https://doi.org/10.51408/issi2025_144

# Regional Patterns of Plagiarism: Evidence from PhD Theses in Russia

Anna Abalkina[1], Alexander Libman[2], Andrey Zayakin[3]

[1]*anna.abalkina@fu-berlin.de*, [2]*alexander.libman@fu-berlin.de*
Freie Universität Berlin (Germany)

[3]*a.zayakin@gmail.com*
The Insider (Latvia)

## Abstract

The paper explains the regional patterns of plagiarism in PhD theses (dissertations) in Russia. Using data from more than 108 thousand dissertations dated between 1996 and 2021, which exhibit a significant amount of text similarities with at least 1,000 6-grams, we explore the regional drivers of scientific misconduct and the science characteristics of each region. Applying two sets of linear regressions, we find a strong negative correlation between the share of dissertations with plagiarism and urbanization, the share of ethnic Russians, and the quality of science in the region. Additionally, we identify a strong positive correlation between plagiarism and the level of corruption within a region.

## Introduction

The discrepancy in the frequency of academic dishonesty and plagiarism can be attributed to a core-periphery pattern, where scholars from the periphery are more prone to engaging in scientific misconduct. The core-periphery pattern is used to explain the differences in academic community structures and dishonesty practices across countries. Honig and Bedi (2012) conducted an analysis of text similarities among papers presented at the Academy of Management conference and observed that authors affiliated with developing countries were more prone to plagiarism compared to authors from Western countries. This finding was further supported by Citron and Ginsparg (2015), who investigated the reuse of texts available in arXiv and discovered that authors from developing or ex-socialist countries were more inclined to reuse text. Macháček and Srholec (2022) discovered cross-country differences in questionable research practices and identified geographical patterns of publications in predatory journals, which are associated with middle-income countries with relatively large research sectors.

The core-periphery model is used to explain university-specific patterns of academic misconduct, such as scholars from first-tier universities being found to be less susceptible to engaging in academic misconduct. The study by Fanelli et al. (2022) analyzed papers with problematic image duplications and found that the probability of scientific misconduct is higher if the author is affiliated with a low-ranking university. Bagues et al. (2019) also found that questionable practices such as publications in predatory journals are negatively correlated with university ranking. Much less attention is paid to the core-periphery pattern within a country context and regional patterns of scientific misconduct.

We explore the regional variation in scientific misconduct in PhD theses (dissertations) in Russia, a country known for its significant levels of plagiarism in scientific works and dissertations (Guba & Tsivinskaya, 2024), as well as for significant extensive investigations and awareness of such misconduct (Abalkina, 2024).

We look at the regional perspective of misconduct and corruption at the regional level rather than the university level for several reasons. First, corruption exhibits deep regional patterns in Russia. Second, the mobility of faculty members in Russia is relatively low, not only between different regions but also between universities (Sivak & Yudkevich, 2015). This is further compounded by high rates of inbreeding, where 64% of faculty members in Russia have studied at the same university where they are currently employed (Yudkevich et al., 2013). As a result, the organization of the scientific community in Russian universities has a significant local effect. Thus, the prevalence of academic misconduct is primarily attributed to local patterns rather than the transmission of dishonest practices through the mobility of faculty members. Third, since the early 1990s, regional-level analysis has been more appropriate because Russia has experienced four waves of university mergers with the primary goals of establishing universities based on specialized institutes during the transition period, optimizing existing institutions, and establishing federal universities as well as flagship universities (Romanenko & Lisyutkin, 2018). These mergers exhibit a strong geographical pattern and were implemented within the same city or federal region. Regional patterns of scientific dishonesty in Russia remain underexplored in the literature.

Our work contributes to the literature in several ways. We provide an explanation of plagiarism variation at a country level across regions. While previous studies looked at the differences in scientific misconduct prevalence across countries (Honig and Bedi, 2012) or universities (Rudakov et al., 2019), we show that local regional patterns in Russia such as urbanization, corruption, and ethnic structure explain the variation in plagiarism frequency across Russian regions.

## Data and methods

The Russian State Library, which by law deposits the texts of dissertations (Ministry of Education and Science of Russia, 2017) contains more than 1,1 million entries of dissertations dated from 1950 to 2021 that were defended in the Soviet Union and Russia. Among these, 508,352 dissertations were defended in Russia between 1996 and 2021, with 486,586 of them having their texts digitized (Russian State Library, n.d.). Dissernet, a network of researchers and journalists dedicated to identifying plagiarism in dissertations and academic papers written in the Russian language, has performed an automated analysis of text reuse among 460 thousand dissertations taken from the Russian State Library (see Figure 1). Dissernet found more than 111 thousand dissertations dated between 1996 and 2021 with the amount of text similarities with at least 1,000 6-grams, i.e. a sequence of six words in the dissertation. To ensure accuracy, commonly used phrases in dissertations, such as “retaining manuscript rights, the work is accomplished in” (“на правах рукописи

работа выполнена в"), were eliminated. Additionally, the reference lists were not taken into account during the analysis.

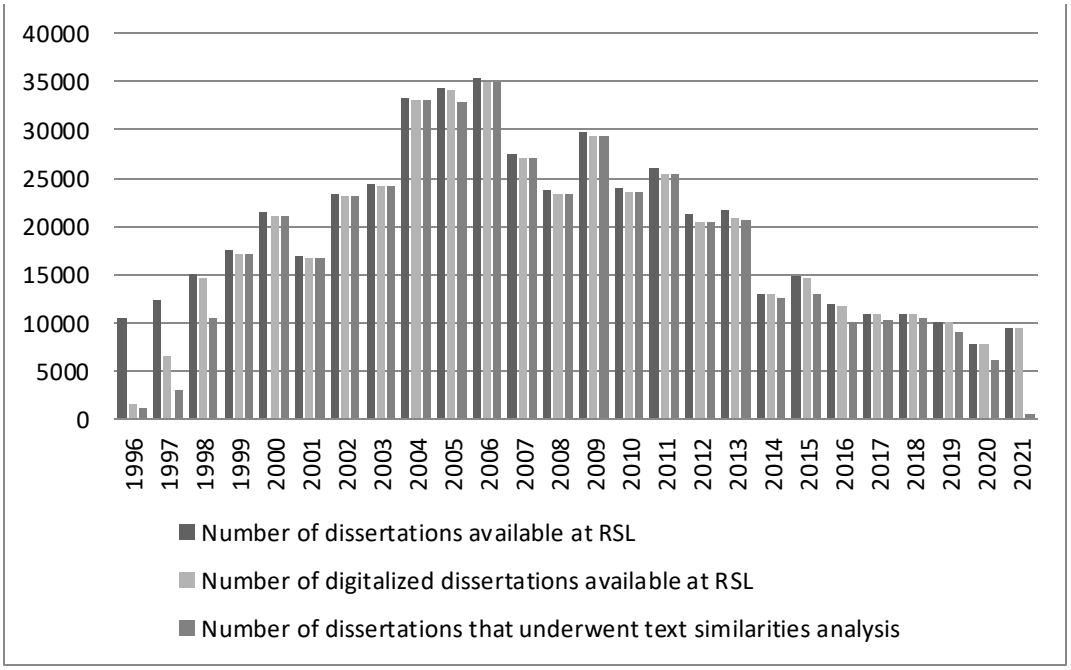


**Figure 1. Number of dissertations available at Russian State Library by year.**

While the literature suggests that manual checks are necessary to qualify text similarities as plagiarism (Weber-Wulff, 2019), the results of this automated analysis did not undergo manual plagiarism checks to avoid false positives. However, we believe that such identified text similarities are highly likely to be instances of plagiarism. First, Dissernet used a conservative approach by setting a threshold of 1,000 6-grams for text similarities, which exceeds the size used in similar studies, such as Citron and Ginsparg (2015), who used 100 7-grams as a threshold. This conservative approach aims to capture large-scale text similarities and takes into account the relatively high tolerance towards plagiarism in Russia (Rudakov et al., 2019). Second, Dissernet manually checked over 12,500 dissertations (Dissernet, n.d.), establishing that dissertations were the primary source of documented cases of plagiarism in dissertations. Third, duplicates with earlier dates are identified as sources, while the subsequent dissertations are recognized as instances of text reuse. The text similarity analysis was performed in 2022. The region of the dissertation was successfully identified for 108 thousand dissertations, which were included in the subsequent econometric analysis.

In order to analyze regional aspects of misconduct, we perform two sets of linear regressions. The first set of regressions deals with the regional drivers of scientific dishonesty, such as the communist legacy measured as the share of members of the Communist Party in the Soviet Union, urbanization, ethnicity (share of ethnic Russians), and corruption. The second set of regressions takes into consideration the science characteristics of the region, such as the growth of dissertations, the quality

of science measured as a share of publications indexed in the Russian Scientific Citation Index, and the quantity of ideological dissertations during Soviet times. We also control for the regional domestic product and the number of dissertations defended in 2000. The dependent variable is the number of dissertations from 1996 to 2020 with automatically detected text similarities of at least 1,000 6-grams, normalized by the number of dissertations in 2000.

**Results**

Tables 1 and 2 present results of multiple linear regressions. Regression results reveal that regional patterns are associated with the share of plagiarized dissertations. Specifically, the findings indicate a statistically significant and negative association between urbanization level and plagiarism, suggesting that plagiarism is more common in rural regions. Furthermore, plagiarism is correlated with the ethnic composition of regions, with ethnic regions exhibiting more widespread plagiarism. Additionally, plagiarism, as a form of corruption, is associated with the overall level of corruption in regions, indicating that corruption extends to universities as well.

In some regression specifications, the share of communists in the Soviet Union is also statistically significant in explaining the variation in plagiarism by region. It is known that the share of communists explains variations in corruption and inequality in Russian regions (Libman & Obydenkova, 2021), which is also indirectly related to differences in the spread of plagiarism in Russian regions.

Variation in plagiarism is also associated with scientific patterns. In particular, plagiarism is less prevalent in regions where science is stronger. At the same time, plagiarism is more common in regions where there was a higher growth in dissertation defenses, which apparently indicates a mechanism for the spread of plagiarism. In other words, the increase in the number of dissertations was directly linked to dishonest defenses.

Another aspect of the development of science in Russia is associated with the adaptation of disciplines, especially social sciences, due to the transition to a market economy. Many social sciences were essentially established from scratch in the 1990s. There is evidence that faculty members who previously taught ideological disciplines, such as, for example, the history of the CPSU and dialectical materialism, among others, helped to organize networks of plagiarized dissertations in the newly established social sciences. However, our analysis showed that the share of ideological dissertations defended during the Soviet period turned out to be statistically insignificant in explaining the variation of plagiarism in Russia.

**Table 1. Regional patterns of plagiarism variation in Russian regions.**

| | (1) | (2) | (3) |
|---|---|---|---|
| **VARIABLES** | **Plagiarism** | **Plagiarism** | **Plagiarism** |
| Urbanization | -0.714*** | -0.527*** | -0.341*** |
| | (0.123) | (0.0894) | (0.0753) |
| Share of communists | 1.114** | 1.487** | 0.534 |
| | (0.487) | (0.576) | (0.409) |
| Share of ethnic Russians | | -16.85*** | -7.635*** |
| | | (3.812) | (2.611) |
| Corruption | | | 10.20** |
| | | | (4.184) |
| Constant | 62.92*** | 59.91*** | 41.99*** |
| | (9.207) | (5.460) | (4.866) |
| Observations | 70 | 70 | 63 |
| R-squared | 0.508 | 0.621 | 0.514 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table 2. Science pattern of plagiarism variation in Russian regions.**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VARIABLES | Plagiarism | Plagiarism | Plagiarism | Plagiarism |
| Urbanization | -0.300*** | -0.201** | -0.0926 | -0.0872 |
| | (0.0832) | (0.0852) | (0.0815) | (0.0819) |
| Share of communists | 0.501 | 0.332 | 0.143 | 0.150 |
| | (0.399) | (0.404) | (0.381) | (0.383) |
| Share of ethnic Russians | -7.872*** | -9.390*** | -9.431*** | -9.013*** |
| | (2.467) | (2.290) | (2.175) | (2.345) |
| Corruption | 9.876** | 12.17*** | 13.54*** | 13.68*** |
| | (4.083) | (4.228) | (3.632) | (3.665) |
| Log Regional domestic product | -1.603 | -2.022 | -1.545 | -1.465 |
| | (1.999) | (2.088) | (1.596) | (1.609) |
| Growth of dissertations | | 0.170** | 0.152*** | 0.148*** |
| | | (0.0660) | (0.0474) | (0.0466) |
| Quality of science (Share of RINC publications) | | | -0.961*** | -0.989*** |
| | | | (0.181) | (0.187) |
| Share of ideological dissertations | | | | 0.0756 |
| | | | | (0.129) |
| Constant | 59.77** | 57.99** | 50.15*** | 48.11** |
| | (22.45) | (23.46) | (17.69) | (18.41) |
| Observations | 63 | 63 | 63 | 63 |
| R-squared | 0.522 | 0.581 | 0.683 | 0.685 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

## Conclusions

This study explores regional patterns of plagiarism in Russian dissertations. The results of the analysis showed that regional characteristics determined the spread of plagiarism. This research also contributes to the understanding of scientific misconduct through the lens of the core-periphery pattern.


## Acknowledgments

Andrey Zayakin thanks Freie Universität Berlin for scholars at risk grant that allowed to design the study.


## Competing interests

Andrey Zayakin is a co-founder of Dissernet.

## Author contributions

AA – conceptualization, formal analysis, methodology, investigation, visualization, writing – original draft, AL – conceptualization, methodology, supervision, AZ – conceptualization, data curation, formal analysis, methodology, investigation, writing – original draft.

## References


Abalkina, A. (2024). Quality and Policies for Academic Integrity: Challenges Faced by Russian Universities. In: Eaton, S.E. (eds) Second Handbook of Academic Integrity. Springer International Handbooks of Education. Springer, Cham. https://doi.org/10.1007/978-3-031-54144-5_174

Bagues, M., Sylos-Labini, M., & Zinovyeva, N. (2019). A walk on the wild side: 'Predatory' journals and information asymmetries in scientific evaluations. *Research Policy*, *48*(2), 462-477. https://doi.org/10.1016/j.respol.2018.04.013

Citron, D. T., & Ginsparg, P. (2015). Patterns of text reuse in a scientific corpus. Proceedings of the National Academy of Sciences, 112(1), 25-30. https://doi.org/10.1073/pnas.1415135111

Dissernet (n.d.). Dissernet in exact numbers. Retrieved on 02.07.2023 from https://web.archive.org/web/20230702195941/https://dissernet.org/

Fanelli, D., Schleicher, M., Fang, F. C., Casadevall, A., & Bik, E. M. (2022). Do individual and institutional predictors of misconduct vary by country? Results of a matched-control analysis of problematic image duplications. *PLOS ONE, 17*(3), e0255334. https://doi.org/10.1371/journal.pone.0255334

Guba, K.S., Tsivinskaya, A.O. (2024). Ambiguity in Ethical Standards: Global Versus Local Science in Explaining Academic Plagiarism. *Science and Engineering Ethics*, **30**, 4. https://doi.org/10.1007/s11948-024-00464-6

Honig, B., & Bedi, A. (2012). The fox in the hen house: A critical examination of plagiarism among members of the Academy of Management. *Academy of Management Learning & Education, 11*(1), 101–123. https://doi.org/10.5465/amle.2010.0084

Libman, A., & Obydenkova, A. V. (2021). *Historical legacies of communism: Modern politics, society, and economic development*. Cambridge University Press.

Macháček, V., & Srholec, M. (2022). Predatory publishing in Scopus: Evidence on cross-country differences. *Quantitative Science Studies, 3*(3), 859-887. https://doi.org/10.1162/qss_a_00213

Ministry of Education and Science of Russia (2017). Order of the Ministry of Education and Science of Russia N 1093 "On approval of the Regulations on the Council for the defense of dissertations for the degree of candidate of science, for the degree of doctor of science", November 10, 2017.

Romanenko, K. & Lisyutkin, M. (2018) University Mergers in Russia, Russian Education & Society, 60:1, 58-73, DOI: 10.1080/10609393.2018.1436295

Rudakov, V., Roshina, Ya., Bitokova, L. (2019). Izmeneniya strategiy, motivatsiy i ekonomicheskogo povedeniya studentov i prepodavateley rossiyskikh vuzov (Changes in strategies, motivations and economic behavior of students and teachers of Russian universities). Higher School of Economics.
URL: https://memo.hse.ru/data/2019/03/05/1196154632/2019_inbul_133(1).pdf

Russian State Library (n.d.). General electronic catalogue. Retrieved on 01.07.2023 from https://search.rsl.ru/ru/search

Sivak, E. & Yudkevich, M. (2015). Academic immobility and inbreeding in Russian University sector. In M. Yudkevich, P. G. Altbach, & L. E. Rumbley (Eds.), Academic inbreeding and mobility in higher education. Global perspectives (pp. 130–155). Basingstoke, Palgrave Macmillan.

Weber-Wulff, D. (2019). Plagiarism detectors are a crutch, and a problem. Nature, 567.

Yudkevich, M., Kozmina, Y., Sivak, E., Bain, O., & Davydova, I. (2013). Changing academic profession: Russia Country report. Moscow: Higher School of Economics.