergis, 2017; Cerutti et al., 2016; Gaganis et al., 2019].

However, the advantages and disadvantages of macroprudential capital control remain primarily theoretical. Microprudential regulations continue to control commercial banks' capital requirements. While most scholars and regulators now agree that micro-prudential capital regulation alone is insufficient to prevent large-scale bank failures, the effects of preserving this regime in the face of systemic risk have not been well investigated. The majority of research on the effectiveness of microprudential capital regulation, which was conducted before the financial crisis, focuses on individual banks rather than the banking industry as a whole [McKeever, 2023]. By evaluating the risk that each financial institution faces as well as the amount of risk derived from each bank institution's performance outcomes, microprudential policy may be understood. The implementation of supervision or microprudential policies can mitigate the risk of financial instability by impeding the creation of a financial institution risk that may lead to the failure or bankruptcy of other financial enterprises [Fajriani, Sudarmawan, 2022].

This study uses microprudential policy, namely the capital adequacy ratio (CAR). Capital adequacy is defined as having enough capital to absorb losses and successfully avert banking business failures. Implementing regulations to guarantee that these institutions have adequate capital to assure the survival of a secure and efficient financial system that can withstand any predicted challenges ultimately protects banks, their clients, the government, and economy. The purpose of capital adequacy requirements is to safeguard depositors and other creditors from loss in the event of liquidation and to create a buffer against losses not covered by current bank revenues [Okon, 2022]. The quantity of capital that can successfully carry out the primary capital function of preventing bank collapse by absorbing losses was considered appropriate capital from a functional standpoint. The riskier the asset composition, the more capital is needed to maintain a certain degree of soundness and, consequently, the larger the capital adequacy needed to maintain solvency, as capital serves as a buffer against which to charge off losses. The capital adequacy ratio indicates that the bank has risky loans, and the bank should raise capital to keep up with the loans as these risk assets increase [Zulkhibri, 2018].

The Asia Pacific region's economic outlook remains strong, and the region continues to be the most dynamic region in the global economy, accompanied by a range of possible risks such as tightening global financial conditions, a shift towards protectionist policies, and increased geopolitical tensions. Given these uncertainties, macroeconomic policies must be conservative and aimed at building cushions and enhancing resilience. Policymakers also need to continue to push for structural reforms to address medium- and long-term challenges, such as population aging and declining productivity, and to ensure that Asia can reap the full benefits of increased digitalization in the global economy.

Banking NPLs in the Asia-Pacific Region tend to increase, which is indicated by the increase in NPLs in the region. In 2018, the average NPL increased by 0.30%. Indonesia, Vietnam, and Thailand have the highest NPLs in Asia-Pacific [Dahl et al., 2019]. Unlike in other countries, the banking system in the Asia Pacific Region has been primarily deregulated and privatized by implementing Basel Accord III rules in the form of tightening capital to reduce credit risk [Andrle et al., 2017; Casimano, Hakura, 2011]. Capital cushioning and liquidity buffers significantly contributed to recovering and maintaining banking activities from risks after the financial crisis.

In contrast to previous studies [Bruno, Shim, Shin, 2017; Cantu, Gambarcota, Shim, 2020; Kim, Mehrota, 2019], this paper extends existing research by analyzing the effectiveness of macroprudential policies in controlling commercial banking credit risk in Asia Pacific countries after the 2008 crisis from 2012 to 2023. Other researchers have not used the capital conservation buffer (CCoB) instrument as a critical determinant in influencing bank credit risk. CCoB and the countercyclical capital buffer in Basel Accord III are essential indicators in the regulation of banking capital buffers. In addition, no similar research covers countries in the Asia-Pacific Region, where most previous researchers only conducted research within the scope of a particular country. The benefits gained from a broader research coverage can test whether macroprudential policies projected by CCoB variables can consistently control credit risk in the Asia-Pacific region.

This research also adds an institutional control variable, namely corruption, to credit risk in a country. The influence of corruption on banks' credit risk, particularly in developing nations where banks face significant non-performing loan (NPL) loads and pervasive corruption, it is higher levels of corruption in the targeted nations result in a higher number of bad loans rather than a decrease in loan defaults in the economies with higher lending rates and quicker growth rates. Because of these nations' existing bureaucratic structures, the public sector's hegemonic status, and the lack of transparency, corruption will likely impact the financial sector. Chen, Jeon, Wang, and Wu (2015) looked into how corruption affected banks' propensity for taking risks and discovered a continuous pattern linking higher levels of corruption to higher bank risk-taking. Additionally, this study demonstrated that monetary policy exacerbates the issue of credit risk and that the indirect effects of corruption also affect banks' risk. It also showed how corrupttion impedes economic growth in developing nations by misallocating loanable funds and has a greater effect on banks' propensity for taking risks as corruption worsens.

This study is divided into the following sections. A review of the literature and the formulation of hypotheses are presented in Section 2. The data and the study's empirical approach are described in Section 3. Section 4 summarizes the study's key findings and outcomes. Section 5 gives the study's conclusion.

Literature Review

This paper relates to several sections of the literature. Macroprudential policies emphasize prudential principles to assist central banks in limiting the presence of systemic risks that would weaken the performance of the financial system and impact the economic system in a country [Dumičić, 2017; Lombardi, Siklos, 2016]. Various macroprudential instruments help address procyclicality and daily exposures, and the macroprudential policy is one of the measures established by a central bank that the central bank uses to help banks manage credit risk. Since the 2000s, macroprudential policies have been implicitly implemented. The purpose of macroprudential policy is to lower expenses arising from interruptions in financial services, including credit, insurance, and other payment services. According to the central bank, controlling systemic risk is the primary goal of macroprudential policy to maintain the financial system's stability as a whole. The Global Financial Crisis (GFC) of 2007–2008 highlighted the need for macroprudential intervention. Systemic risk is dangerous, and the Global Financial Crisis (GFC) of 2007–2008 is a stark reminder of this in a nation's financial system's stability.

Three pillars support the macroprudential policy framework: financial system stability, balanced and high-quality intermediaries, and economic and financial inclusion. Banking, economics, and Sharia finance are the three pillars of economic and financial inclusion. These three dimensions focus on the many tools the central bank uses, which apply to conventional and Sharia banks. The initial pillar centers on promoting equitable and superior intermediation in prospective industries. This pillar contains instruments related to credit, like the capital conservation buffer, which aims to cover losses on risky investments, Loan-to-Value (LTV) for property credit/financing, down payments for motor vehicles, and the Capital Conservation Buffer (CCoB), which is calculated as the weighted average of the buffers in effect in the jurisdictions to which banks have a credit exposure.

The capital conservation buffer, or CCB, ensures that banks accumulate capital reserves outside of stressful periods, which they can use to absorb losses during those times. The requirement is predicated on straightforward capital conservation guidelines meant to prevent minimum capital requirements from being broken. The bank must maintain a minimum Common Equity Tier I (CET1) of 5.5% (8% including CCB) and a minimum Capital to Risk-Weighted Assets Ratio (CRAR) of 9% (11.5% including Capital Conservation Buffer (CCB)). Basel III standards were gradually adopted beginning April 1, 2013. The concept of CCB was first presented in the International Basel III standards. The idea gained prominence after the 2008 financial crisis when big banks saw their capital rapidly depleting due to systemic stress. The Capital Conservation Buffer aims to protect against losses on high-risk investments. The purpose of the countercyclical capital buffer, which must be met, is to shield banks that only deal with the joint sector from potential equity losses. Banks that fail to maintain the countercyclical capital buffer that the capital conservation buffer requirement mandates will be subject to capital payment limits when dividends and share buybacks increase rapidly, thereby depleting the buffer and bonuses.

In addition to the statutory minimum capital requirement of 9%, banks must maintain a capital conservation buffer of 2.5%, made up of Common Equity Tier 1 capital. If the capital level drops below the designated level, banks should not distribute capital – that is, pay dividends or incentives in any manner. But, if losses eat away at their capital conservation buffer, they can still operate normally. As a result, the restrictions only apply to dividend distribution and have nothing to do with how banks operate. Banks should maintain capital buffers above the statutory minimum in good times or outside of stressful periods, which they can use to absorb losses during stressful times. Only in cases of systemic stress on the bank may CCB be withdrawn. Only when the bank has a clear plan to restore capital through internal capital accruals and by lowering the discretionary earnings distribution will drawdowns from the CCB be permitted.

A version of Basel II, known as Basel III, incorporates prudential measures to prevent a banking crisis. Similar to Basel II, Basel III is composed of three main pillars: (1) strengthening banks' capacity to withstand shocks brought on by financial and economic forces, regardless of where they originate; (2) enhancing bank governance and risk management; and (3) boosting transparency and disclosure of banking data. The first pillar is the capital adequacy ratio, or CAR, which better represents and can foresee the numerous risks banks face. Credit risk, operational risk, market risk, and capital sufficiency can all be undermined by various threats. Risk management must become ingrained in the culture of the banking industry. Two primary elements might impact a bank's capital adequacy ratio (CAR): the quantity of capital it possesses and the quantity of risk-weighted assets (RWA) it owns. This is so because the weighted capital asset ratio (RWA) is the foundation for the capital ratio computation. Maintaining capital adequacy helps keep outside parties satisfied with the company's performance and ability to absorb losses. Anisa and Suryandari (2021) assert that a company's increased CAR level can draw in investors, thus impacting demand and driving up prices, ultimately increasing the company's worth.

Perhaps the most popular metric for evaluating credit risk and financial leverage is the loan-to-value (LTV) ratio, computed as the loan amount as a percentage of the transaction price of the collateralized property. Credit risk rises as the LTV ratio increases since the borrower's newly acquired property is a security for the loan. Practical credit risk assessment and control heavily depend on the LTV ratio to accurately reflect genuine financial leverage because of the narrow margin that separates the transaction price from the total amount of debt. In particular, the premise that the transaction price is an objective indicator of a property's collateral value is necessary for the validity of the LTV ratio as a credit risk indicator.

This study uses several macroprudential policy instruments, including CAR, CCoB, and LTV, adopted from other research [Bruno et al., 2017; Cantu et al., 2020; Wijayanti et al., 2020]. Banks must maintain CAR as the minimum capital requirement to cover possible losses due to banking activities. CCoB serves as a capital cushion for banks by supporting the risk of banking assets. CCoB is an additional layer of capital policy used when losses occur. The level of the CCoB ratio applied is Tier 1-based and is constantly recalculated for each capital distribution calculation. CCoB helps banks continue running their businesses when capital constraints and limitations occur. As a macroprudential policy, LTV controls the growth of property loans and the various risks that may arise.

The existence of credit risk disrupts the smooth operation of banks, especially on the liquidity side, which is the primary source of banking as an intermediary institution for creditors and debtors. NPL can be used to measure credit risk that can affect the country's banking system [Alshebami et al., 2020]. A high NPL value indicates that the bank is in an unhealthy condition. The level of NPLs owned by banks is caused by internal factors, namely the management system, and external factors, namely the decline in economic performance, so customers find it challenging to fulfill their installment obligations. Weak credit procedures, high markup spreads, poor lending principles, and a lack of supervision from policymakers can cause high NPL rates.

Nakatami (2020) investigated the impact of macroprudential policy using the LTV instrument on the banking crisis, especially credit risk, in 2008. The author concluded that macroprudential policy effectively reduced the banking crisis stemming from credit growth and LTV. A significant statistical coefficient value indicates this. However, LTV policy is not effective if the country does not have an inflation target and a capital control framework. Mauleman et al. (2020) analyzed the interaction between macroprudential policy and systemic banking risk and showed that macroprudential policy affects banking systemic risk conditions, especially bank credit and property sales.

Quint and Rabanal (2014) found that implementing macroprudential policies helped banks perform better and reduce losses through countercyclical spread loans. Cantu et al. (2020) analyzed the effectiveness of macroprudential policies in the Asia-Pacific region and found that they play an active role in dampening credit growth in the household sector. Implementing macroprudential policies in the Asia-Pacific region based on bank size and liquidity conditions can reduce credit growth. Vatansever and Hepsen (2013) examined the determinants of bank NPLs in Turkey. With profit-based banking activities, encouraging the availability of capital to be used for business expansion encourages more unusual cyclical behavior, which allows banking losses to be more significant in periods of economic depreciation. Furthermore, Dadashova et al. (2018) researched the application of CCoB to macroeconomic conditions in the banking sector in Ukraine. They concluded that the policy could increase banking resilience, especially in financial conditions facing global uncertainty.

Hallissey et al. (2014) researched macroprudential instruments and credit risk and found that limiting property lending through loan-to-value (LTV) can reduce the risk of default. Banks will be encouraged to tighten or reduce LTV provisions by considering the risk of loss on default, which will ultimately dampen the mortgage demand cycle and build resilience in the banking sector and households. CCoB affects bank risk-taking in Canada, as Guidara et al. (2013) found still the compatibility of micro and macroprudential «through-the-cycle» approaches to capital adequacy may explain why Canada performed better during the 2008 global financial crisis. LTV can slow down credit growth but cannot improve procyclicality in Indonesia, but CCoB and RR can [Dana, 2018]. However, LTV has a significant and long-lasting impact on household debt and actual house values in Korea [Jung, Kim, Yang, 2017].

Zhang et al. (2018) have conducted research using the assessment results of individual macroprudential tools and group macroprudential tools, namely CCoB, RR, and LTV, to determine whether bank risk-taking behavior will be reduced by strengthening macroprudential supervision. In addition, the reserve requirement has the most enormous impact, followed by LTV and CCoB. Macroprudential policy is usually implemented according to the credit cycle; we specifically consider the credit cycle's function in the transmission mechanism of macroprudential policy [Bruno et al., 2017]. The credit boom increased the benefits of macroprudential policy on bank risk-taking, proving that macroprudential policy was implemented in China at the right time. To support the sustainability of the banking system, the central bank should further enhance the evaluation and supervision of macroprudential measures and implement appropriate mechanisms for different purposes and times, especially during the credit boom period. In addition, this study found that the Chinese commercial banking system is run with modest NPLs, indicating that most Chinese commercial banks have a low-risk appetite.

Our study also alludes to the growing literature on the effectiveness of the capital adequacy ratio as a capital requirement for non-performing loans. The literature shows that microprudential regulation does not significantly affect credit risk using NPL. Mustafa and Mirza (2022) and Barus and Erick (2016) state that CAR does not affect non-performing loans.

H1. The Capital Adequacy Ratio does not affect credit risk.

CCoB serves as a capital buffer for banks in supporting the risk of banking assets, and literature shows that a Capital Conservation Buffer continuously increasing bank capital buffer will not continuously reduce bank risk-taking [Neef et al., 2023].

H2. Capital conservation buffer does not affect credit risk.

Loan-to-value (LTV) is the ratio between the value of loans/financing granted by traditional and Islamic commercial banks and the value of the collateral in the form of property at the time of granting the loan/financing, based on the results of real estate valuation. Bian et al. (2018), Sasikirono et al. (2019), Behncke (2023), and Kim and Oh (2021) concluded that Loan-To-Value (LTV) affects Non-Performing Loans.

H3. Loan-to-value affects credit risk.

The impact of corruption on banks' credit risk, especially in less developed nations where banks are grappling with large non-performing loans (NPLs) and widespread corruption. Corruption has a significant effect on credit risk [Agarwal et al., 2020; Bahoo et al., 2019; Djalilov, Piesse, 2019; Houston, Jiang, Lin, Ma, 2014; Liu, Li, Guo, 2020; Son et al., 2020; Toader et al., 2017; Weill, 2011].

H4. Corruption affects credit risk.

The central bank uses interest rate increases as one tool to manage inflation. Excessive interest rates will limit demand, which will lower inflation. Low interest rates promote inflation and economic expansion. Interest rates have no impact on the likelihood of problematic financing, which suggests that raising interest rates is not the best way to increase lending.

H5. Interest rate does not affect credit risk.

Methodology

Data

This study uses cross-sectional data from 11 central bank countries in the Asia Pacific and nine time-series data from 2015 to 2023. The year selection is adjusted to the availability of macroprudential data at each bank. Meanwhile, the data collected is annual secondary data obtained from World Bank data, financial statements and reports from the Central Bank, and macroprudential regulations.

Variable Measurement

The prudential macroprudential policy stipulates various provisions regarding bank capital and liquidity, which can be used as buffers to absorb losses. This study uses CAR as a microprudential tool indicator to control banking minimum capital requirements. CCoB as macroprudential policy indicators based on the provisions set by the Basel Committee on Banking Supervision, where capital provisions are standardized for banks to determine how much liquid capital to maintain from their assets [Imani, Pracoyo, 2018; Junos et al., 2021]. LTV has been established as a macroprudential policy instrument based on lending. The LTV policy was the regulator's response to the 2008 crisis, caused by the explosion of property prices in America that brought down the entire economy and banking activities. LTV, as a credit policy, is mainly aimed at property loans. Loan to value in each country varies depending on the condition of public interest in property loans [Bian et al., 2018; Morgan et al., 2015; Sasikirono et al., 2019]. This study uses NPL as a measurement indicator or proxy for credit risk based on the risk-based bank rating method, which describes the value of credit failures caused by the inability of debtors to fulfill payment obligations following predetermined conditions. Therefore, a high NPL level indicates a considerable risk value and poor banking conditions [Imani, Pracoyo, 2018]. This study also looks at institutional variables, such as corruption, which is gauged by the Corruption Perception Index (CPI), which shows that the public disapproves of the government's efforts to combat corruption and vice versa. This study also included macroeconomic indicators as the control variable, which is the interest rate. Interest rates on loans to the finance and credit sectors fluctuate in response to the goals and requirements of monetary policy. The risk the relevant bank assumes increases with the loan amount extended. A bank's capacity to insure against the risk of debtor credit failure can be determined using its non-performing loan (NPL) ratio. The degree of credit risk increases with the NPL level. One barrier to banks extending credit is the amount of non-performing loans (NPLs) [Rahmaningtyas, 2022].

Empirical model

The popularity of the time series model and the development of panel data methods have led to the idea of creating a panel data model by using the dependent variable's lag as a regressor in the regression. By introducing the dependent variable's lag as a regressor into the regression, the results will be biased and inconsistent when calculated using static panel data due to endogeneity issues, either the Random Effect Model (REM) or Fixed Effect Model (FEM) methodology. The Generalized Method of Moments (GMM), often known as the Method of Moments methodology, can be used to tackle this problem.

The System Generalized Method of Moments (SYSGMM) update by Blundell and Bond in 1998 and the First Differences Generalized Method of Moments (FDGMM) by Arellano and Bond are the two estimate techniques that are frequently employed in the GMM method. An effect from the previous year is shown as the explanatory variable in this dynamic panel model. Arellano and Bond developed the FDGMM technique in 1991; this model adds a lag for the dependent variable. When applying this method over a brief or small period of years, consistent estimators will be produced in cases where the number of persons is unlimited. Unlimited number of people during a short or limited period of years. The estimator that is employed may be more biased if there is a connection between the delays or if there are only a few periods. There's a chance the estimator was more biased. One can identify data bias by contrasting the PLS model with the coefficients of the lag variables, models FEM and FDGMM.

This study used the generalized method of moment (GMM) as a data analysis method and the Eview 9 software by Robert Hall. The First Difference Generalized Method of Moment (GMM) is a general technique for estimating parameters in statistical models. GMM is frequently applied to data with little distributional information. A time series that is tagged with forward values by one or more time steps. Using lag 1 in the GMM approach can result in estimators that are consistent. The research model for the influence of macroprudential policy on NPLs can be written as follows:

$$NPL_{i,t} = \beta_1 CAR_{i,t} + \beta_2 LTV_{i,t} + \beta_3 CCoB_{i,t} + \beta_4 Corr + \beta_5 LIR + \varepsilon.$$

NPLs are used as a proxy for measuring credit risk. CAR is used as a microprudential tool. LTV and CCoB are macroprudential policy instruments that maintain banking liquidity conditions and absorb possible risks. This study included the macroeconomic indicator of lending interest rate as the control variable. This study also included institutional variables, namely corruption. GMM is an estimator tool for the dynamic panel research model that includes the condition of the moment, where the moment condition is an expected value of the model parameter. GMM finds the parameters get as close as possible to the weighted sample moment condition [Kim, 2020].

Results and Discussions

Base result

Unit root test for time series data to detect unit root components and random walk trends in time series data. This study uses Levin, Lin, and Chut, and Im, Pesaran, and Shin's W-stat methods to analyze the unit root condition of the data. The stationarity condition of the data is compared with the error degree of 5% (= 0.05).

Suppose the probability value of the variable is above the 5% error degree. Then, the related variable has a unit root, or the data is not stationary. Conversely, if the probability result of the variable is below the 5% error degree, then there is no unit root in the related variable, or the data is stationary. In the first testing stage, all variables are tested at level 1. If one variable is not stationary, the data is tested again in the first or second condition until all variables are stationary. Table 1 describes the unit root test results using Levin, Lin, Chut, Im, Pesaran, and PP Fisher.

Table 1.

Unit root-level test results

Variable	Level		
	Probability Levin, Lin, and Chut	Probability PP-Fisher Chi-Square	Expl
NPL	0.0000	0.0001	Stationary
CAR	0.0002	0.0000	Stationary
CCoB	0.0000	0.0000	Stationary
LTV	0.1269	0.4881	Non-stationary
Corruption	0.0001	0.0168	Stationary
LIR	0.1721	0.1507	Non-stationary

Based on the results of Table 1, the Capital Conservation Buffer (CCoB) and Lending Interest Rate (LIR) were in a non-stationary condition at the level of the research variable. The research was continued by testing the unit root test at different levels. Table 2 describes the unit root test results through the first different levels.

In Table 3, it can be seen that all research variables are stationary with a probability of 0.00. At a significance level of 5%, all variables have passed the unit root problem. The Pedroni test (Pedroni Residual) Cointegration Test is used in this study's panel data cointegration testing, which employs a residual technique. Comparing values is one method of determining whether or not cointegrated data is present in this study. Critical values for the Pedroni test probability are 1%, 5%, and 10%. The data is cointegrated if the Pedroni test probability exceeds the critical threshold. Table 3 describes the cointegration test results.

Table 2.

Unit root 1st different test result

Variable	Level		
	Probability Levin, Lin, and Chut	Probability PP-Fisher Chi-Square	Expl
NPL	0.0000	0.0000	Stationary
CAR	0.0000	0.0000	Stationary
CCoB	0.0000	0.0000	Stationary
LTV	0.0117	0.0307	Stationary
Corruption	0.0000	0.0008	Stationary
LIR	0.0000	0.0000	Stationary

Table 3.

Cointegration test result

Method	Statistics	Probability
Alternative Hypothesis (Ha): Common AR Coefsion (within- dimension)		
Panel PP-Statistic	–12.3382	0.0000
	–3.01908	0.0013
(Ha): Individual AR Coefsion (between dimensions)		
Group PP-Statistic	–7.63521	0.0000

It is known that all indicators of the cointegration evaluation criteria are above the significance level of 5% and 10% based on the table of Pedroni test findings above. This issue demonstrates that the research variables are cointegrated, i.e., there are long-term linkages between the variables that would imply equality and balance. According to research by Saeed Ass Khan and Abbas (2016), data is considered cointegrated when it can be used to conclude. The data is deemed cointegrated if the majority of results (more than 50% of the criteria) from the Pedroni test criteria are significant (less than the critical value). After all the conditions of the research variables were stationary, the generalized method of moment analysis was carried out to see the effect and effectiveness of macroprudential policy in controlling credit risk.

GMM estimator result and discussion

This study used the GMM analysis method to analyze the effect and effectiveness of capital and credit-based prudential policy instruments in controlling the NPL level in the Asia-Pacific Region. Table 4 presents the GMM test results.

Table 4.

GMM test results

Variable	Coefficient	Std. Error	t-Statistic	Prob.
NPL(-1)	0.493936	0.06417	7.697298	0.0000
LTV	0.047493	0.025537	1.859741	0.0661*
LIR	0.383542	0.360181	1.06486	0.2897
CAR	1.040706	1.046395	0.994563	0.3225
COC	8.27E-05	3.28E-05	2.523687	0.0133**
CCoB	3.305189	2.382051	1.387539	0.1686

Note: *– significant in 10%; ** – significant in 5%.

LTV is a macroprudential policy launched in response to the 2008 crisis, aiming to control credit, especially property loans. Based on the results of LTV testing, the value of the NPL variable has a positive and significant influence. The increase in LTV will increase the condition of Non-Performing Loans. The borrower's property acquisition acts as collateral for the loan. Thus, as the LTV ratio rises, so does the credit risk. The likelihood of negative equity – when the property's value falls below the amount owed on the mortgage – increases with a greater LTV ratio. Defaults thus become more probable. Corruption has a positive and significant influence on the value of the NPL variable. The increase in corruption will increase the condition of Non-Performing Loans. The impact of corruption on banks' credit risk, especially in less developed nations where banks are grappling with large non-performing loans (NPLs) and widespread corruption throughout society.

The Arellano Bond test (AB test) and the Sargan test are further tests that must be performed to test the model specifications after the generalized technique of the moment test. The Arellano Bond Test, often known as the AB Test, is used to assess the consistency of the estimated findings before employing the Arellano-Bond statistic to test for autocorrelation. This test's AR(2) probability value must be greater than the 5% (0.05) significance level. To find the number of instrument variables that are more than the number of estimated parameters, the Sargan test is used (overidentifying constraints condition). The significance value of 5% (0.05) must be exceeded by the probability value (J-Statistic).

Table 5.

Arellano-Bond test results

Test order	m-Statistic	Rho	SE(rho)	Prob.
AR(1)	–0.88574	–13.9165	15.71182	0.3758
AR(2)	–0.01287	–0.67837	52.70533	0.9897

Based on the results in Table 5 above, the AR(2) probability value is above 5% (0.05) significance, which is 0.9897, indicating that the estimation results are consistent.

Table 6.

Sargan test results

Mean dependent var	0.010137	S.D. dependent var	0.665711
S.E. of regression	0.8135	Sum squared resid	61.54582
J-statistic	4.398645	Instrument rank	11
Prob(J-statistic)	0.493558		

Based on the results in Table 6 above, the probability (J-statistic) value is above 5% (0.05) significance, which is 0.493558, indicating that the instrument is valid.

Discussion

A greater LTV ratio indicates that, in the event of failure, the collateralized property is less likely to sell for enough money at a foreclosure auction to pay off the outstanding loan total, past-due payments, and other foreclosure-related expenses paid by the lender. Loss as a result of default will thus be larger. Real estate is considered more «collateralizable» than most other asset types, and real estate ownership is frequently accompanied by relatively high LTV ratios [Bian et al., 2018; Ranisavljević, Hadžić, 2016].

The application of LTV has a stronger interest in property loans and affects the risk of banking property loans. LTV policies in each country will differ depending on each property's risks. When tightening the LTV, the bank will provide a lower property valuation so that the debtor provides higher installments and down payments on the property to meet the required criteria. Indirectly, this determination will improve credit quality and suppress NPLs [Kinghan et al., 2022; Sasikirono et al., 2019]. For countries with high interest in property loans, tightening loans to value will be effective when balanced with high economic activity to suppress banking NPLs. The results of this study were in line with the credit default theory, which explains that banks need control tools to understand risk and manage risk levels.

The greatest level accessible in the market at the time of origination significantly impacts the evolution of originating LTV. Recent initial LTV levels have decreased considerably, primarily due to a change from values above 90% to those between 80% and 90%. The comparatively higher percentage of first-time purchasers after 2010 may help to explain this since many mortgage holders suffered negative equity after that year, raising the cost of changing homes. Although there is no universal agreement on what defines an LTV that is «too high» at the time of loan origination (in actuality, this will depend on the specific borrower's circumstances) [Hallissey et al., 2014], the tightening of LTV will increase the level of risk faced by banks because property values will decrease. In this condition, the level of banking losses is even more significant, especially in property assets that are the bank's collateral. The results of this study contradicted the research of Yao and Lu (2019), which concluded that there was a negative relationship between LTV and NPL. Table 6 indicates that the magnitude of the previous year's NPL had a positive and significant influence, indicating a lack of proper banking policies or regulations addressing the previous period's NPLs.

Given that corruption contributes to social and economic issues, social scientists have found corruption to be a significant concern. The government enacts regulations to combat corruption, but creating and enforcing these rules has an accounting cost. Results from such an effort must be positive in the form of socioeconomic progress. Huang (2016) finds that corruption significantly influences Asia-Pacific's economic development. Studies on corruption and economic development have changed over time, with the most recent research focusing on how corruption affects financial development. Economic development is significantly influenced by financial development. Bolarinwa and Soetan (2019) report conflicting results about the effect of corruption on bank profitability in developing nations; however, they affirm that corruption benefits wealthy nations.

Tabish and Jha (2012) examine the connection between fair punishment, standard operating procedures, and corruption-free indicators. The study concludes that corruption decreased as a result of just punishment. One of the important factors contributing to decreased corruption in a community is justice without discretion. Maintaining one's distance from corruption requires adhering to ethical principles. According to Hassan et al. (2021), the public investment sectors of infrastructure and capital-intensive projects are more corrupt than the health and education sectors. Rent-seeking is more prevalent in the defense and infrastructure sectors than in the health and education sectors. To analyze the misallocation of NPL and the connection between these institutions' administrative structures and corruption, it is crucial to examine the organization of financial institutions, particularly the banking industry. After examining the stability of Islamic banks, Bougatef (2015) finds a significant correlation between NPL and corruption. The economy's financial sector ensures that financial resources are allocated effectively within the financial system and that financial institutions meet the needs of both domestic and international investors. The degree of confidence that foreign investors have has disastrous effects, particularly for developing economies where non-performing loans would become a significant source of gambling. The number of non-performing loans can increase or decrease depending on several factors, including information sharing, the bank's ownership structure, corruption, and macroeconomic and bank-specific variables. Among these, corruption is a key element affecting the non-performing loans. Corruption is a significant factor in rising non-performing loans in emerging nations.

Based on this research, it is evident that managing credit risk in a nation involves the banking industry and the collaboration of the government's fiscal sector. The study is significantly impacted by two variables: loan to value and corruption. As a macroprudential policy, loan-to-value is commonly used to evaluate credit risk and financial leverage. It is computed by dividing the loan amount by the collateralized property's transaction price. The borrower's recently purchased home serves as the security for the loan. Thus, as the LTV ratio rises, so does the credit risk. Given the narrow buffer that exists between the transaction price and the total loan amount, the loan-to-value ratio (LTV) is a crucial instrument for appropriately assessing and managing credit risk. The makeup of the administration largely determines the level of bad debts and corruption. Corruption leads to a rise in bank lending, even though potential profits are lower in industrialized and developing nations. Banks operating in severely corrupt situations utilize income-smoothing tactics to control their positive returns. These findings call for more research on the connection between credit risk and insider trading in the banking sector.

Conclusion

Macroprudential policy instruments effectively manage credit risk in banks in the Asia Pacific Region. Based on the study results, it can be concluded that the loan-to-value instruments positively and significantly impact non-performing loans. Increasing the level of loan-to-value will increase public interest in owning property loans, and it will also simultaneously increase risk. The increase in loan value will reduce property prices and increase demand for property. The uncontrolled growth of property loans will cause the bubble price of houses to explode and increase the risk of default. Corruption instruments positively and significantly impact non-performing loans. Increasing corruption problems in a country will increase credit risk. High levels of corruption will reduce public trust in institutions. On the other hand, the government must also maintain the condition of corruption problems in each country because the existence of corruption will hinder economic growth. For the banking industry, which is based on trust in banking consumers, corruption problems will affect banks' trust levels.

This study recommends a need for proactive prudential policy, considering that currently, almost all countries face the COVID-19 pandemic. This pandemic has reduced banking performance, and the risk factors faced are getting bigger, not only on the credit side but also on banking liquidity conditions. In addition, the use of macroprudential policy can accelerate the transmission process in the monetary sector. Banks, especially the central bank that carries out two policies, namely prudential and monetary, need to align the objectives of the two policies in facing economic uncertainty in the era of the COVID-19 pandemic. An increase in the precautionary principle is also needed to deal with various speculations due to procyclical action.

The implementation of macroprudential policy is not only aimed at absorbing losses experienced by banks but also aimed as a form of banking defense and determining the health condition of banks. Banking performance based on trust needs healthy banking conditions regarding capital, liquidity, operations, and internals. The capital adequacy ratio, as an indicator of banking capital, is intended as a capital reserve to cover and absorb losses. On the other hand, this determination maintains banking liquidity conditions by forcing banks not to use all capital for business expansion. Still, it can also be used to finance sound risk management.

The loan-to-value policy will work based on economic conditions and banking liquidity. In tightening LTV, banks give lower property valuations and apply higher installments. With this policy, the down payment on the property will be higher, and the debtor can only fulfill it. This condition will differ from credit easing, reducing property prices, and attracting public interest in housing credit loans. With the aim of money circulation, it is expected to be stable in its original condition. Stable money circulation conditions in all sectors will be positive and run as expected by regulators. Based on the above explanation, the emergence of macroprudential policy has positively influenced the global banking system. Moreover, it can be seen that macroprudential policy has become very effective in absorbing losses. It is also coupled with other policy alignments to absorb and reduce sources of risk.

This research allows the government, politicians, and bank regulators to use this study to help them create pertinent regulations. Bank regulators should provide an effective system for monitoring loan activity to lower credit risk and boost public confidence in banking. Additionally, the reason for the increase in banking performance can be the stability of the financial banking system. By doing this, prudential policy – particularly Basel III – will be established, enabling banks to sustain liquidity conditions despite fluctuations in the overall economy. Government officials must also confront the corruption problem and create anti-corruption measures to win over investors. Banks will be able to draw in more deposits in this way, which will eventually help the banking sector grow. One policy conclusion of funding risk is that macroprudential policy should be under control, and all financial authorities should work together to enhance their effective lending for client deposits to provide stable conditions.

* *
*

References

Aiyar S., Calomiris C.W., Wieladek T. (2014) Does Macro-Prudential Regulation Leak? Evidence from a UK Policy Experiment. *Journal of Money, Credit and Banking*, 46(SUPPL.1), pp. 181–214. https://doi.org/10.1111/jmcb.12086

Alper K., Binici M., Demiralp S., Kara H., Özlü P. (2018) Reserve Requirements, Liquidity Risk, and Bank Lending Behavior. *Journal of Money, Credit and Banking*, 50, 4, pp. 817–827. https://doi.org/10.1111/jmcb.12475

Alshebami A., Thomran M., Hassan Mohammad Adam M., Saleh Alshebmi A., Mustafa A.M., Elsayed Abdelmaksoud Fathelbab O. (2020) Assessing the Non-Performing Loans and their Effect on Banks Profitability: Empirical Evidence from the Saudi Arabia Banking Sector. *International Journal of Innovation, Creativity and Change. www.ijicc.net* (Vol. 11). Available at: www.ijicc.net

Altunbas Y., Binici M., Gambacorta L. (2018) Macroprudential Policy and Bank Risk. *Journal of International Money and Finance*, 81, pp. 203–220. https://doi.org/10.1016/j.jimonfin.2017.11.012

Andrle M., Tomsik V., Vlcek J. (2017) *Banks' Adjustment to Basel III Reform: A Bank-Level Perspective for Emerging Europe.* IMF Working Paper 2017/024.

Anisa H.V., Suryandari D. (2021) The Effect of NPL, GCG Self Assessment, ROA, and CAR on Firm Value (Empirical Study on Commercial Banks listed on the Indonesia Stock Exchange). *Accounting and Finance Studies,* 1, 2, pp. 111–124. Doi: 10.47153/afs12.1352021

Apergis N. (2017) Monetary Policy and Macroprudential Policy: New Evidence from a World Panel of Countries. *Oxford Bulletin of Economics and Statistics*, 79, 3, pp. 395–410. https://doi.org/10.1111/obes.12182

Bahoo S., Alon I., Paltrinieri A. (2019) Corruption in International Business: A Review and Research Agenda*. International Business Review, 101660.* Doi: 10.1016/j.ibusrev.2019.10166

Baker A. (2013) The Gradual Transformation? The Incremental Dynamics of Macroprudential Regulation. *Regulation and Governance*, 7, 4, pp. 417–434. https://doi.org/10.1111/rego.12022

Baker A., Widmaier W. (2015) Macroprudential Ideas and Contested Social Purpose: A Response to Terrence Casey. *British Journal of Politics and International Relations*, 17, 2, pp. 371–380. https://doi.org/10.1111/1467-856X.12058

Barus A.C., Erick (2016) Analysis of Factors Influencing Non-Performing Loans at Commercial Banks in Indonesia*. Journal of Microskil Economic Wira*, 6, 2, pp. 113–122. Available at: https://www.mikroskil.ac.id/ejurnal/index.php/jwem/article/view/325

Behncke S. (2023) Effects of Macroprudential Policies on Bank Lending and Credit Risks. *Journal of Finanancial Service Research*, 63, pp. 175–199.

Benigno G., Chen H., Otrok C., Rebucci A., Young E.R. (2013) Financial Crises and Macro-Prudential Policies. *Journal of International Economics*, 89, 2, pp. 453–470. https://doi.org/10.1016/j.jinteco.2012.06.002

Berger A.N., Bouwman C.H.S. (2013) How Does Capital Affect Bank Performance during Financial Crisesa. *Journal of Financial Economics*, 109, 1, pp. 146–176. https://doi.org/10.1016/j.jfineco.2013.02.008

Bian X., Lin Z., Liu Y. (2018) House Price, Loan-to-Value Ratio and Credit Risk. *Journal of Banking and Finance*, 92, pp. 1–12. https://doi.org/10.1016/j.jbankfin.2018.04.006

Bolarinwa S.T., Soetan F. (2019) The Effect of Corruption on Bank Profitability. *Journal of Financial Crime,* 26, 3, pp. 753–773. https://doi.org/10.1108/JFC-09-2018-0102

Bougatef K. (2015) The Impact of Corruption on the Soundness of Islamic Banks*. Borsa Istanbul Review,* 15, 4, pp. 283–295*.* Doi:10.1016/j.bir.2015.08.001

Bruno V., Shim I., Shin H.S. (2017) Comparative Assessment of Macroprudential Policies. *Journal of Financial Stability*, 28, pp. 183–202. https://doi.org/10.1016/j.jfs.2016.04.001

Bulut U. (2015) The Interest Rate Corridor as a Macroprudential Tool to Mitigate Rapid Growth in Credits: Evidence from Turkey. *Theoretical and Applied Economics*, XXII(4), pp. 133–144. Available at: https://www.researchgate.net/publication/288825691

Bumin M., Taşkın D., Dilvin Taşkın F. (2015) *The Impact of Macro-Prudential Measures on Retail Loans: Evidence from Turkey*. Available at: https://www.researchgate.net/publication/289953323

Buncic D., Melecky M. (2013) Macroprudential Stress Testing of Credit Risk: A Practical Approach for Policy Makers. *Journal of Financial Stability*, 9, 3, pp. 347–370. https://doi.org/10.1016/j.jfs.2012.11.003

Calmès C., Théoret R. (2014) Bank Systemic Risk and Macroeconomic Shocks: Canadian and U.S. Evidence*. Journal of Banking & Finance,* 40, pp. 388–402*.* Doi: 10.1016/j.jbankfin.2013.11.03

Cantu C., Gambarcota L., Shim I. (2020) *How Effective Are Macroprudential Policies in Asiapacific? Evidence from a Meta-Analysis*. BIS Paper 110b.

Casimano T., Hakura D. (2011) *Bank Behavior in Response to Basel III: A Cross-Country Analysis.* IMF Working Paper 119.

Cerutti E., Claessens S., Laeven L. (2016) The Use and Effectiveness of Macroprudential Policies: New Evidence. *Journal of Financial Stability*.

Cerutti Eugenio, Claessens S., Laeven L. (2015) *The Use and Effectiveness of Macroprudential Policies: New Evidence*. IMF Working Paper, 61, 15.

Chaibi H., Ftiti Z. (2015) Credit Risk Determinants: Evidence from a Cross-Country Study. *Research in International Business and Finance*, 33, pp. 1–16. https://doi.org/10.1016/j.ribaf.2014.06.001

Chen M., Jeon B.N., Wang R., Wu J. (2015) Corruption and Bank Risk-Taking: Evidence from Emerging Economies*. Emerging Markets Review,* 24, pp. 122–148*.* Doi:10.1016/j.ememar.2015.05.009

Claessens S., Ghosh S.R., Mihet R. (2013) Macro-Prudential Policies to Mitigate Financial System Vulnerabilities. *Journal of International Money and Finance*, 39, pp. 153–185. https://doi.org/10.1016/j.jimonfin.2013.06.023

Cordella T., Pienknagura S. (2020) Macroprudential Policies from a Microprudential Angle: A Note*. Latin American Journal of Central Banking,* 1(1–4), 100005*.* Doi:10.1016/j.latcb.2020.100005

Dadashova P., Jonsson M., Onyshchenko H. (2018) Macroeconomic Effects of Introducing a Capital Conservation Buffer in the Ukrainian Banking Sector. *Visnyk of the National Bank of Ukraine*, 243, pp. 33–42. https://doi.org/10.26531/vnbu2018.243.031

Dahl K., Licker R., Abatzoglou J.T., Declet-Barreto J. (2019) Increased Frequency of and Population Exposure to Extreme Heat Index Days in the United States during the 21st Century. *Environmental Research Communications*, 1, 7. https://doi.org/10.1088/2515-7620/ab27cf

Dana B.S. (2018) Evaluation of Macroprudential Policy on Credit Growth in Indonesia: Credit Registry Data Approach. *Etikonomi*, 17, 2, pp. 199–212. https://doi.org/10.15408/etk.v17i2.7324

Djalilov K., Piesse J. (2019) Bank Regulation and Efficiency: Evidence from Transition Countries*. International Review of Economics & Finance.* Doi: 10.1016/j.iref.2019.07.003

Dumičić M. (2017) A Brief Introduction to the World of Macroprudential Policy. *Journal of Central Banking Theory and Practice*, 6, 1, pp. 87–109. https://doi.org/10.1515/jcbtp-2017-0005

Ebrahimi Kahou M., Lehar A. (2017) Macroprudential Policy: A Review. *Journal of Financial Stability*, 29, pp. 92–105. Doi: 10.1016/j.jfs.2016.12.005

Fajriani N., Sudarmawan B.N. (2022) Microprudential Policy in Maintaining Bank Stability. *Journal of Economics, Finance and Management Studies*, 05(06). https://doi.org/10.47191/jefms/v5-i6-18

Fendoğlu S. (2017) Credit Cycles and Capital Flows: Effectiveness of the Macroprudential Policy Framework in Emerging Market Economies. *Journal of Banking and Finance*, 79, pp. 110–128. https://doi.org/10.1016/j.jbankfin.2017.03.008

Freixas X., Laeven L., Peydro J.L. (2015) *Systemic Risk, Crises, and Macroprudential Regulation*. Boston: MIT Press.

Freixas X., Perez-Reyna D. (2021) Optimal Macroprudential Policy and Rational Bubbles. *Journal of Financial Intermediation*, 46. https://doi.org/10.1016/j.jfi.2021.100908

Gaganis C., Lozano-Vivas A., Papadimitri P. (2019) Macroprudential Policies, Corporate Governance and Bank Risk: Cross-Country Evidence. *Journal of Economic Behavior and Organization*, November. https://doi.org/10.1016/j.jebo.2019.11.004

Galati G., Moessner R. (2013) Macroprudential Policy – A Literature Review. *Journal of Economic Surveys*, December. https://doi.org/10.1111/j.1467-6419.2012.00729.x

Guidara A., Lai V.S., Soumaré I., Tchana F.T. (2013) *Banks' Capital Buffer, Risk and Performance in the Canadian Banking System: Impact of Business Cycles and Regulatory Changes*. MPRA Paper. Germany: University Library of Munich.

Hallissey N., Kelly R., O'Malley T. (2014) Macro-Prudential Tools and Credit Risk of Property Lending at Irish banks. *Economic Letter Series*, 10, pp. 1–8.

Hassan M.K., Hasan R., Miah M.D., Ashfaq M. (2021) Corruption and Bank Efficiency: Expanding the «Sand or Grease The Wheel Hypothesis» for the Gulf Cooperation Council. *Journal of Public Affairs.* Doi: 10.1002/pa.2737

Houston J.F., Jiang L., Lin C., Ma Y. (2014) Political Connections and the Cost of Bank Loans*. Journal of Accounting Research,* 52, 1, pp. 193–243*.* Doi:10.1111/1475-679x.12038

Huang C.-J. (2016) Is Corruption Bad for Economic Growth? Evidence from Asia-Pacific Countries*. The North American Journal of Economics and Finance*, 35, pp. 247–256. Doi: 10.1016/j.najef.2015.10.013

İlhan A., Özdemir M., Eryiğit K. (2021) Makro İhtiyati Politikaların Kredi Büyümesini Sınırlamadaki Rolü: Türkiye Örneği. *Journal of Yaşar University*, 16, 61, pp. 393–413. https://doi.org/10.19168/jyasar.821930

Imani A., Pracoyo A. (2018) Analysis of the Effect of Capital, Credit Risk, and Liquidity Risk on Profitability in Banks. *Journal Ilmu Manajemen & Ekonomika*, 10, 2, pp. 44–50.

Jiménez G., Ongena S., Peydró J., de España B., Adrian T., Albertazzi U., … Brunnermeier M. (2017) Macroprudential Policy, Countercyclical Bank Capital Buffers, and Credit Supply: Evidence from the Spanish Dynamic Provisioning Experiments. *Journal of Political Economy*, 125. Available at: http://www.journals.uchicago.edu/t-and-c

Jung Y., Kim S., Yang D.Y. (2017) Optimal Macroprudential Policies and House Prices in Korea. *Emering. Markets Finance Trade*, 53, pp. 2419–2439.

Junos S., Main M., Jalil S.A. (2021) Assessing the Impact of Macroeconomic Variables and Macroprudential Policy Elements on the Capital Adequacy Ratio. *International Journal of Academic Research in Business and Social Sciences*, 11, 8. https://doi.org/10.6007/ijarbss/v11-i8/10740

Kara H., Unalmis D., Binici M., Erol H., Özlü P., Ünalmış D. (2013) Interest Rate Corridor: A New Macroprudential Tool? *CBRT Research Notes in Economics*, *20*. https://doi.org/10.13140/RG.2.1.2729.9684

Khan M.A., Abbas F. (2016) The Dynamics of Electricity Demand in Pakistan: A panel Cointegration Analysis*. Renewable and Sustainable Energy Reviews,* 65, pp. 1159–1178*.* Doi: 10.1016/j.rser.2016.06.054

Kim D. (2020) An Alternative Two-Step Generalized Method of Moments Estimator Based on a Reduced Form Model*. Economics Letters,* 109, 184*.* Doi:10.1016/j.econlet.2020.109184

Kim M., Oh Y. (2021) Effects of Easing LTVDTI Regulations on the Debt Structure and Credit Risk of Borrowers. *KDI Journal of Economic Policy*, 43, 3, pp. 1–32.

Kim S., Mehrota A. (2019) *Examining Macroprudential Policy and its Macroeconomic Effects – Some New Evidence.* BIS Working Paper 825.

Kinghan C., McCarthy Y., O'TooleC. (2022) How Do Macroprudential Loan-to-Value Restrictions Impact First Time Home Buyers? A Quasi-Experimental Approach. *Journal of Banking and Finance*, 138. https://doi.org/10.1016/j.jbankfin.2019.105678

Kuttner K.N., Shim I. (2016) Can Non-Interest Rate Policies Stabilize Housing Markets? Evidence from a Panel of 57 Economies. *Journal of Financial Stability*, 26, pp. 31–44. https://doi.org/10.1016/j.jfs.2016.07.014

Liu P., Li H., Guo H. (2020) The Impact of Corruption on Firms' Access to Bank Loans: Evidence from China*. Economic Research-Ekonomska Istraživanja,* 33, 1, pp. 1963–1984*.* Doi: 10.1080/1331677x.2020.1768427

Lombardi D., Siklos P.L. (2016) Benchmarking Macroprudential Policies: An Initial Assessment. *Journal of Financial Stability*, 27, pp. 35–49. https://doi.org/10.1016/j.jfs.2016.08.007

Lopez C., Adams-Kane J., Saeidinezhad E., Wilhelmus J. (2018) *Macroprudential Policyand Financial Stability: Where Do We Stand?* Available at: https://ssrn.com/abstract=3154649

McKeever D. (2023) Microprudential Bank Capital Regulation in A Complex System. *Heliyon*, 9, *3*.

Meuleman E., Vander Vennet R. (2020) Macroprudential Policy and Bank Systemic Risk. *Journal of Financial Stability*, 47. https://doi.org/10.1016/j.jfs.2020.100724

Mustafa A., Mumtaz M.Z. (2022) Examining the Bank-Specific and Macro-Economic Factors that Influence Capital Adequacy in Pakistan. *International Journal of Management Research and Emerging Sciences*, 12, 2. https://doi.org/10.56536/ijmres.v12i2.221

Nakatami R. (2020) Macroprudential Policy and the Probability of a Banking Crisis. *Journal of Policy Modeling*, 6624, pp. 1–18.

Neef H.O., Schandlbauer A., Wittig C. (2023) Countercyclical Capital Buffers and Credit Supply: Evidence from the COVID-19 Crisis. *Journal of Banking & Finance,* 154. https://doi.org/10.1016/j.jbankfin.2023.106930

Okon C.E. (2022) Bank Valuation in Nigeria: Exploring the Role of Capital Adequacy Post-Consolidation. *Top Journal of Economics and Finance*, 7, 6.

Quint D., Rabanal P. (2014) *Monetary and Macroprudential Policy in an Estimated DSGE Model of the Euro Area*. Available at: www.paurabanal.net/research.html.

Rahmaningtyas N.A. (2022) The Effect of Interest Rate and Inflation on Loan Growth and Non Performing Loan (Literature Review of Financial Management). *DIJEMSS ,* 4, 1, pp. 128–137.

Ranisavljević D., Hadžić, M. (2016) Realistic Evaluation of the Ratio: Loan-To-value – The Key to Minimising the Credit Risk. *Economic Themes*, 54, 3, pp. 449–468. https://doi.org/10.1515/ethemes-2016-0022

Rojas D., Vegh C., Vuletin G. (2020) *The Macroeconomic Effects of Macroprudential Policy: Evidence from a Narrative Approach.* Working Paper 27687. NBER.

Sasikirono N., Sumanto S., Sudana I., Meidiaswati H. (2019) Loan-to-Value Policy and Property Loans Risk in Conventional Commercial Banks of Indonesia. *International Research Journal of Business Studies*, XII(03), pp. 267–276.

Son T.H., Liem N.T., Khuong N.V. (2020) Corruption, Nonperforming Loans, and Economic Growth: International Evidence*. Cogent Business & Management,* 7, 1. Doi: 10.1080/23311975.2020.1735691

Spelta A., Pecora N., Rovira Kaltwasser P. (2018) Identifying Systemically Important Banks: A Temporal Approach for Macroprudential Policies. *Journal of Policy Modeling.* Doi: 10.1016/j.jpolmod.2018.06.004

Tabish S.Z.S., Jha K.N. (2012) The Impact of Anti-Corruption Strategies on Corruption Free Performance in Public Construction Projects. *Construction Management and Economics,* 30, 1, pp. 21–35. Doi: 10.1080/01446193.2011.654128

Thiemann M., Aldegwy M., Ibrocevic E. (2018) Understanding the Shift from Micro-to Macro-Prudential Thinking: A Discursive Network Analysis. *Cambridge Journal of Economics*, 42, pp. 935–961. https://doi.org/https://doi.org/10.1093/cje/bex056

Toader T., Onofrei M., Popescu A.-I., Andrieș A.M. (2017) Corruption and Banking Stability: Evidence from Emerging Economies*. Emerging Markets Finance and Trade,* 54, 3, pp. 591–617*.* Doi: 10.1080/1540496x.2017.1411257

Vatansever M., Hepsen A. (2013) Determining Impacts on Non-Performing Loan Ratio in Turkey. *Journal of Finance and Investment Analysis*, 2, 4, pp. 1–7.

Wang B., Sun T. (2013) *How Effective Are Macroprudential Policies in China?* IMF Working Paper no 13/75.

Weill L. (2011) How Corruption Affects Bank Lending in Russia*. Economic Systems,* 35, 2, pp. 230–243*.* Doi: 10.1016/j.ecosys.2010.05

Wijayanti R., Adhi N., Harun C. (2020) *Effectiveness of Macroprudential Policies and Their Interaction with Monetary Policy in Indonesia.* BIS Paper 110d.

Yao F., Lu B. (2019) *The Effectiveness of Loan-to-Value Ratio Policy and its Interaction with Monetary Policy in New Zealand: An Empirical Analysis Using Supervisory Bank-Level Data*. BIS Working Paper no 110.

Zhang L., Zoli E. (2014) *Leaning Against the Wind: Macroprudential Policy in Asia.* IMF Working Paper WP/14/22.

Zhang X., Li F., Li Z., Xu Y. (2018) Macroprudential Policy, Credit Cycle, and Bank Risk-Taking. *Sustainability (Switzerland)*, 10, 10. https://doi.org/10.3390/su10103620

Zulkhibri M. (2018) Macroprudential Policy and Tools in a Dual Banking System: Insights from the Literature*. Borsa Istanbul Review.* Doi:10.1016/j.bir.2018.04.001

Имеет ли макропруденциальная политика значение для управления банковским кредитным риском? Данные коммерческих банков Азиатско-Тихоокеанского региона

Заинури Заинури[1], Вифиндрартин Себастьяна[2], Вилантари Регина Никен[3], Розик Ахмад[4]

[1] Доцент, Экономический департамент, факультет экономики и бизнеса,
Джемберский университет,
Джембер 68121, Индонезия.
E-mail: zainuri.feb@unej.ac.id

[2] Доцент, Экономический департамент, факультет экономики и бизнеса,
Джемберский университет,
Джембер 68121, Индонезия.
E-mail: sebastiana@unej.ac.id

[3] Доцент, Экономический департамент, факультет экономики и бизнеса,
Джемберский университет,
Джембер 68121, Индонезия.
E-mail: reginanikenw.feb@unej.ac.id

[4] Профессор, Департамент бухгалтерского учета, факультет экономики и бизнеса,
Джемберский университет,
Джембер 68121, Индонезия.
E-mail: ahmadroziq.feb@unej.ac.id

Мировой финансовый кризис вызвал дискуссию о плюсах и минусах использования макропруденциальной политики в качестве инструмента пруденциального контроля, который включает резервы капитала или требования для устранения системного риска, финансовых кредитных циклов и целей макроэкономической стабилизации. Макропруденциальная политика в настоящее время прочно утвердилась в качестве области финансовой политики, призванной остановить принятие чрезмерных рисков в финансовом секторе и уменьшить их последствия для реальной экономики в ответ на уроки, извлеченные из мирового финансового кризиса. Целью исследования является проверка эффективности инструментов макропруденциальной политики в контроле общего банковского кредитного риска в Азиатско-Тихоокеанском регионе в период с 2012 по 2019 год. В качестве инструмента анализа в исследовании использовался обобщенный метод момента (GMM). Анализ показал, что инструменты макропруденциальной политики эффективно управляют кредитным риском в банках Азиатско-Тихоокеанского региона. По результатам исследования можно сделать вывод, что инструменты «Коэффициент достаточности капитала» и «Ссуда к стоимости» положительно и существенно влияют на необслуживаемые кредиты, как индивидуально, так и в глобальном масштабе.

Ключевые слова: коэффициент достаточности капитала; буфер консервации капитала; коррупция; стоимость кредита; кредитный риск; GMM.

Подписано в печать 14.03.2025. Формат 70×100 1/16.
Печ. л. 11,0. Тираж 150 экз. Заказ №

Отпечатано в ООО «Фотоэксперт»
109316, Москва, Волгоградский проспект, д. 42

ОБЪЕДИНЕННЫЙ КАТАЛОГ «ПРЕССА РОССИИ»
ПОДПИСНОЙ ИНДЕКС **79264**